Author/s:
Sharma, S;Wilson, R;Pryymachenko, Y;Pathak, A;Chua, J;Gwynne-Jones, D;Bennell, KLL;Metcalf, B;Hinman, RSS;Hunt, MAA;Lim, B-W;Abbott, JH

Title:
Reliability, Validity, Responsiveness, and Minimum Important Change of the Stair Climb Test in Adults With Hip and Knee Osteoarthritis

Date:
2023-05-01

Citation:
Sharma, S., Wilson, R., Pryymachenko, Y., Pathak, A., Chua, J., Gwynne-Jones, D., Bennell, K. L. L., Metcalf, B., Hinman, R. S. S., Hunt, M. A. A., Lim, B. -W. & Abbott, J. H. (2023). Reliability, Validity, Responsiveness, and Minimum Important Change of the Stair Climb Test in Adults With Hip and Knee Osteoarthritis. Arthritis Care and Research, 75 (5), pp.1147-1157. https://doi.org/10.1002/acr.24821.

Sharma Saurab (Orcid ID: 0000-0002-9817-5372)

Wilson Ross (Orcid ID: 0000-0001-8505-8081)

Hunt Michael Anthony (Orcid ID: 0000-0002-8648-1591)

Abbott J. Haxby (Orcid ID: 0000-0001-6468-7284)

TITLE PAGE

Title: Reliability, validity, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Authors: Saurab Sharma PhD [1, 2, 3], Ross Wilson PhD [3], Yana Pryymachenko PhD [3], Anupa Pathak BPT [3], Jason Chua PhD [3], David Gwynne-Jones MA, BM BCh, FRACS (Orth) [3], Kim L. Bennell PhD [4], Ben Metcalf BSc (Hons) [4], Rana S. Hinman PhD [4], Michael A. Hunt PhD [5], Boon-Whatt Lim PhD [6], J. Haxby Abbott PhD [3]

Affiliations:

[1] School of Health Sciences, Faculty of Medicine, University of New South Wales, Sydney, Australia
[2] Centre for Pain IMPACT, Neuroscience Research Australia, Sydney, Australia
[3] Centre for Musculoskeletal Outcomes Research, Department of Surgical Sciences, Otago Medical School, University of Otago, Dunedin, New Zealand
[4] Centre for Health, Exercise and Sports Medicine, Department of Physiotherapy, University of Melbourne, Australia
[5] Department of Physical Therapy, University of British Columbia, Vancouver, Canada
[6] School of Applied Science, Republic Polytechnic, Singapore

Corresponding author:

Saurab Sharma, PhD
Postdoctoral Research Fellow
Centre for Pain IMPACT, Neuroscience Research Australia, Sydney, Australia
139 Barker Street, Randwick, 2031, Sydney, Australia
Email: saurabsharma1@gmail.com
Phone: + 61 403 785 336

Funding sources:

The funding sources do not have any input on the current analysis and reporting of this paper. The trials were supported by the National Health and Medical Research Council (Projects 628644, 628556, 454686) and United Pacific Industries through a grant from the Physiotherapy Research Foundation, Australia. SS is supported by the International Association for the Study of Pain (IASP) John J. Bonica Postdoctoral Fellowship. AP is supported by the University of Otago Doctoral Scholarship. RSH is supported by National Health and Medical Research Council Senior Research Fellowship (#1154217). KLB is supported by a National Health & Medical Research Council Investigator grant (#1174431).

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/acr.24821

Current Manuscript word count: 3789

Running headline: Clinimetrics of the Stair Climb Test

ABSTRACT

Objective: The Osteoarthritis Research Society International (OARSI) recommends assessment of physical function using a performance-based test of stair negotiation, but was unable to recommend any specific test. We assessed the reliability, validity, responsiveness, measurement error, and minimum important change (MIC) of the 6-step timed Stair Climb Test (SCT).

Methods: We used pooled data from 397 participants with hip or knee osteoarthritis (54% women) from four clinical trials (86% retained at 12-week follow-up). Construct validity was assessed by testing six *a priori* hypotheses against other OARSI-recommended physical function measures. A self-reported Global Rating of Change scale was used to classify participants as *worsened*, *improved* and *stable*. Participants who worsened in physical function were excluded from all analyses. Responsiveness and MIC were assessed using multiple anchor-based and distribution-based approaches. Test-retest reliability, standard error of measurement (SEM) and smallest detectable change (SDC) were assessed on *stable* participants.

Results: Five of six hypotheses (83%) for construct validity were met. Test-retest reliability was excellent (intraclass correlation coefficient$_{2,1}$: 0.83; 95% confidence interval: 0.71, 0.90). The SEM and SDC values were 0.44 and 1.21 seconds respectively. We did not find adequate support for responsiveness. The MIC values ranged from 0.78 to 1.95 seconds using different approaches (median=1.37 seconds).

Conclusion: The 6-step timed SCT adequately assesses the construct of physical function in individuals with hip or knee osteoarthritis with excellent 12-week test-retest reliability. However, support for its responsiveness was inadequate to recommend its use as an outcome measure in people with osteoarthritis for research and clinical practice.

Keywords

Osteoarthritis; Physical Function; Clinimetrics; Stair Climb Test; Knee; Hip

SIGNIFICANCE AND INNOVATIONS

- The Stair Climb Test (SCT) is recommended measure of physical function for hip and knee osteoarthritis without sufficient empirical evidence around its measurement properties and minimum important change.
- This study confirmed construct validity and test-retest reliability of the SCT, but could not find support for its responsiveness.
- We reported minimum important change values for the SCT for the first time.

Osteoarthritis is a common and debilitating long-term musculoskeletal condition. Impaired physical functioning is a key contributor to disability in osteoarthritis, and improvement in physical function is a common target of many osteoarthritis interventions. Physical function is thus a recommended outcome domain by the Osteoarthritis Research Society International (OARSI) and the Outcome Measures in Rheumatology (OMERACT), endorsed by all relevant stakeholders including patients, clinicians, and researchers[1,2].

Clinical guidelines on hip and knee osteoarthritis frequently recommend a physical performance test, the Stair Climb Test (SCT)[1,3,4]. The SCT measures the time it takes an individual to climb and descend a set number of stairs. It has numerous variations (e.g., varying the number or height of steps, completing the test in set time or timed test to ascend and descend set number of steps)[5-9]. It is feasible (takes less than five minutes to complete) and demands minimal equipment (staircase and stopwatch)[9]. It can also be tested using a mobile app, adding the potential for use in tele-rehabilitation and 'virtual' remote consultations[10]. It can be a useful test as it tests stair climbing, a common activity of daily living (ADL) that people with hip and knee osteoarthritis report difficulty on which is not measured by other OARSI recommended physical performance tests (e.g., walking tests).

The SCT is increasingly used in clinical trials in hip and knee osteoarthritis[7,11-14] and other populations[15,16] despite limited evidence on its measurement properties. Although it has been described as a test that conceptually measures lower extremity strength, power, and balance (i.e. construct validity) [6,9], the empirical evidence to support this is contradictory[17,18]. Evidence for its reliability and responsiveness in individuals with hip and knee osteoarthritis is mixed, showing (1) unacceptable to excellent reliability[4-6,9,19] and (2) acceptable responsiveness following hip and knee arthroplasty[9] but no evidence for its responsiveness in hip and knee osteoarthritis managed non-surgically based on two relatively small studies[17,18]. Coleman and colleagues, in a recent narrative review, also highlighted the lack of evidence on its minimum important change (MIC; the smallest within-subject change in score that patients consider to be important)[9].

Given the gap and inconsistencies in the literature, we aimed to assess the measurement properties of the SCT in individuals with hip or knee osteoarthritis who received non-surgical interventions. Specifically, we planned to assess its test-retest reliability (including measurement error), construct validity, responsiveness, and MIC.

PARTICIPANTS AND METHODS

Study design and settings

We used data from four completed randomised controlled trials from the same research centre that used the SCT as an outcome alongside additional outcome measures of physical function. All four trials assessed conservative non-drug, non-surgical treatment interventions and were conducted with community-dwelling participants with hip or knee osteoarthritis in Melbourne, Australia. We used data at baseline (T0) and 12-weeks post-treatment follow-up (T1) from each trial. We followed the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) recommendations for assessing all measurement properties[20-22]. Fifty participants are considered adequate to evaluate different measurement properties of outcome measures[23].

All participants provided written informed consent to participate in the respective trials. The University of Melbourne Human Research Ethics Committee approved all studies. All trials were registered with the Australia New Zealand Clinical Trials Registry. The study design and samples have been described previously[7,12-14]. Study participant selection criteria are summarised in **Table 1**.

Measures

The outcome measures used, and the constructs assessed in the study are described below and summarised in **Table 2**. Not all trials collected data on all measures; the specific outcome data collected in each trial are indicated in **Table 2**. All tests have shown to have good to excellent reliability in people with hip and knee osteoarthritis[5,9,14,24-27].

Stair climb test (SCT)

The SCT is a physical performance test in which participants are asked to ascend and descend a flight of stairs[6,9]. In all trials providing data for this study, participants completed a timed SCT with 6 steps, each 17.5cm high[7,12,14], at their own pace[4]. Participants could use handrails if preferred. The total time taken to complete the task was recorded in seconds; longer times indicate poorer physical function[4,6,9].

Physical performance tests for comparison

The 30-second sit-to-stand test assesses the number of completed sit-stand-sit repetitions in 30 seconds[4]. A chair (43.2 cm or 17 inches high) with straight back support placed against a wall was used to perform the tests[1].

The 40m fast-paced walking test (2X20m) assesses ability to walk as quickly as possible over a 20m space, turn, and return to the initial position without overexerting[1]. The average walking speed (in meters per second, m/s) is recorded. The time required to turn is not included in the recorded time. One can use walking aids if preferred.

The step test assesses a person's ability to step rapidly in the forward direction[28]. Participants are required to stand on the study leg, and complete as many steps as possible by the non-study leg on to a 15-cm high step and back to the floor in 15 seconds. The total number of steps completed is recorded.

The 4-square step test assesses the ability to rapidly change direction while stepping forward, backward, and sideways over a low obstacle[25,29]. Time to complete the test is recorded in seconds.

Quadriceps strength was assessed using an isokinetic dynamometer (Kin-Com 125-AP; Chattecx Corporation, Chatanooga, TN, USA)[14]. Isometric quadriceps strength at knee flexion at 60 degrees was recorded in a sitting position. The best score of the three tests was used. Quadriceps strength is recorded as maximum torque normalized by body mass (Nm/kg).

Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)

The WOMAC 17-item Likert version 3.1 physical function subscale was used to assess physical functioning[26,27]. Each item is rated on a score of 0 to 4, with total score thus ranging from 0 (representing no difficulty) to 68. Higher scores indicate lesser physical function.

Global rating of change (GRoC)

The Global Rating of Change (GRoC) scale asked participants to rate their change in physical function since the baseline assessment. The GRoC was used to classify participants as "worsened", "stable" or "improved" at the 12-week follow-up. Two versions of the GRoC were used in three trials: a 7-point GRoC (GRoC-7)[7,13] and a 5-point GRoC (GRoC-5) scale[12]. Response options for each of these GRoC scales are shown in **Supplementary Figure 1**.

Scores from the GRoC-7 were combined with those from the GRoC-5 by collapsing response items 6 ("Moderately better") and 7 ("Much better"). All participants who deteriorated were excluded from the analyses. Previous studies have found that choices of GRoC scales based on response options do not affect the responsiveness of measurement instruments[30]. GRoC scales have shown to have adequate reliability[31].

A priori, we defined GRoC scores to classify participants as stable and improved based on two sets of criteria. For our primary stringent analysis, we categorised participants as stable if they reported "no change" (score of 3 on GRoC-5 and score of 4 on GRoC-7) and as improved if they reported "moderately improved" or "much improved" (score of 5 on GROC-5 or score of 6/7 on GROC-7)[7,13]. In secondary analyses, we categorised participants as stable if they reported

"slightly worse", "no change", or "slightly improved", for both GRoC scales, and as improved if they reported "moderately improved" or "much improved". These more relaxed criteria have been used previously (due largely to the need to obtain a large enough sample of 'stable' participants for analysis)[32,33].

Data analysis

Sociodemographic variables (age, sex, body mass index), and osteoarthritis-related parameters (duration of symptoms, Kellgren-Lawrence radiographic score) were reported using descriptive statistics (count and percentage for discrete variables; mean and standard deviation (SD) for continuous measures). We performed complete case analyses on an analysis-by-analysis basis; that is, we discarded cases with missing responses for the SCT scores and other comparator measures for each analysis.

Reliability

Test-retest reliability at 12 weeks was estimated on the *stable* group using both stringent and relaxed criteria. The intraclass correlation coefficient ($ICC_{2,1}$) with 95% confidence interval (CI) was calculated using the two-way random effects model. ICC ≥ 0.70 was considered sufficient[23].

Measurement error

The standard error of measurement (SEM) was computed as $SD_{change} \times \sqrt{1 - ICC}$ in the *stable* group[34]. $ICC_{2,1}$ calculated above was used in the formula. The smallest detectable change (SDC) was computed as $1.96 \times \sqrt{2} \times SEM$ [23]. Additionally, we constructed Bland-Altman plots to view systematic errors of the two measurements[35].

Construct validity

We assessed construct validity using hypothesis testing as described by COSMIN guidelines[20], comparing the baseline scores of the SCT with the baseline scores of other measures of physical function. We hypothesized that the SCT would strongly correlate with tests

(namely the 30 second sit-to-stand test, the step test, and the 4-square step test) that assess at least two of the four sub-constructs of physical performance assessed by the SCT (strength, speed, power, balance; see **Table 2**). We also hypothesized that the SCT would at least moderately correlate with tests (namely the fast-paced walk test and quadriceps strength) that assess only one of the four sub-constructs of physical function assessed by the SCT, or patient-reported physical function (the WOMAC-function sub-scale). We determined the constructs assessed by each test conceptually and based on previous research[1,5,6,19].

The Pearson or Spearman correlation coefficient was used to assess the strength and magnitude of correlations for all six *a priori* hypotheses for normally distributed and non-normally distributed data respectively. We considered correlation coefficients <0.30 as weak, from 0.30 to 0.49 as moderate, and ≥ 0.50 as strong *a priori*[36]. If at least 75% of hypotheses were met (i.e. 5 of 6 hypotheses)[23], we considered the timed 6-step SCT to be a valid measure of physical function (i.e. as having sufficient construct validity[37]).

Responsiveness

Responsiveness is the ability of an instrument to detect change over time in the construct to be measured. We assessed responsiveness using five approaches proposed in the literature. We considered the SCT to be highly responsive if the hypotheses for four or five of the approaches were met; moderately responsive if three of five criteria were met; and poorly responsive if only one or two criteria were met[37].

The first approach considered was the *construct approach.* It is similar to hypothesis testing for construct validity, the only difference being that the unit of comparison is the *change* score rather than *baseline* score. We tested the same six hypotheses as for construct validity, using change scores. At least 75% hypotheses needed to be met to support responsiveness using this approach.

The second approach was the *criterion approach*. The GRoC was used for this purpose to categorise the sample into *improved* and *stable* groups. Receiver Operating Characteristic (ROC) curves were plotted to assess responsiveness. A ROC curve plots the true positive scores (sensitivity) against false positive scores (1-specificity). An area under the curve (AUC) value of 0.7 was considered as the threshold for the ability of the SCT to discriminate the *improved* and *stable* groups (AUC of 0.5: complete inability to discriminate; AUC of 1.0: perfect ability to discriminate)[23,38].

The third and fourth approaches were the effect size and standardized response mean (SRM) approaches respectively. We hypothesised *a priori* that the effect size and SRM of the SCT would be lie within ±0.10 of the *range* of effect size or SRM of other 6 comparator measures of physical function. The ±0.10 allowed to capture the effect sizes/SRM slightly outside the range available from other tests. Effect sizes were calculated by dividing mean change score by SD of baseline SCT scores and each comparator measure. SRM was computed as mean change score divided by SD of change score. We mapped the effect size and SRM of SCT against their respective *range* of values (±0.10) of the comparator measures. Effect sizes and SRM are considered acceptable methods for assessing responsiveness if *a priori* hypotheses about the expected magnitude of effect size and SRM are used[38].

The final approach was the between-group construct method. We hypothesised that the mean difference in SCT change scores between the stable and improved groups would be greater than the SDC. We also considered the SCT mean change scores for different categories of the GRoC scale (combined to GRoC-5 as described above). We hypothesised that the SCT mean change scores would *monotonically* increase with increasing levels of improvement in the GRoC[38,39]. Both these criteria needed to be met in order to score positively in this category.

Interpretability

We explored data distribution and MIC for interpretability. As SCT scores are continuous and can theoretically range from 0 to infinity (i.e. has no upper limit), we did not assess floor and

ceiling effects – the third recommended approach for the interpretability[38]. Data distribution was inspected visually for the SCT and comparator instruments for the baseline and change scores.

The MIC score is the minimum change score of an instrument that study participants consider (clinically) meaningful. While clinically meaningful deterioration or worsening can be computed, we focused only on meaningful improvement in this paper because of a small number of participants who reported worsening at 12 weeks. We used two anchor-based and two distribution-based approaches to estimate MIC as previously reported[39-42]. Fifty participants are considered adequate to estimate MIC[23].

Anchor-based approaches for MIC. The GRoC in physical function was used as an external criterion as per recommendations for the anchor-based approach[23,38]. The first method to estimate MIC was the ROC curve approach. The estimated MIC was the score of the SCT resulting in minimum misclassification, i.e., the value of the SCT minimising the sum of *false positive* (1-specificity) and *false negative* (1-sensitivity) classifications[38,39]. Primary analysis was the anchor-based approach using the stringent criteria. The second anchor-based approach was using the mean change score in the sub-group of participants who reported "slightly improved" in the GRoC scale[44], as recommended by Jaeschke and colleagues[43].

Distribution-based approaches for MIC. The first of the two distribution-based approaches used was the effect size approach. As proposed by Norman and colleagues[44] and used by others[45,46], we considered a medium effect size (i.e. 0.50) as the MIC, which is 0.50 × SD of baseline scores, Jaeschke and colleagues found that the MIC values roughly coincide with the medium effect size[43]. The second approach was the SRM approach[46]. As for the effect size approach, we considered the MIC as 0.50 × SD of change score. As other authors have used different thresholds of effect sizes and SRMs for estimating MIC, we additionally computed MIC scores using small effect size (0.20) as the cutoff[42] and the mid-point value between small effect size and medium effect size (i.e. 0.35)[47]. The results using different effect sizes and SRM

thresholds were reported. We summarised the results using median as a measure of central tendency.

Sensitivity analyses for MIC. We conducted sensitivity analyses by (1) analysing data separately for the trials that used GRoC-7 and GRoC-5; and (2) computing MIC separately using the stringent and relaxed categorisations of stable and improved groups as described previously.

RESULTS

Out of 398 participants with hip or knee osteoarthritis, 397 (99.7%) completed the SCT at baseline and 342 (86.1%) at 12-week follow-up. Demographic and baseline characteristics are presented in **Table 3**.

Data distribution

The visual inspection of the data using histograms indicated that data are right skewed (**Supplementary Figure 2**). Data distribution for the change scores were all normally distributed (**Supplementary Figure 3**). Skewness and kurtosis values with their standard errors are presented in **Supplementary Table 1**.

Reliability and measurement errors

The results of the 12-week test-retest reliability and measurement errors are presented in **Table 4**. The SCT was found to have sufficient reliability (ICC $\geq$ 0.70) using both the primary (stringent) and secondary (relaxed) criteria to identify stable participants. The Bland-Altman plots are presented in **Supplementary Figure 4**.

Construct validity

The results of hypothesis testing for construct validity, presented along with respective sample sizes in **Table 5,** showed that 5 out of 6 *a priori* hypotheses (83%) were met (based on

Spearman Correlation Coefficient for non-normally distributed baseline SCT scores) supporting its construct validity.

Responsiveness

Overall, we found poor responsiveness for the SCT, with our *a priori* hypotheses met for only two of five approaches (**Table 6**).

Using the construct approach, 4 of 6 hypotheses (67%) were met; this did not meet our *a priori* cut off of 75% of hypotheses within this approach (**Table 6**). Within the criterion-based approaches, 3 of 5 hypotheses (60%) were met. AUC value was 0.628 (95% CI: 0.523, 0,734) using stringent criteria (see **Tables 4 and 6**). The ROC curves are presented in **Supplementary Figure 5**. The effect sizes and SRM approaches met our *a priori* hypotheses for responsiveness (see **Supplementary Table 2**). Next, although the mean change of the SCT scores was statistically significantly different for the stable and improved groups (Approach 5a, **Table 6**), these values were within the measurement error of SCT scores (i.e. smaller than the SDC values). Finally, the SCT mean change scores monotonically increased corresponding to the improvement indicated by the GRoC categories (**Supplementary Figure 6**).

Interpretability

Minimum important change (MIC)

The MIC values for the SCT ranged from 0.44 to 1.95 seconds using different approaches (**Supplementary Table 3**). The median MIC value using different approaches was 1.37 seconds.

DISCUSSION

This study aimed to assess the measurement properties of the SCT in individuals with hip and knee osteoarthritis – using data from four trials. We found that the SCT had sufficient test-retest reliability and construct validity to assess physical function in hip and knee osteoarthritis, but inadequate responsiveness based on overall *a priori* hypothesis-based criteria for responsiveness.

These results should be viewed considering the limitations of the study. First, these results are based on data from a self-paced 6-step SCT in a research laboratory, and may not be generalisable to other types of SCTs or other settings. Second, as with all studies of measurement properties, there is no true gold standard approach to assess responsiveness and estimate the MIC. To avoid putting greater weight on one approach than others, we assessed responsiveness and MIC using multiple distribution- and anchor-based approaches methods[39-42,45,46]. The use of GRoC to categorise stable and improve groups is prone to recall bias[31,42], especially at a 12-week long follow-up, although it is commonly used and recommended[42,48]. The distribution-based approaches are limited as they do not provide a sense of the clinical relevance of changes and do not take patients' voice on what meaningful change is into account[42], but they are not affected by recall bias and complement anchor-based approaches.

The study also has several strengths. We used current recommendations for assessing measurement properties of outcome measures[20-22]. We used GRoC scales that asked about the change in physical functioning (the construct SCT captures) rather than change in general health[38]. We also supplemented the results using multiple anchor- and distribution-based methods[39,42,45]. With a substantially larger sample size than previous studies, we were able to use more stringent criteria to classify stable and improved groups for reliability and responsiveness assessment and MIC estimation. We place greater confidence in our findings using the stringent criteria because it excludes the "slightly improved" response from the stable category, providing a more homogeneous stable group and more distinct stable and improved groups.

Based on our findings and in light of existing literature[4,5,17-19], the SCT have adequate test-retest reliability when assessed in short term (e.g., after 30 minutes rest after the baseline assessment)[17,18]. Our results support the reliability for longer follow-up (i.e. at 12-week). Between- and within-rater reliability of the 6-step SCT is unknown.

The results support the notion that the SCT measures physical function[6] but is not adequately responsive. Two recent studies did not find evidence for the construct validity and responsiveness of the 10-step SCT[17,18]. Unlike these previous studies, our *a priori* hypotheses for construct validity were based on theoretical considerations; that is, the magnitude of correlation between the SCT and comparable instruments depended on the overlap between the constructs assessed by these instruments. As opposed to previous studies[17,18] that used a single construct approach only to assess responsiveness, we used multiple approaches to triangulate our findings. Our results for construct validity and responsiveness were also based on a larger sample size than previous studies.

The inadequate evidence for responsiveness of the SCT could be due to several reasons. First, responsiveness results depends on the approach used to assess it (we found support for responsiveness using the effect size and SRM approaches but not using anchor-based approaches). Second, responsiveness could also have been influenced by the right skewed SCT baseline data (i.e., a significant proportion of study participants had good baseline physical function), limiting the ability to improve meaningfully post-intervention, especially because participants were instructed to complete the task at their own preferred pace[4]. This may have affected the SCT scores by factors other than ability, such as habit, comfort, caution and safety. This is also supported by the fact that the SCT has been found to be responsive in post-surgical patients following hip and knee arthroplasty, who have much more severe baseline impairment and thus greater scope for improvement[9]. The literature indicates that fast-paced walk tests have greater responsiveness than self-paced walk tests[1,17].

Our study also adds novel findings related to the MIC of the 6-step timed SCT. Using the anchor-based approach and our primary stringent criteria, we found that the MIC value (1.60 seconds) was larger than the SDC value (1.21 seconds) suggesting that changes considered important to patients can feasibly be detected over and above measurement error. The median MIC value (1.37) derived using 0.35 effect size approach also exceeded the SDC value.

However, the range of MIC values obtained by different methods was wide (range 0.44 to 1.95), with several approaches providing a value less than the SDC.

Although the SCT intuitively seems to be a useful test in clinical practice, as it directly assesses the ADL of stair climbing and is inexpensive and feasible to complete, we do not recommend using the 6-step SCT to track improvement in physical function over time in individuals with hip or knee osteoarthritis due to its inadequate responsiveness. Future studies may explore the responsiveness of the SCT on a subset of participants with more debilitating hip or knee osteoarthritis (i.e., on participants with greater room for improvement). Future studies may also examine the measurement properties of the SCT in other populations (e.g., neurological disorders) or settings (e.g., community or telehealth settings), and responsiveness and MIC related to worsening (or deterioration).

The 6-step timed SCT is a reliable and valid physical performance test to assess physical function in individuals with hip and knee osteoarthritis assessed in outpatient physiotherapy settings; however, there is inadequate support for its responsiveness. We do not recommend its use as an outcome measure to track improvement in physical function in individuals with hip or knee osteoarthritis.

Author contributions

SS conducted the data analysis, led the interpretation of findings, and drafted the manuscript. SS and JHA conceived and designed the study. SS, RW, AP, YP, JC and JHA participated in study design, interpretation of findings, revision and final approval of the manuscript. KLB, RSH and BM acquired the data, participated in interpretation of findings, revision and final approval of the manuscript.

Conflict of interest

Authors have no conflict of interests to declare.

REFERENCES

1. 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.
2. Smith TO, Hawker GA, Hunter DJ, et al. The OMERACT-OARSI Core Domain Set for Measurement in Clinical Trials of Hip and/or Knee Osteoarthritis. *J Rheumatol.* 2019.
3. Logerstedt DS, Scalzitti DA, Bennell KL, et al. Knee Pain and Mobility Impairments: Meniscal and Articular Cartilage Lesions Revision 2018. *J Orthop Sports Phys Ther.* 2018;48(2):A1-A50.
4. Rejeski WJ, Ettinger WH, Jr., Schumaker S, James P, Burns R, Elam JT. Assessing performance-related disability in patients with knee osteoarthritis. *Osteoarthritis Cartilage.* 1995;3(3):157-167.
5. Dobson F, Hinman RS, Hall M, et al. Reliability and measurement error of the Osteoarthritis Research Society International (OARSI) recommended performance-based tests of physical function in people with hip and knee osteoarthritis. *Osteoarthritis Cartilage.* 2017;25(11):1792-1796.
6. 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:S350-370.
7. Bennell KL, Egerton T, Martin J, et al. Effect of physical therapy on pain and function in patients with hip osteoarthritis: a randomized clinical trial. *JAMA.* 2014;311(19):1987-1997.

8. Ni M, Brown LG, Lawler D, Bean JF. Reliability, Validity, and Minimal Detectable Change of Four-Step Stair Climb Power Test in Community-Dwelling Older Adults. *Phys Ther.* 2017;97(7):767-773.

9. Coleman G, Dobson F, Hinman RS, Bennell K, White DK. Measures of Physical Performance. *Arthritis Care & Research.* 2020;72(S10):452-485.

10. Adusumilli G, Joseph SE, Samaan MA, et al. iPhone Sensors in Tracking Outcome Variables of the 30-Second Chair Stand Test and Stair Climb Test to Evaluate Disability: Cross-Sectional Pilot Study. *JMIR mHealth uHealth.* 2017;5(10):e166.

11. Liao CD, Lin LF, Huang YC, Huang SW, Chou LC, Liou TH. Functional outcomes of outpatient balance training following total knee replacement in patients with knee osteoarthritis: a randomized controlled trial. *Clin Rehabil.* 2015;29(9):855-867.

12. Bennell KL, Hunt MA, Wrigley TV, et al. Hip strengthening reduces symptoms but not knee load in people with medial knee osteoarthritis and varus malalignment: a randomised controlled trial. *Osteoarthritis Cartilage.* 2010;18(5):621-628.

13. Bennell KL, Kyriakides M, Metcalf B, et al. Neuromuscular Versus Quadriceps Strengthening Exercise in Patients With Medial Knee Osteoarthritis and Varus Malalignment. *Arthritis & Rheumatology.* 2014;66(4):950-959.

14. Lim BW, Hinman RS, Wrigley TV, Sharma L, Bennell KL. Does knee malalignment mediate the effects of quadriceps strengthening on knee adduction moment, pain, and function in medial knee osteoarthritis? A randomized controlled trial. *Arthritis Rheum-Arthritis Care Res.* 2008;59(7):943-951.

15. Mavros Y, O'Neill E, Connerty M, et al. Oxandrolone Augmentation of Resistance Training in Older Women: A Randomized Trial. *Med Sci Sports Exerc.* 2015;47(11):2257-2267.

16. Murphy M, Journeaux S, Hides J, Russell T. Does flexion of the femoral implant in total knee arthroplasty increase knee flexion: a randomised controlled trial. *Knee.* 2014;21(1):257-263.

17. Tolk JJ, Janssen RPA, Prinsen C, van der Steen MMC, Bierma Zeinstra SMA, Reijman M. Measurement properties of the OARSI core set of performance-based measures for hip osteoarthritis: a prospective cohort study on reliability, construct validity and responsiveness in 90 hip osteo-arthritis patients. *Acta Orthop.* 2019;90(1):15-20.

18. 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(9):2898-2909.

19. 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(12):1548-1562.

20. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol.* 2010;10:22.

21. Mokkink LB, De Vet HC, Prinsen CA, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research.* 2018;27(5):1171-1179.

22. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27(5):1147-1157.

23. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34-42.

24. Choi YM, Dobson F, Martin J, Bennell KL, Hinman RS. Interrater and intrarater reliability of common clinical standing balance tests for people with hip osteoarthritis. *Phys Ther.* 2014;94(5):696-704.

25. Moore M, Barker K. The validity and reliability of the four square step test in different adult populations: a systematic review. *Syst Rev.* 2017;6(1):187.

26. 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):1833-1840.

27. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum.* 2001;45(5):453-461.

28. Hill KD, Bernhardt J, McGann AM, Maltese D, Berkovits D. A New Test of Dynamic Standing Balance for Stroke Patients: Reliability, Validity and Comparison with Healthy Elderly. *Physiotherapy Canada.* 1996;48(4):257-262.

29. Dite W, Temple VA. A clinical test of stepping and change of direction to identify multiple falling older adults. *Arch Phys Med Rehabil.* 2002;83(11):1566-1571.

30. Lauridsen HH, Hartvigsen J, Korsholm L, Grunnet-Nilsson N, Manniche C. Choice of external criteria in back pain research: Does it matter? Recommendations based on analysis of responsiveness. *Pain.* 2007;131(1-2):112-120.

31. 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.

32. Sharma S, Palanchoke J, Reed D, Haxby Abbott J. Translation, cross-cultural adaptation and psychometric properties of the Nepali versions of numerical pain rating scale and global rating of change. *Health Qual Life Outcomes.* 2017;15(1):236.

33. Sharma S, Palanchoke J, Abbott JH. Cross-cultural Adaptation and Validation of the Nepali Translation of the Patient-Specific Functional Scale. *J Orthop Sports Phys Ther.* 2018;48(8):659-664.

34. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use.* Oxford University Press, USA; 2015.

35. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-310.

36. Cohen J. A Power Primer. *Psychological Bulletin.* 1992;112(1):155-159.

37. de Boer MR, Moll AC, de Vet HC, Terwee CB, Volker-Dieben HJ, van Rens GH. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt.* 2004;24(4):257-273.

38. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine.* Cambridge, UK: Cambridge University Press; 2011.

39. de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res.* 2007;16(1):131-142.

40. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61(2):102-109.

41. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Out.* 2006;4.

42. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56(5):395-407.

43. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10(4):407-415.

44. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life - The remarkable universality of half a standard deviation. *Med Care.* 2003;41(5):582-592.

45. 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.

46. Gilbert AL, Song J, Cella D, Chang RW, Dunlop DD. What is an important difference in gait speed in adults with knee osteoarthritis? *Arthritis Care Res (Hoboken).* 2020.

47. Chen CX, Kroenke K, Stump TE, et al. Estimating minimally important differences for the PROMIS pain interference scales: results from 3 randomized clinical trials. *Pain.* 2018;159(4):775-782.

48. Liang MH. Evaluating measurement responsiveness. *J Rheumatol.* 1995;22(6):1191-1192.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: Saurab Sharma

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
Time frame: Since the initial planning of the work			
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☒ None	
			Click the tab key to add additional rows.
Time frame: past 36 months			
2	Grants or contracts from any entity (if not indicated in item #1 above).	☐ **None**	
		International Association for the Study of Pain (IASP)	I received John J Bonica Postdoctoral Fellowship from the IASP in 2020 which covers part of my salary.
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☒ **None**	
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☐ **None**	
		My travel and accommodation to Amsterdam for the inaugural in person meeting of Global Alliance for Partners for Pain Advocacy was supported by the convening organization, the International Association for the Study of Pain in December 2019.	My accommodation and travel fares were paid.
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☐ **None**	
		I am a board member for an International Association for the Study of Pain Special Interest Group on Pain Mind and Movement (IASP)	This is unpaid work
		I am also a member of Global Alliance for Partners for Pain Advocacy	This is unpaid work

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: Ross Wilson

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
Time frame: Since the initial planning of the work			
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☒ **None**	
			Click the tab key to add additional rows.
Time frame: past 36 months			
2	Grants or contracts from any entity (if not indicated in item #1 above).	☒ **None**	
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☒ **None**	
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: Yana Pryymachenko

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
Time frame: Since the initial planning of the work			
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☒ None	
			Click the tab key to add additional rows.
Time frame: past 36 months			
2	Grants or contracts from any entity (if not indicated in item #1 above).	☒ **None**	
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☒ **None**	
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: Anupa Pathak

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
Time frame: Since the initial planning of the work			
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☒ None	
			Click the tab key to add additional rows.
Time frame: past 36 months			
2	Grants or contracts from any entity (if not indicated in item #1 above).	☐ **None**	
		University of Otago Doctoral Scholarship	This supported me for my PhD scholarship and provided stipend.
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☒ **None**	
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: Jason Chua

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
		Time frame: Since the initial planning of the work	
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☒ **None**	
			Click the tab key to add additional rows.
		Time frame: past 36 months	
2	Grants or contracts from any entity (if not indicated in item #1 above).	☒ **None**	
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☒ **None**	
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: David Gwynne-Jones

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
		Time frame: Since the initial planning of the work	
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☒ **None**	
			Click the tab key to add additional rows.
		Time frame: past 36 months	
2	Grants or contracts from any entity (if not indicated in item #1 above).	☒ **None**	
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☒ **None**	
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: Professor Kim Bennell

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
Time frame: Since the initial planning of the work			
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☐ **None**	
		National Health and Medical Research Council funded the research trials	NHMRC funded the trials through grants
			Click the tab key to add additional rows.
Time frame: past 36 months			
2	Grants or contracts from any entity (if not indicated in item #1 above).	☒ **None**	
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☐ **None**	
		Wolters Kluwer	Provided personal fees for the production of 'UptoDate' knee osteoarthritis guidelines. This is outside the submitted work
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: Ben Metcalf

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
		Time frame: Since the initial planning of the work	
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☒ **None**	
			Click the tab key to add additional rows.
		Time frame: past 36 months	
2	Grants or contracts from any entity (if not indicated in item #1 above).	☒ **None**	
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☒ **None**	
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: Rana Hinman

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
Time frame: Since the initial planning of the work			
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☐ **None**	
		National Health and Medical Research Council	Provided funding for some of the original trials, and currently funds my salary through a Fellowship.
			Click the tab key to add additional rows.
Time frame: past 36 months			
2	Grants or contracts from any entity (if not indicated in item #1 above).	☒ **None**	
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☒ **None**	
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: Michael Hunt

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
Time frame: Since the initial planning of the work			
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☒ **None**	
			Click the tab key to add additional rows.
Time frame: past 36 months			
2	Grants or contracts from any entity (if not indicated in item #1 above).	☒ **None**	
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☒ **None**	
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: Lim Boon Whatt

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
		Time frame: Since the initial planning of the work	
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☒ **None**	
			Click the tab key to add additional rows.
		Time frame: past 36 months	
2	Grants or contracts from any entity (if not indicated in item #1 above).	☒ **None**	
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☒ **None**	
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

ICMJE DISCLOSURE FORM

Date: 10/1/2021

Your Name: J. Haxby Abbott

Manuscript Title: Validity, reliability, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis

Manuscript Number (if known): ACR-21-0446

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
Time frame: Since the initial planning of the work			
1	All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	☒ None	
			Click the tab key to add additional rows.
Time frame: past 36 months			
2	Grants or contracts from any entity (if not indicated in item #1 above).	☐ **None**	
		Health Research Council of New Zealand	Funding through the University of Otago (no funds received personally)
3	Royalties or licenses	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
4	Consulting fees	☐ **None**	
		Allen + Clarke Regulatory and Policy Consultants	Funding through the University of Otago (no funds received personally)
5	Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events	☒ **None**	
6	Payment for expert testimony	☒ **None**	
7	Support for attending meetings and/or travel	☒ **None**	
8	Patents planned, issued or pending	☒ **None**	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ **None**	
10	Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid	☐ **None**	
		Osteoarthritis Aotearoa New Zealand	unpaid
		The Australian and New Zealand Musculoskeletal Clinical Trials Network	unpaid

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
11	Stock or stock options	☒ **None**	
12	Receipt of equipment, materials, drugs, medical writing, gifts or other services	☒ **None**	
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

Table 2. Measures used and constructs assessed.

Tests	Construct assessed					Data available from			
	Strength	**Speed**	**Power**	**Balance**	**Physical function**	**Bennell 2014a [13] (n=100)**	**Bennell 2014b [7] (n=102)**	**Bennell 2010 [12] (n=89)**	**Lim 2008 [14] (n=107)**
Stair climb test*	√	√	√	√	√	√	√	√	√
30 second sit-to-stand test*	√	√	√	-	√	√	√	-	-
40m fast-paced walking test*	-	√	-	-	√	-	√	-	-
Step test*	√	√	√	√	√	√	√	√	√
4-square step test*	-	√	-	√	√	√	√	-	-
Quadriceps strength*	√	-	-	-	√	√	√	√	√
WOMAC-function**	-	-	-	-	√	√	√	√	√

*Physical performance tests; **Self-reported measure.
Abbreviation: WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

Table 3. Descriptive statistics

Variables	N (%)	Mean (SD)
Participants from		
Sex		
Men	182 (46%)	
Women	215 (54%)	
KL Grade		
Grade 2	137 (35%)	
Grade 3	126 (32%)	
Grade 4	134 (34%)	
Age (years)	397	63.76 (7.97)
Duration of symptoms in (years)	397	6.17 (5.63)
BMI (kg/m^2)	397	28.98 (4.51)
Timed Stair Climb Test Baseline score (s)	397	9.24 (3.90)
Timed Stair Climb Test Follow-up score (s)	342	8.48 (3.34)
WOMAC Function Baseline score	397	26.82 (10.56)
Quadriceps Strength Baseline (Nm/kg)	390	1.37 (0.50)
30s Sit-to-Stand Test Baseline (number of repetitions)	201	10.17 (2.70)
40m Fast-paced Walking Test Baseline (m/s)	100	1.63 (0.34)
Step Test Baseline (number of steps)	397	13.67 (3.64)
4-Square Step Test Baseline (s)	200	9.24 (2.61)

Abbreviations: KL, Kellgren Lawrence Scale; BMI, Body Mass Index, WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index

Table 5. Hypothesis testing for construct validity

Number	Test	Interpretation of score	*A priori* hypothesis	N	*rho*	Was hypothesis met?
1	WOMAC physical function score	Higher scores indicate poorer physical function	Positive moderate or strong correlation	397	0.267*	No
2	Quadriceps muscle strength test	Higher scores indicate better physical function	Negative moderate or strong correlation	390	-0.502**	Yes
3	30 second sit-to-stand test	Higher scores indicate better physical function	Negative strong correlation	201	-0.683**	Yes
4	40m fast-paced walking test	Higher scores indicate better physical function	Negative moderate or strong correlation	100	-0.832**	Yes
5	Step-test	Higher scores indicate better physical function	Negative strong correlation	397	-0.547**	Yes
6	4-square step test	Higher scores indicate poorer physical function	Positive strong correlation	200	0.728**	Yes

Note: All correlations tested were using baseline data from all four trials whenever the data was available.
rho, Spearman Correlation Coefficient for non-normally distributed Stair Climb Test data; n, sample;
*Correlation is significant at the 0.05 level (2-tailed); ** Correlation is significant at the 0.01 level (2-tailed).

Table 1. Study characteristics

Trial	OA site	N	Trial arms	Inclusion criteria	Exclusion criteria
Bennell 2014a [13]	Medial knee osteoarthritis	100	Arm 1: Neuromuscular exercise Arm 2: Quadriceps exercises	1. Age ≥50 years 2. Average knee pain greater than or equal to 25 on a 100mm VAS 3. Pain and tenderness predominantly over the medial knee region, 4. Radiographic medial tibiofemoral osteoarthritis.	1. Knee intra-articular injection or surgery in the past 6 months 2. Corticosteroids orally in the past month 3. Systemic arthritis 4. Knee joint replacement or tibial osteotomy surgeries 5. Other non-pharmacological interventions in the past six months 6. BMI of >36 kg/m^2.
Bennell 2014b [7]	Hip osteoarthritis	102	Arm 1: Physiotherapy including education, advice, manual therapy, home exercise, gait aid if appropriate. Arm 2: Sham treatment (inactive ultrasound)	1. Age ≥50 years 2. Hip osteoarthritis meeting the ACR criteria of pain and radiographic changes [25] 3. Pain in groin or hip for more than 3 months 4. Average pain intensity in past week of at least 40 on a 100mm VAS and 5. At least moderate difficulty with daily activities.	1. A prior hip or knee replacement 2. Other lower limb surgeries or other interventions (physiotherapy or chiropractic treatment) for hip or lumbar spine or both in the past 6 months 3. Walking 30 minutes continuously daily 4. Underwent regular exercises more than once weekly.
Bennell 2010 [12]	Medial knee osteoarthritis	89	Arm 1: Hip strengthening exercises Arm 2: Control group (no intervention)	1. Age >50 years 2. Osteoarthritis of one or both knees 3. Met the ACR criteria for osteoarthritis of knee [26] 4. Average knee pain on walking >3 on am 11-point scale 5. Medial knee pain and medial compartment osteophytes or medial joint space narrowing to ensure medial compartment osteoarthritis 6. Knee alignment ≤182 degrees on a standardized semi-flexed PA X-ray.	1. No or doubtful radiographic OA (KL grade 0 or 1) 2. Knee intra-articular injection or surgery in the past 6 months 3. Corticosteroids orally currently or in the past month 4. Systemic inflammatory arthritis 5. A prior hip or knee replacement or tibial osteotomy 6. Intention to start of currently participating in supervised lower limb strengthening programme. 7. BMI of >35 kg/m^2 8. Medical condition that precludes from participation in exercise programme 9. Unable to walk without gait aid.
Lim 2008 [14]	Medial knee osteoarthritis	107	Arm 1: Supervised home based quadriceps strengthening exercises Arm 2: Control group (no intervention)	1. Osteoarthritis of one or both knees fulfilling the criteria from the ACR [27]. 2. To ensure medial knee osteoarthritis, participants were required to report medial knee pain, and show medial joint space narrowing greater than lateral joint space narrowing.	1. History of lower limb joint replacement 2. Knee surgery within the previous 6 months 3. Intraarticular steroid or hylan G-F 20 injection within the previous 6 months 4. Systematic arthritic condition 5. More than 5 degrees of valgus malalignment on radiograph 6. Intention to start or current participation in physiotherapy for knee OA or lower limb strengthening programme 7. Presence of severe medical condition that preclude safe participation in exercise programme.

Abbreviations: OA, Osteoarthritis; VAS, Visual Analog Scale; PA, posteroanterior; BMI, Body Mass Index; KL, Kellgren Lawrence Scale; N, Total number of participants recruited (one participant in Bennell 2014b did not complete the SCT at baseline and was therefore excluded from the analysis).

Table 4. Hypothesis testing for construct validity

Number	Test	Interpretation of score	*A priori* hypothesis	N	*rho*	Was hypothesis met?
1	WOMAC physical function score	Higher scores indicate poorer physical function	Positive moderate or strong correlation	397	0.267*	No
2	Quadriceps muscle strength test	Higher scores indicate better physical function	Negative moderate or strong correlation	390	-0.502**	Yes
3	30 second sit-to-stand test	Higher scores indicate better physical function	Negative strong correlation	201	-0.683**	Yes
4	40m fast-paced walking test	Higher scores indicate better physical function	Negative moderate or strong correlation	100	-0.832**	Yes
5	Step-test	Higher scores indicate better physical function	Negative strong correlation	397	-0.547**	Yes
6	4-square step test	Higher scores indicate poorer physical function	Positive strong correlation	200	0.728**	Yes

Note: All correlations tested were using baseline data from all four trials whenever the data was available.
rho, Spearman Correlation Coefficient for non-normally distributed Stair Climb Test data; n, sample;
*Correlation is significant at the 0.05 level (2-tailed); ** Correlation is significant at the 0.01 level (2-tailed).

Table 6. Responsiveness hypotheses and results

Approach	Hypothesis	N	Results	Was hypothesis met?	Summary of hypothesis testing results
Approach 1: Construct approach using correlations of the SCT change score with change scores of comparator measures	Positive moderate or strong correlation with WOMAC physical function score	342	r = 0.321**	Yes	Hypothesis not met (only 4 of 6 hypotheses (67%) were met)
	Negative moderate or strong correlation with Quadriceps muscle strength test	332	r = -0.251**	No	
	Negative strong correlation with 30 second sit-to-stand test	169	r = -0.507**	Yes	
	Negative moderate or strong correlation with 40m fast-paced walking velocity test	89	r = -0.649**	Yes	
	Negative strong correlation with Step test	341	r = -0.370**	No	
	Positive strong correlation with 4-square step test	166	r = 0.533**	Yes	
Approach 2: Using ROC curve method	AUC at least 0.70 based on criterion approach	108	**Primary analysis** (using stringent criteria): AUC = 0.628 (95% CI: 0.523, 0.734)	No	Hypothesis not met
		239	**Sensitivity analysis** (using relaxed criteria): AUC = 0.587 (95% CI: 0.498, 0.675)		
Approach 3: Using Effect sizes (ES) †	ES of the SCT would lie ± 0.10 of the range of ES of the 6 comparator measures of physical function	89 to 368	ES of the SCT was 0.20 (range of ESs of other 6 tests 0.20 to 0.53).	Yes	Hypothesis met
Approach 4: Using SRM approach †	SRM of the SCT would lie ± 0.10 of the range of SRM of other 6 measures of physical function	89 to 368	SRM of the SCT was 0.35 (range of ESs of other 6 tests 0.31 to 0.59)	Yes	Hypothesis met
Approach 5. Difference between the stable and improved groups and the SCT mean changes for different GRoC responses	**Approach 5a**. Mean difference between stable and improved groups would be greater than SDC values.	108	**Stringent criteria**: MD= -1.00 (95%CI: -1.69, -0.32)	No	Hypothesis not met
		239	**Relaxed criteria**: MD= -0.83 (95%: -1.49, -0.16)		
	Approach 5b. The SCT mean change scores would be more for greater improvement reflected in the GRoC responses.	239	The SCT mean change scores increased monotonically from slightly worse to much better (**see Online Figure 6**).	Yes	
Summary results for responsiveness					**Only 2 of 5 hypotheses (40%) met**

Note: Baseline and follow-up data from Bennell 2014a [13], Bennell 2014b [7], and Bennell 2010 [12] were used. Lim 2008 [14] data were excluded for approach 2 because they did not include a GRoC scale.
Abbreviations: CI: Confidence Interval; WOMAC, Western Ontario & Macmaster osteoarthritis index; ES, Effect sizes; r, Correlation Coefficient (Pearson or Spearman); n, sample; ROC, Receiver Operating Characteristic; AUC, Area Under the Curve; SRM, Standardised Response Mean; MD, mean difference; SDC, Smallest Detectable Change.
Correlations (absolute value) between 0.30 and 0.49 were considered 'moderate', and 0.50 and higher 'strong'.
** Correlation is significant at the 0.01 level (2-tailed).
† Full results presented in Supplementary Table 2.