The Spine Journal 19 (2019) 1276–1293

Review Article

Objective measures of functional impairment for degenerative diseases of the lumbar spine: a systematic review of the literature

Martin N. Stienen, MD/FEBNS[a,b,*], Allen L. Ho, MD[a], Victor E. Staartjes, BSc[b], Nicolai Maldaner, MD[b,c], Anand Veeravagu, MD[a], Atman Desai, MD[a], Oliver P. Gautschi, MD[d], David Bellut, MD[b], Luca Regli, MD[b], John K. Ratliff, MD[a], Jon Park, MD/FRCSC[a]

[a] *Department of Neurosurgery, Stanford University Hospital and Clinics, Stanford, CA, USA*
[b] *Department of Neurosurgery & Clinical Neuroscience Center, University Hospital Zurich, University of Zurich, Zurich, Switzerland*
[c] *Department of Neurosurgery, Kantonsspital St.Gallen, St Gallen, Switzerland*
[d] *Neuro- und Wirbelsäulenzentrum, Hirslanden Klinik St Anna, Lucerne, Switzerland*

Received 25 November 2018; revised 24 February 2019; accepted 24 February 2019

Abstract

BACKGROUND CONTEXT: The accurate determination of a patient's functional status is necessary for therapeutic decision-making and to critically appraise treatment efficacy. Current subjective patient-reported outcome measure (PROM)–based assessments have limitations and can be complimented by objective measures of function.

PURPOSE: To systematically review the literature and provide an overview on the available objective measures of function for patients with degenerative diseases of the lumbar spine.

STUDY DESIGN/SETTING: Systematic review of the literature.

METHODS: The Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines were followed. Two reviewers independently searched the PubMed, Web of Science, EMBASE, and SCOPUS databases for permutations of the words "objective," "assessment," "function," "lumbar," and "spine" including articles on human subjects with degenerative diseases of the lumbar spine that reported on objective measures of function, published until September 2018. Risk of bias was not assessed. No funding was received. The authors report no conflicts of interest.

RESULTS: Of 2,389 identified articles, 82 were included in the final analysis. There was a significant increase of 0.12 per year in the number of publications dealing with objective measures of function since 1989 (95% CI 0.08–0.16, p<.001). Some publications studied multiple diagnoses and objective measures. The United States was the leading nation in terms of scientific output for objective outcome measures (n=21; 25.6%), followed by Switzerland (n=17; 20.7%), Canada, Germany, and the United Kingdom (each n=6; 7.3%). Our search revealed 21 different types of objective measures, predominantly applied to patients with lumbar spinal stenosis (n=67 publications; 81.7%), chronic/unspecific low back pain (n=28; 34.2%) and lumbar disc herniation (n=22; 26.8%). The Timed-Up-and-Go test was the most frequently applied measure (n=26 publications; 31.7%; cumulative number of reported subjects: 5,181), followed by the Motorized Treadmill Test (n=25 publications; 30.5%, 1,499 subjects) and with each n=9 publications (11.0%) the Five-Repetition Sit-To-Stand test (955 subjects), as well as accelerometry analyses (336 subjects). The reliability and validity of many of the less-applied objective measures was uncertain. There was profound heterogeneity in their application and interpretation of results.

FDA Device/Drug Note: Not applicable.

Author disclosures: ***MNS:*** Nothing to disclose; ***ALH:*** Nothing to disclose; ***VES:*** Nothing to disclose; ***NM:*** Nothing to disclose; ***AV:*** Nothing to disclose; ***AD:*** Nothing to disclose; ***OPG:*** Nothing to disclose; ***DB:*** Nothing to disclose; ***LR:*** Nothing to disclose; ***JKR:*** Nothing to disclose; ***JP:*** Nothing to disclose.

* Corresponding author. Department of Neurosurgery, Stanford University, 300 Pasteur Drive, Edwards Bldg., Rm R209, Stanford, CA 94305-5327, USA. Tel.: (+1)-650-382-7158

E-mail address: mnstienen@gmail.com (M.N. Stienen).

https://doi.org/10.1016/j.spinee.2019.02.014

CONCLUSIONS: Clinical studies on patients with lumbar degenerative diseases increasingly employ objective measures of function, which offer high potential for improving the quality of outcome measurement in patient-care and research. This review provides an overview on available options. Our findings call for an agreement and standardization in terms of test selection, conduction and analysis to facilitate comparison of results across cohorts.
PROSPERO REGISTRATION NUMBER: CRD42019122622

Keywords: Objective functional impairment; Disability; Objective outcome measure; Physical function; Systematic review; Functional test

Introduction

The goals of surgical interventions for degenerative diseases of the spine are relieving pain, and improving function and health-related quality of life (hrQoL) [1]. Choice of surgical intervention is complex and depends on many factors. Knowledge of disease natural history is required, since pain (and even motor deficit) may respond to conservative therapy [2]. It is essential to assess pain, functional limitations, and reduction of hrQoL as accurately as possible, since this information serves as a basis for decision-making for or against surgical treatment. Baseline functional status may be used as a reference, against which the success or failure of any treatment will be measured.

An important and necessary evolution has taken place in the last decades, away from the subjective assessment of the treating physician toward a more patient-centered approach [3]. Focus is now on subjective patient-reported generic or disease-specific outcome measures (PROMs) for disability and hrQoL, such as for example, the Oswestry disability index, the Roland-Morris disability index (RMDI) or the Short-Form 12/36 (SF-12/SF-36). Furthermore, generic and disease-specific objective measures of function are gaining increasing attention, adding a further dimension to the comprehensive patient evaluation. The possibilities of broadly-available new technologies such as smartphones equipped with accelerometers or global positioning systems (GPS) have opened additional venues for disability and outcome measurement in research and healthcare.

As the number of reports pertaining to potential objective measures of function continues to grow, the aim of this systematic literature review was to provide an overview on currently available objective measures of function, applicable to patients suffering from degenerative pathologies of the lumbar spine.

Material and methods

The guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) were followed for conducting this systematic review [4]. It was registered under https://www.crd.york.ac.uk/prospero/ (Identifyer CRD42019122622).

Study selection criteria

We included articles of human subjects written in English, German or French that met the following criteria: reporting of one or several objective measures of function, applied to human patients with degenerative diseases of the lumbar spine. We defined objective measures of function as being (1) based on a task to be performed by the patient, (2) evaluated using an objective assessment of the patients' performance on that task (ie, time taken, repetitions, etc.), (3) rated by an observer and/or machine instead of the patient him/herself, and (4) based on a standardized testing protocol. We did not consider widespread objective methods used in orthopedics that measure only certain aspects of the human body, for example, joint mobility with a goniometer, muscle strength with the help of a Newton meter, or radiological parameters (eg, Cobb angle for scoliosis, parameters of sagittal balance, diameter of the spinal canal in the axial magnetic resonance imaging [MRI]). Furthermore, the search was focused strictly on outcome measures for patients with degenerative diseases of the lumbar spine; those applied for trauma (eg, spinal cord injury), spinal oncology, degenerative cervical pathologies or cranial neurosurgery were not included.

Database search and study extraction

A systematic literature was conducted in PubMed, Web of Science, EMBASE, and SCOPUS databases, including articles published until September 2018. We searched for permutations of the words "objective," "assessment," "function," "lumbar," and "spine" in each database's search engine (see Appendices C-F). Full-text papers of which the title and abstract met the eligibility criteria (Table 1) were rigorously assessed to determine inclusion. References from each full-text article were similarly reviewed for inclusion eligibility. The study screening and data extraction were independently performed by two reviewers (M.N.S. & A.L.H.), and any discrepancies were resolved by discussion between those two, or with the entire research group.

Data collection

Reference data such as the study objective, number of included subjects, cohort and disease type studied in general, as well as specifically for each type of applied objective

measure were extracted from the selected articles, together with the study design, year of publication, country of origin (or country of main data generation in case of international collaborations), journal name and the journal's 2017 impact factor (IF, as provided by Thomson-Reuter, whenever available). The latter was done to estimate the scientific robustness and value of each outcome measure. We extracted the method of application, as well as any information regarding its test qualities. The primary objective of each study was characterized as either a study dedicated to (1) exploring qualities of the objective measure (eg, reliability, validity, responsiveness, minimum clinically important difference [MCID], satisfaction), (2) characterizing a certain disease by means of the objective measure (eg, comparing the functional status of patients with or without spondylolisthesis) or (3) investigating a therapeutic effect (applying the objective measure to compare outcomes between treatment groups).

Quality assessment of selected studies and establishment of level of evidence

As we did not intend to carry out a meta-analysis and no valid tools were available to evaluate objective functional tests, we desisted from systematically evaluating quality, level of evidence, and risk of bias of each included study and/or functional test.

Analysis

Quantitative statistical analysis was only possible to a limited extent, due to the significant heterogeneity in included studies' aim, design and type of objective test. Whenever feasible, categorical variables were analyzed by Chi-square and continuous variables by two-sample t-tests. Time trends were analyzed by Poisson regression, allowing calculation of robust standard errors as recommended by Cameron and Trivedi [5]. All analyses were conducted using Stata v14.2 (College Station, TX, USA). p Values <.05 on two-tailed hypotheses were considered statistically significant.

Results

Our database search initially yielded 2,389 articles. After title and abstract screening, 2,301 articles were excluded because they did not meet the inclusion criteria. Eighty-eight potentially eligible articles remained, of which 29 duplicates were removed. Further 73 citations were added through backward and forward citation and hand searching. Thus, 132 articles were retrieved for full-text analysis, of which 50 were subsequently excluded because they were irrelevant to this study. Ultimately, 82 citations were included in this study (Fig. 1). A comprehensive overview on all 82 articles is provided in Supplementary Table 1.

Disease types

Lumbar spinal stenosis (LSS) was by far the disease most frequently studied by objective measures of function (n=67 publications; 81.7%), followed by chronic and/or unspecific low back pain (LBP; n=28; 34.2%), lumbar disc herniation (LDH; n=22; 26.8%), spondylolisthesis (n=18;

Table 1
Table detailing the inclusion and exclusion criteria, according to the PICOS (participants, interventions, comparators, outcomes, and study design) approach detailed in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement. An additional category, "Publications," was added to primarily encompass the language restrictions

	Inclusion criteria	Exclusion criteria
Participants	• Human subjects/patients with lumbar degenerative disc disease • Clinical setting	• Animal subjects • Laboratory setting
Interventions	No intervention required	No intervention required
Comparators	No comparator required	No comparator required
Outcomes	• Objective measure, reflecting functional (dis)ability of a patient/human subject Reporting at least one of the following: • Test quality or feature (Agreement, reliability, validity, minimum clinically important difference, etc.) • Correlation with any subjective outcome measure • Satisfaction with outcome measure • Objective outcome measure used to determine therapeutic effect of a health-care intervention	Either of the following: • No report of any objective test of patient/human subject function • Report of radiological outcomes, electrophysiological or kinematic function of the spine (eg, electromyography or range of motion) only • Outcome data not sufficiently presented or provided upon request from the authors
Study design	Either of the following: • Randomized controlled trial • Quasi-experimental study • Observational study	Either of the following: • Study protocols • Secondary research (review or meta-analysis of primary research)
Publications	Either of the following: • English language • German language • French language	Either of the following: • Conference abstract • Letter, comment or note

22.0%), spinal deformity (n=4; 4.9%), vertebral compression fracture (VCF; n=1; 1.2%) or other types (n=4; 4.9%).

Time-trend in reporting objective measures of function

There was a profound and significant increase of 0.12 scientific papers per year that included an objective measure of function across the last decades (95% CI 0.08–0.16, p<.001; Fig. 2).

Reporting of objective measures of function per country

The United States was the leading nation in terms of overall number of publications that included an objective measure of function (n=21; 25.6%), followed by Switzerland (n=17; 20.7%), Canada, Germany, and the United Kingdom (each n=6; 7.3%). A comprehensive overview of the absolute and relative frequency of publications employing an objective measure of function per country is provided in Fig. 3.

Reporting of objective measures of function per journal

SPINE was the leading journal in terms of overall number of publications that included an objective measure of function (n=19; 23.2%), followed by The Spine Journal (n=8; 9.8%), the European Spine Journal (n=7; 8.5%), Archives of Physical and Medical Rehabilitation (n=4; 4.9%) and Acta Neurochirurgica, the Journal of Neurosurgery: Spine and World Neurosurgery with 3 articles each (3.7%; Supplemental Figure 1).

Objective tests

Our search revealed 21 different types of objective measures of function, for which a comprehensive overview is provided in Table 2, including the absolute and relative frequency of application, study type and disease type for which the measure was applied. The table also summarizes the cumulative and mean number of reported participants per objective measure. The scientific value of each measure is estimated by providing the cumulative and mean IF of the journals that have published articles of each measure. In the table, a brief description of each objective measure is provided. However, many measures were not performed according to uniform and standardized protocols, and instructions given to participants, test protocols and analysis of outcomes profoundly varied across studies for many identified objective measures of function.

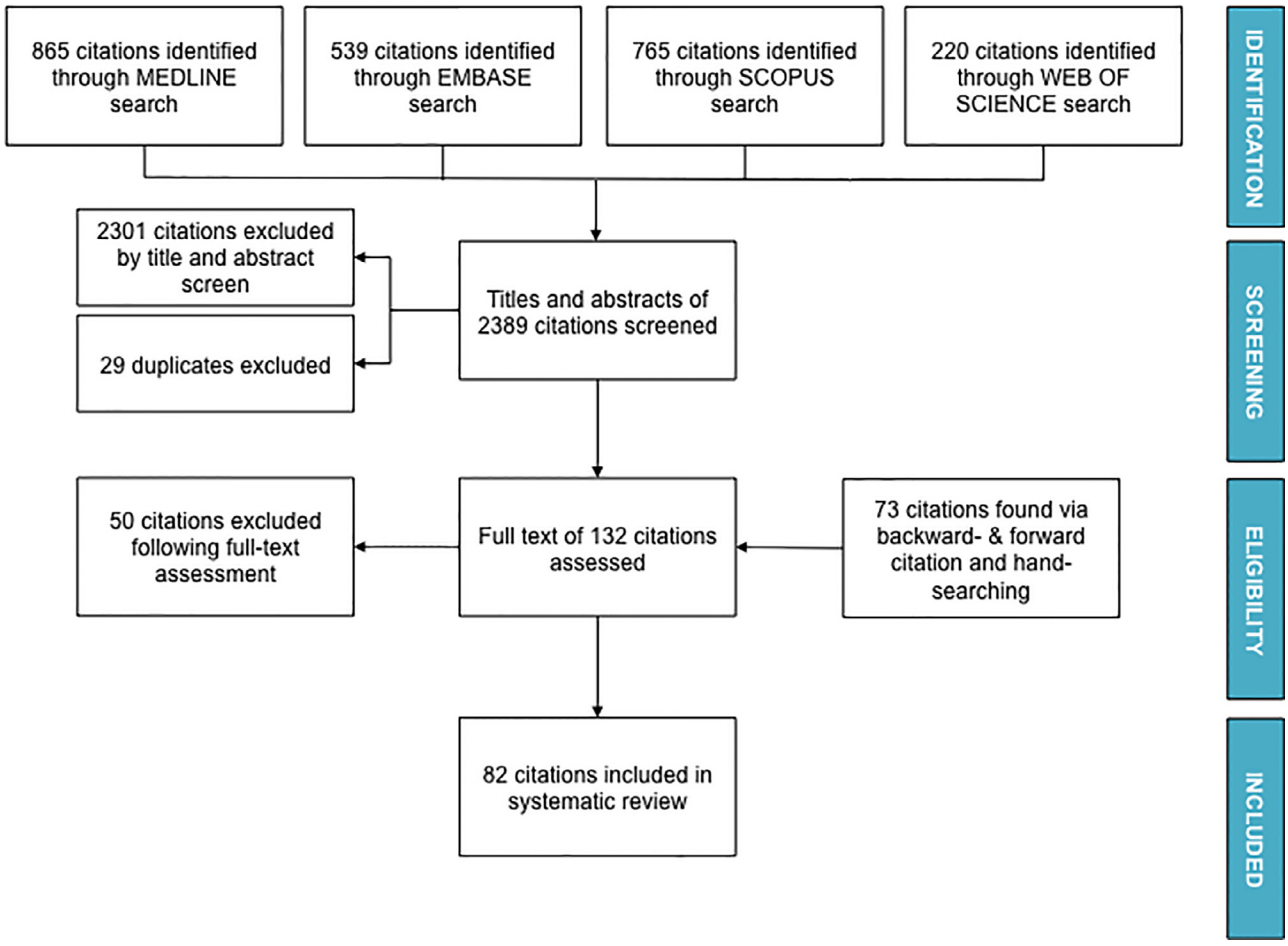

Fig. 1. PRISMA flowchart detailing the process for the selection of papers.

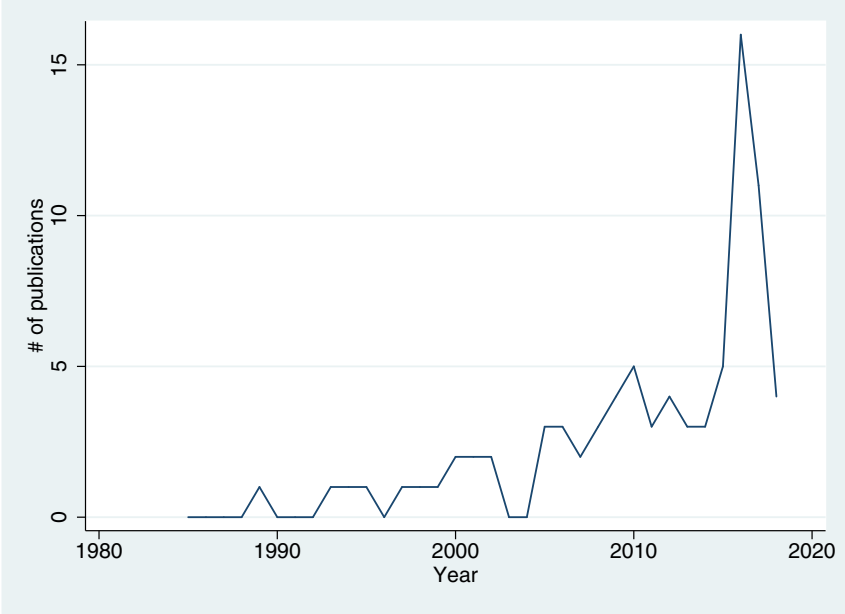

Fig. 2. Line graph highlighting the significant (p<.001) annual increase in the number (#) of publications (y-axis) over the last decades (x-axis).

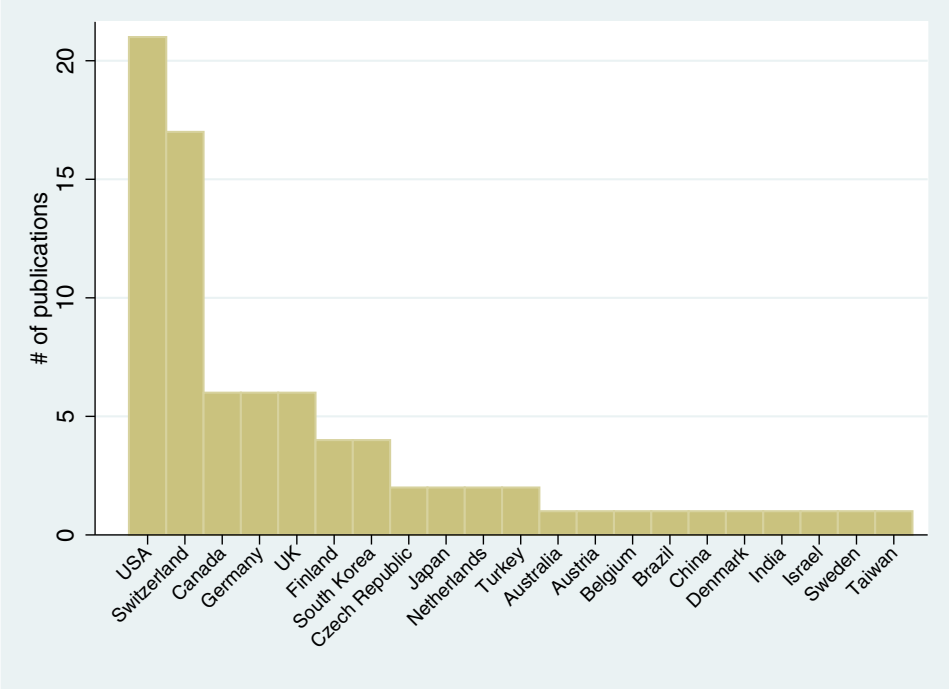

Fig. 3. Histogram indicating the number (#) of publications (y-axis) that employed an objective measure of function per country (x-axis).

The most frequently applied objective measure was the Timed-Up and Go (TUG) test (n=26 publications; 31.7%) with a cumulative number of 5,181 reported subjects. This measure also applied for the widest range of disease types: LSS, LDH, LBP, spondylolisthesis, spinal deformity, VCF, and others. We identified 10 articles focusing primarily on characteristics and qualities of the TUG test, 11 articles applying the TUG test to study a disease and/or condition, and five articles that applied the TUG test to compare outcomes between two different treatment regimes (Table 2). The TUG test was followed in frequency by the Motorized Treadmill Test (MTT; n=25 publications; 30.5%; 1,499

Table 2
Comprehensive list of the objective tests that were applied, together with a brief description, the disease type, study type and objective, number of reported patients and scientific value (estimated by the cumulative impact factor (IF) of publications

No.	Name of objective test	Absolute and relative frequency	Brief description	Study types	Disease types studied	Study objective	Number of reported subjects* (cumulative; mean (SD))	Journal IF (cumulative; mean (SD))	References
1	TUG test	N=26; 31.7%	Participants begin with sitting on a chair. On the word "Go," they get up and walk as fast as possible to a marked line on the floor at 3m distance. At this line, patients turn around, return to the chair and sit down again as quickly as possible. The test result is the time between getting up and sitting down again (s), using a stopwatch or the smartphone "TUG app" (see Appendix A). Transformation of raw test values into age- and sex-standardized T-scores to determine OFI is recommended. [6, 8]	RCT (n=2), prospective observational (n=23), retrospective (n=1)	LSS (n=19); LDH (n=16); LBP (n=17); listhesis (n=14); deformity (n=2); VCF (n=1); other (n=2)	Test characteristics (n=10); disease characteristics (n=11); outcome measure (n=5)	5181; 199 (141)	69.55; 2.78 (1.11)	[6−21,49,50,52, 67−73]
2	MTT	N=25; 30.5%	Patients walk on a treadmill, usually at a predefined protocol. Different studies have proposed different protocols in terms of speed, time or incline and there is no clearly superior or "gold standard" program (see article text). Test results are the time of onset or significant increase in symptoms (s), the total ambulation time (s), the total distance walked (m), as well as the maximum walking speed (m/s) for protocols that allow individual speed selection.	RCT (n=5); prospective observational (n=19), retrospective (n=1)	LSS (n=24); LDH (n=1); LBP (n=3); listhesis (n=1); other (n=1)	Test characteristics (n=6); disease characteristics (n=12); outcome measure (n=6)	1499; 60 (42)	65.21; 2.61 (1.52)	[22−37, 39−46,74]
3	5R-STS test	N=9; 11.0%	Participants sit down on an armless chair (standard height) with a hard seat, firmly placed against the wall. With arms folded across the chest and feet kept flat on the ground participants then stand up fully and sit back down again without using the upper limbs. [18,48] The test result is the time needed until the complete standing position is reached (s). In order to increase discriminative capacity, previous researchers usually asked patients to perform five repetitions of the test, measuring the overall time to complete, with a maximum of 30 seconds (5R-STS). [18,48−51]	RCT (n=2); prospective observational (n=6), retrospective (n=1)	LSS (n=8); LDH (n=2); LBP (n=3); listhesis (n=1)	Test characteristics (n=4); disease characteristics (n=3); outcome measure (n=2)	955; 106 (75)	24.94; 2.77 (1.08)	[18,28,42, 48−53]

Table 2 (***Continued***)

No.	Name of objective test	Absolute and relative frequency	Brief description	Study types	Disease types studied	Study objective	Number of reported subjects* (cumulative; mean (SD))	Journal IF (cumulative; mean (SD))	References
4	Accelerometry analysis	N=9; 11.0%	A number of studies have applied various wearable devices on the body (usually throughout the day only) that measure acceleration and filter these raw acceleration data into a metric known as activity counts, representing the intensity of physical activity. Some devices include further functions such as altimeters. Depending on the device, the number of steps taken, distance walked (m), or calories expended can be calculated. There is a body of literature supporting that accelerometers are usually reliable and provide a valid indicator of overall physical activity in adults.	Prospective observational (n=7), retrospective (n=2)	LSS (n=8); LDH (n=2); LBP (n=2)	Test characteristics (n=2); disease characteristics (n=5); outcome measure (n=2)	336; 37 (23)	22.22; 2.47 (1.13)	[36,54,75−81]
5	SPWT	N=8; 9.8%	Patients walk continuously at their own pace around a 200 m track, until they have to stop for back-related symptoms (or other reasons). Time is kept with a stop-watch and distance measured via a distance wheel or similar device. The main test result is the total walking distance (m), further results include TAT (s), DTFS and walking speed (m/s). [46]	Prospective observational (n=8)	LSS (n=8); deformity (n=1)	Test characteristics (n=4); disease characteristics (n=2); outcome measure (n=2)	388; 49 (55)	22.45; 2.81 (0.74)	[23,46,54,55,79,80,82,83]
6	Gait analysis	N=7; 8.5%	Gait analyses have been performed using walkways containing pressure sensors, [53] reflective markers on participants and infrared cameras, [40,72,84] infrared-emitting diodes on participants captured by motion analysis systems, [85] inertial sensors [62] or sensor-equipped smart shoes [86] to calculate spatiotemporal parameters, such as walking velocity, stride length, step width, gait cycle times (on defined gait cycles) among others. Usually several barefoot gait cycles are performed per participant. The systems were reported as reliable and valid for spatiotemporal parameters. [53,62]	Prospective observational (n=7)	LSS (n=5); LDH (n=3); LBP (n=3); listhesis (n=2); other (n=3)	Test characteristics (n=3); disease characteristics (n=1); outcome measure (n=3)	293; 42 (16)	14.89; 2.13 (0.70)	[40,53,62,72,81,84,85]

Table 2 (***Continued***)

No.	Name of objective test	Absolute and relative frequency	Brief description	Study types	Disease types studied	Study objective	Number of reported subjects* (cumulative; mean (SD))	Journal IF (cumulative; mean (SD))	References
7	10m walking test	N=6; 7.3%	For the 10- or 15 m walking test, participants are instructed to walk (at a comfortable [87] or at maximum speed [51,88,89]) on a flat, straight 10- or 15 m walkway. [17] Most groups have used a 10 m distance; the 15 m distance was used once. [51] The test result is the time to complete the selected distance (s). [17,88,89] One group evaluated patients by their ability to run rather than walk the distance of 10 m.	Prospective observational (n=6)	LSS (n=4); LBP (n=1); deformity (n=1)	Test characteristics (n=1); disease characteristics (n=4); outcome measure (n=1)	250; 42 (14)	16.62; 2.77 (0.63)	[17,51,86−89]
8	SWT	N=5; 6.1%	Participants walk a 10 m course on level ground and marked with cones at each end to complete one shuttle. Assistive devices are allowed if the participant normally uses them. The walking pace is monitored by a predetermined set of beeps from a sound-emitting device, which indicate the amount of time allowed to walk one shuttle. The evaluation is progressive in that the time allowed between beeps for one shuttle gradually decreases. All participants are eventually unable to complete a shuttle in the allowed time. The test includes a maximum of 14 transits in 12 min, with a maximum total distance of 1020 m. [56] The assessor counts the number of completed shuttles and the test result is the walking distance (m; number of completed shuttles multiplied by 10).	RCT (n=2), prospective observational (n=3)	LSS (n=3); LBP (n=2)	Test characteristics (n=2); disease characteristics (n=1); outcome measure (n=2)	954; 191 (199)	31.96; 6.39 (9.53)	[47,56,57,90,91]
9	6WT	N=5; 6.1%	Participants walk as fast as possible back and forth along a flat hallway for 6 minutes. They are informed of the time and encouraged each minute. The main result of the test is the 6WD (m), [17, 60−62] traditionally documented by recording complete laps and using walkway marks for incomplete laps. [60,62] A smartphone application (see Appendix B) has been programmed to measure the 6WD, as well as TTFS and DTFS more conveniently using GPS-coordinates. [3]	RCTs (n=1), prospective observational (n=4)	LSS (n=2); listhesis (n=1); deformity (n=2)	Test characteristics (n=1); disease characteristics (n=1); outcome measure (n=3)	518; 104 (88)	90.73; 18.15 (34.16)	[17,51,60−62]

Table 2 (***Continued***)

No.	Name of objective test	Absolute and relative frequency	Brief description	Study types	Disease types studied	Study objective	Number of reported subjects* (cumulative; mean (SD))	Journal IF (cumulative; mean (SD))	References
10	Bicycle ergometer test	N=3; 3.7%	Participants sit in their preferred posture on a stationary bicycle ergometer, holding the handlebars with both hands. Throughout the entire test, they are instructed to continue at a constant pedaling speed of 50−60 rpm. No resistance is added for the first minute, but resistance is increased to 20 W (≈150 kpm/m) for the second, and to 50 W (≈300 kpm/m) for additional eight minutes. The total maximum test time is 10 min, if the patient does not have to stop earlier. The test result is the time that the patient pedaled (s), as well as the total distance (m). Pain and/or paresthesia can be measured before and after the test; the TTFS can also be monitored.	Prospective observational (n=3)	LSS (n=3)	Test characteristics (n=1); disease characteristics (n=2)	124; 41 (18)	7.63; 2.54 (2.43)	[26,27,29]
11	6m walking test	N=3; 3.7%	Participants complete timed walks over a 6 m walkway at their preferred speed. Having ample space before and after the walking space is required to ensure that walking speed is constant. [18] The main test results is the time (s) taken to complete the walk (single trial [18,50] or mean of six trials [92]), whereas number of steps, walking velocity and cadence have also been analyzed [92].	Prospective observational (n=2), retrospective (n=1)	LSS (n=2); LBP (n=1)	Disease characteristics (n=2); outcome measure (n=1)	256; 85 (38)	10.42; 3.47 (1.18)	[18,50,92]
12	AST	N=2; 2.4%	For the alternative step test (AST), the entire left and right foot (shoes removed) alternatively have to be placed as fast as possible onto a step with a distinct height (eg 18 cm) and depth (eg 40 cm). The time taken to take eight steps comprises the test measure (s). The AST is used to evaluate a participant's ability to maintain standing balance while performing a potentially destabilizing activity, such as standing on one leg while stepping.	Prospective observational (n=1), retrospective (n=1)	LSS (n=2)	Disease characteristics (n=2)	206; 103 (32)	7.63; 3.82 (1.45)	[18,50]

Table 2 (***Continued***)

No.	Name of objective test	Absolute and relative frequency	Brief description	Study types	Disease types studied	Study objective	Number of reported subjects* (cumulative; mean (SD))	Journal IF (cumulative; mean (SD))	References
13	WC test	N=2; 2.4%	For the weight carrying (WC) test participants walk 20 m as fast as possible while carrying 10% of their body weight evenly distributed in hand-held weights. The test result is the time needed to complete the distance (s).	RCT (n=1), prospective observational (n=1)	LSS (n=2)	Test characteristics (n=1); outcome measure (n=1)	182; 91 (88)	5.87; 2.93 (0.20)	[28,42]
14	Single leg balance	N=2; 2.4%	Participants maintain single-leg balance, unsupported, for as long as possible (maximum of 30 s). The test result is the time until failing to keep balance (s).	Prospective observational (n=2)	LSS (n=1); deformity (n=1)	Test characteristics (n=1); disease characteristics (n=1)	180; 90 (49)	3.98; 1.99 (1.18)	[14,70]
15	GPS-based assessment	N=2; 2.4%	GPS is used to track position- and movement data of participants during the day in intervals of about 10 s. Outcomes include total distance walked, average distance, walking speed and total walking duration per day. Precision of measurements of about ±1.5 m outside (optimal conditions) have been reported. [93] The data have to be preprocessed using complex algorithms and checked for plausibility.	Prospective observational (n=1)	LSS (n=1); LDH (n=1); LBP (n=1)	Test characteristics (n=1); disease characteristics (n=1)	6; 3 (1)	5.43; 2.71 (0.11)	[93,94]
16	Balance test	N=1; 1.2%	This test requires an industrial force plate balance platform, designed for testing postural stability/trace length, indicating how far the participant shifts from the center of pressure over a 20 s period while performing balance tasks. The test result is the participant's shift (mm^2).	Prospective observational (n=1)	LSS (n=1)	Disease characteristics (n=1)	10	3.12	[80]
17	Fast stair descent	N=1; 1.2%	Participants are timed as they descent twelve steps with a defined depth (eg 28 cm) and height (eg 17 cm) "as quickly and as safely as possible". The test result is the time (s) and an average of two trials is calculated.	Prospective observational (n=1)	LBP (n=1)	Disease characteristics (n=1)	106	3.08	[13]
18	Gait speed	N=1; 1.2%	Participants walk 2.44 m at their usual (self-selected) pace, providing space for acceleration and deceleration.	Prospective observational (n=1)	LBP (n=1)	Disease characteristics (n=1)	106	3.08	[13]

Table 2 (***Continued***)

No.	Name of objective test	Absolute and relative frequency	Brief description	Study types	Disease types studied	Study objective	Number of reported subjects* (cumulative; mean (SD))	Journal IF (cumulative; mean (SD))	References
19	Sitting and standing time	N=1; 1.2%	Participants sit and stand as long as possible. The test result is the maximum duration (min) for sitting (mean: 122−130 min) and standing (mean: 10−20 min).	Prospective observational (n=1)	LSS (n=1); deformity (n=1)	Outcome measure (n=2)	179	3.12	[82]
20	One minute stair climbing	N=1; 1.2%	Participants walk up and down a staircase with five stairs for 1 min. The test result is the number of stairs climbed during the time period.	Prospective observational (n=1)	LBP (n=1)	Test characteristics (n=1)	53	2.93	[51]
21	PILE	N=1; 1.2%	For the progressive isoinertial lifting evaluation (PILE) participants lift a box with a weight 4 times within 20 s from the floor up to a 75 cm-high table. Starting weights and incremental weights are different for men and women. The starting weight for women is 3.6 kg and 5.85 kg for men (weight of box included). After each completed lifting cycle, the weight for women is increased by 2.25 kg and for men by 4.5 kg. The test stops when the participant cannot lift the box 4 times within 20 s, the participant decides to stop, the heart rate exceeds 85% of the maximal heart rate, the maximal amount of the weight that could safely be lifted is reached (60% of participant's body weight), or the test observer considers further lifting unsafe. The test result is the number of fully completed lifting stages.	Prospective observational (n=1)	LBP (n=1)	Test characteristics (n=1)	53	2.93	[51]

RCT, randomized controlled trial; LSS, lumbar spinal stenosis; LDH, lumbar disc herniation; LBP, low back pain; VCF, vertebral compression fracture; OFI, objective functional impairment; TAT, total ambulation time; DTFS, distance to first symptoms; TTFS, time to first symptoms; GPS, Global Positioning System.

* Subjects include both patients and controls.

reported subjects) and both the Five-Repetition Sit-To-Stand test (5R-STS; n=9 publications; 11.0%; 955 reported subjects), as well as accelerometry analyses (n=9 publications; 11.0%; 336 reported subjects).

Reports applying the 6-minute walking test (6WT) had the highest cumulative IF (90.73), followed by those applying the TUG test (69.55) and the MTT (65.21).

A comprehensive overview on all metrics for identified objective measures of function is provided in Table 2. The most frequently applied, reproducible, reliable, and validated objective measures of function are described in more detail.

The TUG test

The TUG is a simple test that does not require any special equipment except for a chair and 3 m of walking space. It has frequently been applied in patients harboring a multitude of degenerative conditions of the lumbar spine. Here, patients sit on a chair and lean back, with arms resting on the armrests. On the word "Go," they are asked to get up and walk as fast as possible to a marked line on the floor at 3 m distance. At this line, patients turn around (180°), return to the chair and sit back down, as quickly as possible. The time between getting up and sitting down again is recorded in seconds using a stopwatch [6−8]. Besides interpreting raw test times (in seconds), categorizing patients into those with no, mild, moderate or severe "objective functional impairment" (OFI) is possible using age- and sex-standardized cut-off values [6,8]. Moreover, the calculation of standardized OFI T-scores allows for exact determination of a patient's functional condition as a deviation from the normal population mean [6−10]. Working with OFI rather than TUG test raw values prevents bias naturally introduced by the high influence of the variables age and sex on the TUG test result [10−12]. A free smartphone app can be utilized for both TUG measurement and automatic OFI calculation (more information in Appendix A).

The TUG test had excellent intra- (intraclass correlation coefficient [ICC] 0.97) and inter-rater reliability (ICC 0.99), with a standard error of measurement of 0.21 and 0.23 seconds, respectively [6]. It was shown to discriminate between disability in patients with or without chronic LBP [13]. Among a set of clinical variables, the TUG test result was the one that showed the highest correlation with disability and walking capacity [14]. The convergent validity with PROMs, such as visual analog scale back (r=0.25) and leg pain (r=0.29), RMDI (r=0.38) and ODI (r=0.34), as well as SF-12 physical component summary (PCS; r=−0.32) and EQ-5D (r=−0.28) was demonstrated [6]. In surgical candidates with lumbar degenerative disc disease (DDD), convergent validity of the TUG test with PROMs of pain intensity, functional impairment, and QoL was even higher after as compared to before the surgical intervention [9]. Various studies demonstrate that the TUG test is sensitive to a patient's postoperative change in function [7,9,15]. A change in the TUG test of at least 3.4 seconds is considered a clinically meaningful change in function (MCID) for patients with lumbar DDD [7]. For single, but especially for repetitive evaluations, patients preferred the TUG test over questionnaire-based assessments [16].

Considering its high intrarater reliability, a single trial would be sufficient to measure a participants level of impairment [6−8], but some studies preferred to calculate the mean of two or three TUG trials [13,17]. While one study suggested that a patients' body mass index (BMI) might adversely effects the performance of functional mobility tests [18], a dedicated report did not find a significant influence of the BMI on the TUG test [19]. Further research indicated little or no influence of a patients' smoking and of the mental health status on the TUG test result [20,21], making this test a particularly interesting option for the functional assessment of patients with psychiatric comorbidities that often interfere with PROM-based assessments [20].

The MTT

For the MTT, patients are instructed to walk on a calibrated treadmill, usually starting on a level surface (0% grade) and at an established protocol speed and time. Participants should not place both hands at the handrails for support, as this can improve their walking capacity by bending forward; [22−25] holding one handrail for balance purposed is usually allowed. Pain and/or paresthesia can be measured before and after the test; the time of symptom onset (TTFS=time to first symptoms; minutes and seconds) can also be monitored.

Prior studies have proposed to start with 10 minutes at 2 mph, increase to 2.5 mph for the next 5 minutes, then to 3 mph for additional 5 minutes (total of 20 minutes) [26,27], or to remain at a constant speed of 2−2.5 mph for the complete duration of 15 or 30 minutes [28−34]. Other groups had participants walk at maximum, individually selected speed for up to 15 or 30 minutes [22,23,35−37]. According to the modified Bruce protocol, two warm-up stages of 3 minutes are followed by incremental increase in speed and gradient [38−40]. Further, individualized protocols have been used [25,41−43]. If a participant is unable to tolerate the standard speed and distance, the speed is reduced or the test is ended, if necessary. The test is also stopped when subjects reach a safety endpoint, for example, 85% of predicted maximal heart rate (220−age) [39].

Raw test results are the time of onset or significant increase in symptoms (TTFS = time to first symptoms; minutes and seconds), the total ambulation time (minutes and seconds), the total distance walked (m), as well as the maximum walking speed (m/s) for protocols that allow individual speed selection [26,28,35,43]. To the best of the authors' knowledge, no studies have interpreted test results in a standardized fashion.

The authors are also not aware of any study determining the optimal protocol for the MTT in patients with degenerative disease of the lumbar spine. Studies do

suggest, however, that the additional information gained after 15 minutes of walking time is negligible [25,41,44]. The intrarater reliability of the MTT was high to excellent for both TTFS (ICC 0.90–0.98) and total ambulation time (0.89–0.96) at 1.2 mph or at an individually selected speed [45]. For an individual protocol with a gradual increase in walking speed, intrarater reliability was equally high (ICC 0.83) [42]. As the MTT protocols differed between studies, reliability is unclear for other protocols. The total distance walked was significantly less in patients with LSS (mean 292±21 m) than in a healthy control group (409±16 m; p<.01) [42]. Convergent validity of the MTT was otherwise demonstrated with the self-paced walking test (SPWT; r=0.88) [46], self-reported walking distance (r=0.62) [33], as well as with self-reported symptoms of neurogenic claudication (r=0.88) [26]. Other studies indicated a weak to moderate correlation between the objectively measured walking distance on the MTT with the walking distance that patients reported being able to walk [23,35]. The MTT was shown to be sensitive to change in the postoperative setting [44].

The MTT has been primarily studied in patients with LSS so far, and it was applied as an objective outcome measure in a number of randomized controlled trials and observational studies (Table 2). Despite a similar number of publications reporting on the MTT and the TUG test, the number of reported subjects was by far less for the MTT. In direct comparison to the SPWT, the MTT showed poorer internal responsiveness for LSS patients and patients consistently walked further in the SPWT [23,46]. Also, a distinct drawback of the MTT is that special equipment (motorized treadmill) and trained personal is required, whereas other tests (eg, TUG test, 6WT) can be performed without additional resources. The potential risks of frightening or even injuring patients on a motorized treadmill must also be considered, especially when examining the elderly [47].

The 5R-STS

For this test, participants are asked to sit down on an armless chair of standard height (48 cm) and with a hard seat, firmly placed against the wall. With arms folded across the chest and feet kept flat on the ground (wearing stable footwear) participants are asked to stand up fully and sit back down again without using the upper limbs and as fast as possible [48]. In order to increase discriminative capacity, most previous researchers have asked participants to perform five repetitions of the test, measuring the overall time to complete, with a maximum of 30 seconds [18,48–51].

The test result usually is the time to perform the five trials. Besides reporting raw values (in seconds), the 5R-STS was standardized and cut-off values have been proposed to discriminate between patients with lumbar DDD and no (≤10.4 seconds), mild (10.5–15.2 seconds), moderate (15.3–22.0 seconds), or severe OFI (>22.0 seconds) [48]. One study asked participants to perform as many repetitions of the STS test as possible within 30 seconds; the test result being the total number of repetitions [52]. Other groups only measured the time required to rise from the chair (chair rise time), without sitting back down [42,51,53].

The 5R-STS' intrarater reliability was found to be high for a single (ICC 0.84) [42] and excellent for five repetitive trials (ICC 0.95–0.98) [18,48]. The test time for a single trial was significantly longer in patients with LSS (mean 0.99±0.16 sec) than in a healthy control group (0.57±1.72 seconds; p<.01) [42]. For logarithmic 5R-STS test results, moderate convergent validity was reported in a cohort of n=157 patients with lumbar DDD in terms of RMDI (r=0.49), ODI (r=0.44), visual analog scale back pain (r=0.31), and the EQ-5D index (r=-0.41; all p<.001) [48]. Age, body weight, and the BMI were shown to influence the result of the 5R-STS test. A patient's expected "normal" test time (or "targeted 5R-STS performance" after successful treatment) can be predicted by the formula $t_a = 0.03$ age + 0.15 BMI + 1.7 [48].

The SPWT

For the SPWT, patients are instructed to walk continuously and at their own pace around an indoor 200 m track, until they have to stop for back-related symptoms (or other reasons). A maximum walking time limit of 30 minutes has been proposed previously for patients that are little or asymptomatic [23,54,55]. Time is kept with a stop-watch and distance measured via a distance wheel or similar device. The main test result is the total walking distance (m), further results include total walking time (s), distance to first symptoms (DTFS) and walking speed (m/s). The intrarater reliability was excellent for total walking distance (ICC=0.98), DTFS (ICC 0.94), and walking speed (ICC 0.80) [46,55]. In patients with LSS, total walking distance ranged from 60 to 2065 m (mean 776±726 m, SD) and 67–1800 seconds (mean 840±690 seconds, SD) [54]. The standard error of measurement and MCID of the SPWT have been reported to be 131 and 363 m, respectively, in a small sample of 26 LSS patients [55]. The convergent validity with the MTT, self-estimated walking time and distance, as well as with symptoms of neurogenic claudication (back and leg pain, paresthesia, leg weakness, unsteadiness, ODI, SF-36 PCS and Swiss Spinal Stenosis Questionnaire) were moderate-to-high [23,54]. The SPWT outperformed the MTT in terms of internal (post-therapeutic) responsiveness, whereas external responsiveness (concordance with the patient's subjective perception of change in clinical status) was relatively poor for both tests [23]. Comparative studies between the two tests indicated that LSS patients walked a higher absolute distance in the SPWT (mean 987±914 m) as compared to the MTT (mean 611±666 m; p<.05), probably as the SPWT allows for greater (self-selected) speed [46]. The SPWT also showed higher correlation with self-reported measures of pain, functional impairment and hrQoL than a digital activity monitor [54].

The Shuttle Walking Test

For the Shuttle Walking Test (SWT), participants are asked to walk a 10 m course (32 ft, 81 in) on level ground and marked with cones at each end to complete one shuttle. Assistive devices (eg, canes or walkers) are allowed if the participant normally uses them. The walking pace is monitored by a predetermined set of beeps from a sound-emitting device (CD-player, mp3-player, etc.), which indicate the amount of time allowed to walk one shuttle. The evaluation is progressive in that the time allowed between beeps for one shuttle gradually decreases. The test is maximal in that all participants are eventually unable to complete a shuttle in the allowed time, either for being short of breath or having too much pain or discomfort to continue. During the first minute of the test, beeps sound each 20 seconds, and three shuttles (30 m) are completed. During the second minute, four shuttles are completed; during the third minute five shuttles are completed; and so on up to 14 transits in 12 minutes, with a maximum total distance of 1,020 m [56]. The assessor counts the number of completed shuttles and the test result is the walking distance in meters (number of completed shuttles multiplied by 10).

The main test result is the total walking distance (m), for which excellent intrarater reliability was reported (ICC 0.92–0.99) [56,57]. The SWT also demonstrated substantial changes in the functional status before and after surgery for LSS [56,57]. For 95% certainty of change between two assessments in a single patient, the SWT should change by at least 76 m [56]. In direct comparison with the MTT, the SWT exhibited similar test qualities for the assessment of patients with LSS, while evoking a lower level of cardiovascular stress [58].

The 6WT

The 6WT is typically performed on a 3 m wide and 30 m long well illuminated flat hallway, according to the American Thoracic Society guidelines [59]. Patients are instructed to walk as fast as possible back and forth along the course for 6 minutes. Each minute, they are informed of the time and encouraged to continue. The main result of the test is the 6-minute walking distance (6WD) [17,60–62], traditionally documented by recording complete laps and using additional walkway marks every 3 m for incomplete laps [60,62]. Modifications with 5 minutes walking time have been proposed [51], but the majority of studies agree in the 6 minutes assessment. Recently, a free smartphone application has been programmed to allow measuring the 6WD, as well as DTFS (m) and time to first symptoms (TTFS; s) in the patients home environment by GPS-coordinates (more information in Appendix B) [3].

The 6WT is less explored than the SPWT, MTT or the SWT in the context of lumbar DDD. A previous study found the 6WD to range around 357±107 m in n=29 LSS patients (mean ODI of 30.7±16.3), with a similar 6WD in n=27 healthy control subjects (mean 408±73 m) [62]. The authors noticed a 6WD increase by 21 m around 10 weeks and by 26 m around one year postoperatively, but the result did not differ significantly from the baseline assessment. In a Swedish multicenter randomized controlled trial, mean 6WD in surgical candidates with LSS with or without spondylolisthesis was in the range of 309–331 m and improved by 70–80 m at 2 years postoperatively [61]. More available literature on the 6WT derives from other medical fields. In populations with various chronic cardiopulmonary diseases, the MCID for the 6WD ranged between 14.0–30.5 m [63]. The MCID currently remains to be determined for lumbar DDD and in particular for LSS.

The 6WT appears useful in particular for its ease of administration using smartphone apps, but also because it closely resembles ambulatory activities in which patients with lumbar DDD are limited [17].

Discussion

This article provides an overview of currently available objective measures of function, applied to patients with degenerative diseases of the lumbar spine. The systematic review of the available literature yields some interesting findings.

First, there was a significant and gradual increase in the reporting of objective measures of function over the last three decades. Second, there were number of countries and scientific journals, which appeared to be particularly interested in publishing research that employed objective measures of function. Third, and perhaps most important, we found that there was uncertainty pertaining to the reliability and validity of many of the objective measures applied in clinical studies. There was profound heterogeneity concerning the types of objective measure, their method of application, as well as regarding the definition of their main test results. Reporting of raw test values dominated the available literature and only few studies so far interpreted the results in a standardized fashion, adjusting for potential confounders such as age, BMI or gender. Given this variability across studies, comparison of cohorts in terms of OFI is currently limited.

Is there a current "gold standard"?

Based on the literature research there is no single "gold standard" for objective functional testing. Each physician and researcher must consider the type of function and impairment that is inherent to the patient he/she is going to examine. The TUG test, possibly combined with the 5R-STS test appears to be a reasonable choice, given both tests' ease of administration. They only require a chair and a stopwatch, allowing them to be performed spontaneously, for example in case OFI is suspected during an outpatient consultation in clinics. Both tests were found to be reliable and valid for patients with lumbar DDD [6,8,48]. The TUG test was shown to be particularly sensitive in patients with predominant lumboradicular pain (eg, LDH) [8,9,15], whereas the 5R-STS test was more adequate in patients

with predominant LBP [48]. Longer and more challenging reliable and validated tests such as the SPWT, MTT or 6WT may be chosen for LSS, considering that neurogenic claudication may not clinically manifest during examination with the shorter tests. For those planning to employ objective measures of function for research or clinical care, Table 2 summarizes existing options.

Opportunities for future research

There are some potential advantages of including objective measures of function in patient-care and research. Some of them, in particular the modern motion-sensor or smartphone-/GPS-based evaluations are a venue for passive and unobtrusive acquisition of longitudinal data, which could help overcome weaknesses inherent to current data collection such as missing data and loss of follow-up. Smartphones are integrated virtually in every aspect of our lives, having become a mirror of our behavior and likely very directly reflect change in behavior and loss of function, respectively. Further advantages include the usually high reliability versus high inter- and intraobserver variability of physician- and patient-rated measures, misinterpretation of questionnaire items and differences in the subjective PROM scoring for educational, cultural, and motivational reasons [1]. In contrast to subjective PROMs, objective outcome measures are applicable in foreign-language patients and illiterates. Presenting test-results as Z- or T-scores—expressing the patient's deviation from the healthy population norm—enables comparison between different tests and across studies and/or cohorts [6]. While PROM results are usually difficult to interpret for nonmedical personnel such as the patient, relatives or the public, result interpretation is more obvious for objective tests. Objective outcome measures comply with the modern trend of patient empowerment and patient-centered healthcare and research [64]. Lastly, objective measures of function are well accepted by patients [16]. Convergent validity between objective outcome measures and PROMs was consistently weak to moderate, indicating that objective measures cannot replace PROMs. However, these measures may add an important, further dimension to the comprehensive patient evaluation [8,20].

Need for standardization

This review revealed a broad variety of available assessments. Most of the authors provided no reliability and validity measures. Even between studies that agreed on a similar type of objective measure, differences existed pertaining to the test protocol, definition of main outcome, and analytical approach. Also, objective tests of function can be heavily influenced by further neurological and/or orthopedic comorbidities (eg, Parkinson's disease, previous stroke, hip and/or knee osteoarthritis), and not all prior studies accounted for this. Deyo et al. recommended the introduction of uniform standards for measuring PROM-based outcome about 20 years ago [65]. This review now indicates a need for agreement in terms of objective test selection, conduction and analysis, which should facilitate future comparison of study results across cohorts, studies, and countries.

Strengths and limitations

To the best of the author's knowledge, there is no prior work that summarized currently available objective measures of function using a systematic approach. As such, this review may be a valuable resource for physicians when choosing one or several tests for patient care or research. Notwithstanding the systematic approach, additional articles may exist that we failed to identify. Furthermore, one may argue that excluding tests that measure only certain aspects of the human body, such as range of motion, might be a weakness. However, this would have exceeded the scope of this article, and such a review was recently published [66]. Several studies included relatively low numbers of patients and/or subjects and more data on the objective measures of function will further increase our understanding of their specific value. Lastly, we were unable to perform a systematic assessment of the risk of bias in individual studies, since no validated tools to assess bias in systematic reviews of functional tests were available.

Conclusions

Clinical studies of patients with lumbar degenerative diseases increasingly employ objective measures of function, which offer high potential for patient-care and research. This review provides an overview on available options. Our findings call for an agreement and standardization in terms of test selection, conduction, and analysis to facilitate comparison of results across cohorts.

Acknowledgment

There was no conflict of interest and no funding was received for this study.

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.spinee.2019.02.014.

Appendix

A. The "TUG" app (webgearing ag, Switzerland) is available free of charge in multiple languages at the Apple app store and Google Play.
B. The "6WT" app (webgearing ag, Switzerland) is available free of charge in multiple languages at the Apple app store or Google Play.
C. Medline (Pubmed) search terms: ("goals"[MeSH Terms] OR "goals"[All Fields] OR "objective"[All Fields]) AND ("Assessment"[Journal] OR "assessment"[All Fields]) AND ("physiology"[Subheading] OR "physiology"[All Fields] OR "function"[All Fields] OR "physiology"[MeSH Terms] OR "function"[All Fields]) AND ("lumbar vertebrae"

[MeSH Terms] OR ("lumbar"[All Fields] AND "vertebrae"[All Fields]) OR "lumbar vertebrae"[All Fields] OR ("lumbar"[All Fields] AND "spine"[All Fields]) OR "lumbar spine"[All Fields])

D. SCOPUS search terms: objective AND assessment AND function AND lumbar AND spine
E. EMBASE search terms: 'objective assessment function lumbar spine' OR (objective AND ('assessment'/exp OR assessment) AND ('function'/exp OR function) AND lumbar AND ('spine'/exp OR spine))
F. Web of Science search terms: TOPIC: (objective assessment function lumbar spine)

References

[1] Gautschi OP, Corniola MV, Schaller K, Smoll NR, Stienen MN. The need for an objective outcome measurement in spine surgery—the timed-up-and-go test. Spine J 2014;14:2521–2.
[2] Peul WC, van Houwelingen HC, van den Hout WB, et al. Surgery versus prolonged conservative treatment for sciatica. N Engl J Med 2007;356:2245–56.
[3] Stienen MN, Bellut D, Regli L, Hausmann ON, Gautschi OP. Functional assessment of patients with lumbar degenerative disc disease: who is right − the doctor, the patient or the objective test? Praxis 2017;106:1041–52.
[4] Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700.
[5] Cameron AC, Trivedi PK. Microeconometrics using Stata. Revised Edition ed. College Station, TX: Stata Press; 2010.
[6] Gautschi OP, Smoll NR, Corniola MV, et al. Validity and reliability of a measurement of objective functional impairment in lumbar degenerative disc disease: the timed up and go (TUG) test. Neurosurgery 2016;79:270–8.
[7] Gautschi OP, Stienen MN, Corniola MV, et al. Assessment of the minimum clinically important difference in the timed up and go test after surgery for lumbar degenerative disc disease. Neurosurgery 2017;80:380–5.
[8] Stienen MN, Smoll NR, Joswig H, et al. Validation of the baseline severity stratification of objective functional impairment in lumbar degenerative disc disease. J Neurosurg Spine 2017;26:598–604.
[9] Gautschi OP, Joswig H, Corniola MV, et al. Pre- and postoperative correlation of patient-reported outcome measures with standardized Timed Up and Go (TUG) test results in lumbar degenerative disc disease. Acta neurochirurgica 2016;158:1875–81.
[10] Gautschi OP, Smoll NR, Joswig H, et al. Influence of age on pain intensity, functional impairment and health-related quality of life before and after surgery for lumbar degenerative disc disease. Clin Neurol Neurosurg 2016;150:33–9.
[11] Gautschi OP, Corniola MV, Smoll NR, et al. Sex differences in subjective and objective measures of pain, functional impairment, and health-related quality of life in patients with lumbar degenerative disc disease. Pain 2016;157:1065–71.
[12] Gautschi OP, Smoll NR, Corniola MV, et al. Sex differences in lumbar degenerative disc disease. Clin Neurol Neurosurg 2016;145:52–7.
[13] Sions JM, Coyle PC, Velasco TO, Elliott JM, Hicks GE. Multifidi muscle characteristics and physical function among older adults with and without chronic low back pain. Arch Phys Med Rehabil 2017;98:51–7.
[14] Lin SI, Lin RM. Disability and walking capacity in patients with lumbar spinal stenosis: association with sensorimotor function, balance, and functional performance. J Orthop Sports Phys Ther 2005;35:220–6.
[15] Gautschi OP, Corniola MV, Joswig H, et al. The timed up and go test for lumbar degenerative disc disease. J Clin Neurosci 2015;22:1943–8.
[16] Joswig H, Stienen MN, Smoll NR, et al. Patients' preference of the timed up and go test or patient-reported outcome measures before and after surgery for lumbar degenerative disk disease. World Neurosurg 2017;99:26–30.
[17] Kondo R, Yamato Y, Nagafusa T, et al. Effect of corrective long spinal fusion to the ilium on physical function in patients with adult spinal deformity. Eur Spine J 2017;26:2138–45.
[18] Kim HJ, Chun HJ, Han CD, et al. The risk assessment of a fall in patients with lumbar spinal stenosis. Spine 2011;36:E588–92.
[19] Stienen MN, Joswig H, Smoll NR, et al. Influence of body mass index on subjective and objective measures of pain, functional impairment, and health-related quality of life in lumbar degenerative disc disease. World Neurosurg 2016;96. 570-7 e1.
[20] Stienen MN, Smoll NR, Joswig H, et al. Influence of the mental health status on a new measure of objective functional impairment in lumbar degenerative disc disease. Spine J 2017;17:807–13.
[21] Joswig H, Stienen MN, Smoll NR, et al. Effects of smoking on subjective and objective measures of pain intensity, functional impairment, and health-related quality of life in lumbar degenerative disk disease. World Neurosurg 2017;99:6–13.
[22] Gulbahar S, Berk H, Pehlivan E, et al. [The relationship between objective and subjective evaluation criteria in lumbar spinal stenosis]. Acta Orthop Traumatologica Turc 2006;40:111–6.
[23] Rainville J, Childs LA, Pena EB, et al. Quantification of walking ability in subjects with neurogenic claudication from lumbar spinal stenosis—a comparative study. Spine J 2012;12:101–9.
[24] Markman JD, Gewandter JS, Frazer ME, et al. A randomized, double-blind, placebo-controlled crossover trial of oxymorphone hydrochloride and propoxyphene/acetaminophen combination for the treatment of neurogenic claudication associated with lumbar spinal stenosis. Spine 2015;40:684–91.
[25] Deen Jr. HG, Zimmerman RS, Lyons MK, McPhee MC, Verheijde JL, Lemens SM. Measurement of exercise tolerance on the treadmill in patients with symptomatic lumbar spinal stenosis: a useful indicator of functional status and surgical outcome. J Neurosurg 1995;83:27–30.
[26] Tenhula J, Lenke LG, Bridwell KH, Gupta P, Riew D. Prospective functional evaluation of the surgical treatment of neurogenic claudication in patients with lumbar spinal stenosis. J Spinal Disord 2000;13:276–82.
[27] Yukawa Y, Lenke LG, Tenhula J, Bridwell KH, Riew KD, Blanke K. A comprehensive study of patients with surgically treated lumbar spinal stenosis with neurogenic claudication. J Bone Joint Surg Am Vol 2002;84-A:1954–9.
[28] Koc Z, Ozcakir S, Sivrioglu K, Gurbet A, Kucukoglu S. Effectiveness of physical therapy and epidural steroid injections in lumbar spinal stenosis. Spine 2009;34:985–9.
[29] Dong G, Porter RW. Walking and cycling tests in neurogenic and intermittent claudication. Spine 1989;14:965–9.
[30] Malmivaara A, Slatis P, Heliovaara M, et al. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. Spine 2007;32:1–8.
[31] Herno A, Airaksinen O, Saari T. Computed tomography after laminectomy for lumbar spinal stenosis. Patients' pain patterns, walking capacity, and subjective disability had no correlation with computed tomography findings. Spine 1994;19:1975–8.
[32] Herno A, Airaksinen O, Saari T, Pitkanen M, Manninen H, Suomalainen O. Computed tomography findings 4 years after surgical management of lumbar spinal stenosis. No correlation with clinical outcome. Spine 1999;24:2234–9.
[33] Barz T, Melloh M, Staub L, et al. The diagnostic value of a treadmill test in predicting lumbar spinal stenosis. Eur Spine J Soc 2008;17:686–90.
[34] Hurri H, Sainio P, Kinnunen H, et al. Walking distance as a measure of disability in lumbar spinal stenosis. Suomen Ortopedia ja Traumatologia 2008;31:254–6.
[35] Zeifang F, Schiltenwolf M, Abel R, Moradi B. Gait analysis does not correlate with clinical and MR imaging parameters in patients with symptomatic lumbar spinal stenosis. BMC Musculoskelet Disord 2008;9:89.
[36] Geisser ME, Haig AJ, Tong HC, et al. Spinal canal size and clinical symptoms among persons diagnosed with lumbar spinal stenosis. Clin J Pain 2007;23:780–5.

[37] Whitman JM, Flynn TW, Childs JD, et al. A comparison between two physical therapy treatment programs for patients with lumbar spinal stenosis: a randomized clinical trial. Spine 2006;31:2541–9.
[38] McInnis KJ, Balady GJ, Weiner DA, Ryan TJ. Comparison of ischemic and physiologic responses during exercise tests in men using the standard and modified Bruce protocols. Am J Cardiol 1992;69:84–9.
[39] Henchoz Y, de Goumoens P, Norberg M, Paillex R, So AK. Role of physical exercise in low back pain rehabilitation: a randomized controlled trial of a three-month exercise program in patients who have completed multidisciplinary rehabilitation. Spine 2010;35:1192–9.
[40] Stief F, Meurer A, Wienand J, Rauschmann M, Rickert M. Has a mono- or bisegmental lumbar spinal fusion surgery an influence on self-assessed quality of life, trunk range of motion, and gait performance? Spine 2015;40:E618–26.
[41] Fritz JM, Erhard RE, Delitto A, Welch WC, Nowakowski PE. Preliminary results of the use of a two-stage treadmill test as a clinical diagnostic tool in the differential diagnosis of lumbar spinal stenosis. Journal of spinal disorders 1997;10:410–6.
[42] Whitehurst M, Brown LE, Eidelson SG, D'Angelo A. Functional mobility performance in an elderly population with lumbar spinal stenosis. Arch Phys Med Rehabil 2001;82:464–7.
[43] Moon ES, Kim HS, Park JO, et al. Comparison of the predictive value of myelography, computed tomography and MRI on the treadmill test in lumbar spinal stenosis. Yonsei med J 2005;46:806–11.
[44] Deen HG, Zimmerman RS, Lyons MK, McPhee MC, Verheijde JL, Lemens SM. Use of the exercise treadmill to measure baseline functional status and surgical outcome in patients with severe lumbar spinal stenosis. Spine 1998;23:244–8.
[45] 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.
[46] 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 2009;34:2444–9.
[47] Drury T, Ames SE, Costi K, Beynnon B, Hall J. Degenerative spondylolisthesis in patients with neurogenic claudication effects functional performance and self-reported quality of life. Spine 2009;34:2812–7.
[48] Staartjes VE, Schroder ML. The five-repetition sit-to-stand test: evaluation of a simple and objective tool for the assessment of degenerative pathologies of the lumbar spine. J Neurosurg Spine 2018: 1–8.
[49] Nielsen PR, Jorgensen LD, Dahl B, Pedersen T, Tonnesen H. Prehabilitation and early rehabilitation after spinal surgery: randomized clinical trial. Clin Rehabil 2010;24:137–48.
[50] Lee BH, Kim TH, Park MS, et al. Comparison of effects of nonoperative treatment and decompression surgery on risk of patients with lumbar spinal stenosis falling: evaluation with functional mobility tests. J Bone Joint Surg Am Vol 2014;96:e110.
[51] Smeets RJ, Hijdra HJ, Kester AD, Hitters MW, Knottnerus JA. The usability of six physical performance tasks in a rehabilitation population with chronic low back pain. Clin Rehabil 2006;20:989–97.
[52] Park S, Han HS, Kim GU, et al. Relationships among disability, quality of life, and physical fitness in lumbar spinal stenosis: an investigation of elderly Korean women. Asian Spine J 2017;11:256–63.
[53] Conrad BP, Shokat MS, Abbasi AZ, Vincent HK, Seay A, Kennedy DJ. Associations of self-report measures with gait, range of motion and proprioception in patients with lumbar spinal stenosis. Gait Posture 2013;38:987–92.
[54] 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.
[55] 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.
[56] 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 2002;27:84–91.
[57] Taylor S, Frost H, Taylor A, Barker K. Reliability and responsiveness of the shuttle walking test in patients with chronic low back pain. Physiother Res Int 2001;6:170–8.
[58] Zwierska I, Nawaz S, Walker RD, Wood RF, Pockley AG, Saxton JM. Treadmill versus shuttle walk tests of walking ability in intermittent claudication. Med Sci Sports Exerc 2004;36:1835–40.
[59] ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med 2002;166:111–7.
[60] Alves VL, Avanzi O. Objective assessment of the cardiorespiratory function of adolescents with idiopathic scoliosis through the six-minute walk test. Spine 2009;34:E926–9.
[61] Forsth P, Olafsson G, Carlsson T, et al. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. N Engl J Med 2016;374:1413–23.
[62] Loske S, Nuesch C, Byrnes KS, et al. Decompression surgery improves gait quality in patients with symptomatic lumbar spinal stenosis. Spine J 2018;18:2195–204.
[63] Bohannon RW, Crouch R. Minimal clinically important difference for change in 6-minute walk test distance of adults with pathology: a systematic review. J Eval Clin Prac 2017;23:377–81.
[64] Stienen MN, Smoll NR, Fung C, et al. Home-time as a surrogate marker for functional outcome after aneurysmal subarachnoid hemorrhage. Stroke 2018;49:3081–4.
[65] Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. Spine. 1998;23:2003–13.
[66] Essendrop M, Maul I, Laubli T, Riihimaki H, Schibye B. Measures of low back function: a review of reproducibility studies. Clin Biomech 2002;17:235–49.
[67] Corniola MV, Stienen MN, Joswig H, et al. Correlation of pain, functional impairment, and health-related quality of life with radiological grading scales of lumbar degenerative disc disease. Acta Neurochir 2016;158:499–505.
[68] Gautschi OP, Stienen MN, Joswig H, Smoll NR, Schaller K, Corniola MV. The usefulness of radiological grading scales to predict pain intensity, functional impairment, and health-related quality of life after surgery for lumbar degenerative disc disease. Acta Neurochir 2017;159:271–9.
[69] Hartmann S, Hegewald AA, Tschugg A, Neururer S, Abenhardt M, Thome C. Analysis of a performance-based functional test in comparison with the visual analog scale for postoperative outcome assessment after lumbar spondylodesis. Eur Spine J 2016;25:1620–6.
[70] Muramoto A, Imagama S, Ito Z, Hirano K, Ishiguro N, Hasegawa Y. Spinal sagittal balance substantially influences locomotive syndrome and physical performance in community-living middle-aged and elderly women. J Orthop Sci 2016;21:216–21.
[71] Stienen MN, Joswig H, Chau I, et al. Efficacy of intraoperative epidural triamcinolone application in lumbar microdiscectomy: a matched-control study. J Neurosurg Spine 2018;28:291–9.
[72] Toosizadeh N, Harati H, Yen TC, et al. Paravertebral spinal injection for the treatment of patients with degenerative facet osteoarthropathy: evidence of motor performance improvements based on objective assessments. Clin Biomech 2016;39:100–8.
[73] Van Meirhaeghe J, Bastian L, Boonen S, et al. A randomized trial of balloon kyphoplasty and nonsurgical management for treating acute vertebral compression fractures: vertebral body kyphosis correction and surgical parameters. Spine 2013;38:971–83.
[74] Prasad BC, Ramesh Chandra VV, Devi BV, Chivukula SS, Pundarikakshaiah K. Clinical, radiological, and functional evaluation of surgical treatment in degenerative lumbar canal stenosis. Neurol India 2016;64:677–83.
[75] Mobbs RJ, Phan K, Maharaj M, Rao PJ. Physical activity measured with accelerometer and self-rated disability in lumbar spine surgery: a prospective study. Global Spine J 2016;6:459–64.

[76] Norden J, Smuck M, Sinha A, Hu R, Tomkins-Lane C. Objective measurement of free-living physical activity (performance) in lumbar spinal stenosis: are physical activity guidelines being met? Spine J 2017;17:26–33.
[77] Pryce R, Johnson M, Goytan M, Passmore S, Berrington N, Kriellaars D. Relationship between ambulatory performance and self-rated disability in patients with lumbar spinal stenosis. Spine 2012;37:1316–23.
[78] Schulte TL, Schubert T, Winter C, et al. Step activity monitoring in lumbar stenosis patients undergoing decompressive surgery. Eur Spine J 2010;19:1855–64.
[79] Tomkins-Lane CC, Conway J, Hepler C, Haig AJ. Changes in objectively measured physical activity (performance) after epidural steroid injection for lumbar spinal stenosis. Arch Phys Med Rehabil 2012;93:2008–14.
[80] Tomkins-Lane CC, Lafave LM, Parnell JA, et al. The spinal stenosis pedometer and nutrition lifestyle intervention (SSPANLI): development and pilot. Spine J 2015;15:577–86.
[81] Zheng CF, Liu YC, Hu YC, et al. Correlations of Japanese Orthopaedic Association Scoring Systems with gait parameters in patients with degenerative spinal diseases. Orthop Surg 2016;8:447–53.
[82] Rolfe KW, Zucherman JF, Kondrashov DG, Hsu KY, Nosova E. Scoliosis and interspinous decompression with the X-STOP: prospective minimum 1-year outcomes in lumbar spinal stenosis. Spine J 2010;10:972–8.
[83] Tomkins-Lane CC, Battie MC. Predictors of objectively measured walking capacity in people with degenerative lumbar spinal stenosis. J Back Musculoskelet Rehabil 2013;26:345–52.
[84] Stief F, Meurer A, Wienand J, Rauschmann M, Rickert M. Effect of lumbar spinal fusion surgery on the association of self-report measures with objective measures of physical function. Gait Posture 2018;61:7–12.
[85] Passmore SR, Johnson M, Pelleck V, Ramos E, Amad Y, Glazebrook CM. Lumbar spinal stenosis and lower extremity motor control: the impact of walking-induced strain on a performance-based outcome measure. J Manipulat Physiol Therap 2014;37:602–9.
[86] Lee SI, Park E, Huang A, et al. Objectively quantifying walking ability in degenerative spinal disorder patients using sensor equipped smart shoes. Medical Eng Phys 2016;38:442–9.
[87] Lee SI, Campion A, Huang A, et al. Identifying predictors for postoperative clinical outcome in lumbar spinal stenosis patients using smart-shoe technology. Journal Neuroeng Rehabil 2017;14:77.
[88] Adamova B, Vohanka S, Dusek L, Jarkovsky J, Chaloupka R, Bednarik J. Outcomes and their predictors in lumbar spinal stenosis: a 12-year follow-up. Eur Spine J 2015;24:369–80.
[89] Micankova Adamova B, Vohanka S, Dusek L, Jarkovsky J, Bednarik J. Prediction of long-term clinical outcome in patients with lumbar spinal stenosis. Eur Spine J 2012;21:2611–9.
[90] Comer CM, Johnson MI, Marchant PR, Redmond AC, Bird HA, Conaghan PG. The effectiveness of walking stick use for neurogenic claudication: results from a randomized trial and the effects on walking tolerance and posture. Arch Phys Med Rehabil 2010;91:15–9.
[91] Fairbank J, Frost H, Wilson-MacDonald J, et al. Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. Bmj 2005;330:1233.
[92] Khodadadeh S, Eisenstein SM. Gait analysis of patients with low back pain before and after surgery. Spine 1993;18:1451–5.
[93] Bostelmann R, Schneller S, Cornelius JF, Steiger HJ, Fischer I. A new possibility to assess the perioperative walking capacity using a global positioning system in neurosurgical spine patients: a feasibility study. Eur Spine J 2016;25:963–8.
[94] Barzilay Y, Noam S, Meir L, et al. Assessing the outcomes of spine surgery using global positioning systems. Spine 2011;36:E263–7.