



**Universiteit
Leiden**
The Netherlands

Normative data of a smartphone app-based 6-minute walking test, test-retest reliability, and content validity with patient-reported outcome measures

Tosic, L.; Goldberger, E.; Maldaner, N.; Sosnova, M.; Zeitlberger, A.M.; Staartjes, V.E.; ... ; Stienen, M.N.

Citation

Tosic, L., Goldberger, E., Maldaner, N., Sosnova, M., Zeitlberger, A. M., Staartjes, V. E., ... Stienen, M. N. (2020). Normative data of a smartphone app-based 6-minute walking test, test-retest reliability, and content validity with patient-reported outcome measures. *Journal Of Neurosurgery: Spine*, 33(4), 480-489. doi:10.3171/2020.3.SPINE2084

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/3182250>

Note: To cite this publication please use the final published version (if applicable).

Normative data of a smartphone app–based 6-minute walking test, test-retest reliability, and content validity with patient-reported outcome measures

*Lazar Tasic, MD,¹ Elior Goldberger, BSc,¹ Nicolai Maldaner, MD,² Marketa Sosnova, MD,² Anna M. Zeitberger, MD,² Victor E. Staartjes, BSc,¹ Pravesh S. Gadraj, MD,³ Hubert A. J. Eversdijk, BSc,⁴ Ayesha Quddusi, MBBS,⁵ Maria L. Gandía-González, MD,⁶ Jamasb Joshua Sayadi, BA,⁷ Atman Desai, MD,⁷ Luca Regli, MD,¹ Oliver P. Gautschi, MD,⁸ and Martin N. Stienen, MD^{1,7}

¹Department of Neurosurgery, University Hospital Zurich and Clinical Neuroscience Center, University of Zurich, Switzerland; ²Department of Neurosurgery, Cantonal Hospital St. Gallen, St. Gallen, Switzerland; ³Department of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands; ⁴Department of Neurosurgery, Bergman Clinics, Amsterdam, The Netherlands; ⁵Center for Neuroscience, Queens University, Kingston, Ontario, Canada; ⁶Department of Neurosurgery, Hospital Universitario La Paz, Madrid, Spain; ⁷Department of Neurosurgery, Stanford University Hospital and Clinics, Stanford, California; and ⁸Neuro and Spine Center, Hirslanden Clinic St. Anna, Lucerne, Switzerland

OBJECTIVE The 6-minute walking test (6WT) is used to determine restrictions in a subject's 6-minute walking distance (6WD) due to lumbar degenerative disc disease. To facilitate simple and convenient patient self-measurement, a free and reliable smartphone app using Global Positioning System coordinates was previously designed. The authors aimed to determine normative values for app-based 6WD measurements.

METHODS The maximum 6WD was determined three times using app-based measurement in a sample of 330 volunteers without previous spine surgery or current spine-related disability, recruited at 8 centers in 5 countries (mean subject age 44.2 years, range 16–91 years; 48.5% male; mean BMI 24.6 kg/m², range 16.3–40.2 kg/m²; 67.9% working; 14.2% smokers). Subjects provided basic demographic information, including comorbidities and patient-reported outcome measures (PROMs): visual analog scale (VAS) for both low-back and lower-extremity pain, Core Outcome Measures Index (COMI), Zurich Claudication Questionnaire (ZCQ), and subjective walking distance and duration. The authors determined the test-retest reliability across three measurements (intraclass correlation coefficient [ICC], standard error of measurement [SEM], and mean 6WD [95% CI]) stratified for age and sex, and content validity (linear regression coefficients) between 6WD and PROMs.

RESULTS The ICC for repeated app-based 6WD measurements was 0.89 (95% CI 0.87–0.91, $p < 0.001$) and the SEM was 34 meters. The overall mean 6WD was 585.9 meters (95% CI 574.7–597.0 meters), with significant differences across age categories ($p < 0.001$). The 6WD was on average about 32 meters less in females (570.5 vs 602.2 meters, $p = 0.005$). There were linear correlations between average 6WD and VAS back pain, VAS leg pain, COMI Back and COMI subscores of pain intensity and disability, ZCQ symptom severity, ZCQ physical function, and ZCQ pain and neuroischemic symptoms subscores, as well as with subjective walking distance and duration, indicating that subjects with higher pain, higher disability, and lower subjective walking capacity had significantly lower 6WD (all $p < 0.001$).

CONCLUSIONS This study provides normative data for app-based 6WD measurements in a multicenter sample from 8 institutions and 5 countries. These values can now be used as reference to compare 6WT results and quantify objective functional impairment in patients with degenerative diseases of the spine using z-scores. The authors found a good to

ABBREVIATIONS 6WD = 6-minute walking distance; 6WT = 6-minute walking test; ASA = American Society of Anesthesiologists; COMI = Core Outcome Measures Index; COPD = chronic obstructive pulmonary disease; DDD = degenerative disc disease; GPS = Global Positioning System; HRQoL = health-related quality of life; ICC = intraclass correlation coefficient; LDH = lumbar disc herniation; LSS = lumbar spinal stenosis; MRC = Medical Research Council; OFI = objective functional impairment; PROM = patient-reported outcome measure; SEM = standard error of measurement; TUG = Timed Up and Go; VAS = visual analog scale; ZCQ = Zurich Claudication Questionnaire.

SUBMITTED January 18, 2020. **ACCEPTED** March 30, 2020.

INCLUDE WHEN CITING Published online May 29, 2020; DOI: 10.3171/2020.3.SPINE2084.

* L.T. and E.G. contributed equally to this work and share first authorship.

excellent test-retest reliability of the 6WT app, a low area of uncertainty, and high content validity of the average 6WD with commonly used PROMs.

<https://thejns.org/doi/abs/10.3171/2020.3.SPINE2084>

KEYWORDS 6-minute walking test; normative data; reliability; validity; test qualities; spinal stenosis; neurogenic claudication; degenerative disc disease; spine surgery; objective functional impairment

DEGENERATIVE diseases of the lumbar spine typically present with varying degrees of pain, functional disability, and reduction of health-related quality of life (HRQoL). As the degree of suffering is one of the key factors to determine whether or not surgical treatment is required, the accurate determination of a patient's functional status is of paramount importance in daily clinical practice.¹

Traditionally, the patient's functional status is estimated by use of scales (such as the visual analog scale [VAS] for pain) or by handing out questionnaires for patient self-assessment, so-called patient-reported outcome measures (PROMs). Those measures serve to "objectify" a patient's subjective degree of suffering and the complex behavioral patterns behind pain avoidance and disability. More recently, however, objective functional tests are increasingly used to support the standardized assessment of patients with lumbar degenerative disc disease (DDD).^{2,3}

While the Timed Up and Go (TUG) test is the currently best-explored test to quantify objective functional impairment (OFI) in patients with lumbar radicular pain (e.g., lumbar disc herniation [LDH]),^{2,4-7} tests incorporating longer ambulation time may be particularly helpful to determine OFI in patients with neurogenic claudication or lumbar radicular pain that does not manifest in the TUG test's 2 × 3-meter walking distance (e.g., lumbar spinal stenosis [LSS]).^{8,9} The 6-minute walking test (6WT) has gained popularity for the evaluation of LSS patients in the last few years.^{2,10-12} Its main outcome is the maximum walking distance over 6 minutes (6-minute walking distance [6WD], in meters), traditionally recorded in a resource-intensive manner by trained physiotherapy personnel. New smartphone technology allows for Global Positioning System (GPS)-based determination of the 6WD using a spine-specific, free, and accurate smartphone app (Fig. 1).¹³

The aim of the current study was to determine normative data for the 6WT by use of this app, and to determine its reliability and content validity to support interpretation of app-based 6WD measurements in patients with degenerative diseases of the lumbar spine.

Methods

Subject Identification and Inclusion/Exclusion Criteria

We recruited volunteers—visitors, family members, friends, and next of kin of patients with lumbar DDD—treated at 8 neurosurgical departments in Switzerland, the Netherlands, Spain, Canada, and the US. Those subjects were chosen to obtain a sample that resembles patients with degenerative spine disease in terms of important epidemiological characteristics (e.g., age, social class, smoking status, etc.).

To be included, participating subjects had to fulfill the following inclusion criteria: 1) voluntary participation and written informed consent; 2) age ≥ 16 years; 3) possessing a smartphone (iOS or Android); and 4) able to understand and follow the instructions of the 6WT app. The following exclusion criteria were defined: 1) any known history of degenerative spine disease; 2) any other disease or illness likely to influence the walking ability (e.g., osteoarthritis of the hip/knee, Parkinson's disease, congestive heart failure ≥ New York Heart Association class III, chronic obstructive pulmonary disease [COPD] ≥ Global Initiative for Chronic Obstructive Lung Disease stage III, lung cancer, etc.); or 3) pregnancy.

Data Collection and PROMs

A standardized data sheet was filled out for each subject, including the variables age (years), sex (male/female), height (cm), weight (kg), working status (full-time, part-time, retired, disabled/invalid, other), smoking status (smoker, nonsmoker), previous spine surgery (yes/no), lowest muscle strength in any lower-extremity index muscle (according to the British Medical Research Council [MRC] paresis grading from 0/5 [no strength] to 5/5 [full strength]), and comorbidities including arterial hypertension, diabetes, COPD, other respiratory disease, any cardiac disease, any neurological disease, any gastrointestinal disease, depression or other mental health disorder, osteoarthritis or other joint disease, and any other disease. Moreover, the American Society of Anesthesiologists (ASA) grading scale was estimated, ranging from 1 (normal, healthy patient) to 5 (moribund patient). To estimate spine-related disability, the following PROMs were filled out by participating subjects.

VAS

Pain intensity was measured for low-back pain and lower-extremity (radicular leg) pain on a VAS ranging from 0 (none) to 10 (severe pain).

Zurich Claudication Questionnaire

The Zurich Claudication Questionnaire (ZCQ),¹⁴ as the recommended "gold standard" for neurogenic claudication,¹⁵ was completed with its two main scores and two subscores. The first main score was the ZCQ symptom severity, ranging from 7 (none) to 33 (severe symptoms). The two subscores included ZCQ pain (range 3 [none] to 15 [severe pain]) and ZCQ neuroischemic symptoms (range 4 [none] to 18 [severe neuroischemic symptoms]). The second main score was ZCQ physical function, ranging from 5 (none) to 20 (severe disability).

Core Outcome Measures Index

The Core Outcome Measures Index (COMI) Back, a

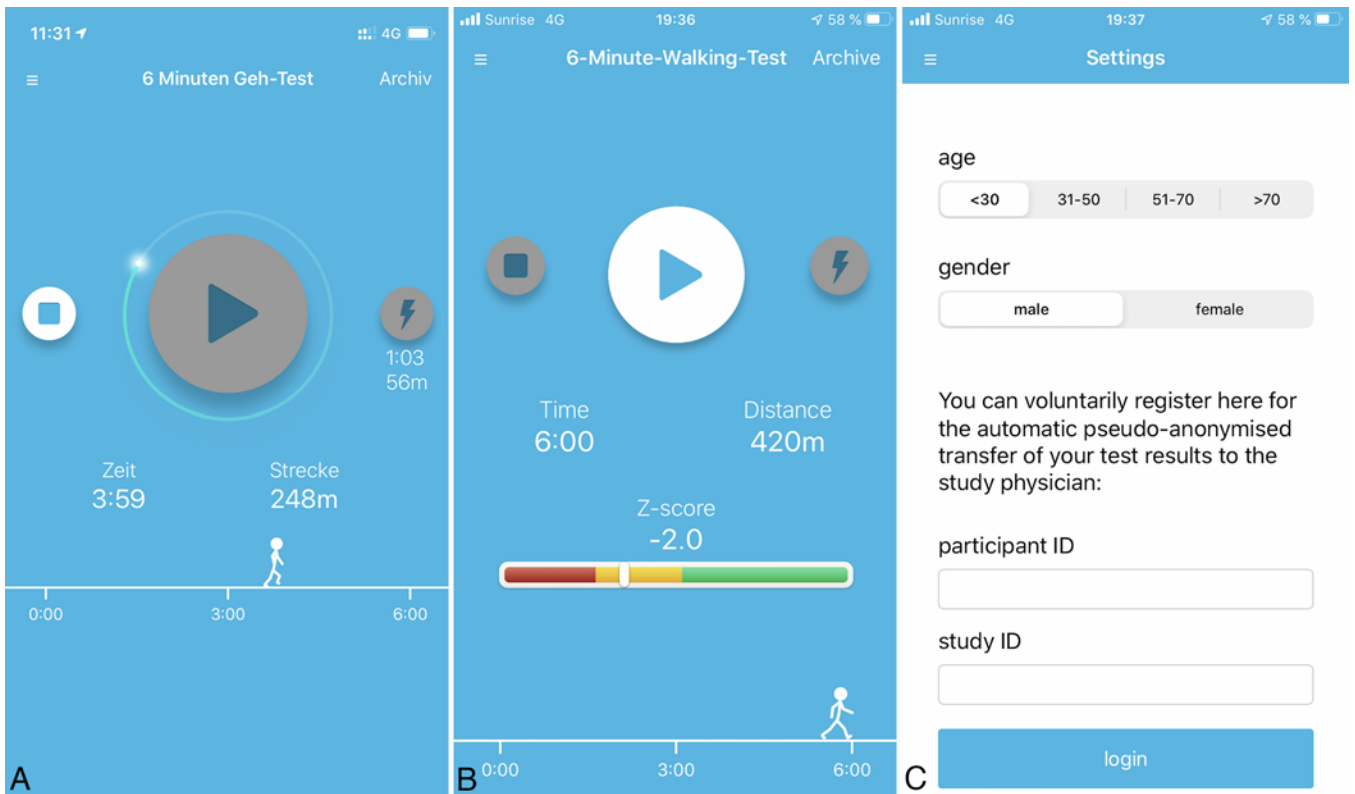


FIG. 1. Screenshot of the user interfaces of the 6WT smartphone app. **A:** In this example, the measurement is at 3:59 minutes and 248 meters and ongoing. Patients can push the “flash” button (lightning bolt) when they experience clinically relevant symptoms (e.g., back pain, neurogenic claudication, sciatica). The app will register the time to first symptoms (1:03 minutes in this example) and distance to first symptoms (56 meters here) as well as the 6WD as the main outcome. **B:** In this example, the measurement is finished; the 29-year-old male patient walked a total of 420 meters. The app uses the normative data to convert the test result into an age- and sex-adjusted z-score $[(420 - 626 \text{ m})/101 \text{ m} = -2.0]$. Here, the test result is 2 standard deviations below the normative population value. The color code expresses this deviation (*green* = z-score > -0.5 , *yellow* = z-scores between -0.5 and -3.0 , *red* = z-scores < -3.0). **C:** In the “settings” interface the patient can indicate his or her age and sex, with stratifications based on Table 3. Copyright Martin N. Stienen. Published with permission. Figure is available in color online only.

short, multidimensional, and thoroughly validated instrument to record pain intensity, back-related function, symptom-specific well-being, HRQoL, and social/work disability,¹⁶ was completed. The COMI consisted of the main score (the COMI Back, ranging from 0 [none] to 10 [severe back-related disability]) and its two subscores: COMI pain intensity (range 0 [none] to 10 [severe pain]) and COMI disability (range 0 [none] to 10 [severe disability]).

Maximum Walking Distance and Duration

The maximum walking distance, ranging from 0 to 1000 meters, was expressed on a VAS of 10 cm, with each millimeter corresponding to a 10-meter walking distance. The maximum walking duration, ranging from 0 to 60 minutes, was expressed on a VAS of 6 cm, with each millimeter corresponding to a 1-minute walking duration.

Performance and Scoring of the 6WT

Participating subjects were instructed on how to download and install the free 6WT app (Fig. 1), using standardized information sheets in plain language and augmented by figures/illustrations (Supplemental Fig. 2). Those sheets

also contained instructions on the optimal testing environment, being a sufficiently long, straight, and relatively level path without high-rises (or other objects blocking the GPS signal) and without obstacles (such as red traffic lights), in agreement with prior research.¹³ All subjects performed the 6WT three times and raw test results for each trial were recorded. A person and contact number were provided to help resolve questions, should any occur.

Statistical Considerations

The main outcome was the 6WD of spine-healthy subjects, stratified according to variables that are likely to influence the 6WD. First, the effect sizes of the influence of baseline variables on 6WD were determined by calculating η^2 (range 0–1). After Cohen’s rule of thumb, η^2 values around 0.01, 0.06, and > 0.14 are considered small, medium, and large effect sizes, respectively.¹⁷ We then selected age and sex as two main categories for stratification, considering their impact and practical availability. Next, the average of the three 6WT trials was taken and reported as mean, standard deviation, and 95% confidence interval (CI). No outliers were excluded from the calculations, and we planned on enrolling at least 10 subjects per category

a priori. Providing normative data (mean and SD) for different age and sex categories allows for the calculation of standardized z-scores using the following formula: $z = (x - \mu)/\sigma$, in which x = observed test result, μ = normal population mean test result, and σ = normal population standard deviation of test result.

Rather than reporting raw test results (6WD), z-scores express the number of SDs patients deviate from the mean of the normal population and hereby help to accurately quantify the degree of OFI. The required values for μ and σ are provided below (see Table 3). Because these calculations may be impractical for use in daily patient care, the normative values reported in this article have been implemented in the free 6WT app, which automatically performs the z-score calculations (see Fig. 1B and the *Appendix* for more information on the app).

Reliability was determined to verify the independence of the cutoff point for 6WT scores indicating no OFI versus scores indicating OFI. The intraclass correlation coefficient (ICC) was determined according to McGraw and Wong¹⁸ and Shrout and Fleiss¹⁹ to ascertain consistency and agreement between the three measures. Good (ICC between 0.75 and 0.9) or excellent (ICC > 0.9) agreement implies that the difference between the absolute values of the three measurements is very small.²⁰ The latter (determination of consistency and agreement) was only used for intrarater reliability because the determination of interrater agreement is not required for app-based outcome measurement. Moreover, the standard error of measurement (SEM) as a “gray zone” of uncertainty between patient scores indicating the presence or absence of OFI was determined. In this gray zone, the physician must use clinical judgment to determine whether or not a patient’s test result represents OFI.⁴ The SEM is equal to the SD multiplied by the square root of 1 minus the intrarater ICC, as demonstrated by Stratford and Goldsmith.²¹

Content validity was expressed as the relationship between average 6WD and the subscores of the subjective PROMs (listed above), using linear regression. All analyses were conducted using Stata (version 14.2, StataCorp LLC) for Mac. A p value < 0.05 was considered statistically significant.

Ethical Considerations

This project enrolled no patients and no approval by an IRB is required for this type of study. Subjects were included on a voluntary basis after written informed consent. A convenient sample of 6 consecutive patients with degenerative diseases of the lumbar spine from an ongoing IRB-approved prospective study (<http://clinicaltrials.gov/identifier/NCT03977961>) were included to illustrate the applicability of the 6WT app in clinical practice. Patients were also included on a voluntary basis after written informed consent. No payments to participating subjects or patients were made.

Results

Study Sample

A total of 330 subjects with a mean age of 44.2 ± 18.6 years (range 16–91 years) were enrolled, of whom 160

TABLE 1. Baseline demographic information of the sample of spine-healthy volunteers

Variable	Value
Mean age (SD), range, yrs	44.2 (18.6), 16–91
Sex, n (%)	
Male	160 (48.5)
Female	170 (51.5)
Mean body dimensions (SD), range	
Height, cm	172.9 (9.6), 152–199
Weight, kg	74.0 (15.8), 44–130
BMI, kg/m ²	24.6 (4.2), 16.3–40.2
Working status, n (%)	
Working, full-time	158 (47.9)
Working, part-time	66 (20.0)
Retired	46 (13.9)
Other	60 (18.2)
Smoking status, n (%)	
Nonsmoker	283 (85.8)
Smoker	47 (14.2)
Comorbidities, n (%)	
Arterial hypertension	51 (15.5)
Diabetes	17 (5.2)
COPD	4 (1.2)
Other respiratory disorder	9 (2.7)
Any cardiac disease	17 (5.2)
Any neurological disease	3 (0.9)
Any gastrointestinal disease	6 (1.8)
Any rheumatic disease	15 (4.6)
Depression/mental health disease	6 (1.8)
Any other disease*	24 (7.3)
ASA risk scale grade, n (%)	
1	297 (90.0)
2	31 (9.4)
3	2 (0.6)

* Not belonging to any of the previously listed categories.

(48.5%) were male. Table 1 summarizes the basic demographic and medical information of the sample. None of the subjects reported prior spine surgery and all 330 (100%) reported full muscle strength (MRC 5/5) in both lower extremities. Supplemental Table 1 contains the effect sizes of the demographic and medical variables on the subject’s average 6WD.

Table 2 contains the pain, disability, and HRQoL scores of the subjective PROMs. The sample scored low in terms of mean VAS back pain (0.52), VAS leg pain (0.24), ZCQ symptom severity (7.97), ZCQ physical function (5.25), and COMI Back score (0.44). Average walking distance and walking duration were almost unrestricted, with values of 989 meters and 58.7 minutes, respectively (Table 2).

Normative Data

The overall mean 6WD was 586 meters (95% CI 575–597 meters), with significant differences between age categories ($p < 0.001$). The 6WD was on average about 32 meters less in females (570.5 vs 602.2 meters, $p = 0.005$). Normative values (mean \pm SD, 95% CI) for the total co-

TABLE 2. Pain and disability scores obtained in a sample of 330 spine-healthy volunteers

Scale	Mean (SD), Range
VAS	
Back pain	0.52 (1.22), 0–8
Leg pain	0.24 (0.87), 0–8
ZCQ	
Symptom severity	7.97 (2.07), 7–21
Pain subscore	3.64 (1.38), 3–11
Neuroischemic symptoms subscore	4.32 (0.82), 4–10
Physical function	5.25 (0.82), 5–12
COMI	
Pain intensity	0.58 (1.32), 0–8
Disability	0.04 (0.34), 0–4
Back score	0.44 (1.00), 0–6.54
Subjective walking restriction	
Walking distance, m	989.0 (82.4), 100–1000
Walking duration, mins	58.7 (6.1), 9–60

hort, as well as for subcategories of age and sex, are listed in Table 3.

Reliability Measurements

The ICC for repeated app-based 6WD measurements was good to excellent ($\beta = 0.89$, 95% CI 0.87–0.91, $p < 0.001$) and the SEM was 34.1 meters (Table 4).

Content Validity

Figure 2 illustrates examples of the linear relationships between average 6WD and the PROMs, including VAS back pain (β coefficient -22.0 , 95% CI -30.8 to -13.1 , $p < 0.001$; Fig. 2A), COMI Back (β coefficient -30.9 , 95% CI -41.6 to -20.3 , $p < 0.001$; Fig. 2B), ZCQ symptom severity (β coefficient -15.3 , 95% CI -20.4 to -10.1 , $p < 0.001$; Fig. 2C), and the subjective walking duration (β coefficient 5.7 , 95% CI 4.0 – 7.4 , $p < 0.001$; Fig. 2D).

Linear relationships were likewise significant for all other PROMs, including VAS leg pain (β coefficient -23.6 , 95% CI -36.2 to -10.9 , $p < 0.001$; Supplemental Fig. 1A), the COMI subscores for pain intensity (β coefficient -21.6 , 95% CI -29.7 to -13.4 , $p < 0.001$; Supplemental Fig. 1B) and disability (β coefficient -77.4 , 95% CI -109.4 to -45.3 , $p < 0.001$; Supplemental Fig. 1C), and the ZCQ subscores for pain (β coefficient -20.7 , 95% CI -28.5 to -13.0 , $p < 0.001$; Supplemental Fig. 1D), neuroischemic symptoms (β coefficient -30.8 , 95% CI -44.0

TABLE 4. Intrarater measurements of reliability (ICC) and measurement error (SEM)

Measurement	Intrarater Value
ICC (range)	0.89 (0.87–0.91)
SEM, m	34.1

to -17.6 , $p < 0.001$; Supplemental Fig. 1E), and physical function (β coefficient -41.2 , 95% CI -54.0 to -28.4 , $p < 0.001$; Supplemental Fig. 1F), as well as for subjective walking distance (β coefficient 0.34 , 95% CI 0.21 – 0.47 , $p < 0.001$; Supplemental Fig. 1G).

Case Illustrations

Table 5 contains 6WT and PROM data (ZCQ and COMI Back) from 6 consecutive patients with degenerative diseases of the lumbar spine at baseline (preoperative) and 6 weeks and 3 months postoperatively.

Discussion

In this paper we present normative data for the 6WT, conveniently measured by a previously validated GPS-based smartphone app (6WT app),¹³ as well as for a set of commonly used PROMs of 330 subjects recruited at 1 of 8 centers from 5 countries. We furthermore provide reliability measures and report content validity for the 6WT in this population. In summary, the average 6WD was 586 meters with 95% of measurements ranging between 575 and 597 meters. The ICC of repeated measurements was good to excellent (0.89, 95% CI 0.87–0.91), indicating that one single measurement per patient and time point is sufficiently accurate. The SEM was 34 meters, indicating the “gray zone” of uncertainty for each measurement. Even in this sample of spine-healthy subjects, we found highly significant linear relationships with all applied PROMs, indicating that content validity of the app-based 6WT is given. The information reported in this article is essential for clinicians when interpreting results obtained in patients with degenerative diseases of the spine.

How to Use Normative Data for the Patient Evaluation?

The 95% CIs reported in Table 3 can be used as references to determine whether or not a patient with lumbar DDD is functionally impaired. We can use patient 1 from Table 5 as an example: a 41-year woman with a left paramedian LDH at L4–5 (Fig. 3), who reported mild to

TABLE 3. Normative values for the 6WD in a sample of spine-healthy volunteers, stratified according to age and sex categories

Age (yrs)	Males	Females	Total
<30, n = 118	626 ± 101 (601–651)	576 ± 70 (556–596)	604 ± 92 (587–621)
31–50, n = 70	649 ± 89 (615–682)	605 ± 75 (581–629)	624 ± 83 (604–643)
51–70, n = 120	585 ± 103 (555–615)	566 ± 85 (546–585)	573 ± 92 (556–590)
>70, n = 22	468 ± 117 (406–530)	360 ± 138 (215–505)	438 ± 130 (381–496)
Total	602 ± 112 (585–620)	571 ± 91 (557–584)	586 ± 103 (575–597)

Results are presented as mean ± SD (95% CI).

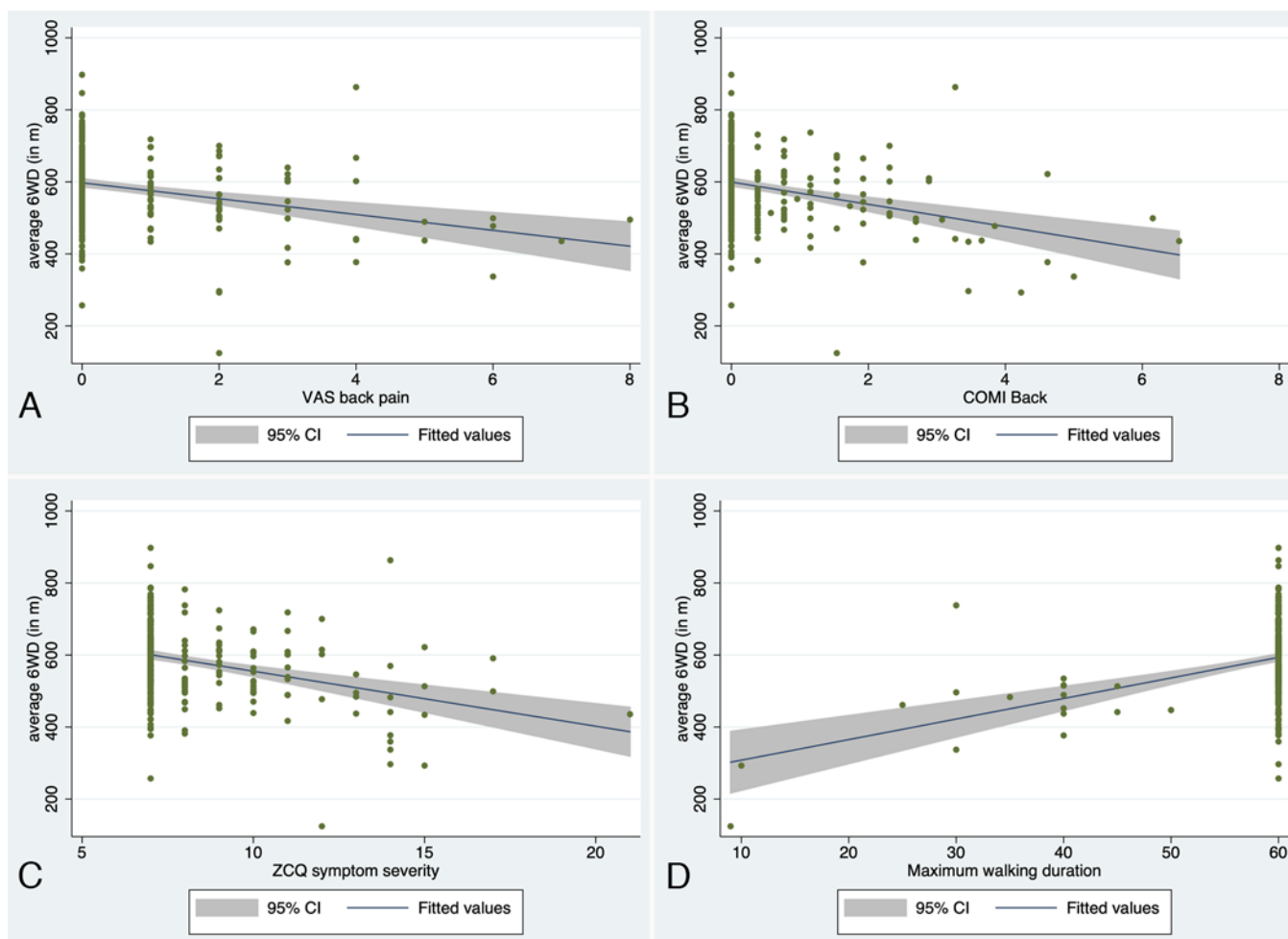


FIG. 2. Linear prediction plot (blue) with 95% CIs (gray) and overlaid scatterplot (green) illustrating agreement between PROMs (x-axis) and the average from three trials of app-based 6WD measurements (y-axis). **A:** For each 1-step increase in VAS back pain, subjects walked about 22 meters less (β coefficient -22.0 , 95% CI -30.8 to -13.1 , $p < 0.001$). **B:** For each 1-step increase in the COMI Back score, subjects walked about 31 meters less (β coefficient -30.9 , 95% CI -41.6 to -20.3 , $p < 0.001$). **C:** For each 1-step increase in the ZCQ symptom severity score, subjects walked about 15 meters less (β coefficient -15.3 , 95% CI -20.4 to -10.1 , $p < 0.001$). **D:** For each 1-minute increase in maximum walking duration, subjects walked about 5.7 meters more (β coefficient 5.7 , 95% CI 4.0 – 7.4 , $p < 0.001$). Figure is available in color online only.

moderate left radicular pain but severe disability related to a foot drop (BMRC grade 1/5). Her 6WD is 504 meters. With the 6WT's SEM of 34 meters, we know that her true 6WD ranges around 538–470 meters. As evident from Table 3, she scores well below the 95% CI of her age- and sex-adjusted normal population range (581–629 meters). We can be 95% certain that this patient has OFI.

Why Express OFI by z-Scores?

The concept of z- or t-scores helps to quantify the degree of OFI and to express disability in a standardized way and to adjust for factors that naturally influence the test result.^{4,7} Whereas z-scores express the number of SDs patients deviate from the mean of a defined population, t-scores are transformations of z-scores to numbers that are more convenient to calculate or easier to compare with other tests. A prominent example for their use is bone

density testing, in which test results are not expressed as “raw x-ray absorption” but as z-/t-scores. Here, the z-score shows how much a patient deviates from his or her age- and sex-adjusted reference population, and the t-score shows how much a patient deviates from the bone density of an average healthy 30-year-old adult. Both z- and t-scores are also routinely used in the evaluation of neuropsychological test results, where test results can be adjusted for variables such as a patient's age, sex, and the educational level. Moreover, when several tests (evaluating different domains of outcome) are conducted in the same patient, the use of z-scores allows for easy direct comparison of test results and for identification of deficits in a certain trait (e.g., memory, attention, executive functions, etc.).

Also, in the field of spine care, the standardized evaluation can help to reduce the bias introduced by age and sex.^{22,23} Unfortunately, however, the practical use of z-/t-scores is not yet established in the evaluation of patients

TABLE 5. Objective (6WT) raw results, OFI z-scores, and subjective data (PROMs, ZCQ, and COMI Back) from 6 consecutive patients with degenerative diseases of the lumbar spine (LDH or LSS) at baseline (preoperative) as well as 6 weeks and 3 months postoperatively

Patient No.	Age (yrs)/Sex/ Pathology	Preop	6 Wks Postop	3 Mos Postop
1	41/F/LDH	6WT: 504 m; OFI z-score: -1.35; ZCQ symptom severity: 24; ZCQ physical function: 14; COMI Back: 6.45	6WT: 532 m; OFI z-score: -0.98; ZCQ symptom severity: 14; ZCQ physical function: 6; COMI Back: 6.5	6WT: 580 m; OFI z-score: -0.33; ZCQ symptom severity: 14; ZCQ physical function: 5; COMI Back: 3.25
2	72/F/LSS	6WT: 281 m; OFI z-score: -0.57; ZCQ symptom severity: 32; ZCQ physical function: 14; COMI Back: 8.0	6WT: 384 m; OFI z-score: 0.18; ZCQ symptom severity: 9; ZCQ physical function: 6; COMI Back: 0.95	6WT: 399 m; OFI z-score: 0.28; ZCQ symptom severity: 15; ZCQ physical function: 9; COMI Back: 2.1
3	62/F/LDH	6WT: 524 m; OFI z-score: -0.49; ZCQ symptom severity: 19; ZCQ physical function: 9; COMI Back: 6.2	6WT: 639 m; OFI z-score: 0.86; ZCQ symptom severity: 19; ZCQ physical function: 5; COMI Back: 0.6	6WT: 623 m; OFI z-score: 0.68; ZCQ symptom severity: 15; ZCQ physical function: 8; COMI Back: 5.3
4	73/M/LSS	6WT: 268 m; OFI z-score: -1.71; ZCQ symptom severity: 19; ZCQ physical function: 15; COMI Back: 4	6WT: 579 m; OFI z-score: 0.95; ZCQ symptom severity: 16; ZCQ physical function: 9; COMI Back: 4.6	6WT: 614 m; OFI z-score: 1.25; ZCQ symptom severity: 19; ZCQ physical function: 13; COMI Back: 3.9
5	74/M/LSS	6WT: 513 m; OFI z-score: 0.38; ZCQ symptom severity: 19; ZCQ physical function: 9; COMI Back: 6.5	6WT: 501 m; OFI z-score: 0.28; ZCQ symptom severity: 15; ZCQ physical function: 7; COMI Back: 3.6	6WT: 551 m; OFI z-score: 0.71; ZCQ symptom severity: 16; ZCQ physical function: 7; COMI Back: 2.4
6	60/M/LSS	6WT: 552 m; OFI z-score: -0.32; ZCQ symptom severity: 17; ZCQ physical function: 9; COMI Back: 5.1	6WT: 750 m; OFI z-score: 1.60; ZCQ symptom severity: 13; ZCQ physical function: 6; COMI Back: 2.1	6WT: 770 m; OFI z-score: 1.79; ZCQ symptom severity: 9; ZCQ physical function: 5; COMI Back: 1.5

with spinal diseases, mostly because normative data and tools for calculation are scarce.^{4,7} As an example, again consider the 41-year-old woman with LDH (Table 5, patient 1). She walked 504 meters at the preoperative assessment, which is difficult to interpret without normative data. From the raw result, it essentially remains unclear whether it is indicative of disability. With the required values for μ and σ provided in Table 3, the z-score can be calculated as $(504 - 605)/75 = -1.35$ using the formula provided in the *Methods* section. We now understand that the patient's test result lies 1.35 SDs below her age- and sex-adjusted reference population. Even 6 weeks after surgery, she still scores about 1 SD below the mean test result of her reference population (Tables 3 and 5), which is likely related to the persistent motor deficit at that time point (Fig. 3). At 3 months postoperatively, her 6WT result falls within the 95% CI of her reference population and her z-score of -0.33 confirms her trend toward normalization (Tables 3 and 5). At this time, her foot drop has partially recovered (Fig. 3).

Another scenario can be used to underscore the value of z-scores: an imaginary 85-year-old female patient presents to a clinic with an MR image demonstrating severe concentric LSS at L3-4. Her 6WD is also 504 meters, similar to the result scored by the 41-year-old LDH patient above. While in the younger patient a raw test result of 504 meters was indicative of significant morbidity (z-score = -1.35), the same 504 meters represents a test result that lies 1 SD above the reference population [z-score = $(504 - 360)/138$, or 1.04]. This 85-year-old patient has no signs of OFI and should possibly be motivated to continue conservative treatment.

Furthermore, compared to the standardized z-scores, it is more difficult to comprehend a patient's well-being by raw test result of the PROMs (ZCQ and COMI Back;

Table 5). Use of the free 6WT app obviates the need to manually calculate z-scores when a patient's age and sex are indicated (Fig. 1B and C). A further case vignette is presented in Fig. 4.

Why Use OFI in Daily Clinical Patient Care?

We and other groups have repeatedly demonstrated that objective outcome measures should rather complement than substitute for PROMs, because they provide complementary information.^{2,7-9,12,24,25} However, there are certain clear advantages of objective outcome measures. First, the use of objective outcome measures is on the increase in both clinical application and research.^{2,3,12,26-28} Second, patients consistently prefer physical tests over questionnaire-based evaluations, particularly if repeated assessments are required.²⁹ Third, self-measurements are possible in the patient's home environment, thereby empowering patients to monitor their own outcomes and be active partners in healthcare.^{3,30} It was reported that, compared to PROMs, objective functional tests such as the TUG test better account for disability resulting from lower-extremity motor deficits.³¹ Also, we noticed that patients consistently overrate their functional recovery on subjective PROMs in the early postoperative period, whereas the objective assessment lags behind and may reflect the situation with approach-related muscle and wound pain more realistically.^{3,7,32}

While PROMs are influenced by a patient's mental health condition, objective functional tests such as the TUG test seem to be more independent from depressive comorbidity.³³ As evident in Supplemental Table 1, depressive comorbidity showed no significant influence on the 6WT ($\eta^2 = 0.005$, $p = 0.206$). There are clinical scenarios in which the repeated use of objective functional tests can be helpful to demonstrate recovery, which may not be evi-



FIG. 3. Case vignette of a 41-year-old female nonsmoking patient (BMI 30.4 kg/m²), who presented to the hospital with a history of left L5 radicular pain for 2 weeks, which had resolved almost completely (VAS score 1/10) at the time of presentation, but with the appearance of a severe motor deficit (MRC grade 1/5) of the left foot dorsiflexion. Her disability measures were high, related to the paresis (COMI Back score 6.45, ZCQ symptom severity score 24, ZCQ physical function score 14). The average 6WD in three trials of the app-based 6WT was 504 meters, translating into an age- and sex-adjusted OFI z-score of -1.35. The lumbar MRI (*upper*: sagittal T2-weighted; *lower*: axial T2-weighted) revealed a large left paramedian disc herniation at the L4–5 level with compression of the left L5 nerve root. The patient underwent microsurgical sequestrectomy without complications. At 6 weeks postoperatively, she had a VAS score of 0/10 for back and leg pain, but persistent disability related to the MRC grade 1/5 left-foot paresis, which had not improved (COMI Back score 6.5, ZCQ symptom severity score 14, ZCQ physical function score 6). At this time, the change in 6WD was within the range of the SEM (532 meters, OFI z-score -0.98). At 12 weeks postoperatively, some improvement in the paresis grade was evident (MRC 3/5) and the 6WD improved to 580 meters (OFI z-score -0.33). The PROMs improved to COMI Back score 3.25, ZCQ symptom severity score 14, ZCQ physical function score 5.

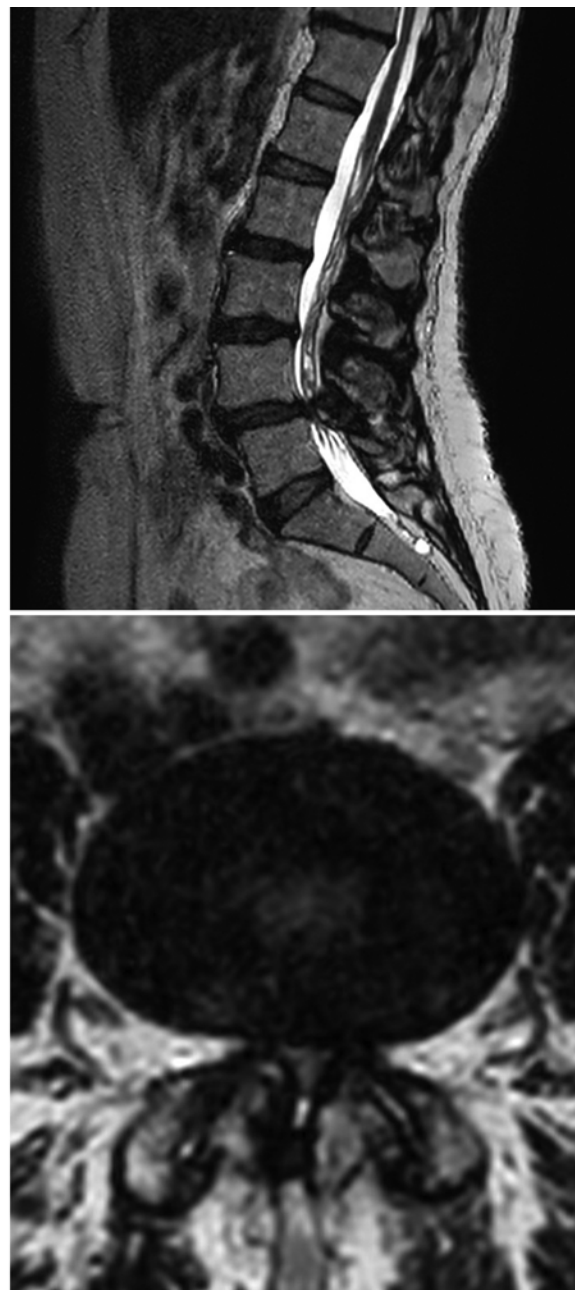


FIG. 4. Case vignette of a 59-year-old male nonsmoking patient (BMI 25.0 kg/m²), who reported progressive neurogenic claudication over the last 6 months. He had full strength in all lower-extremity myotomes. His back pain was 2/10 and his leg pain 3/10 on the VAS. His disability measures were COMI Back score 5.1, ZCQ symptom severity score 17, and ZCQ physical function score 9. The average 6WD in three trials of the app-based 6WT was 552 meters, translating into an age- and sex-adjusted OFI z-score of -0.32. The lumbar MRI (*upper*: sagittal T2-weighted; *lower*: axial T2-weighted) revealed a concentric spinal stenosis at the L4–5 level with compression of both L5 nerve roots. The patient underwent microscopic lumbar decompression at the L4–5 level without complications. Around 6 weeks postoperatively, he had almost no pain (VAS back score 1/10, VAS leg score 0/10) and his PROM values improved (see Table 5, patient 6). His walking ability had improved beyond the normative population values (raw 6WD 750 meters, OFI z-score 1.60) with slight further improvement until the 3-month follow-up (raw 6WD 770 meters, OFI z-score 1.79).

dent to patients and physicians on purely subjective PROM metrics.³³ For example, in patient 4 (Table 5) no significant improvement on either the ZCQ or the COMI Back was evident when comparing preoperative and 3-month postoperative PROM results. However, the objective functional 6WT assessment demonstrated a continuous and substantial improvement in walking function from preoperative (6WD 268 meters, z-score -1.71) over 6 weeks (6WD 579 meters, z-score 0.95) to 3 months postoperatively (6WD 614 meters, z-score 1.25).

Strengths and Limitations

We used a predefined protocol to collect normative population data in a sample of 330 subjects, enrolled at 8 sites in 5 countries, which is likely to allow for better generalization of the result to other linguistic, ethnic, and cultural settings. By including spine-healthy visitors, family members, and next of kin at the hospital, we were able to include a sample that shares important epidemiological features with our spine patients.

Even though we arrived at a sample size that exceeds the ones used in previous similar studies by far,^{4,7,12,22,25,33} including even more subjects—in particular, elderly subjects > 70 years of age—would have been desired to allow for even more accurate interpretation of test results in elderly patients. As evident from Supplemental Table 1, further variables including body dimensions and comorbidities may significantly influence the 6WD. Because of collinearity between age and some significant variables (e.g., working status, comorbidity) and for practical reasons (variables not always available), we decided not to adjust for each of those, which would also require a much greater sample. Moreover, walking speed and thus 6WD may be influenced by ethnicity and culture. While our results are most representative for Europe and North America, more normative data from Australasia, South America, and Africa are warranted. We used clinical examination but no imaging-based protocols to exclude subjects with comorbidities and diseases affecting their ability to walk, such as knee/hip osteoarthritis, and cervical and/or thoracic myelopathy. Lastly, we determined content validity in spine-healthy subjects, but further validation in the target population with lumbar DDD is required. Research is ongoing in this direction (<http://clinicaltrials.gov> identifiers NCT03977961 and NCT04062942).

Conclusions

In this study we provided normative data for app-based 6WD measurements in a multicenter sample from 8 institutions and 5 countries. These values can now be used as reference to compare 6WT results and quantify objective functional impairment in patients with degenerative diseases of the spine using z-scores. We found good to excellent test-retest reliability of the 6WT app, a low area of uncertainty, and high content validity of the average 6WD with commonly used PROMs.

Acknowledgments

We would like to thank all subjects who agreed to participate and provide data for this research project. Dr. Stienen thanks the

AO Spine Foundation for the AOSEU Young Researcher Award that helped cover some of this project's expenses.

Appendix

The 6WT app is available in the English, French, and German languages. It can be downloaded free of charge for iOS (Apple App Store) and Android (Google Play Store) smartphones. Installation on a patient's smartphone takes less than 1 minute (feasible during outpatient clinic visits).

References

- Gautschi OP, Corniola MV, Schaller K, et al. The need for an objective outcome measurement in spine surgery—the timed-up-and-go test. *Spine J*. 2014;14(10):2521–2522.
- Stienen MN, Ho AL, Staartjes VE, et al. Objective measures of functional impairment for degenerative diseases of the lumbar spine: a systematic review of the literature. *Spine J*. 2019;19(7):1276–1293.
- Stienen MN, Rezaii PG, Ho AL, et al. Objective activity tracking in spine surgery: a prospective feasibility study with a low-cost consumer grade wearable accelerometer. *Sci Rep*. 2020;10(1):4939.
- 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(2):270–278.
- 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(3):380–385.
- Stienen MN, Maldaner N, Joswig H, et al. Objective functional assessment using the “Timed Up and Go” test in patients with lumbar spinal stenosis. *Neurosurg Focus*. 2019;46(5):E4.
- 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(5):598–604.
- Anderson DB, Mathieson S, Eyles J, et al. Measurement properties of walking outcome measures for neurogenic claudication: a systematic review and meta analysis. *Spine J*. 2019;19(8):1378–1396.
- Smuck M, Muaremi A, Zheng P, 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(1):15–21.
- Alves VL, Avanzi O. Objective assessment of the cardiorespiratory function of adolescents with idiopathic scoliosis through the six-minute walk test. *Spine (Phila Pa 1976)*. 2009;34(25):E926–E929.
- Försth P, Ólafsson G, Carlsson T, et al. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. *N Engl J Med*. 2016;374(15):1413–1423.
- Loske S, Nüesch C, Byrnes KS, et al. Decompression surgery improves gait quality in patients with symptomatic lumbar spinal stenosis. *Spine J*. 2018;18(12):2195–2204.
- Stienen MN, Gautschi OP, Staartjes VE, et al. Reliability of the 6-minute walking test smartphone application. *J Neurosurg Spine*. 2019;31(6):786–793.
- Stucki G, Daltroy L, Liang MH, et al. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. *Spine (Phila Pa 1976)*. 1996;21(7):796–803.
- NASS Evidence-Based Clinical Guidelines Committee. *Diagnosis and Treatment of Degenerative Lumbar Spinal Stenosis*. North American Spine Society; 2011.
- Mannion AF, Elfering A, Staerke R, et al. Outcome assessment in low back pain: how low can you go? *Eur Spine J*. 2005;14(10):1014–1026.

17. Pierce CA, Block RA, Aguinis H. Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educ Psychol Meas*. 2004;64(6):916–924.
18. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30–46.
19. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428.
20. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163.
21. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther*. 1997;77(7):745–750.
22. 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(5):1065–1071.
23. 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;150:33–39.
24. Gautschi OP, Corniola MV, Joswig H, et al. The timed up and go test for lumbar degenerative disc disease. *J Clin Neurosci*. 2015;22(12):1943–1948.
25. Staartjes VE, Schröder 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;29(4):380–387.
26. Bostelmann R, Schneller S, Cornelius JF, et al. 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(3):963–968.
27. Conway J, Tomkins CC, Haig AJ. Walking assessment in people with lumbar spinal stenosis: capacity, performance, and self-report measures. *Spine J*. 2011;11(9):816–823.
28. Smeets RJ, Hijdra HJ, Kester AD, et al. The usability of six physical performance tasks in a rehabilitation population with chronic low back pain. *Clin Rehabil*. 2006;20(11):989–997.
29. 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.
30. Staartjes VE, Beusekamp F, Schröder ML. Can objective functional impairment in lumbar degenerative disease be reliably assessed at home using the five-repetition sit-to-stand test? A prospective study. *Eur Spine J*. 2019;28(4):665–673.
31. Stienen MN, Maldaner N, Sosnova M, et al. Lower extremity motor deficits are underappreciated in patient-reported outcome measures: added value of objective outcome measures. *Neurospine*. 2020;17(1):270–280.
32. Gautschi OP, Joswig H, Corniola MV, et al. Pre- and post-operative correlation of patient-reported outcome measures with standardized Timed Up and Go (TUG) test results in lumbar degenerative disc disease. *Acta Neurochir (Wien)*. 2016;158(10):1875–1881.
33. 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(6):807–813.

Disclosures

Dr. Desai reports being a consultant to DePuy Synthes and Stryker.

Author Contributions

Conception and design: Stienen, Regli. Acquisition of data: Stienen, Tosic, Goldberger, Maldaner, Sosnova, Zeitlberger, Staartjes, Gadradj, Eversdijk, Quddusi, Gandía-González, Sayadi, Gautschi. Analysis and interpretation of data: Stienen, Tosic, Goldberger, Maldaner. Drafting the article: Stienen, Tosic, Goldberger. Critically revising the article: Maldaner, Sosnova, Zeitlberger, Staartjes, Gadradj, Eversdijk, Quddusi, Gandía-González, Sayadi, Desai, Regli, Gautschi. Approved the final version of the manuscript on behalf of all authors: Stienen. Statistical analysis: Stienen. Administrative/technical/material support: Stienen, Regli, Gautschi. Study supervision: Stienen, Regli, Gautschi.

Supplemental Information

Online-Only Content

Supplemental material is available with the online version of the article.

Supplemental Table 1 and Figs. 1 and 2. <https://thejns.org/doi/suppl/10.3171/2020.3.SPINE2084>.

Correspondence

Martin N. Stienen: University Hospital Zurich & Clinical Neuroscience Center, University of Zurich, Switzerland. mnstienen@gmail.com.