The Spine Journal 19 (2019) 1378–1396

Systematic Review

Measurement properties of walking outcome measures for neurogenic claudication: a systematic review and meta analysis

David B. Anderson, MSc[a,*], Stephanie Mathieson, PhD[b], Jillian Eyles, PhD[a], Christopher G. Maher, PhD[b], James M. Van Gelder, MBBS, MD[c,g], Christy C. Tomkins-Lane, PhD[d], Carlo Ammendolia, PhD[e], Vida Bella, MD[f], Manuela L. Ferreira, PhD[a]

[a] *Institute of Bone and Joint Research, the Kolling Institute, Faculty of Medicine and Health, The University of Sydney, Australia*
[b] *Institute for Musculoskeletal Health, Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Australia*
[c] *Department of Neurosurgery, Liverpool Hospital, University of New South Wales, Sydney, New South Wales, Australia*
[d] *Mount Royal University, Department of Health and Physical Education, Calgary, Canada*
[e] *University of Toronto, Faculty of Medicine, Toronto, Canada*
[f] *Westmead Hospital, Sydney, New South Wales, Australia*
[g] *Sydney Spine Institute, Burwood, New South Wales, Australia*

Received 1 January 2019; revised 5 April 2019; accepted 5 April 2019

Abstract

BACKGROUND CONTEXT: Selecting a walking outcome measure for neurogenic claudication requires knowledge of its measurement properties.

PURPOSE: To systematically review and appraise the literature on the measurement properties of walking outcome measures for patients with neurogenic claudication.

STUDY DESIGN: A systematic review and meta-analysis.

METHODS: A systematic search was conducted on the following seven databases: PubMed, PsychINFO, Web of Science, Embase, CINAHL, MEDLINE, and Cochrane Central Register of Controlled Trials. Clinical studies that assessed a measurement property of a walking outcome measure for patients with neurogenic claudication were selected. The methodological quality of studies was assessed using the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist. Measurement property results were assessed using the adapted criteria from Terwee et al. (2007).

RESULTS: Twelve studies that evaluated 15 separate walking outcome measures were included. Out of the 12 studies included, half had poor methodological quality. Four measures had acceptable test-retest reliability: the self-paced walking test (intraclass correlation coefficient, or ICC was 0.98, 95% CI: 0.95–0.99), Physical Function Scale (PFS) (pooled analysis ICC = 0.79, 95% CI: 0.77–0.89), PFS walk item (ICC = 0.81, 95% CI: 0.68–0.89), and Oswestry Disability Index (ODI) walk item (ICC = 0.86, 95% CI: 0.76–0.92). Responsiveness was assessed on five walking outcome measures, and three had adequate responsiveness: the ODI walk item (Area under the Curve, or AUC, was 0.76, SD 0.15), Treadmill test (AUC = 0.70), and PFS (AUC = 0.77, SD 0.14). A meta-analysis demonstrated the PFS had adequate test retest reliability (pooled ICC = 0.79, 95% CI: 0.77–0.89) and internal consistency (pooled Cronbach's αlpha (α) = 0.84, 95% CI:

FDA device/drug status: Not applicable.

Author disclosure: ***DBA:*** Nothing to disclose; ***SM:*** Nothing to disclose; ***JE:*** Nothing to disclose; ***CGM:*** Nothing to disclose; ***JG:*** Nothing to disclose; ***CCT:*** Nothing to disclose; ***CA:*** Grant: Canadian Chiropractic Research Foundation-Salary Grant (F); Speaking and/or Teaching Arrangements: Ontario/Canadian/Saskatchawan chiropractic Associations (B); ***VB:*** Nothing to disclose; ***MLF:*** Nothing to disclose.

* Corresponding author. Institute of Bone and Joint Research, Kolling Institute, Royal North Shore Hospital, Pacific Highway, St Leonards, New South Wales, 2065, Australia.

E-mail address: dand3231@uni.sydney.edu.au (D.B. Anderson).

https://doi.org/10.1016/j.spinee.2019.04.004

0.81−0.86), but not criterion validity (pooled correlation coefficient = -0.59, 95% CI: -0.71, -0.45). Measures that recorded adequate criterion validity were the ODI walk item (pooled correlation coefficient = -0.71, 95% CI: -0.80, -0.58), Treadmill test (pooled correlation coefficient = 0.86, 95% CI: 0.78−0.91), and self predicted walking item (pooled correlation coefficient = 0.74, 95% CI: 0.63−0.82).

CONCLUSIONS: The results of our systematic review demonstrated that high-quality studies that asses the measurement properties of walking outcome measures for patients with neurogenic claudication are lacking. There was only limited evidence available for each walking measure, which prevented any single outcome from being confirmed as the gold standard measure of neurogenic claudication. Clinicians and researchers are recommended to use the self-paced walking test and ODI walk item until further evidence is available. Future research should focus on producing high-quality studies with excellent methodology and larger sample sizes.

Keywords: Lumbar vertebrae; Spinal stenosis; Walking test; Psychometric; Self-report questionnaire; Walking distance

Introduction

Neurogenic claudication is a leading cause of pain and disability in people over the age of 65 [1,2]. This condition is attributed to age-related degenerative changes of the spine which lead to spinal canal narrowing, compression, and diminished blood flow to the nerve roots [3,4]. People with neurogenic claudication typically present with gluteal and/or leg, pain, numbness, weakness or fatigue with walking or standing, that is alleviated with sitting and/or lumbar flexion [1]. Reduced walking ability is considered the most important functional impairment by people with neurogenic claudication [5] and negatively impacts activities of daily living (ADL), work capacity, and quality of life [6].

For people with neurogenic claudication, treatments aim to increase walking ability via reducing pain, disability, and leg symptoms [6]. Current treatment options for neurogenic claudication include nonoperative therapies, such as exercise (eg, lumbar flexion exercises, cycling), manual therapy, epidural injections, nonsteroidal anti-inflammatories, gabapentinoids, opioid analgesic medicines, and surgical treatments such as decompressive spinal surgery with or without lumbar fusion [7−9]. These treatment options differ greatly in terms of costs and associated risk of adverse events [4,10]. For example, the average total cost of nonoperative treatment ranges from USD 2,105 to USD 13,000 [11,12] and present with a low risk of adverse events [11]. In contrast, the average costs of surgery range from USD 12,700 to USD 25,000 [11,12] with a 5.4% to 24% rate of an adverse event [11,13]. Surgery is also associated with a greater risk of serious adverse events (eg, death) than with nonoperative care [14]. Further, there is insufficient evidence that surgical treatments are more efficacious in reducing pain and increasing walking compared with the less expensive and safer nonoperative treatments [4]. For instance, two Cochrane systematic reviews [13,8] that compared nonoperative care to surgical care concluded there was low-quality evidence that surgery resulted in an improvement over nonoperative care for function at 12 months (mean difference on a (+/-) 100 point scale = -6.18, 95% CI: -15.03 to 2.66) and 24 months (mean difference on a (+/-) 100 point scale = -4.43, 95% CI: -7.91 to 0.96). One reason for the inconclusive results may be related to the choice of outcome measures used to assess participants with neurogenic claudication.

When assessing neurogenic claudication, outcome measures need to not only assess the relevant constructs of the condition (ie, walking, leg symptoms), but to possess adequate measurement properties. The three key measurement properties are validity, reliability, and responsiveness [16]. Validity indicates how well an outcome measure is at measuring the construct it is purported to measure [2,17]. Reliability reflects the ability of the measure to provide similar scores under consistent conditions [2,17]. Responsiveness measures the ability of the outcome to detect true changes over time [2,17]. If an outcome measure does not possess these three key measurement properties, clinicians and researchers may be unable to rely upon its findings. Therefore, this study will provide the first known systematic review of the measurement properties of walking outcome measures for neurogenic claudication, as well as the levels of evidence and study quality behind each measure.

Methods

The protocol for this review was devised in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement guidelines [19] and was published on the 5th Decemeber 2017 (CRD42017077326).

Eligibility

Studies were included in the systematic review if they met the following criteria with regard to participants, outcome measures, study type, language, and setting.

Study types: Quantitative studies including randomized controlled studies and observational studies (eg, cohort,

cross-sectional studies). Only full text studies published as original articles were included.

Participants: Studies of adults over 45 years diagnosed with neurogenic claudication of any duration. Neurogenic claudication was defined as gluteal or leg, numbness, weakness, fatigue, or pain, which develops with walking or standing, but is alleviated with sitting or lumbar flexion positions [1]. Studies were excluded if the participants presented with malignancy, tumors, or vascular claudication.

Outcome measures: Studies that assessed any measurement properties of a walking outcome measure for patients with neurogenic claudication. Outcome measures were self-reported or objective measures.

Language: There was no restriction on publication date or language.

Setting: Any setting (eg, outpatient clinic).

A manual search for eligible papers was also conducted. The first selection of papers was performed by two author based on titles and abstracts, then two independent authors assessed the full text of articles. Consensus was used to resolve any disagreement. When there were several articles reporting outcomes from a single study, that is, with the same authors, location, patient population, and recruitment dates, we only included the original article, unless additional measures/data was contained in the subsequent articles.

Search strategy

The search strategy was developed and informed by those used in previous systematic reviews of measurement properties [20,21], and from recommendations from COSMIN's guidelines for systematic reviews [22]. An electronic database search was performed on the following databases: PubMed (via NLM database) Medline (via OvidSP), EMBASE (via OvidSP), CINAHL (via Ebsco), PsychINFO (via CSA), Web of Science (via Thomson Reuters), and Cochrane Central Register of Controlled Trials (via CENTRAL) from inception to October 11, 2017. An updated search was completed on the October 16, 2018. A search was completed by grouping search terms of our target population, walking outcomes measures and measurement properties. Details of the search strategy for PubMed and Medline are in Appendix 1, with the strategy adapted for the search in the other databases.

Data extraction

Using a standardized, piloted data extraction form, the following information was extracted from each paper: bibliometric (authors, title, year of publication, language, funding sources); study characteristics (country, sample size, study design type, participant recruitment source, intervention details [type, number of groups], follow-up timepoints); participants (age, gender, symptom duration, eligibility criteria (inclusion and exclusion), diagnostic criteria, comorbidities [diabetes, knee osteoarthritis]); outcomes (measurement properties [type, evaluation criteria and results]); and data completeness (percentage of missing data, how missing data were handled). Measurement properties were defined as per Terwee et al. [2,17], and are outlined in Appendix 2.

Data were independently extracted from the included studies by two authors (DA and one of SM, JE, CM, CL, CA, JVG). Disagreements were resolved by discussion first, then arbitration by an independent third review author (MF).

Evaluation of the measurement property result

Each measurement property was assessed using adapted criteria from Terwee et al. [17,20] as detailed in Appendix 3. For each measurement property, a score of either positive, negative, or indeterminate was awarded by two separate authors during data extraction, based upon the criteria. A third author resolved any disagreements that occurred between the two separate authors.

Evaluation of the methodological quality of the studies

The methodological quality of each included study was assessed using the COSMIN tool [2]. The tool consists of a section for each measurement property, and a checklist of items to be reviewed. The answer for each question was graded on a four point scale as either: poor, fair, good, or excellent. The overall score for each measurement property was determined by the lowest score awarded for that measurement property, as per COSMIN's protocol [2]. Given that the COSMIN criteria were designed for questionnaires, and not functional measures, a second COSMIN score was reported on, that removed the sample size criteria. This was consistent with the methodology used in a previous systematic review by Dobson et al. [20], who consulted with the COSMIN authors as part of their design. Methodological quality was evaluated by two authors (DA and one of SM, JE, CM, CL, CA, JVG). Any disagreements were resolved by an independent third review author (MF) if an agreement between two authors was not obtained.

Handling of missing data

In the presence of missing data, study authors were contacted to provide further information on participant data, study methods, or results. If the study included data of patients with mixed disease populations and the data from participants with neurogenic claudication were not reported independently, the authors were contacted to provide relevant data on these patients.

Synthesis of results

For all outcome measures where multiple studies reported comparable data, a pooled analysis was completed (Table 5). When a single study reported more than one result for a measurement property (eg, two reliability scores for Self Paced Walking Test [SPWT]), all results were

pooled, and one overall mean result was reported. For construct validity, studies could only be pooled if the study used a comparable criterion measure.

A synthesis of overall levels of evidence was conducted for each walking outcome measure. The overall levels of evidence combined (1) the methodological quality of the study, using the COSMIN tool and (2) the evaluation of the measurement property result (Appendix 3), and (3) the criteria proposed by Terwee et al. (Appendix 4).

Results

Twelve studies [23–33] were found to meet the criteria for inclusion (Table 1). The flowchart of included studies is outlined in the Figure. In total, 15 separate walking outcome measures were reported on, with four being objective measures (SPWT, Shuttle Walking Test (SWT), Treadmill test and Activity monitor). Of the 11 subjective outcome measures, the Physical Function Scale (PFS) was the most assessed, having been assessed in eight of the 12 studies. Full details of the PFS and all other included walking outcome measures can be found in Table 2. Data were found on the following measurement properties: construct validity, criterion validity, reliability, internal consistency, floor and ceiling effects, and responsiveness. No walking outcome measures reported on content validity or measurement error.

A total of 853 participants were included in this review, with a mean age that ranged from 65.8 to 73.8 years of age. In six of the studies, there was a higher percentage of women than men. Neurogenic claudication was diagnosed via both clinical symptoms of neurogenic claudication and confirmation on diagnostic imaging (eg, magnetic resonance imaging) in nine of the 11 studies. In two of the studies, subjects were only required to have clinical symptoms suggestive of neurogenic claudication to be included. Full details of each study is included in Table 1.

Internal consistency

Two outcome measures were assessed for internal consistency seen in Table 3. The PFS was assessed in the studies by [25,27,29,31,32], and found to be adequate with a correlation that ranged from 0.80 to 0.89 and a pooled correlation of 0.84 (95% CI: 0.81–0.86; p = < .01). The internal consistency of the Oxford Claudication Scale Physical Function Domain (OCS-pfD) was also assessed [31]. There was poor evidence that the OCS-pfD's internal consistency ranged from inadequate (0.68) to adequate (0.74).

Test-retest reliability

Seven outcome measures were assessed for test-retest reliability [24,25,28,31,33] seen in Table 3. Adequate test-retest reliability of poor to fair evidence was found for the SPWT (ICC = 0.98), SWT (ICC = 0.92), Treadmill test (CCC = 0.96), Oswestry Disability Index (ODI) walk item (ICC = 0.86), PFS walk item (ICC = 0.81), and PFS (pooled analysis 0.79, 95% CI: 0.69–0.86; p < .01). Only the self-predicted walking item had fair evidence of inadequate test-retest reliability (ICC = 0.65).

Responsiveness

A total of five outcome measures were assessed for responsiveness [26,33,34] seen in Table 4. There was poor to fair evidence of the PFS's responsiveness (Guyatt Responsiveness Index or GRI range was 1.45–2.74). The evidence on responsiveness of the PFS walk item (GRI 1.04 84% CI: 0.36–1.73), and SPWT (GRI 0.96 84% CI: 0.44–1.48) was also poor. Responsiveness using the area under the curve was adequate for the ODI walk item (0.76) Treadmill test (0.70) and PFS (0.77), and inadequate for the SPWT (0.56) and PFS walk item (0.64) with evidence ranging from poor to fair.

Criterion validity

Eleven outcome measures were assessed against the SPWT for criterion validity [23–25,34] (Table 5), with six measures included in the meta-analysis. Three of the measures recorded adequate criterion validity during the meta-analysis, the Treadmill test (pooled analysis 0.86, 95% CI: 0.78–0.91; p < .01), ODI walk item (pooled analysis -0.71, 95% CI:-0.80, -0.58; p < .01) and self-predicted walk item (pooled analysis 0.74, 95% CI: 0.63–0.82; p < .01). The three other measures included in the meta-analysis, the PFS (pooled analysis -0.59, 95% CI: -0.70, -0.45; p < .01), PFS walk item (pooled analysis -0.67, 95% CI: -0.79, -0.50; p < .01) and walking capacity item (pooled analysis 0.66, 95% CI: 0.48–0.78; p < .01) recorded inadequate criterion validity.

Construct validity

Six outcome measures were assessed for construct validity [27,30,34] (Table 5). In the two studies [30,34] that used the Treadmill test as the criterion measure, fair evidence of acceptable construct validity was present for the ODI walk item (-0.54) and self-predicted walk item (0.72). There was good evidence for acceptable construct validity in the ODI walk item (0.80), OCS speed item (0.60), and HUI3 ambulation item (-0.62). Only the PFS was included in the meta-analysis, demonstrating an inadequate construct validity (pooled analysis -0.37, 95% CI: -0.47, -0.27; p < .01).

Floor and ceiling effects

Three outcome measures were assessed for floor and ceiling effects [24,25,28,32] seen in Table 4. The SPWT was found to have an unacceptably high ceiling effect [24,25,34] which ranged from 22.4% to 24.2%. The Treadmill test was also found to have a ceiling effect [24,28,34] which ranged from an acceptable 11.1% to an unacceptable 26.6%. The PFS possessed an acceptable result with no floor and ceiling effects present [32].

Table 1
Articles included in the review with key criteria

Author	Age in y	Males (%)	Properties	Inclusion criteria	Exclusion criteria	Symptom duration
Comer et al. 2011 [29]	Mean 72.0 (range 55-89)	43.2	Internal consistency	Patients referred for surgical opinion for NC symptoms	Not meeting inclusion criteria	Not reported
Conway et al. 2011 [23]	Mean 66.3 (+/-9.8)	75	Construct validity	Age >45 y; clinical diagnosis of LSS (NC symptoms and MRI confirming anatomical stenosis) by either a surgeon or physiatrist	Disc herniation and/or any comorbidities that could limit walking capacity (eg, knee OA)	Mean duration of 4.7 y (SD 1.5)
Deen et al. 2000 [28]	Mean 73.8 (range 57-91)	60.7	Test re-test reproducibility	Intractable NC; severe LSS confirmed on either MRI or CT AND myelography	PVD	Not reported
Markman et al. 2015 [30]	Median 69.0 (IQR 60-76)	55	Construct validity	NC confirmed via treadmill test	Acute musculoskeletal injury unrelated to LSS; comorbid medical problem that affected their walking capacity; SOB with walking and PVD	>3 mo as per inclusion criteria; no mean calculated
Pratt et al. 2002 [31]	Mean 69.0 (range 49-82)	58.6	Internal consistency Test re-test reliability	Clinical history of NC and confirmation on MRI	Patients with comorbid states that limited walking distance;	Not reported
Tomkins et al. 2007 [27]	Mean 69.5 (SD: 11.27)	48.7	Construct validity Internal consistency	Age >50; clinical presentation of NC and referred to one of four adult imaging centers in Canada; confirmed LSS on imaging	Not meeting inclusion criteria	Mean leg pain duration of 7.46 (SD 8.97)
Tomkins et al. 2009 [24]	Mean 66.9 (+/- 9.6)	42.2	Criterion validity	>45 y; Referred to 1 of the 5 study surgeons with suspected LSS; dominant central canal stenosis confirmed on imaging (MRI/CT) by specialist spine surgeon	Dominant lateral or foraminal stenosis confirmed by the study surgeon; Comorbid condition that would limit walking capacity of make exercise medically inadvisable as judged by patients physician; surgery for LSS in the past y	Mean leg pain duration of 7 +/- 9 y
Tomkins-Lane et al. 2010 [25]	Mean 65.8 (+/- 10.0)	51	Criterion Validity Reproducibility (Reliability) Internal consistency	>45 y; Referred to 1 of the 5 study surgeons with suspected LSS; dominant central canal stenosis confirmed on imaging (MRI/CT) by specialist spine surgeon	Dominant lateral or foraminal stenosis confirmed by the study surgeon; Comorbid condition that would limit walking capacity of make exercise medically inadvisable as judged by patients physician; surgery for LSS in the past y	6 +/- 9 y
Tomkins-Lane et al. 2014 [26]	Mean 68.5 (SD: 9.2)	34.6	Responsiveness	>45 y; Referred to 1 of the 5 study surgeons with suspected LSS; dominant central canal stenosis confirmed on imaging (MRI/CT) by specialist spine surgeon	Surgery for LSS within the last 12 mo or any comorbid condition that would limit walking capacity or make a SPWT medically inadvisable, as judged by the subjects' physician (eg, PVD)	Mean leg pain duration of 9.0 y (SD 11.1)
Rainville et al. 2012 [34]	Mean 68.0 (SD 7.9)	58.0	Construct validity Criterion validity Responsiveness	Degenerative LSS documented by lumbar MRI or CT; walking-induced NC with or without concurrent neurologic symptoms of weakness, sensory loss, or impaired balance; self-reported walking ability limited by NC to	symptoms of buttock and leg pain not aggravated by walking; concurrent acute disc herniation as the probable cause of symptoms; spinal stenosis caused by non-degenerative spinal disorders (neoplasm, metabolic bone	Mean NC duration of 18 mo (SD 21)

Table 1 (***Continued***)

Author	Age in y	Males (%)	Properties	Inclusion criteria	Exclusion criteria	Symptom duration
				30 minutes or less; duration of NC of at least 3 mo; and scheduled to undergo physical therapy, spinal injections, or surgery as treatment for NC.	disease, and vertebral fracture); non-palpable dorsalis pedis and posterior tibial pulses in the symptomatic leg(s), suggesting possible peripheral vascular insufficiency; symptomatic arthritis of the hip, knee, ankle, or foot causing limitation in walking; neurologic disease affecting ambulation (Parkinson disease, myelopathy, and stroke); cardiac or pulmonary disease that limits walking; severe cognitive difficulties; and general frailty that would make participating in the walking tests unsafe.	
Stucki et al. 1995 [32]	Median 68.0 (SD 9.1)	35.4	Responsiveness	>50 y; pain in low back, buttock, and lower extremity exacerbated by lumbar extension and evidence of central or central-lateral stenosis by degenerative lesion of the facet joint, disk, or ligamentum flavum on CT, MRI or myelography	Unable to complete questionnaires because of cognitive or language difficulties	Not reported
Stucki et al. 1996 [33]	Median 68.0 (SD 9.1)	33.7	Internal consistency Convergent validity Reproducibility (Reliability) Responsiveness	>50 y; pain in low back, buttock, and lower extremity exacerbated by lumbar extension and evidence of central or central-lateral stenosis by degenerative lesion of the facet joint, disk, or ligamentum flavum on CT, MRI or myelography	Unable to complete questionnaires because of cognitive or language difficulties	Not reported

CT = computed tomography; IQR = interquartile range; LSS = Lumbar spinal stenosis; MRI = Magnetic Resonance Imaging; NC = Neurogenic claudication; OA = Osteoarthritis; PVD = Peripheral Vascular disease; SD = standard deviation; SOB = Shortness of Breath.

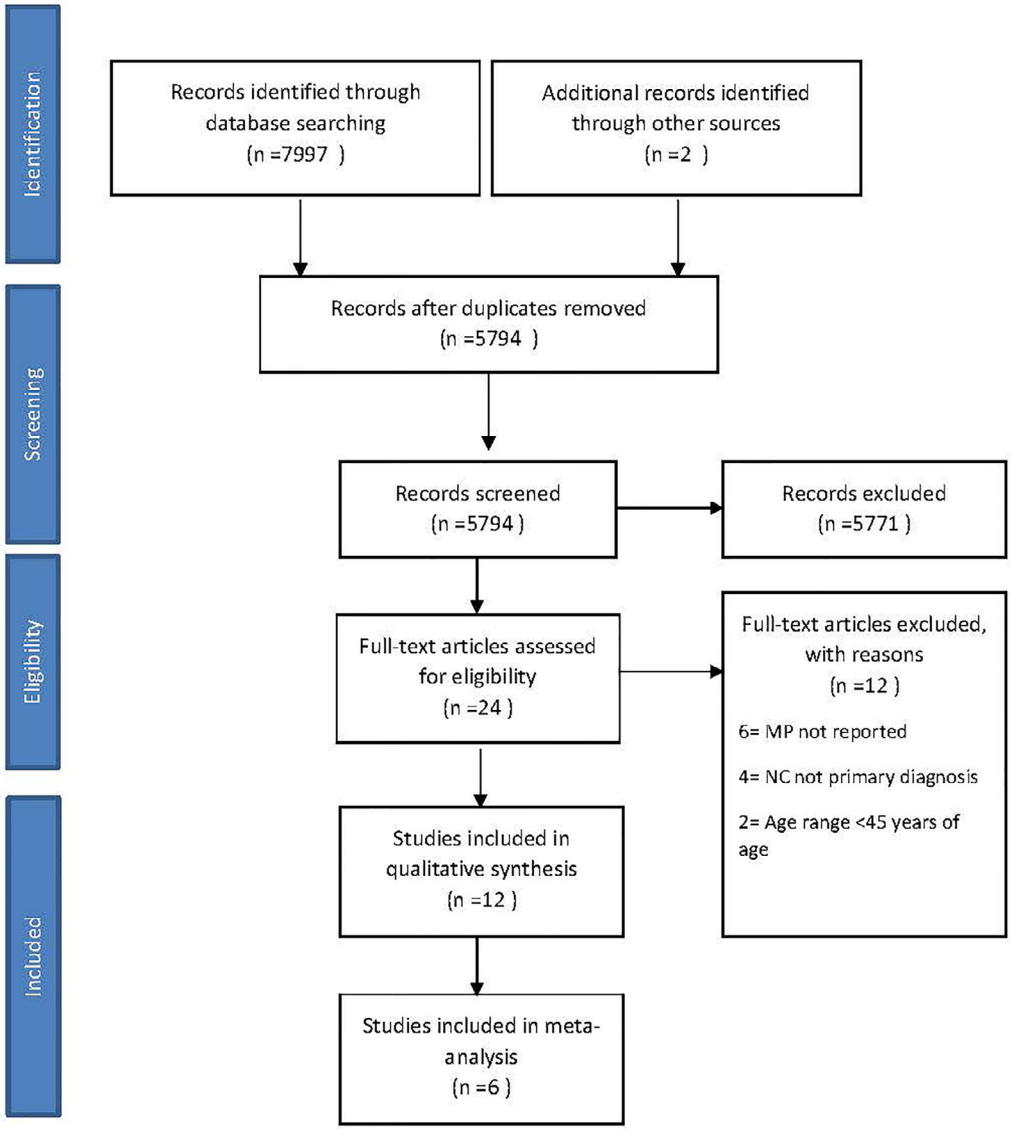


Fig. Flow chart outlining the process in this systematic review. Based upon the PRISMA preferred reporting for systematic reviews [15].

Methodological evaluation of measurement properties

Using the COSMIN criteria, only one study [33] was found to have "excellent" methodological quality for a measurement property. When the modified COSMIN criteria was used (where the sample size is removed), six studies [23–25,31,33,34] were found to have "excellent" methodological quality. Table 7 lists all the measurement properties with both the original and modified COSMIN criteria, where the sample size item is removed.

Best evidence synthesis: overall levels of evidence

The overall levels of evidence for each of the 15 walking outcome measures is summarized in Table 8. The overall levels of evidence were a combination of (1) the methodological quality of the study, using the COSMIN tool and (2) the evaluation of the measurement property result (Appendix 3) and (3) the criteria proposed by Terwee et al. (Appendix 4).

Discussion

To the best of our knowledge, this is the first systematic review of the literature on the measurement properties of walking outcome measures for subjects with neurogenic claudication. Twelve studies were included containing 15 different walking outcome measures. Tests with adequate construct validity were the ODI walk item, self-predicted walking distance item, HUI3 ambulation item, and OCS speed item. Adequate criterion validity was present for the Treadmill test and HUI3 ambulation item. Tests with adequate test-retest reliability were the SPWT, ODI walk item, PFS, and PFS walk item. Only the Treadmill test had adequate responsiveness.

Table 2
Details of walking outcome measures for neurogenic claudication

Test	Specifications of the test	Strengths	Limitations
Objective tests			
Self-paced walking test	The SPWT was conducted by asking participants to walk continuously at their own pace around a flat indoor track until they had to stop, caused by neurogenic claudication, or until the time limit of 30 minutes had been reached. The test was terminated if the participant stopped for more than 3 s. An assessor followed behind the participant with a distance measurement device and a stopwatch. At the termination of the test, the total distance was recorded by the assessor.	The objective measurement of walking distance and a simple protocol	The time demand on participants and assessors
Treadmill test	The Treadmill test required participants walk continuously for a maximum of 30 minutes, or until they had to stop due to neurogenic claudication. For two [24,28] of the three studies that assessed the Treadmill test, the speed was set at 1.2 miles per hour initially, with participants then able to adjust the speed to their preference. In the other study [32], participants completed a pre-test walk on the treadmill to determine preferred walking speed.	Limited space was required to complete	The cost of equipment
Shuttle walking test	Before the start of the test, participants listened to a prerecorded audio tape, which explained the protocol of the test. Participants were required to walk between two points, 10 meters apart. When the test started, participants were required to complete the 10 meters within the specified time, made clear by an audible beep sound. During the first minute of the test, beeps sound every 20 s, equating to 30 meters. The time between beeps decreased every minute until a maximum of 14 shuttles or 1020 meters had been completed.	Limited space was required to complete	The total distance result was dependent upon walking speed.
Activity monitor	The activity monitor included in this review was an Actigraph accelerometer. Participants wore the monitor inside an elastic band around their waist, at the height of their hip bone. Participants were instructed to remove the monitor when sleeping, showering and swimming.	Ability to capture walking data throughout the day, not at a single time point	Inconvenience to participants of having to wear an item to the participants
Subjective tests			
Physical function scale (PFS)	The PFS is one component of the larger, three-part Swiss Spinal Stenosis Questionnaire. The PFS asked participants five questions related to walking capacity and performance. The questions are listed below: How far have you been able to walk? (PFS walk item) (i) More than 2 miles (ii) More than 2 blocks, but less than 2 miles (iii) More than 50 feet, but less than 2 blocks (iv) Less than 50 feet Have you taken walks outdoors or around the shops for pleasure? (i) Yes, comfortably (ii) Yes, but sometimes with pain (iii) Yes, but always with pain (iv) No Have you been shopping for groceries or other items? (i) Yes, comfortably (ii) Yes, but sometimes with pain (iii) Yes, but always with pain (iv) No	Measured different aspects of walking. Quick and simple administration that could be completed online, over telephone etc.	Unable to identify the participants specific walking distance

Table 2 (***Continued***)

Test	Specifications of the test	Strengths	Limitations
	Have you walked around the different rooms in your house or apartment? (i) Yes, comfortably (ii) Yes, but sometimes with pain (iii) Yes, but always with the pain (iv) No Have you walked from your bedroom to the bathroom? (i) Yes, comfortably (ii) Yes, but sometimes with pain (iii) Yes, but always with pain (iv) No		
ODI walk item	ODI walk item is one of the items that make-up the ten-part ODI. The ODI walk item asked participants to select one of the following answers to their walking: (i) Pain does not prevent me walking any distance (ii) Pain prevents me from walking more than 1 mile (iii) Pain prevents me from walking more than $^{1}/_{2}$ mile (iv) Pain prevents me from walking more than 100 yards (v) I can only walk using a stick or crutches (vi) I am in bed most of the time	Easy to administer, could be completed online, over telephone etc.	Unable to identify the participants specific walking distance
Self-predicted walking capacity	Asked participants to answer the question ""if you were to go for a walk today, how far would you be able to walk before being forced to stop, caused by symptoms of Lumbar Spinal Stenosis (miles/kilometers)"	Simple, specific question of walking capacity, could be completed online, over telephone etc.	Difficult for most participants to know their specific walking capacity
Estimated walking distance	Asked participants to respond to the question "I can walk __ feet without a rest before I MUST stop"	Provided a specific distance for the participants walking capacity	Difficult to interpret at which time point to respond
Oxford claudication scale – physical function domain	The Physical function domain are the three-items within the larger, ten-part Oxford Claudication Scale. The Physical function included the following items: Walking distance On the average, in the past month, when you go for a walk, how far are you able to walk before your back or leg troubles you? (i) More than 2 miles or no limit (ii) More than 1/4 mile, but less than 2 miles (iii) More than 100 yards, but less than 1/4 mile (iv) More than 50 feet, but less than 100 yards (v) Less than 50 feet (vi) Not at all Walking ability On the average, which statement best describes your walking ability over the past month? (i) There is no limit to my walking ability. (ii) I can walk far enough to do everything I want to do. (iii) I am able to walk comfortably from home to the shops or my transport. (iv) I am able to walk comfortably around the house. (v) I am able to walk only from the bedroom to the bathroom or kitchen.	Measured different aspects of walking, simple administration that could be completed online, over telephone etc.	Unable to identify the participants specific walking distance

Table 2 (***Continued***)

Test	Specifications of the test	Strengths	Limitations
	(vi) I am not able to walk at all. Walking speed On the average, which statement best describes your walking over the past month? (i) I am able to walk at a normal speed. (ii) I walk slowly standing upright. (iii) I walk slowly bent forward. (iv) I have to stop and stand still when I walk. (v) I have to stop and sit down when I walk. (vi) I cannot walk at all.		
HUI3 amb	The HUI3 Single Attribute Score for Ambulation asked participants to select which response from the following list represented their walking capacity: (i) Able to walk around their neighbourhood without difficulty, and without walking equipment (ii) Able to walk around the neighbourhood with difficulty; but does not require walking equipment or the help of another person (iii) Able to walk around the neighbourhood with walking equipment; but without the help of another person (iv) Able to walk short distances with walking equipment, and requires a wheelchair to get around the neighborhood. (v) Unable to walk alone, even with walking equipment. Able to walk short distances with the help of another person, and requires a wheelchair to get around the neighborhood (vi) Cannot walk at all	Provided specific information on walking difficulty, which can be used to assess treatment effectiveness, could be completed online, over telephone etc.	Does not measure walking distance
Walking capacity	Asked participants "How would you say your walking capacity is today, on a scale of 0 to 10, with 0 being worst day, and 10 being your best day?"	Provided specific information on participant's walking capacity, which can be used to assess treatment effectiveness, could be completed online, over telephone etc.	Does not measure walking distance
Walking difference	Asked participants "on a 100-mm line, indicate how much difficulty walking you have had during the past week, with 0 denoting no difficulty and 100 denoting maximum difficulty".	Provided specific information on walking difficulty, which can be used to assess treatment effectiveness, could be completed online	Does not measure walking distance
QBPDS walk	Asked participants to respond to the question "How difficult is it for you to perform the task of walking several miles, from 0 (not at all difficult) to 5 (unable to do)"	Provided specific information on walking difficulty, which can be used to assess treatment effectiveness, could be completed online, over telephone etc.	Does not measure walking distance

SPWT, self-paced walking test; PFS, physical function scale; ODI walk item, walk item from the Oswestry Disability Index; HUI3 amb, health utilities index single attribute utility score for ambulation; QBPDS walk, Quebec Back Pain Disability scale walk item.

Table 3
Reliability of the walking outcome measures

	Internal consistency					Reliability (test retest)				
Test	Result (95% CI)	Design	Quality	n	COSMIN	Result (95% CI)	Design	Quality	n	COSMIN
PFS [29]	0.825 0.726‡	PSI	N/A	92	Good					
PFS [31]	0.87 0.89	α α	+ +	29	Poor*	0.82	ICC	+	29	Poor*
PFS [27]	0.88 (0.83–0.92)	α	+	72	Fair					
PFS [33]	0.82	α	+	193	Excellent	0.94	SRCC	+	23	Poor†
PFS [25]	0.80 0.87	α	+	49	Fair	0.77 (0.61–0.86)	ICC	+	49	Fair†
PFS walk item [25]						0.81 (0.68–0.89)	ICC	+	49	Fair†
ODI walk [25]						0.86 (0.76–0.92)	ICC	+	49	Fair†
Self-predicted walk [25]						0.65 (0.46–0.79)	ICC	-	49	Fair†
Treadmill test [28]						0.96	CCC	+	28	Poor
OCS-pfD [31]	0.68 0.74	α	+	29	Poor*					
SWT [31]						0.92	ICC	+	29	Poor*
SPWT [24]						0.98 (0.95–0.99)	ICC	+	33	Fair*

α, Cronbach's αlpha; ICC, intraclass correlation coefficient; AUCs, area under curve; SRCC, Spearman rank correlation coefficient; SRM, standardised response mean; PSI, Person separation index; CCC, concordance correlation coefficient; n, sample size; PFS, physical function scale; PFS walk item, walking item from physical function scale; ODI walk, walking item from Oswestry Disability Index; Self-predicted walk, self-predicted walking distance; OCS-pfD, Oxford claudication scale physical function domain; SWT, shuttle walking test; SPWT, self paced walking test.

* Rating of "Excellent" if using the modified COSMIN tool;

† Rating of "Good" if using the modified COSMIN tool.Rating of each measurement property quality using the COSMIN criteria where "+" is positive, "?" is indeterminate and "-" is negative.

‡ item 11 was removed from the analysis

Table 4
Responsiveness and floor and ceiling effects of walking outcome measures

	Responsiveness					Floor and ceiling effects			
Test	Result (84% CI)	Design	Quality	n	COSMIN	Result	Type	Rating	COSMIN
PFS [32]	1.07	SRM	N/A	130	Fair				
PFS [32]	2.74	Guyatt	N/A	193	Fair	0%	Floor/Ceiling	+	N/A
PFS [26]	1.45 (0.74, 2.16)	Guyatt	N/A	26	Poor***				
PFS [26]	0.77 (SD, 0.14)	AUCs	+	26	Poor***				
PFS walk item [26]	0.64 (SD, 0.17)	AUCs	-	26	Poor***				
PFS walk item [26]	1.04 (0.36, 1.73)	Guyatt	N/A	26	Poor***				
ODI walk [26]	0.76 (SD, 0.15)	AUCs	+	26	Poor***				
ODI walk [26]	1.23 (0.60, 1.86)	Guyatt	N/A	26	Poor***				
Treadmill test [28]						12.5% (1.2mph) 11.1% (preferred speed)	Ceiling	+	N/A
Treadmill test [24]						22%	Ceiling	-	N/A
Treadmill test [34]	0.70†	AUCs	+	39	Fair	26.6%	Ceiling	-	N/A
SPWT [24]						24.2%	Ceiling	-	N/A
SPWT [34]	0.56	AUCs	-	39	Fair	24%	Ceiling	-	N/A
SPWT [26]	0.96 (0.44-1.48)	Guyatt	N/A	26	Poor***				
SPWT [25]						22.4%	Ceiling	-	N/A

AUCs, area under curve; Guyatt, guyatt responsiveness index; SRM, standardised response mean; †, not statistically significant; PFS, physical function scale; PFS walk item, walking item from physical function scale; ODI walk, walking item from Oswestry Disability Index; SPWT, self paced walking test.

Rating of "Excellent" if sample size removed from COSMIN tool.

Rating of "Good" if sample size removed from COSMIN tool.

***Rating of "Fair" if sample size removed from COSMIN tool.

Rating of each measurement property's quality using the COSMIN criteria where "+" is positive, "?" is indeterminate and "-" is negative.

Table 5
Validity of walking outcome measure for neurogenic claudication

	Validity										
	Criterion					Construct					
Test	Result	Design	Quality	n	COSMIN‡	Result	Design	Quality	Correlation with	n	COSMIN
Treadmill test [24]	0.88	Pearson	+	33	Fair						
Treadmill test [34]	0.84	ICC	+	45	Fair						
ODI walk [27]						0.797*	Spearman	+	PFS	72	Good
ODI walk [25]	-0.83	Spearman	+	49	Fair						
ODI walk [34]	-0.49	Pearson	-	45	Fair						
ODI walk [34]						-0.54*		+	Treadmill Test	45	Fair
PFS [25]	-0.62	Spearman	-	49	Fair						
PFS [34]	-0.55	Pearson	-	45	Fair						
PFS [27]						0.797*		+	ODI walk	72	Good
PFS [34]						-0.45*		-	Treadmill Test	45	Fair
PFS [30]						-0.361†		-	Treadmill Test	269	Poor
PFS [23]	-0.610	Spearman	-	12	Poor						
PFS walk item [25]	-0.66	Spearman	-	49	Fair						
PFS walk item [23]	-0.715	Spearman	+	12	Poor						
PFS walk item [30]						-0.443†		-	Treadmill Test	269	Poor
Self-predicted walk [25]	0.80	Spearman	+	49	Fair						
Self-predicted walk [34]	0.65	Pearson	-	45	Fair						
Self-predicted walk [34]						0.72*		+	Treadmill Test	45	Fair
HUI3 Amb [25]	-0.71	Spearman	+	49	Fair						
HUI3 Amb [27]						-0.616*	Spearman	+	PFS	72	Good
OCS speed [27]						0.601*	Spearman	+	PFS	72	Good
Walk cap [25]	0.65	Spearman	-	49	Fair						
Walk cap [23]	0.682	Spearman	-	12	Poor						
Estimated walking [23]	0.886	Spearman	+	12	Poor						
Walk diff [25]	0.47	Spearman	-	49	Fair						
QBPS walk [23]	-0.755	Spearman	+	12	Poor						
Activity monitor [23]	0.629	Pearson	-	12	Poor						

n, sample size; PFS, physical function scale; PFS walk item, walking item from physical function scale; ODI walk, walking item from Oswestry Disability Index; Self-predicted walk, self-predicted walking distance; OCS-speed, Oxford claudication scale item 10; SPWT, self-paced walking test; QBPS walk, walk item from the Quebec Back Pain Disability Scale; HUI3 AMb, ambulation item from the Health Utilities Index 3; Walk diff, walking difference, which asked "on a 100-mm line, indicate how much difficulty walking you have had during the past week, with 0 denoting no difficulty and 100 denoting maximum difficulty". Walk cap, walking capacity, which asked "How would you say your walking capacity is today, on a scale of 0 to 10, with 0 being worst day, and 10 being your best day?".

* =p value of <0.01.

† =p value of <0.05.

‡ Rating of "Excellent" if sample size removed from COSMIN tool. Rating of each measurement properties quality using the COSMIN criteria where "+" is positive, "?" is indeterminate and "-" is negative.

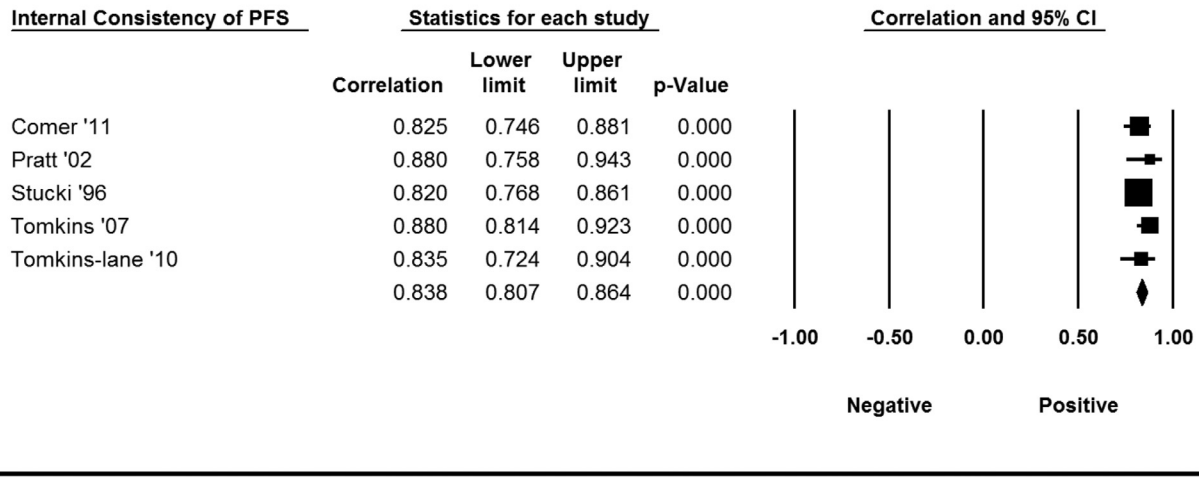
The SPWT was selected as the criterion measure for half of the walking outcome measures, confirming its use among researchers as the gold standard [23,25] measure for walking assessment in neurogenic claudication. This review demonstrates that further high-quality studies are required however, before a single walking outcome measure can be relied upon as the gold standard for the assessment of walking in neurogenic claudication.

Quality of evidence

All included studies and measurement properties were assessed for their methodological quality using the COSMIN tool. Methodological quality ranged from poor to excellent, but limitations were noted when using the COSMIN criteria. For example, as the overall methodological score of each property is based on the "worst score counts," measurement properties were frequently scored "poor" as a result of one question irrespective of all other study methodology considerations. This was particularly notable in the sample size criteria whereby studies with samples sizes below 30 were rated as "poor" for methodological quality. Given that functional outcome measure studies often have lower sample sizes [20,35], seven of the 22 measurement property results were rated as "poor." For the benefit of the reader, we included two COSMIN methodological quality results, one with the strict implementation of the COSMIN method, and the other, in which the sample size item was removed (Tables 3, 4, 5 and 7). The reader can then observe what impact the sample size item had on the overall methodological score.

Table 6
Pooled meta-analysis

Walking Test	Measurement Property	Pooled effect (95% CI)	P value
Physical Function Scale	Internal consistency	0.838 (0.807, 0.864)	p<0.01
Physical Function Scale	Reliability	0.789 (0.685, 0.862)	p<0.01
Physical Function Scale	Construct validity	-0.374 (-0.465, -0.274)	p<0.01
Physical Function Scale	Criterion validity	-0.590 (-0.705, -0.445)	p<0.01
PFS walk item	Criterion validity	-0.670 (-0.791, -0.497)	p<0.01
ODI walk item	Criterion validity	-0.705 (-0.795, -0.584)	p<0.01
Treadmill test	Criterion validity	0.858 (0.784, 0.908)	p<0.01
Self-predicted walk	Criterion validity	0.737 (0.626, 0.819)	p<0.01
Walking capacity	Criterion validity	0.655 (0.478, 0.781)	p<0.01



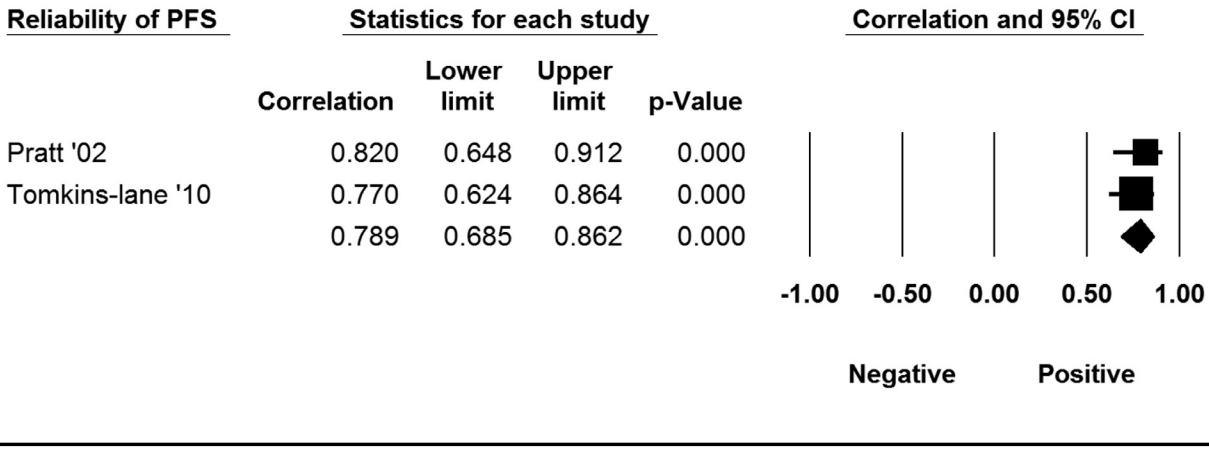
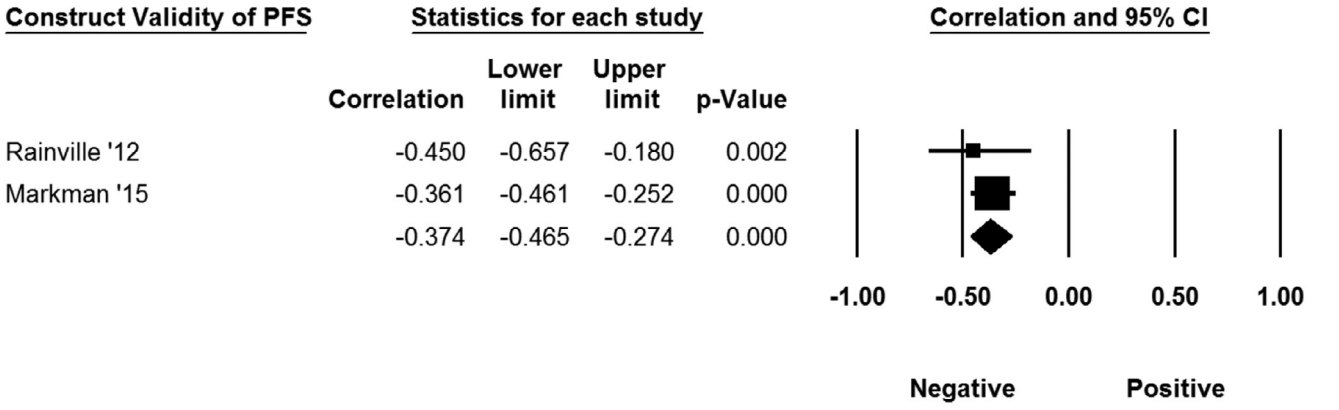
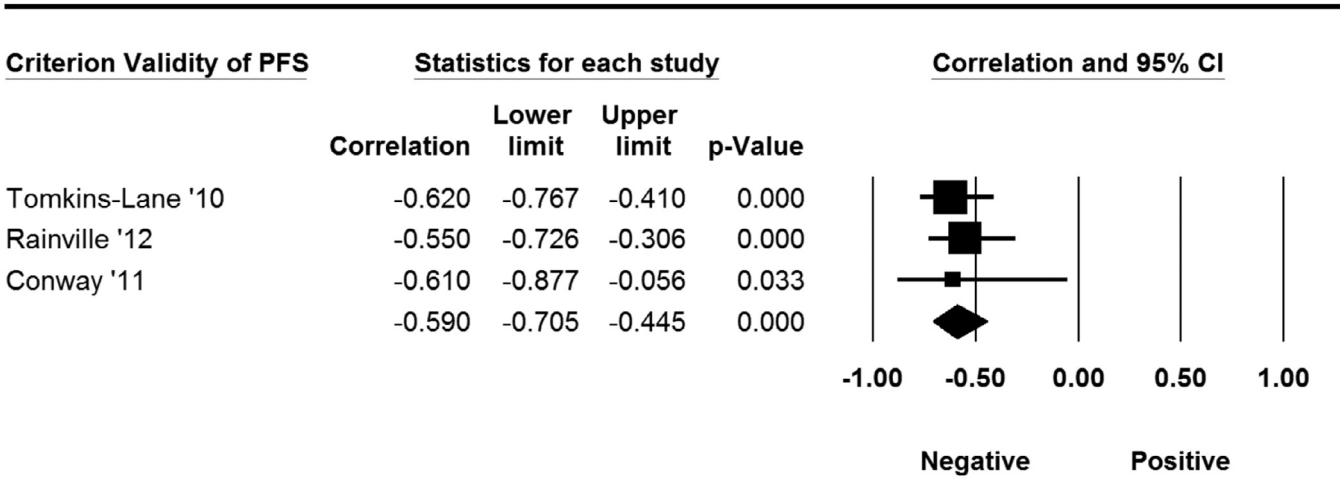


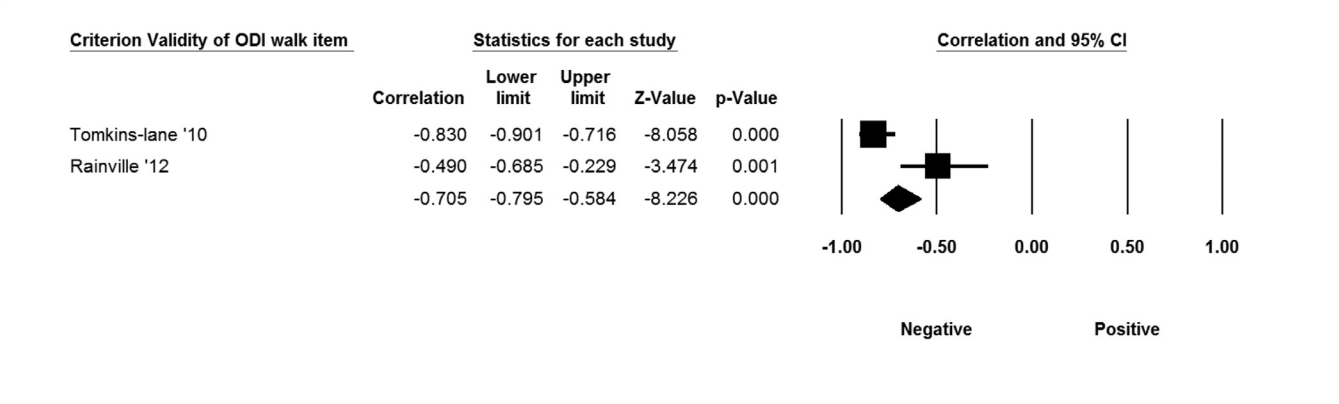
Table 6 (***Continued***)

Walking Test	Measurement Property	Pooled effect (95% CI)	P value





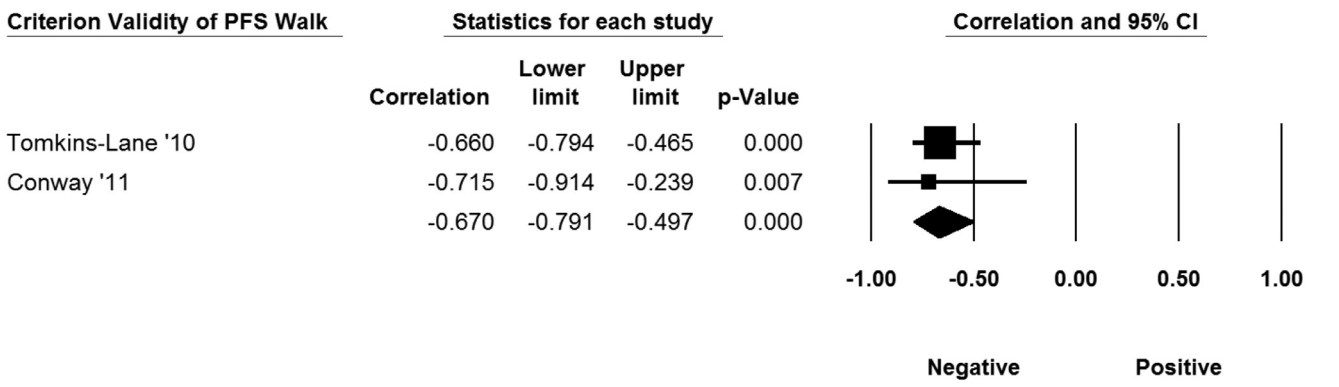
Meta Analysis



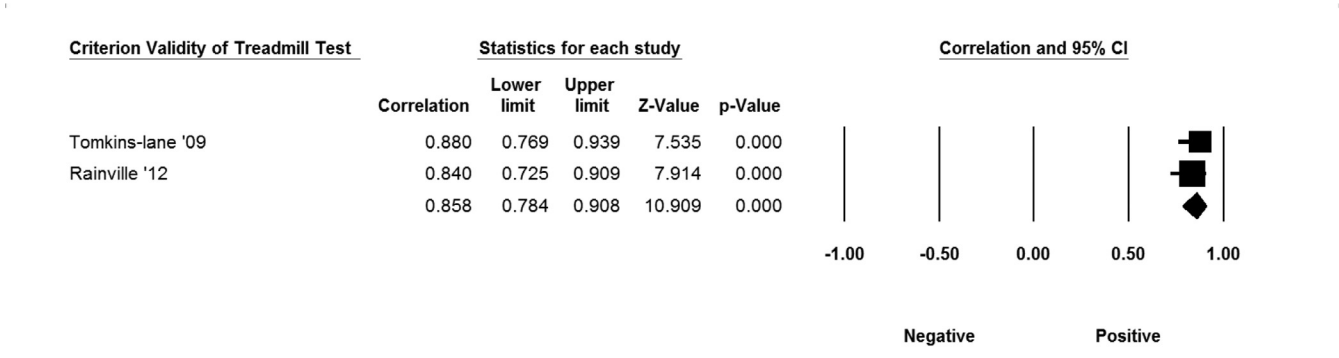
Meta Analysis

Table 6 (***Continued***)

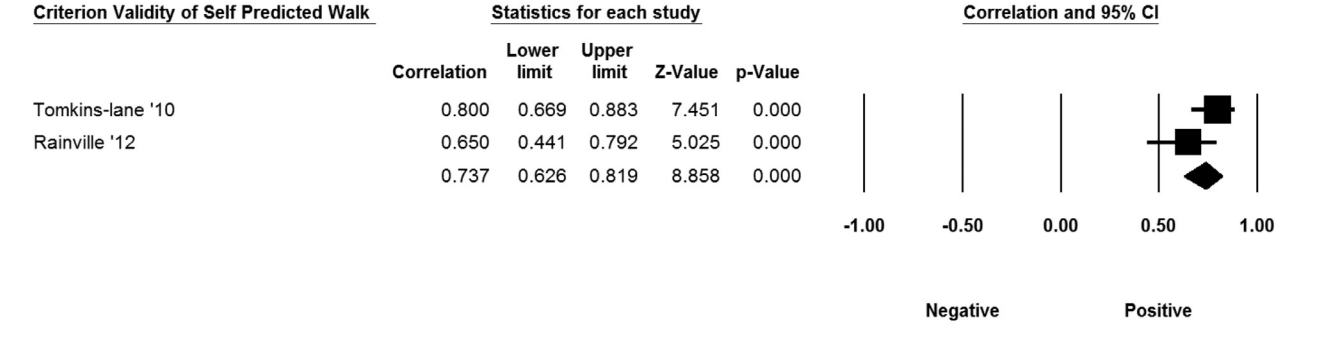
Walking Test	Measurement Property	Pooled effect (95% CI)	P value



Meta Analysis



Meta Analysis



Meta Analysis

Table 6 (***Continued***)

Walking Test	Measurement Property	Pooled effect (95% CI)	P value
Criterion Validity of Walking Capacity	**Statistics for each study**	**Correlation and 95% CI**	
	Correlation **Lower limit** **Upper limit** **Z-Value** **p-Value**		
Tomkins-lane '10	0.650 0.451 0.787 5.258 0.000		
Conway '11	0.682 0.178 0.903 2.499 0.012		
	0.655 0.478 0.781 5.820 0.000		
		-1.00 -0.50 0.00 0.50 1.00	
		Negative Positive	
Meta Analysis			

Clinical implication

Assessment is integral to the evaluation of a patient's condition. All treating clinicians and researchers are therefore encouraged to select the outcome measure which possesses the highest overall validity, reliability, and responsiveness. In the case of neurogenic claudication, the limited evidence available made nomination of any one measure difficult. Although on balance, the SPWT and ODI walk item were nominated as the most appropriate of the current walking measures for neurogenic claudication. The ODI walk item has the additional benefit of being low cost, requiring minimal time, and no equipment. A limitation of the ODI walk item (compared with the SPWT), is that, like the PFS, it is limited in its capacity to quantify walking distance.

Table 7
Articles included in the review with both original and modified COSMIN methodological evaluation scores

	Properties	COSMIN	Modified COSMIN*
Comer et al. 2011 [29]	Internal consistency	Good	Good
Conway et al. 2011 [23]	Criterion validity	Poor	Excellent
Deen et al. 2000 [28]	Test retest reproducibility	Poor	Poor
Markman et al. 2015 [30]	Construct validity	Poor	Poor
Pratt et al. 2002 [31]	Internal consistency	Poor	Poor
	Test retest reliability	Poor	Excellent
Tomkins et al. 2007 [27]	Construct validity	Good	Good
	Internal consistency	Fair	Fair
Tomkins-lane et al. 2009 [24]	Criterion validity	Fair	Excellent
	Reliability	Fair	Excellent
Tomkins-Lane et al. 2010 [25]	Criterion validity	Fair	Excellent
	Reliability	Fair	Good
	Internal consistency	Fair	Excellent †
Tomkins-lane. 2014 [26]	Responsiveness	Poor	Fair
Rainville. 2012 [34]	Construct validity	Fair	Fair
	Criterion validity	Fair	Excellent
	Responsiveness	Fair	Fair
Stucki et al. 1995 [32]	Responsiveness	Fair	Fair
Stucki et al. 1996 [33]	Internal consistency	Excellent	Excellent
	Construct validity	Excellent	Excellent
	Reliability	Poor	Good
	Responsiveness	Fair	Fair

* If sample size criteria is removed from COSMIN.
† Item that assessed "total items within measure" removed for the SPWT.

Table 8
Summary of measurement properties

Test	Criterion	Construct	Internal consistency	Reliability	Responsiveness
Objective tests					
SPWT	O	O	O	+	-
Treadmill test	++	O	O	?	+
SWT	O	O	O	?	O
Activity monitor	O	?	O	O	O
Subjective tests					
ODI walk	±	++	O	+	?
PFS	--	±	+++	+	?
PFS walk item	±	O	O	+	?
Self-predicted walking distance	±	+	O	-	O
Estimated walking distance	O	?	O	O	O
HUI3 Amb	+	++	O	O	O
OCS speed	O	++	O	O	O
OCS-pfd	O	O	?	O	O
Walking capacity	--	O	O	O	O
Walking difference	-	O	O	O	O
QBPS walk	?	O	O	O	O

SPWT, self-paced walking test; SWT, shuttle walk test; PFS, physical function scale; PFS walk item, walking item from physical function scale; ODI walk, walking item from Oswestry Disability Index; Self-predicted walking distance, response to question "if you were to go for a walk today, how far would you be able to walk before being forced to stop due symptoms of LSS (m/km)"; Estimated walking distance, response to the question "I can walk __ feet without a rest before I MUST stop"; OCS-speed, Oxford claudication scale item 10; OCS-pfs, physical function domain from the Oxford Claudication Scale; QBPS walk, walk item from the Quebec Back Pain Disability Scale; HUI3 AMb, ambulation item from the Health Utilities Index 3; Walking capacity, response to the question "How would you say your walking capacity is today, on a scale of 0 to 10, with 0 being worst day, and 10 being your best day?"; Walking difference, response to "on a 100-mm line, indicate how much difficulty walking you have had during the past week, with 0 denoting no difficulty and 100 denoting maximum difficulty."

+++= consistent findings of positive rating in multiple studies of "good" methodological quality OR in one study of "excellent" methodological quality; ++= consistent findings of positive rating in multiple studies of "fair" methodological quality OR in one study of "good" methodological quality; += positive result in one study of "fair" methodological quality; ±= conflicting findings; ?= only studies of "poor" methodological quality; - - -= consistent findings of negative rating in multiple studies of "good" methodological quality OR in one study of "excellent" methodological quality; - -= consistent findings of negative rating in multiple studies of "fair" methodological quality OR in one study of "good" methodological quality; -= negative result in one study of "fair" methodological quality; O= no evidence.

We would encourage clinicians to select both the ODI walk item and the SPWT if adequate space and time is available in their clinics. Participants would need the capacity to continuously walk on a flat, stable surface, without obstacles for up to 30 minutes. The SPWT and ODI walk item should be prescribed on initial assessment, and then repeated over time, particularly after interventions to determine any change in the participants condition. If adequate space or time is not possible, the ODI walk item can be administered individually. Based on the available evidence, clinicians should exercise caution when using the PFS, estimated walking distance questions, or the SWT as walking tests for neurogenic claudication; although this may change with future high-quality studies.

Direction of future research

Future studies are required to increase the amount of evidence on measurement properties of walking outcome measures for neurogenic claudication. Particularly needed are more studies assessing the SPWT and the newer forms of activity monitor such as wearable accelerometers and walking apps. Recently, researchers have begun to use activity monitors and technology to evaluate outcomes of treatment for neurogenic claudication [36,37]. This type of monitoring shows promise in that it allows an objective measurement of change in walking behavior and function. Specifically, it assesses multiple dimensions of function, which include both capacity and performance. It is possible that in the future, accelerometry data could be used as a gold standard measure of walking. Before this, however, additional studies are needed that assess the measurement properties of these accelerometers for participants with neurogenic claudication.

Limitations

There were some limitations present in this review. First, the process of determining the criterion validity of a walking outcome measure relied upon the comparison walking measure (criterion) being considered a gold standard measure of walking. In the case of lumbar spinal stenosis with neurogenic claudication, the SPWT was considered the gold standard but was not validated until 2009 [24]. This

resulted in studies before 2009 using other criterion measures such as the PFS and the Treadmill test, which have been shown to underestimate walking capacity [24]. To address this, only outcome measures compared with the SPWT could be considered for criterion validity, with all other comparators (ie, PFS and Treadmill test) considered for construct validity.

Another limitation of this review were the results of the COSMIN methodological quality evaluation tool. The COSMIN tool was designed for questionnaires, and has a component that assesses the appropriateness of power calculation and sample sizes. The sample size when evaluating functional outcome measures (ie, SPWT or Treadmill test) is frequently lower, given the logistical difficulties of objective measures. As a result, a large number of studies were given a "poor" COSMIN score for the sample size item despite scoring excellent in every other category. This limitation has been observed in previous systematic reviews of the measurement properties of functional outcome measures [eg, 18] where an overall methodological score has been adjusted to address the sample size limitation. We have chosen to instead report both the original COSMIN methodological score (with the sample size item score included) as well as the methodological score with the sample size item removed [20].

Conclusions

The results of our systematic review demonstrated that high-quality studies that assess the measurement properties of walking outcome measures for patients with neurogenic claudication are lacking. There was only limited evidence available for each walking measure, which prevented any single outcome from being supported as the gold standard measure of neurogenic claudication. Clinicians and researchers are recommended to use the SPWT and ODI walk item until further evidence is available. Future research should focus on producing high-quality studies with excellent methodology and larger sample sizes.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/10.1016/j.spinee.2019.04.004.

References

[1] Ammendolia C, Chow N. Clinical outcomes for neurogenic claudication using a multimodal program for lumbar spinal stenosis: a retrospective study. J Manipulative Physiol Ther 2015;38:188–94.

[2] Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651–7.

[3] Hida S, Naito M, Kubo M. Intraoperative measurements of nerve root blood flow during discectomy for lumbar disc herniation. Spine (Phila Pa 1976) 2003;28:85–90.

[4] Ammendolia C, Stuber K, de Bruin LK, Furlan AD, Kennedy CA, Rampersaud YR, et al. Nonoperative treatment of lumbar spinal stenosis with neurogenic claudication: a systematic review. Spine (Phila Pa 1976) 2012;37:E609–16.

[5] Fritz JM, Delitto A, Welch WC, Erhard RE. Lumbar spinal stenosis: a review of current concepts in evaluation, management, and outcome measurements. Arch Phys Med Rehabil 1998;79:700–8.

[6] Ammendolia C, Schneider M, Williams K, Zickmund S, Hamm M, Stuber K, et al. The physical and psychological impact of neurogenic claudication: the patients' perspectives. J Can Chiropr Assoc 2017;61:18–31.

[7] NASS Evidence-Based Clinical Guidelines Committee, Diagnosis and treatment of degenerative lumbar spinal stenosis, in Evidence-based clinical guidelines for multidisciplinary spine care. 2011: Illinois, USA.

[8] Ammendolia C, Stuber KJ, Rok E, Rampersaud R, Kennedy CA, Pennick V, et al. Nonoperative treatment for lumbar spinal stenosis with neurogenic claudication. Cochrane Database Syst Rev 2013:CD010712.

[9] Ammendolia C, Stuber K, Tomkins-Lane C, Schneider M, Rampersaud YR, Furlan AD, et al. What interventions improve walking ability in neurogenic claudication with lumbar spinal stenosis? A systematic review. Eur Spine J 2014;23:1282–301.

[10] Machado GC, Ferreira PH, Yoo RI, Harris IA, Pinheiro MB, Koes BW, et al. Surgical options for lumbar spinal stenosis. Cochrane Database Syst Rev 2016;11:CD012421.

[11] Parker SL, Anderson LH, Nelson T, Patel VV. Cost-effectiveness of three treatment strategies for lumbar spinal stenosis: conservative care, laminectomy, and the superion interspinous spacer. Int J Spine Surg 2015;9:28.

[12] Aichmair A, Burgstaller JM, Schwenkglenks M, Steurer J, Porchet F, Brunner F, et al. Cost-effectiveness of conservative versus surgical treatment strategies of lumbar spinal stenosis in the Swiss setting: analysis of the prospective multicenter Lumbar Stenosis Outcome Study (LSOS). Eur Spine J 2017;26:501–9.

[13] Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus non-surgical treatment for lumbar spinal stenosis. Cochrane Database Syst Rev 2016:CD010264.

[14] Machado GC, Maher CG, Ferreira PH, Harris IA, Deyo RA, McKay D, et al. Trends, complications, and costs for hospital admission and surgery for lumbar spinal stenosis. Spine 2017;42:1737–43.

[15] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.

[16] Turner JA, Ersek M, Herron L, Deyo R. Surgery for lumbar spinal stenosis. Attempted meta-analysis of the literature. Spine (Phila Pa 1976) 1992;17:1–8.

[17] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.

[18] Wessberg P, Frennered K. Central lumbar spinal stenosis: natural history of non-surgical patients. Eur Spine J 2017;26 (10):2536–42.

[19] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009;62 (10):e1–34.

[20] Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthritis Cartilage 2012;20:1548–62.

[21] Kroman SL, Roos EM, Bennell KL, Hinman RS, Dobson F. Measurement properties of performance-based outcome measures to assess physical function in young and middle-aged people known to be at

high risk of hip and/or knee osteoarthritis: a systematic review. Osteoarthritis Cartilage 2014;22:26–39.
[22] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737–45.
[23] Conway J, Tomkins CC, Haig AJ. Walking assessment in people with lumbar spinal stenosis: capacity, performance, and self-report measures. Spine J 2011;11:816–23.
[24] Tomkins CC, Battie MC, Rogers T, Jiang H, Petersen S. A criterion measure of walking capacity in lumbar spinal stenosis and its comparison with a treadmill protocol. Spine (Phila Pa 1976) 2009;34:2444–9.
[25] Tomkins-Lane CC, Battie MC. Validity and reproducibility of self-report measures of walking capacity in lumbar spinal stenosis. Spine (Phila Pa 1976) 2010;35:2097–102.
[26] Tomkins-Lane CC, Battie MC, Macedo LG. Longitudinal construct validity and responsiveness of measures of walking capacity in individuals with lumbar spinal stenosis. Spine J 2014;14:1936–43.
[27] Tomkins CC, Battie MC, Hu R. Construct validity of the physical function scale of the Swiss Spinal Stenosis Questionnaire for the measurement of walking capacity. Spine (Phila Pa 1976) 2007;32:1896–901.
[28] Deen Jr HG, Zimmerman RS, Lyons MK, McPhee MC, Verheijde JL, Lemens SM. Test-retest reproducibility of the exercise treadmill examination in lumbar spinal stenosis. Mayo Clin Proc 2000;75:1002–7.
[29] Comer CM, Conaghan PG, Tennant A. Internal construct validity of the Swiss Spinal Stenosis questionnaire: rasch analysis of a disease-specific outcome measure for lumbar spinal stenosis. Spine (Phila Pa 1976) 2011;36:1969–76.
[30] Markman JD, Gewandter JS, Frazer ME, Pittman C, Cai X, Patel KV, et al. Evaluation of outcome measures for neurogenic claudication: a patient-centered approach. Neurology 2015;85:1250–6.
[31] Pratt RK, Fairbank JC, Virr A. The reliability of the Shuttle Walking Test, the Swiss Spinal Stenosis Questionnaire, the Oxford Spinal Stenosis Score, and the Oswestry Disability Index in the assessment of patients with lumbar spinal stenosis. Spine (Phila Pa 1976) 2002;27:84–91.
[32] Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. J Clin Epidemiol 1995;48:1369–78.
[33] Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossel AH, Katz JN. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. Spine (Phila Pa 1976) 1996;21:796–803.
[34] Rainville J, Childs LA, Pena EB, Suri P, Limke JC, Jouve C, et al. Quantification of walking ability in subjects with neurogenic claudication from lumbar spinal stenosis—a comparative study. Spine J 2012;12:101–9.
[35] Denteneer L, Van Daele U, Truijen S, De Hertogh W, Meirte J, Stassijns G. Reliability of physical functioning tests in patients with low back pain: a systematic review. Spine J 2018;18:190–207.
[36] Smuck M, Tomkins-Lane C, Ith MA, Jarosz R, Kao MJ. Physical performance analysis: a new approach to assessing free-living physical activity in musculoskeletal pain and mobility-limited populations. PLoS One 2017;12:e0172804.
[37] Smuck M, Muaremi A, Zheng P, Norden J, Sinha A, Hu R, et al. Objective measurement of function following lumbar spinal stenosis decompression reveals improved functional capacity with stagnant real-life physical activity. Spine J 2018;18:15–21.