Eur Spine J (2017) 26:450–461
DOI 10.1007/s00586-016-4587-0

ORIGINAL ARTICLE

A comparison between different outcome measures based on "meaningful important differences" in patients with lumbar spinal stenosis

Maria M. Wertli[1,2,3] · Franziska Christina Buletti[1] · Ulrike Held[1] · Eva Rasmussen-Barr[2,4] · Sherri Weiser[2] · Jakob M. Burgstaller[1] · Johann Steurer[1]

Received: 13 October 2015 / Revised: 27 April 2016 / Accepted: 27 April 2016 / Published online: 13 May 2016

Abstract
Purpose Patient-reported outcome measures (PROM) are used to measure treatment efficacy in clinical trials. The impact of the choice of a PROM and the cut-off values for 'meaningful important differences' (MID) on the study results in patients with lumbar spinal stenosis (LSS) is unclear.
Objective The objective is to study the consequences of applying different PROMs and values for MID for pain and disability on the proportions of patients with improvement.
Design Prospective multi-center cohort study.
Methods Proportions of patients with improvement using established MID cut-off values were calculated and compared for PROMs for pain and disability.
Results 466 patients with LSS completed a baseline and 6-month follow-up assessment and were analyzed. Treatment modalities included surgery (65 %), epidural steroid injections (15 %), or conservative care (20 %). The prevalence of patients fulfilling the criteria for MID ranged from 40 to 70 % across all outcome measures and cut-offs. The agreement of the spinal stenosis outcome measure (SSM) symptom subscale with other pain scales, and the SSM function subscale with other function scales was fair to moderate (Cohen's κ value between 0.24 and 0.5). Disagreement in the assessment of MID (MID reported by patients in one scale but not the other) was found in at least one-third of the patients.
Conclusion The MID in outcome scores for this population varied from 40 to 70 %, depending on the measure or cut-off score used. Further, the disagreement between domain specific measures indicates that differences between studies may be also related to the choice of an outcome measures. An international consensus on the use and reporting of outcome measures in studies on lumbar spinal stenosis is needed.

Keywords Spinal stenosis · Lumbar spinal stenosis · Outcome measures · Patient-reported outcome measures

Introduction

Patients and physicians often have difficulties determining the clinical relevance of a treatment effect quantified as a patient-reported outcome measure (PROM). The clinical relevance, for example, of a mean increase of five points in quality of life after surgery on a scale from 0 to 100 is difficult to interpret. To facilitate the interpretation of PROM, the concept of "minimal clinically important difference" (MCID), also known as "minimal important difference" (MID), has been introduced [1]. The minimal important difference refers to "the smallest amount of benefit a patient can recognize and value" [1]. By applying MID to the analysis of outcome data from a clinical trial, the proportion of patients with an improvement—at least in the magnitude of MID—can be calculated. Studies

✉ Maria M. Wertli
Maria.Wertli@usz.ch

1 Horten Centre for Patient Oriented Research and Knowledge Transfer, Department of Internal Medicine, University of Zurich, Pestalozzistrasse 24, 8032 Zurich, Switzerland

2 NYU Hospital for Joint Diseases, Occupational and Industrial Orthopaedic Center (OIOC), New York University, 63 Downing Street, New York, NY 10014, USA

3 Department of General Internal Medicine, Inselspital, Bern University Hospital, 3010 Bern, Switzerland

4 Institute of Environmental Medicine, Karolinska Institutet, Box 210, 17177 Stockholm, Sweden

reporting the proportion of meaningfully improved patients in the treatment group, compared to placebo or other treatment, are valuable for informing patients about treatment effects.

In patients with spinal lumbar stenosis, various outcome measures are used to quantify treatment efficacy. For example, the spinal stenosis measure (SSM) includes two subscales, one to quantify pain and the other physical function [2]. Other commonly used instruments to measure outcomes in patients with lumber stenosis are: the numeric rating scale (NRS), the Roland Morris questionnaire (RMQ) [3], the Oswestry disability index (ODI) [4], the Oxford spinal stenosis score [4], and the lumbar spinal stenosis-specific symptom scale [5], and health measures including the EuroQol [6]. One would expect that two instruments that are valid to measure pain would be similarly sensitive to change and a high agreement between the proportions of patients with MID can be found for instruments that measure the same domain. Today, there is insufficient evidence available as to whether this is the case. The reported proportion of patients with MID may vary according to the outcome measure utilized. Furthermore, different methods are used to establish MID cut-off values for a scale. The impact of various cut-off values—to categorize patients with meaningful improvement (MID) or no meaningful improvement—on study outcomes is unclear. To date, there is no consensus on how to assess and report treatment outcome in patients undergoing treatment for lumbar spinal stenosis. Furthermore, it is unclear whether the studies that use different outcome measures can be compared.

The objective of this study was to assess the agreement of domain specific outcome measures for pain and disability in patients undergoing treatment for lumbar spinal stenosis. Further, we assessed the impact of different cut-off values proposed for MID on the proportion of improved patients treated for lumbar spinal stenosis.

Method

This research is part of a multi-center prospective cohort study in Switzerland investigating the prognosis of patients with lumbar spinal stenosis treated with or without surgery [7]. The study was approved by the local ethical committee and conducted in accordance with the Declaration of Helsinki [8]. All patients received written and oral information about the study and gave their written consent to participate.

Eligibility criteria and patients

Patients were recruited during consultations in the Rheumatology and Spine Surgery Units in eight hospitals located in the Cantons of Zurich and Lucerne, Switzerland. Inclusion criteria were: (1) age $\geq$50 years; (2) uni- or bilateral neurogenic claudication (defined by pain in the buttocks and/or lower extremities provoked by walking or extended standing and relieved by rest and/or bending forward); (3) verified spinal stenosis (central or lateral verified by magnetic resonance imaging or computer tomography); (4) anticipated life expectancy more than 1 year; (5) able to give informed consent; (6) available for follow-up; and (7) able to complete questionnaires in German. For the current study, all consecutive patients who completed data at baseline and 6 months' follow-up were included.

Exclusion criteria were: the presence of red flags (e.g., cauda equina syndrome, infection), current vertebral fracture, significant deformity (>15° lumbar scoliosis), or clinically relevant peripheral arterial disease (confirmed by a vascular specialist).

Procedure and measurements

All patients participating in the prospective cohort study received a set of questionnaires after agreeing to participate and signing the informed consent. They completed self-reported baseline information about socio-demographic characteristics, symptoms and returned the questionnaires by mail. After inclusion in the study, the patients were contacted by the study coordinator for a clinical examination, an interview on comorbidities and previous treatments for lumbar spinal stenosis received within the previous 6 months. The treatment received was at the discretion of the treating physician. After 6 months a set of questionnaires was sent for the follow-up evaluation and returned by mail.

PROM for pain

The spinal stenosis measure (SSM) measure is a disease-specific evaluation tool for patients with lumbar spinal stenosis that assesses symptoms and physical function [2, 4]. The scale is also known as the spinal stenosis measure, the Zurich Claudication questionnaire, or the Brigham spinal stenosis questionnaire. It is a self-administered, reliable, valid, and internally consistent questionnaire that is responsive to clinical change and has been validated in English [2, 4] and other languages [9–11]. The German version has been shown to be reliable and valid with a Cronbach's alpha for the SSM Sy of 0.83, the SSM F 0.86, and the SSM Sat. 0.87 [11]. The three subscales are the SSM symptom scale (SSM Sy, seven items), the SSM physical function scale (SSM F, five items), and the SSM satisfaction scale (SSM Sat., six items). Each item is rated on a Likert scale. The SSM Sy measures pain over the prior

month including pain location, pain frequency, and neurological disturbances on a scale from 1 (no) to 5 (very severe symptoms) points.

The numeric rating scale (NRS) measures pain intensity during the previous 7 days on a scale of 0 (no pain) to 10 (worst possible pain) [12, 13]. The question was framed as follows: during the past 7 days on average, how strong was your pain? How intensive was your back or leg pain during walking and activity?

The feeling thermometer (range 0–100) assesses the impact of the current complaint over the previous 7 days. The question is framed so that it is ambiguous whether pain or function is being assessed [14]. It was, therefore, compared to scales measuring pain, disability, and the SSM sum score.

PROM for disability

The SSM F subscale measures physical function during the prior month in six items on a scale from 1 (yes, comfortably) to 5 (no, could not perform). The item addresses walking distance in different settings including walk for pleasure, for shopping, and for getting around the house. The physical function scale score was calculated from the unweighted mean of all answered items with the possible range of scores between 1 and 4.

The Roland Morris disability questionnaire (RMQ, range 0–24) was developed to assess functional disability in back pain patients [15–17]. The RMQ is frequently used to assess outcome in lumbar spinal stenosis studies [18, 19]. A limited range of physical functions, including walking, bending over, sitting, lying down, dressing, sleeping, self-care, and daily activities are assessed and the RMQ correlates well with measures of physical function [20]. The questionnaire assesses disabilities because of the back problem, e.g., "Because of the pain in my back, I lie down to rest more often" or "I get dressed more slowly than usual because of the pain in my back". The RMQ has been shown to be reliable and consistent in assessing back pain populations with a high Cronbach's alpha between 0.84 and 0.93 [20].

General PROM

We compared MID in both the SSM Sy and SSM F subscale [2, 4] to MID in the feeling thermometer. The questions posed by the feeling thermometer are framed in such a way that it is open whether pain or function is being assessed. We analyzed the agreement between the feeling thermometer and a combination of the SSM F and SSM Sy.

Additional measures

The SSM Sat subscale measures satisfaction with the operation, with pain relief, with the ability to walk and perform everyday activities, and assesses neurological improvement on a 1 (very satisfied) to 4 point scale (very dissatisfied). We calculated the percentage of satisfied patients for each SSM subscale. Stucki et al. [2] used the SSM Stat subscale to derive the MID cut-off for the SSM Sy and SSM F (described below). Patients were classified as "satisfied" when they reported 1.0–2.0 points. Patients with >2 points in the SSM Sat. subscale were considered "not satisfied".

Meaningful important difference

Two studies evaluated the values for MID in the SSM. Stucki et al. [2] used an anchor-based approach using the SSM Sat subscale. MID in the SSM Sy and SSM F were based on the difference in mean change in the patients who were satisfied (1.0–2.0 points on a 1.0–4.0 point scale) and patients who were somewhat/not satisfied (SSM Sat >2.0–4.0 points). MID derived with this approach was

Table 1 Baseline characteristics

	All	Conservative	Injection	Surgery
n (%)	466 (100)	93 (20)	71 (15)	302 (65)
Gender: male/female	223/243	40/53	29/42	154/148
Age: median (IQR)	75 (67–80)	75 (69–81)	75 (67–80)	74 (67–79)
SSM Sy: median (IQR)	3.1 (2.7–3.6)	3.0 (2.4–3.8)	2.9 (2.6–3.5)	3.1 (2.7–3.6)
Neuroischemic: median (IQR)	2.5 (2.0–3.3)	2.5 (1.8–3.0)	2.5 (2.0–3.1)	2.75 (2.25–3.25)
Pain: median (IQR)	3.7 (3.3–4.3)	3.8 (3.0–4.3)	3.7 (3.3–4.3)	4.0 (3.3–4)
SSM F: median (IQR)	2.2 (1.8–2.8)	2.2 (1.8–3.2)	2.0 (1.5–2.6)	2.4 (1.8–2.8)
NRS: median (IQR)	7.0 (5.0–8.0)	6.0 (3.0–8.0)	6.0 (4.0–8.0)	7.0 (5.0–8.0)
RMQ: median (IQR)	13.0 (8.0–16.0)	10 (6–15)	11.0 (5.5–14.0)	14 (9–16)
FT: median (IQR)	65.0 (50.0–80.0)	50 (35–80)	58 (44–75)	70 (50–80)

SSM Sy Spinal Stenosis Measure subscale [range 1 (none)–5 (very severe)], *SSM F* SSM function subscale (range 1–4), *NRS* numeric rating scale (range 0–10), *RMQ* Roland Morris Disability Questionnaire (range 0–24), *FT* feeling thermometer (range 0–100)

Table 2 Meaningful important differences (MID) for all scales

Scale (MID)	Yes	No
SSM Sy (0.48)[a]	266 (57)	200 (43)
SSM Sy (0.36)[b]	303 (65)	163 (35)
SSM Sy (30 %)	185 (40)	281 (60)
SSM F (0.52)[a]	232 (50)	234 (50)
SSM F (0.1)[b]	328 (70)	138 (30)
SSM F (30 %)	191 (41)	275 (59)
NRS (≥2 points decrease; $n = 449$)	280 (63)	169 (37)
NRS (30 %)	259 (56)	207 (44)
FT (≥15 points decrease; $n = 449$)	278 (62)	171 (38)
FT (30 %)	257 (55)	209 (45)
RMQ (≥5 points; $n = 417$)	167 (40)	250 (60)
RMQ (30 %)	205 (44)	261 (56)

MID for the NRS, the feeling thermometer (FT) and RMQ according to Ostelo et al. [17]

SSM scale Spinal Stenosis Measure, *SSM Sy* SSM symptom subscale, *SSM F* SSM function subscale, *MID* meaningful important differences

[a] MID SSM Sy and function subscale according to Stucki et al. [2]: SSM F 0.52 points, SSM Sy 0.48 points

[b] MID according to Cleland et al. [3]: SSM F of 0.1 points, SSM Sy 0.36 points

in the SSM F 0.52 points and in the SSM Sy 0.48 points [3]. Using a ROC where the curve nearest the upper left-hand corner was used as cut-off value, Cleland et al. [3] developed MID in the SSM F of 0.1 points and in the SSM Sy of 0.36 points.

To demonstrate the impact of different MID on treatment success, we report both MID values for SSM Sy and SSM F. We also report results for an MID of 30 % change corresponding to the proposed 30 % change for the RMQ, NRS, and feeling thermometer. MID for the NRS, the feeling thermometer, and the RMQ were based on the work of Ostelo et al. [17]. We used the proposed 30 % change for all three outcome parameters. Further, results were reported for the absolute MID of 2 points in the NRS, 15 points in the feeling thermometer, and 5 points in the RMQ [17].

Statistics

For continuous data, median and interquartile ranges are given. Proportions of MID at 6 months' follow-up were calculated for all scales and compared between scales. To quantify the magnitude of agreement between reported MID for different scales, we used the kappa (κ) statistic.

Table 3 Agreement between MID in the SSM symptom subscale and in other pain measures

Scale 1 (MID)/scale 2 (MID)	Prevalence MID scale 1 %	Yes/yes: *n* (%)	Yes/no: *n* (%)	No/yes: *n* (%)	No/no: *n* (%)	κ
SSM Sy (0.48)/NRS (30 %)	57	205 (44)	61 (13)	54 (12)	146 (31)	0.5
SSM Sy (0.48)/NRS (≥2 points decrease; $n = 449$)[a]		213 (47)	49 (11)	67 (15)	120 (27)	0.46
SSM Sy (0.48)/FT (30 %)		202 (43)	64 (14)	55 (12)	145 (31)	0.48
SSM Sy (0.48)/FT (≥15 points decrease; $n = 449$)[b]		208 (46)	50 (11)	70 (16)	121 (27)	0.45
SSM Sy (0.48)/satisfied patients ($n = 459$)[c]		213 (47)	50 (11)	98 (21)	98 (21)	0.32
SSM Sy (0.36)/NRS (30 %)	65	216 (46)	87 (19)	43 (9)	120 (26)	0.42
SSM Sy (0.36)/NRS (≥2 points decrease; $n = 449$)[a]		227 (51)	71 (16)	53 (12)	98 (22)	0.4
SSM Sy (0.36)/FT (30 %)		213 (46)	90 (19)	44 (9)	119 (26)	0.41
SSM Sy (0.36)/FT (≥15 points decrease; $n = 449$)[b]		222 (49)	72 (16)	56 (13)	99 (22)	0.38
SSM Sy (0.36)/satisfied patients ($n = 459$)[c]		229 (50)	71 (15)	82 (18)	77 (17)	0.25
SSM Sy (30 %)/NRS (30 %)	40	164 (35)	21 (5)	95 (20)	186 (40)	0.51
SSM Sy (30 %)/NRS (≥2 points decrease; $n = 449$)[a]		164 (36)	17 (4)	116 (26)	152 (34)	0.44
SSM Sy (30 %)/FT (30 %)		162 (35)	23 (5)	95 (20)	186 (40)	0.50
SSM Sy (30 %)/FT (≥15 points decrease; $n = 449$)[b]		158 (35)	19 (4)	120 (27)	152 (34)	0.41
SSM Sy (30 %)/satisfied patients ($n = 459$)[c]		168 (37)	16 (3)	143 (31)	132 (29)	0.35

SSM scale Spinal Stenosis Measure, *SSM Sy* SSM symptom subscale, *SSM F* SSM function subscale, *MID* meaningful important differences, *FT* feeling thermometer, *NRS* numeric rating scale

[a] NRS 17 patients excluded from the analysis because the baseline NRS was <2

[b] FT 17 patients excluded from the analysis because the baseline FT was <15 points)

[c] Satisfied patients defined according to the definition of Stucki et al. SSM satisfaction 1.0–2.0 points [2], not satisfied include all patients reporting in the SSM satisfaction of >2.0 points (range 1–4 points)

The κ statistic indicates the proportion of agreement beyond that expected by chance [21]. Agreement is: less than chance (<0), slight (0.01–0.20), fair (0.21– 0.40), moderate (0.41–0.60), substantial (0.61–0.80), or (almost) perfect (0.81–1.00). High and low prevalence of positive findings may influence the κ value [22]. The influence of the prevalence on the κ values was low (correlation = −0.39). The prevalence rates may influence the κ value [23]. Therefore, the influence of the prevalence on the κ values was assessed visually and using correlation analysis. The correlation was low ($r = -0.39$), and therefore, did not require addition adjustment. This may be because the observed prevalence was close to 50 % [23].

All analyses were conducted with the statistical software R [24].

Ethics

This cohort study was conducted in compliance with all international laws and regulations as well as any applicable guidelines. The study was approved by the independent Ethics Committee of the Canton Zurich (KEK-ZH-NR: 2010-0395/0).

Results

The LSOS is an ongoing observational cohort study in patients with symptomatic lumbar spinal stenosis. For this analysis, patients with a complete set of baseline and follow-up questionnaire at 6 months were included. By December 2014, 1315 patients were potentially eligible and 704 patients (100 %) agreed to participate in the study ("Appendix 1" study flow). For the current analysis, a full set of questionnaires at baseline and 6-month follow-up was available in 466 patients (66 %) and included in this analysis. The reasons for non-inclusion were lost to follow-up (48 patients, 7 %), not completed 6 months' follow-up (135 patients, 19 %), and incomplete questionnaires (55 patients, 8 %). During the 6-month treatment period, 302 patients (65 %) received surgical treatment, 71 patients (15 %) had epidural steroid injections, and 93 patients (20 %) received conservative care. Table 1 summarizes the baseline characteristics of the study population. The median age was 75 years (IQR 67–80), and 52 % were women.

Table 2 summarizes the percentage of patients with MID for each scale after 6 months. Depending on the cut-off values used, the proportions of patients fulfilling the MID criteria ranged in the different scales from 40 to 70 %. In the SSM Sy scale, for example, the proportions of patients fulfilling MID criteria were highest for the cut-off value of 0.36 points (by Cleland et al. [3], 65 %) and decreased with the cut-off of 0.48 (by Stucki et al. [2], 57 %), and the 30 % (relative improvement) cut-off (40 %). When comparing the RMQ and the SSM F, MID criteria in the SSM F was fulfilled depending on the cut-offs used in between 41 and 71 % and in the RMQ, 40 and 44 %. Comparing the proportion of patients fulfilling MID criteria based on absolute values to MID cut-off criteria of 30 %, a lower proportion fulfilled MID criteria across all scales. Out of 466 patients, between 39 % (cut-off by Stucki et al. [2]) and 56 % (cut-off by Cleland [3]) reported MID in both the SSM Sy and SSM F scale (a summary of the results for the SSM Sy and SSM F is included in "Appendix 2").

The agreement between MID for scales assessing pain is summarized in Table 3. The agreement between the pain measures NRS and SSM Sy was moderate (κ value between 0.4 and 0.51). The proportion of patients with MID reported in the SSM Sy and NRS ranged from 35 to 51 %. Disagreement in MID between the SSM Sy and NRS was found in about one-third of patients (MID was reported in one scale but not the other). Figure 1 depicts how often the estimated results are consistent. The agreement

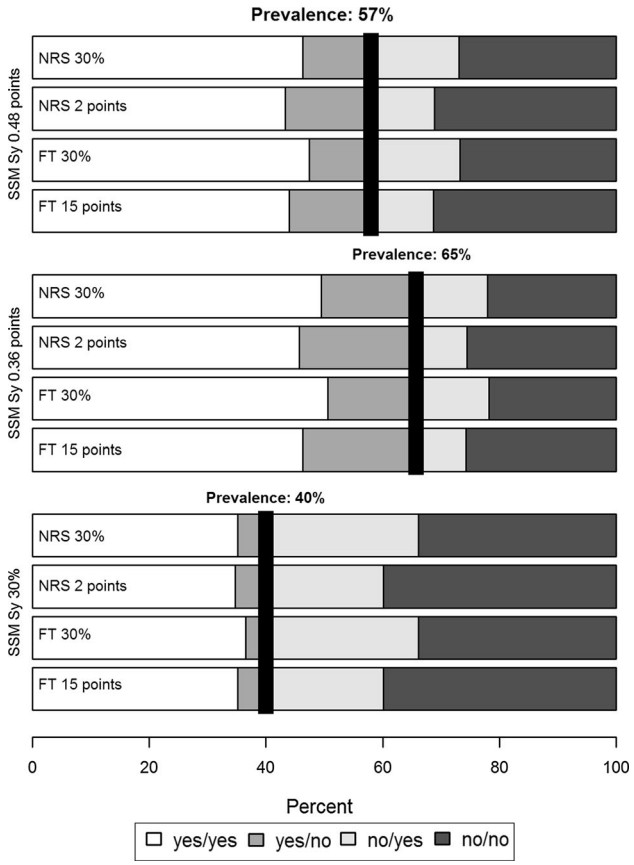

Fig. 1 Agreement between MIDs of scales assessing pain in patients with lumbar spinal stenosis. *Prevalence value* indicates the proportion of patients with MID in the SSM Sy scale; *yes/yes* and *no/no* represent the proportion of agreement on a "true" status (MID and no MID). No/yes and yes/no are subjects with a true positive or negative status in one scale but not in the other and the "true" status is difficult to determine

Table 4 Agreement between MID in the SSM function and MID in other disability measures

Scales (MID)	Prevalence MID scale 1 %	Yes/yes: *n* (%)	Yes/no: *n* (%)	No/yes: *n* (%)	No/no: *n* (%)	κ
SSM F (0.52)/FT (30 %)	50	173 (37)	59 (13)	86 (18)	148 (32)	0.38
SSM F (0.52)/FT (≥15 points; $n = 449$)[a]		183 (41)	42 (9)	95 (21)	129 (29)	0.39
SSM F (0.52)/RMQ (30 %)		146 (31)	86 (18)	59 (13)	175 (38)	0.38
SSM F (0.52)/RMQ (≥5 points; $n = 417$)		125 (30)	92 (22)	42 (10)	158 (38)	0.36
SSM F (0.52)/satisfied patients ($n = 459$)[b]		185 (41)	47 (10)	126 (27)	101 (22)	0.24
SSM F (0.1)/FT (30 %)	70	215 (46)	113 (24)	42 (9)	96 (21)	0.31
SSM F (0.1)/FT (≥15 points; $n = 449$)[a]		228 (51)	92 (20)	50 (11)	79 (18)	0.3
SSM F (0.1)/RMQ (30 %)		175 (38)	153 (33)	30 (6)	108 (23)	0.25
SSM F (0.1)/RMQ (≥5 points; $n = 417$)		147 (35)	154 (37)	20 (5)	96 (23)	0.23
SSM F (0.1)/satisfied patients ($n = 459$)[b]		245 (55)	82 (18)	66 (14)	66 (14)	0.24
SSM F (30 %)/FT (30 %)	41	158 (34)	33 (7)	99 (21)	176 (38)	0.44
SSM F (30 %)/FT (≥15 points; $n = 449$)[a]		162 (36)	22 (5)	116 (26)	149 (33)	0.41
SSM F (30 %)/RMQ (30 %)		134 (29)	57 (12)	71 (15)	204 (44)	0.44
SSM F (30 %)/RMQ: (≥5 points; $n = 417$)		114 (27)	63 (15)	53 (13)	187 (45)	0.43
SSM F (30 %)/satisfied patients ($n = 459$)[b]		169 (37)	22 (5)	142 (31)	126 (27)	0.33

SSM F SSM function subscale, *MID* meaningful important differences, *RMQ* Roland Morris Questionnaire, *FT* feeling thermometer

[a] FT 17 patients excluded from the analysis because the baseline FT was <15 points)

[b] Satisfied patients defined according to the definition of Stucki et al. SSM Satisfaction 1.0–2.0 points [2], not satisfied include all patients reporting in the SSM satisfaction of >2.0 points (range 1–4 points)

between the feeling thermometer and the SSM Sy was moderate according to the kappa values (κ value between 0.38– 0.5). MID in the SSM Sy and feeling thermometer was reported in 35 to 49 %. A disagreement in MID was found in one-third of the patients (i.e., reported MID in feeling thermometer but not in SSM Sy or vice versa).

The agreement between MID in the disability scales is summarized in Table 4. The agreement between the SSM F and the RMQ was fair to moderate (κ values between 0.23 and 0.44). MID in both scales was reported in 27–38 %. The disagreement in MID between the scales was found in at least one-third of the patients. Figure 2 depicts how often the estimated results are consistent. Similarly, the agreement between the feeling thermometer and the SSM F was fair to moderate (κ values 0.3–0.4) and disagreement between MID of the feeling thermometer and the SSM F was found in at least one-third of the patients. The agreement between the MID SSM F and satisfaction was fair (κ value between 0.24 and 0.33).

The agreement of scales that assess both pain and disability is given in Table 5. The agreement between the feeling thermometer and a combination of SSM Sy and SSM F was fair to moderate (κ value between 0.34 and 0.45). MID in both scales were reported in 27–45 %, disagreements between MID of the feeling thermometer and the SSM in one-third of patients.

When using the external criteria satisfaction (SSM Sat. 1.0–2.0 points), the agreement with the SSM Sy, the SSM F, and the SSM Sy + F was fair to moderate (κ values between 0.25 and 0.35). For the SSM Sy, MID cut-off of 0.48 points, 32 % of the patients reported MID without satisfaction or were satisfied without MID in the SSM Sy. For the SSM F MID cut-off of 0.52 points, 37 % of the patients reported MID without satisfaction or were satisfied without MID in the SSM F. Forty percent of the patients reported being satisfied without reporting MID in both scales.

Discussion

In 466 patients treated for symptomatic lumbar spinal stenosis, we demonstrated that the proportions of improved patients after treatment varied considerably, depending on the choice of the outcome measures and MID cut-off values. The proportion of patients fulfilling the criteria for a 'minimal important difference' ranged for pain measures from 40 to 65 % and for disability measures from 40 to 70 %. Disagreement (i.e., MID reported by patients in one scale but not in the other and vice versa) between pain scales was found in 25–31 % and between disability scales in 27–39 %. The kappa statistics for pain measures showed moderate agreement (Cohen's κ value between 0.38 and 0.51) and for disability measures fair to moderate agreement (Cohen's κ value between 0.24 and 0.44).

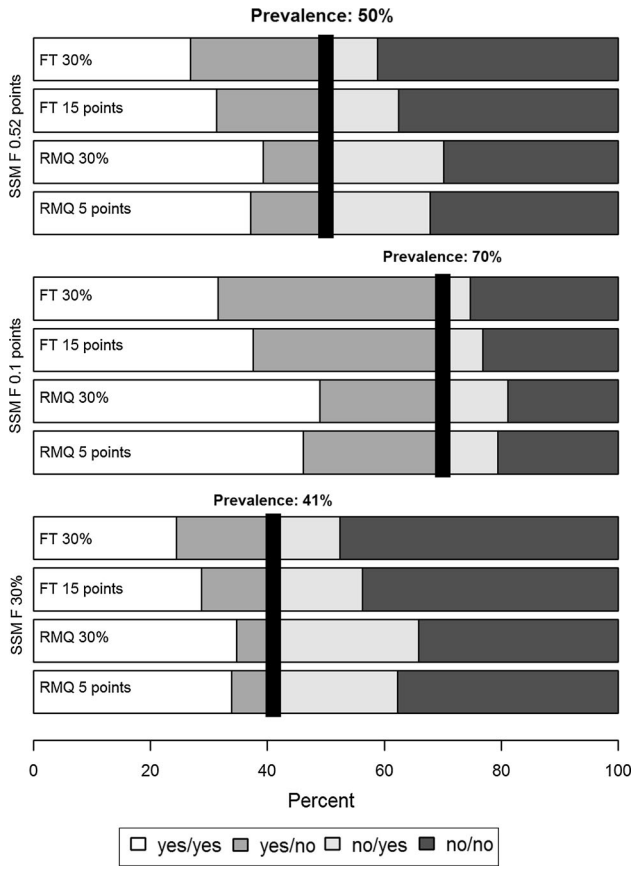

Fig. 2 Agreement between MID of scales assessing disability in patients with lumbar spinal stenosis. *Prevalence value* indicates the proportion of patients with MID in the SSM F scale; *yes/yes* and *no/no* represent the proportion of agreement on a "true" status (MID and no MID). *No/yes* and *yes/no* are subjects with a true positive or negative status in one scale but not in the other and the "true" status is difficult to determine

Results compared to the literature

To the best of our knowledge, no study has addressed the consequences of applying different patient-reported outcome measures (PROMs) and cut-off values for MID in patients with lumber spinal stenosis. PROMs are well established for the assessment of pain and disability in back pain-related disorders. Not all PROMs used in low back pain patients reliably measure disability related to spinal stenosis. Our study showed that the Roland Morris questionnaire (RMQ) was less likely to detect MID in spinal stenosis patients compared to the disease-specific SSM. One explanation of this finding is that the RMQ questions focus on the back pain which is not the primary complaint in many patients with lumbar spinal stenosis. This finding highlights the importance of validation studies in specific patient populations. Most studies focus on the reliability and responsiveness of different scales [3] and the establishment of minimal important differences (MID). MID focuses on the decrease of pain or disability in "the smallest amount of benefit a patient can recognize and value" [1]. MID values for a scale can be developed by various approaches (e.g., anchor-based or sensitivity- and specificity-based approach) and a broad variation in the resulting MID cut-offs has been shown [25]. In addition, different patient populations may respond differently. For example, the SSM was developed for LSS patients undergoing surgery [2]. The validation study of the SSM conducted by Cleland et al. was done in patients mainly undergoing conservative treatment [3]. Therefore, the differences of the MID cut-off values may also be explained by the different patient population under investigation. It is a logical consequence that using lower MID cut-off criteria results in a higher proportion of MID, and therefore, a higher success rate for an intervention. In an effort to reach an international consensus in the interpretation of changes in scores in low back pain, Ostelo et al. proposed a 30 % improvement as a general guide [17]. When applying the 30 % improvement to the current study population instead of using absolute cut-off scores, the success rates were lower for all measures. The decrease was most pronounced for the SSM Sy and SSM F subscales. In the SSM F subscale, the proportion of patients reporting MID dropped from 70 % (cut-off 0.1 point) to 40 % (cut-off of 30 %). Clinical registries offer the advantage that participating centers collect the same core set of measures. For example, analyses based on the international Spine Tango Registry or the National Swedish Register for Spine Surgery (Swespine) includes data on patients undergoing surgery for LSS from many different centers [26, 27]. Though registries permit comparisons on MID scores, different registries do not use uniform outcome measures. The Spine Tango Registry uses the self-reported Core Outcome Measure Index (COMI) questionnaire whereas the National Swedish Registry uses the Oswestry disability index, both validated PROMs in LSS patients [3, 27].

Further, one-third of the patients reported an MID in one domain specific scale but not in the other scale. In scales assessing the same domain (e.g., pain), convergent validity is assumed, and therefore, the efficacies of studies that use different measures are compared by clinicians and in meta-analyses. While we recently reported a high correlation between the change scores in the SSM Sy and the NRS (correlation coefficient 0.64), the correlation between the

Table 5 Agreement between MID in the SSM symptom and function and the generic feeling thermometer

Scale 1 (MID)/scale 2 (MID)	Prevalence MID scale 1 %	Yes/yes: *n* (%)	Yes/no: *n* (%)	No/yes: *n* (%)	No/no: *n* (%)	κ
SSM Sy (0.48) + SSM F (0.52)/FT (30 %)	39	154 (33)	28 (6)	103 (22)	181 (39)	0.45
SSM Sy (0.48) + SSM F (0.52)/FT (≥15 points ($n = 449$)[a]		159 (35)	18 (4)	119 (27)	153 (34)	0.42
SSM Sy (0.48) + SSM F (0.52)/satisfied patients ($n = 459$)[b]		158 (34)	24 (5)	153 (33)	124 (28)	0.28
SSM Sy (0.36) + SSM F (0.1)/FT (30 %)	56	194 (42)	66 (14)	63 (14)	143 (30)	0.44
SSM Sy (0.36) + SSM F (0.1)/FT: MID ≥15 points ($n = 449$)[a]		200 (45)	53 (12)	78 (17)	118 (26)	0.4
SSM Sy (0.36) + SSM F (0.1)/satisfied patients ($n = 459$)[b]		204 (44)	55 (12)	107 (23)	93 (20)	0.26
SSM Sy (30 %) + SSM F (30 %)/FT (30 %)	29	125 (27)	10 (2)	132 (28)	199 (43)	0.42
SSM Sy (30 %) + SSM F (30 %)/FT: MID ≥15 points ($n = 449$)[a]		123 (27)	7 (1)	155 (35)	164 (37)	0.34
SSM Sy (30 %) + SSM F (30 %)/satisfied patients ($n = 459$)[b]		129 (28)	6 (1)	182 (40)	142 (31)	0.29

SSM Sy SSM symptom subscale, *SSM F* SSM function subscale, *MID* meaningful important differences, *RMQ* Roland Morris Questionnaire, *FT* feeling thermometer

[a] 17 patients excluded because of baseline values of less than <15 points in the feeling thermometer (FT)

[b] Satisfied patients defined according to the definition of Stucki et al. SSM satisfaction 1.0–2.0 points [2], not satisfied include all patients reporting in the SSM satisfaction of >2.0 points (range 1–4 points)

SSM F and RMQ was lower (correlation coefficient 0.39) [27]. The sensitivity and specificity (ROC analysis) for an external criterion of clinical change showed a higher responsiveness for the SSM Sy plus SSM F (AUC 0.832) than for the RMQ (AUC 0.631) [27].

The nature of the disease leads to a different presentation of patients with lumbar spinal stenosis compared to patients with low back pain. While the symptom back pain is less prominent, symptoms including neuroischemic pain and disability during walking are frequent. Therefore, it is possible to assume that general back pain measures may not be sensitive to these complaints and underestimate the disability in spinal stenosis patients. It may, therefore, be hypothesized that studies with RMQ as a primary outcome report less favorable results than studies that use the SSM function scale. For example, Friedli et al. found in a study on the efficacy of epidural steroid injections in lumbar spinal stenosis a proportion fulfilling the MID criteria of 30 % decrease in the RMQ in 37.3 and 31.6 % [18], which was comparable to our study. Had another scale been used (e.g., the SSM) the efficacy may have been higher. In a recent meta-analysis on treatment efficacy of surgery in lumbar spinal stenosis, 17 studies were pooled that reported disability on six different outcome scales [28]. According to our findings, it is questionable that all of these study results are comparable.

Reporting the efficacies of interventions in a mode that clinicians and researchers alike can understand is important. To establish clinical guidance on the use of effective interventions in patients with LSS the results of clinical studies need to be reported so that the results can be compared. For low back pain, an international collaboration achieved a consensus on a core set of recommended measures for future research in an effort to improve the quality of prospective studies [29]. No consensus on which outcome data to collect and report is available for studies on lumbar spinal stenosis. Different PROMs are used in spinal stenosis studies and many are derived from back pain research [28] but may not be sensitive to the complaints in this population. There is a need for a consensus in definitions not only in PROMs but also in disease-specific definitions [30].

Strength and limitations

The strength of this study was that a broad variety of patients with symptomatic lumbar spinal stenosis were included in this analysis. Great care was taken to ensure high quality data collection and handling. Up-to-date methods were used for the analysis.

The main limitation of this study is that only two disability measures were available for the analysis. Therefore, the study results cannot be extrapolated to other PROMs

including the frequently used Oswestry disability index. However, it is reasonable to expect wide variations on MID proportions when different cut-off values are used. Therefore, researchers and clinicians should engage in a discussion on patient relevant outcome measures, including cut-off values for meaningful improvement, and a consensus on reporting outcome measure will greatly improve comparability of study results.

Implications for research

Future research in patients with lumbar spinal stenosis should report minimal important changes in PROMs for different cut-off values until an international consensus on the use of outcome measures and MID criteria is achieved. It should be further noted that MID does not measure deterioration and only a few studies have addressed the potential for deterioration [31]. Further research should also address the potential for deterioration despite a treatment.

Implications for clinical practice

In patients with lumbar spinal stenosis, their expectations and preferences have to be taken into account in choosing treatment. Detailed information on the benefits and harms of different treatments may influence and solidify patients' opinion of their treatment choice [32, 33]. Depending on the selected MID-value to estimate treatment efficacy of surgery, epidural steroid injections, or conservative treatment, information on efficacy varies considerably. This uncertainty about treatment efficacy should be reflected in informing patients about their options. The variation between outcome measures reported in this study may explain to certain extent conflicting information about the treatment efficacy described in lumbar spinal stenosis studies. However, the arguments surgeons put forward for a specific choice of surgical treatment are not well explained [30]. It is, therefore, difficult to inform patients about the expected treatment efficacy.

Conclusion

The MID in outcome scores for this population varied from 40 to 70 %, depending on the measure or cut-off score used. Further, the disagreement between domain specific measures indicates that differences between studies may be also related to the choice of an outcome measures. An international consensus on the use and reporting of outcome measures in studies on lumbar spinal stenosis is needed.

Acknowledgments We thank our research staff at all studies sites for their continuous meticulous work to include and follow-up with patients. In particular, we thank MM and GP for their efforts to assure high quality data intake and database handling.

Compliance with ethical standards

Funding sources and role of sponsors The lumbar stenosis outcome study (LSOS) was funded by the Helmut Horten Foundation, OPO-Foundations, Symphasis Foundation, Baugarten Foundation, and the Pfizer-Foundation for geriatrics and research in geriatrics. The funders had no influence on the study design, on the data collection, data management, statistical analysis, the interpretation of the data, the content of the manuscript, and the decision to submit the paper for publication.

Conflict of interest The authors declare no financial interests or affiliations with institutions, organizations, or companies relevant to the manuscript. All authors had full access to the data, interpreted the analysis and commented on the final manuscript.

Ethical statement This cohort study was conducted in compliance with all international laws and regulations as well as any applicable guidelines. The study was approved by the independent Ethics Committee of the Canton Zurich (KEK-ZH-NR: 2010-0395/0). A study protocol was published: Steurer et al. [7].

Appendix 1

See Fig. 3.

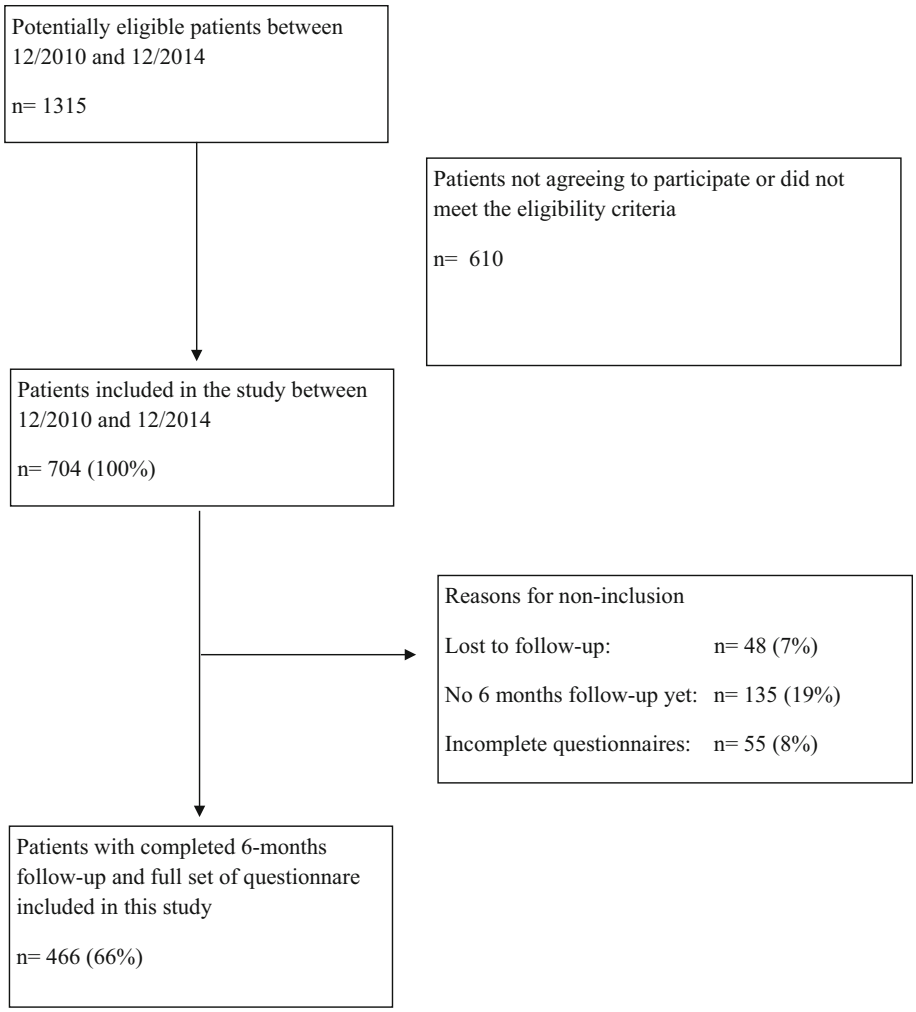

Fig. 3 Study flow

Appendix 2

See Table 6.

Table 6 Meaningful important difference (MID) for the SSM subscales

Scale 1 (MID)/scale 2 (MID)	Prevalence MID scale 1 %	Yes/yes	Yes/no	No/yes	No/no
SSM Sy (0.48)/SSM F (0.52)	57	182 (39)	84 (18)	50 (11)	150 (32)
SSM Sy (0.48)/SSM F (0.1)		231 (50)	35 (7)	97 (21)	103 (22)
SSM Sy (0.48)/satisfied[a]		213 (46)	50 (11)	98 (21)	98 (21)
SSM Sy (0.36)/SSM F (0.52)	65	197 (42)	106 (24)	35 (7)	128 (27)
SSM Sy (0.36)/SSM F (0.1)		260 (56)	43 (9)	68 (15)	95 (20)
SSM Sy (0.36)/satisfied[a]		229 (50)	71 (15)	82 (18)	77 (17)

SSM Scale Spinal Stenosis Measure, *SSM Sy* SSM symptom subscale, *SSM F* SSM function subscale, *MID* meaningful important difference

[a] Satisfied patients defined according to the definition of Stucki et al. SSM satisfaction 1.0–2.0 points [2], not satisfied include all patients reporting in the SSM satisfaction of >2.0 points (range 1–4 points)

References

1. Barrett B, Brown D, Mundt M, Brown R (2005) Sufficiently important difference: expanding the framework of clinical significance. Med Decis Making 25:250–261. doi:10.1177/0272989X05276863
2. Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossel AH, Katz JN (1996) Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. Spine (Phila Pa 1976) 21:796–803
3. Cleland J, Whitman J, Houser J, Wainner R, Childs J (2012) Psychometric properties of selected tests in patients with lumbar spinal stenosis. The spine journal 12:921–931
4. Pratt RK, Fairbank JC, Virr A (2002) The reliability of the Shuttle Walking Test, the Swiss Spinal Stenosis Questionnaire, the Oxford Spinal Stenosis Score, and the Oswestry Disability Index in the assessment of patients with lumbar spinal stenosis. Spine (Phila Pa 1976) 27:84–91
5. Sekiguchi M, Wakita T, Otani K, Onishi Y, Fukuhara S, Kikuchi S, Konno S (2012) Development and validation of a symptom scale for lumbar spinal stenosis. Spine (Phila Pa 1976) 37:232–239. doi:10.1097/BRS.0b013e318216afb4
6. EuroQol Group (1990) EuroQol—a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy 16:199–208
7. Steurer J, Nydegger A, Held U, Brunner F, Hodler J, Porchet F, Min K, Mannion AF, Michel B (2010) LumbSten: the lumbar spinal stenosis outcome study. BMC Musculoskelet Disord 11:254. doi:10.1186/1471-2474-11-254
8. World Medical Assembly (WMA) (2006) WMA Declaration of Helsinki-Ethical principles for medical research involving human subjects. Adopted by the 18th WMA general assembly, Helsinki, Finland, June 1964 and amended by the 64th WMA General Assembly, Fortaleza, Brazil, October 2013. http://www.wma.net/en/30publications/30ethicsmanual/index.html
9. Fokter SK, Yerby SA (2006) Patient-based outcomes for the operative treatment of degenerative lumbar spinal stenosis. Eur Spine J 15:1661–1669. doi:10.1007/s00586-005-0033-4
10. Thornes E, Grotle M (2008) Cross-cultural adaptation of the Norwegian version of the spinal stenosis measure. Eur Spine J 17:456–462. doi:10.1007/s00586-007-0576-7
11. Wertli MM, Steurer J, Wildi LM, Held U (2014) Cross-cultural adaptation of the German version of the spinal stenosis measure. Eur Spine J 23:1309–1319. doi:10.1007/s00586-014-3245-7
12. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA (1978) Studies with pain rating scales. Ann Rheum Dis 37:378–381. doi:10.1136/ard.37.4.378
13. Kremer E, Atkinson JH, Ignelzi RJ (1981) Measurement of pain: patient preference does not confound pain measurement. Pain 10:241–248
14. Choinière M, Amsel R (1996) A visual analogue thermometer for measuring pain intensity. J Pain Symptom Manag 11:299–311
15. Roland M, Morris R (1983) A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine (Philadelphia, Pa 1976) 8:141–144
16. Exner V, Keel P (2000) Measuring disability of patients with low-back pain—validation of a German version of the Roland and Morris disability questionnaire. Schmerz 14:392–400
17. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, de Vet HC (2008) Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 33:90–94. doi:10.1097/BRS.0b013e31815e3a10
18. Friedly JL, Comstock BA, Turner JA, Heagerty PJ, Deyo RA, Sullivan SD, Bauer Z, Bresnahan BW, Avins AL, Nedeljkovic SS, Nerenz DR, Standaert C, Kessler L, Akuthota V, Annaswamy T, Chen A, Diehn F, Firtch W, Gerges FJ, Gilligan C, Goldberg H, Kennedy DJ, Mandel S, Tyburski M, Sanders W, Sibell D, Smuck M, Wasan A, Won L, Jarvik JG (2014) A randomized trial of epidural glucocorticoid injections for spinal stenosis. N Engl J Med 371:11–21. doi:10.1056/NEJMoa1313265
19. Minamide A, Yoshida M, Yamada H, Nakagawa Y, Hashizume H, Iwasaki H, Tsutsui S (2015) Clinical outcomes after microendoscopic laminotomy for lumbar spinal stenosis: a 5-year follow-up study. Eur Spine J 24:396–403. doi:10.1007/s00586-014-3599-x
20. Roland M, Fairbank J (2000) The Roland–Morris disability questionnaire and the Oswestry disability questionnaire. Spine (Phila Pa 1976) 25:3115–3124
21. Viera AJ, Garrett JM (2005) Understanding interobserver agreement: the kappa statistic. Fam Med 37:360–363
22. Sim J, Wright CC (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 85:257–268
23. Vach W (2005) The dependence of Cohen's kappa on the prevalence does not matter. J Clin Epidemiol 58:655–661. doi:10.1016/j.jclinepi.2004.02.021
24. Core Team R (2013) R: A language and environment for statistical computing. In: Core Team R (ed) R Foundation for statistical computing. R Core Team, Vienna
25. Parker SL, Adogwa O, Mendenhall SK, Shau DN, Anderson WN, Cheng JS, Devin CJ, McGirt MJ (2012) Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis. Spine J 12:1122–1128. doi:10.1016/j.spinee.2012.10.006
26. Munting E, Roder C, Sobottke R, Dietrich D, Aghayev E, Spine Tango C (2015) Patient outcomes after laminotomy, hemilaminectomy, laminectomy and laminectomy with instrumented fusion for spinal canal stenosis: a propensity score-based study from the Spine Tango registry. Eur Spine J 24:358–368. doi:10.1007/s00586-014-3349-0
27. Mannion AF, Fekete TF, Wertli MM, Mattle M, Nauer S, Kleinstuck FS, Jeszenszky D, Haschtmann D, Becker HJ, Porchet F, Lumbar Spinal Stenosis Outcome Study Group (2015) Could less be more when assessing patient-rated outcome in spinal stenosis? Spine (Phila Pa 1976) 40:710–718. doi:10.1097/BRS.0000000000000751
28. Machado GC, Ferreira PH, Harris IA, Pinheiro MB, Koes BW, van Tulder M, Rzewuska M, Maher CG, Ferreira ML (2015) Effectiveness of surgery for lumbar spinal stenosis: a systematic review and meta-analysis. PLoS ONE 10:e0122800. doi:10.1371/journal.pone.0122800
29. Pincus T, Santos R, Breen A, Burton AK, Underwood M, Multinational Musculoskeletal Inception Cohort Study Collaboration (2008) A review and proposal for a core set of factors for prospective cohorts in low back pain: a consensus statement. Arthritis Rheum 59:14–24. doi:10.1002/art.23251
30. Burgstaller JM, Porchet F, Steurer J, Wertli MM (2015) Arguments for the choice of surgical treatments in patients with lumbar spinal stenosis—a systematic appraisal of randomized controlled trials. BMC Musculoskelet Disord 16:96. doi:10.1186/s12891-015-0548-8
31. Gum JL, Glassman SD, Carreon LY (2013) Clinically important deterioration in patients undergoing lumbar spine surgery: a choice of evaluation methods using the Oswestry Disability Index, 36-Item Short Form Health Survey, and pain scales: clinical article. J Neurosurg Spine 19:564–568. doi:10.3171/2013.8.spine12804

32. Lurie JD, Spratt KF, Blood EA, Tosteson TD, Tosteson AN, Weinstein JN (2011) Effects of viewing an evidence-based video decision aid on patients' treatment preferences for spine surgery. Spine (Phila Pa 1976) 36:1501–1504. doi:10.1097/BRS.0b013e3182055c1e

33. Stacey D, Legare F, Col NF, Bennett CL, Barry MJ, Eden KB, Holmes-Rovner M, Llewellyn-Thomas H, Lyddiatt A, Thomson R, Trevena L, Wu JH (2014) Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 1:Cd001431. doi:10.1002/14651858.CD001431.pub4