

Clinical Study

Determination of the Oswestry Disability Index score equivalent to a “satisfactory symptom state” in patients undergoing surgery for degenerative disorders of the lumbar spine—a Spine Tango registry-based study

Miranda L. van Hooff, MSc^{a,*}, Anne F. Mannion, PhD^b, Lukas P. Staub, MD, PhD^c,
Raymond W.J.G. Ostelo, PT, PhD^d, Jeremy C.T. Fairbank, MA, MD, FRCS^e

^aDepartment of Research, Sint Maartenskliniek, PO Box 9011, 6500 GM Nijmegen, The Netherlands

^bDepartment of Teaching, Research and Development, Spine Center Division, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland

^cInstitute for Evaluative Research in Medicine, University of Bern, Stauffacherstrasse 78, 3014 Bern, Switzerland

^dDepartment of Health Sciences & VU Medical Center, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University, PO Box 7057, 1007 MB Amsterdam, The Netherlands

^eNuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Science NDORMS, Nuffield Orthopaedic Centre, University of Oxford, Windmill Rd, Oxford OX3 7HE, United Kingdom

Abstract

BACKGROUND CONTEXT: The achievement of a given change score on a valid outcome instrument is commonly used to indicate whether a clinically relevant change has occurred after spine surgery. However, the achievement of such a change score can be dependent on baseline values and does not necessarily indicate whether the patient is satisfied with the current state. The achievement of an absolute score equivalent to a patient acceptable symptom state (PASS) may be a more stringent measure to indicate treatment success.

PURPOSE: This study aimed to estimate the score on the Oswestry Disability Index (ODI, version 2.1a; 0–100) corresponding to a PASS in patients who had undergone surgery for degenerative disorders of the lumbar spine.

STUDY DESIGN/SETTING: This is a cross-sectional study of diagnostic accuracy using follow-up data from an international spine surgery registry.

PATIENT SAMPLE: The sample includes 1,288 patients with degenerative lumbar spine disorders who had undergone elective spine surgery, registered in the EUROSPINE Spine Tango Spine Surgery Registry.

OUTCOME MEASURES: The main outcome measure was the ODI (version 2.1a).

METHODS: Surgical data and data from the ODI and Core Outcome Measures Index (COMI) were included to determine the ODI threshold equivalent to PASS at 1 year (± 1.5 months; $n=780$) and 2 years (± 2 months; $n=508$) postoperatively. The symptom-specific well-being item of the COMI was used as the external criterion in the receiver operating characteristic (ROC) analysis to determine the ODI threshold equivalent to PASS. Separate sensitivity analyses were performed based on the different definitions of an “acceptable state” and for subgroups of patients. JF is a copyright holder of the ODI.

FDA device/drug status: Not applicable.

Author disclosures: **MLvH:** Nothing to disclose. **AFM:** Nothing to disclose. **LPS:** Nothing to disclose. **RWJGO:** Grants: WCF (E, Paid to the institution), EUROSPINE (E, Paid to the institution), zorginnovatiefonds (F, Paid to the institution), NWO (G, Paid to the institution), ECU (D, Paid to the institution), outside the submitted work.

* Corresponding author. Department of Research, Sint Maartenskliniek, PO Box 9011, 6500 GM Nijmegen, The Netherlands. Tel.: +31 (0) 24-365-9912; fax: +31 (0) 24-365-9154.

E-mail address: m.vanhooff@maartenskliniek.nl (M.L. van Hooff)

RESULTS: The ODI threshold for PASS was 22, irrespective of the time of follow-up (area under the curve [AUC]: 0.89 [sensitivity {Se}: 78.3%, specificity {Sp}: 82.1%] and AUC: 0.91 [Se: 80.7%, Sp: 85.6] for the 1- and 2-year follow-ups, respectively). Sensitivity analyses showed that the absolute ODI-22 threshold for the two follow-up time-points were robust. A stricter definition of PASS resulted in lower ODI thresholds, varying from 16 (AUC=0.89; Se: 80.2%, Sp: 82.0%) to 18 (AUC=0.90; Se: 82.4%, Sp: 80.4%) depending on the time of follow-up.

CONCLUSIONS: An ODI score ≤ 22 indicates the achievement of an acceptable symptom state and can hence be used as a criterion of treatment success alongside the commonly used change score measures. At the individual level, the threshold could be used to indicate whether or not a patient with a lumbar spine disorder is a “responder” after elective surgery.

Oswestry Disability Index; Patient acceptable symptom state; Patient-reported outcome; Satisfaction; Sensitivity; Specificity; Spine; Success; Surgery

Keywords:

Introduction

In Western societies, low back pain (LBP) has the largest disease burden [1]. It is associated with a substantial amount of morbidity, and complaints are multidimensional. Functional status is an important patient-related outcome when evaluating surgical and non-surgical interventions for LBP. One important feature of outcome instruments measuring functional status is their ability to detect meaningful change from the patient’s perspective. In the absence of appropriate objective clinical outcome measures, the use of patient-reported outcome measures (PROM) to assess treatment outcome is commonly accepted [2]. The Oswestry Disability Index (ODI) [3], and the ODI version 2.1a [4,5] in particular, is widely accepted and recommended as a condition-specific PROM in interventional studies [6]. As such, medical decision-making increasingly relies on this measure. Although most clinicians and researchers agree that the success of any intervention should be judged from the patient’s perspective, to date no consensus exists for criteria indicating “success.”

In health services research, it is important to define clear criteria for treatment “success.” Success can be conceptualized in two ways: (1) relevant change or improvement, and (2) achievement of an acceptable state. With the first concept, the emphasis is on whether or not an individual has improved after an intervention [7], whereas with the second the emphasis is on whether or not the achieved outcome is acceptable from the patient’s perspective [7]. The concept of change (minimum clinically important difference or change) is frequently used in spine research to assess treatment success. In relation to this, it is important to specify whether the observed change in an individual’s scores is merely the result of measurement error or whether it constitutes a real change, and whether that change is also clinically relevant to the patient [8]. However, it is difficult to measure what is clinically relevant to patients [9], and methodological issues such as population dependency and baseline dependency [10] are encountered. Moreover, assessment of change does not indicate whether a “normal” or “healthy” symptom state is reached. For these reasons, we have previously used a more stringent

definition of success based on achievement of values seen in “normal,” healthy populations [11]. The threshold used was the achievement of an ODI value, derived from “normal” subjects with little or no back pain, of ≤ 22 [4,11]. The use of “normal, healthy population” values as the reference might, however, be criticized as the ODI is a condition-specific instrument.

An alternative approach to measuring success is to identify the value beyond which patients consider themselves well or consider their health state to be acceptable, ie, the concept of the patient acceptable symptom state (PASS) [12,13]. Determination of the absolute cutoff value (threshold) at follow-up, equivalent to achievement of a PASS, would assist in interpreting scores at the individual level and would allow determination of the proportion of patients within a group who achieve this level, when evaluating the effectiveness or success of interventions. Achievement of this threshold might be more important than the achievement of a given change value, and it probably reflects the ultimate goal of treatment from the patient’s perspective [12,14]. The concepts of “feeling better” and “feeling good” are complementary but distinctly different; a patient’s condition can be markedly improved by the intervention but can still be suboptimal [7,12,13].

The purpose of the present study was to estimate the score on the ODI (version 2.1a) corresponding to a “patient acceptable symptom state” in patients undergoing surgery for degenerative disorders of the lumbar spine. To assess the robustness of the findings, we performed sensitivity analyses with different definitions for “acceptable state” and in various subgroups of patients.

Materials and methods

Study population

This cross-sectional study was performed using postoperative data from the Spine Tango Spine Surgery Registry of EUROSPINE, the Spine Society of Europe (SSE) [15,16], and according to the STARD statement for reporting studies of diagnostic accuracy [17]. The study dataset was prepared in August 2014 by linking the surgical data, recorded on the

Outcome measure

The ODI (version 2.1a) [5] was the main outcome measure. The ODI (10 items) measures the impact of LBP on patients' functional ability in 10 aspects of daily life. The total ODI score is a sum score and ranges from 0 to 100; higher scores indicate greater disability.

External criterion (anchor)

The anchor used to assess the success of surgery, ie, achievement of an acceptable symptom state at follow-up, was the response on the symptom-specific well-being (SSWB) item of the COMI [19,20]: "If you had to spend the rest of your life with the symptoms you have now, how would you feel about it?" The responses are given on a 5-point Likert scale: "very satisfied," "somewhat satisfied," "neither satisfied nor dissatisfied," "somewhat dissatisfied," and "very dissatisfied."

Values for success (PASS)

For the analyses, the 5-point Likert scale of the COMI SSWB item was collapsed to a dichotomous outcome variable (1=acceptable; 0=unacceptable). We performed two separate analyses. The main analysis considered patients who reported feeling "very satisfied" and "somewhat satisfied" as an acceptable symptom state (PASS 1), whereas the sensitivity analysis included only those who reported feeling "very satisfied" (PASS 2). All other categories were considered to represent "unacceptable." To determine PASS, we defined two study samples: those with a 1-year follow-up (± 1.5 months) and those with a 2-year follow-up assessment (± 2 months).

Statistical analyses

Baseline demographic and perioperative data of patients in both study samples were described as means and standard deviations for continuous variables and as counts (and percentages) for categorical variables. Differences between samples were examined with independent Student *t* tests for continuous variables and with Pearson chi-square tests for categorical variables. Proportions (%) are presented for both the distribution of responses on the SSWB item and the % categorized as "acceptable" and "unacceptable."

The ODI threshold for the patient acceptable symptom state (PASS 1) was determined using receiver operating characteristics (ROC) analyses. The corresponding ROC curve is a plot for each cutoff value that represents the relation between the proportion of patients who were correctly classified in the "acceptable" group, based on the COMI SSWB score (sensitivity [true-positive]; y-axis) and the proportion of patients who were incorrectly classified in the "acceptable" group (1-specificity [false-positive]; x-axis). The corresponding ROC tables show for each ODI cutoff value the relation between sensitivity, specificity, and the percentage correctly classified. The ODI "acceptable" threshold is the value that provides the best balance between sensitivity and specificity, as it

SSE Spine Tango Surgery 2006 form (ST-2006-form), with the last available follow-up ODI and Core Outcome Measures Index (COMI) [18–20], which had to have been completed on the same day (Fig. 1). The surgical inclusion criteria were based on the data documented using the registry's ST-2006-form: lumbar, lumbosacral, sacral, or coccyx as the level of procedure, and degenerative disease (black disc or disc degeneration, spondylosis, spondylarthrosis, adjacent segment degeneration, spinal stenosis, disc herniation) or spondylolisthesis (degenerative) as main pathology. The data from 2,530 patients satisfied these inclusion criteria.

The registry's ST-2006-form was also used to derive descriptive information on patient samples to perform sensitivity analyses. Data for the following variables were extracted: date of birth (age), gender, (number of) previous spine surgeries, morbidity state (American Society of Anesthesiologists physical status score), surgical measures (type of surgery), and surgical complications. Patient-reported complications recorded on the Spine Tango patient follow-up questionnaire were also used.

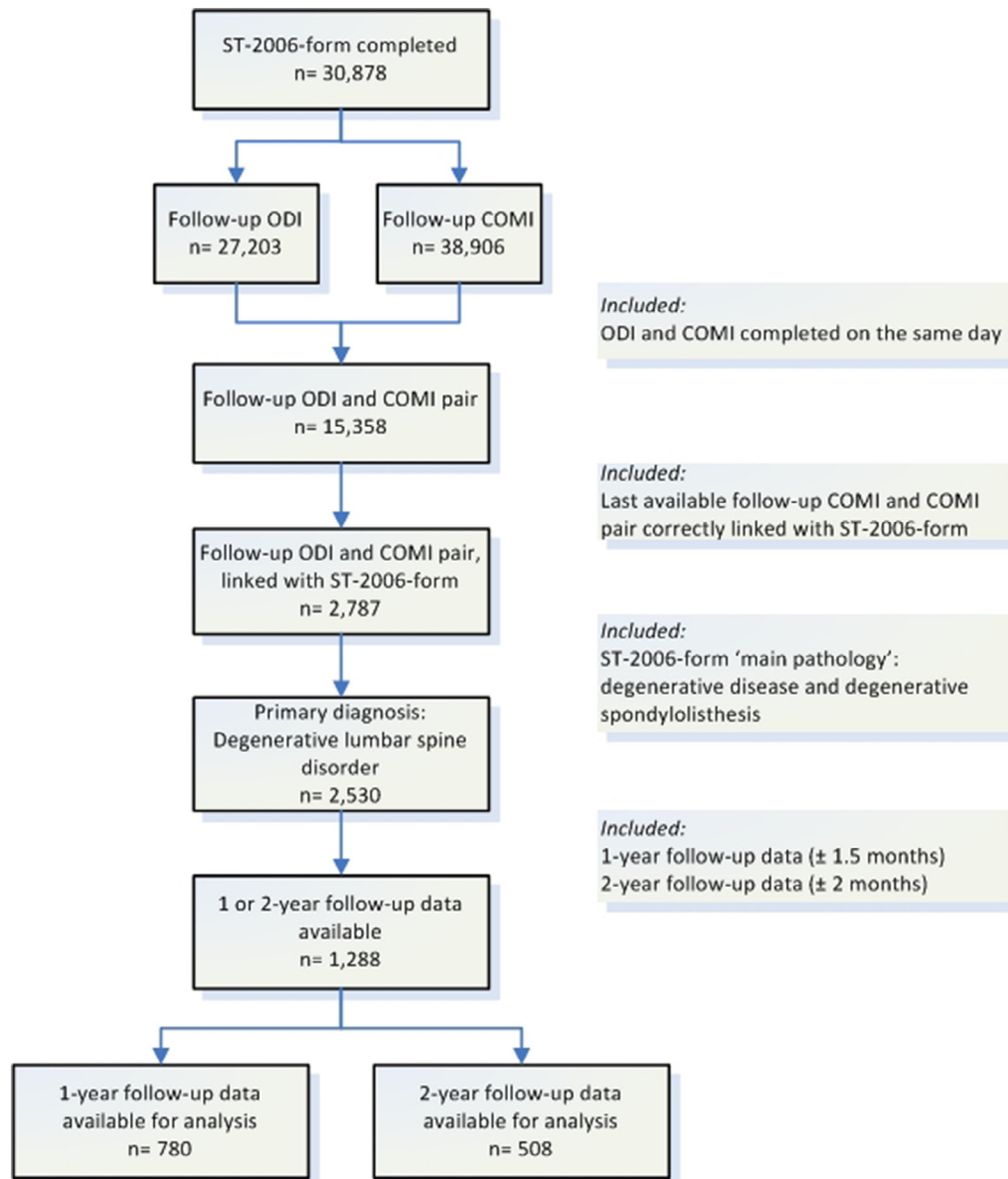


Fig. 1. Data flow. ST-2006-form, SSE Spine Tango 2006 surgery form; COMI, Core Outcome Measure Index; ODI, Oswestry Disability Index.

represents the lowest overall misclassification (ie, minimum of false-positives and false-negatives, or the maximum sum of specificity and sensitivity [21]). The area under the curve (AUC) indicates the probability of correctly differentiating between an “acceptable” and an “unacceptable” state, and an AUC value of >0.7 is considered satisfactory [22]. For each PASS, the AUCs at both follow-up assessments were compared for the discriminative ability of the ODI. When satisfactory, the ROC analyses were subsequently used to determine the ODI cutoff values.

Separate sensitivity analyses were performed based on the second definition of “acceptable” (PASS 2) and for subgroups of patients based on factors possibly influencing the surgical outcome. These factors were gender (male; female), age group (<65 years; ≥ 65 years), previous surgery (yes; no),

surgical procedure (decompression; fusion [ie, fusion alone, decompression and fusion with or without stabilization]), surgical complications (yes; no [wrong level, nerve root damage, cauda equina damage, spinal cord damage, bleeding in spinal canal, bleeding outside spinal canal, malposition of implant, dural lesion, wound infection, implant failure, other]), and patient-reported complications (yes; no).

The statistical analyses were conducted using STATA version 12 for Windows (StataCorp, College Station, Texas, USA). Statistical significance was accepted at $p < .05$.

Results

To determine the ODI threshold corresponding to PASS at the different predefined follow-ups, the data of 1,288

Table 1
Characteristics of the two study samples

Characteristics	Total patient sample 1-year follow-up (n=780)	Total patient sample 2-year follow-up (n=508)	p-value [‡]
Gender, n (%), male : female	373 (47.8) : 407 (52.2)	243 (47.8) : 265 (52.2)	.996*
Age, mean±SD (range, min–max) (y)	58.1±15.1 (19–87)	59.1±15.2 (19–87)	.481 [†]
Age group			.588*
<65, n (%)	475 (60.9)	317 (62.4)	
≥65, n (%)	305 (39.1)	191 (37.6)	
Previous surgery			.698*
n (%) yes : no	180 (23.1) : 600 (76.9)	122 (24.0) : 386 (76.0)	
Number of previous surgeries, n (%)			.732*
1	137 (76.1)	96 (78.7)	
2	29 (16.2)	18 (14.8)	
3	8 (4.4)	4 (3.3)	
4	4 (2.2)	2 (1.6)	
5	2 (1.1)	2 (1.6)	
Previous surgery same level			.233*
n (%) yes : no : partially	107 (59.4) : 42 (23.3) : 31 (17.3)	60 (49.2) : 40 (32.8) : 22 (18.0)	
Morbidity ASA classification, n (%)			.452*
ASA 1	193 (24.7)	108 (21.3)	
ASA 2	308 (39.5)	205 (40.4)	
ASA 3	108 (13.9)	72 (14.2)	
ASA 4	1 (0.1)	3 (0.5)	
ASA 5	–	–	
Unknown	170 (21.8)	120 (23.6)	
Surgical procedure, n (%)			.201*
Decompression	487 (62.4)	335 (65.9)	
Fusion	293 (37.6)	173 (34.1)	
Complications			
Surgical, n (%) yes : no : missing	36 (4.6) : 736 (94.4) : 8 (1.0)	31 (6.1) : 472 (92.9) : 5 (1.0)	.241*
Patient-reported, n (%) yes : no	192 (24.6) : 588 (75.4)	133 (26.2) : 375 (73.8)	.527*

ASA, American Society of Anesthesiologists; SD, standard deviation.

Notes: *Decompression* included both anterior and posterior decompression. *Fusion* included fusion alone; decompression and fusion with or without stabilization (anterior or posterior). *Surgical complications* included wrong level, nerve root damage, cauda equina damage, spinal cord damage, bleeding in spinal canal, bleeding outside spinal canal, malposition of implant, dura lesion, wound infection, implant failure, and other.

Patient-reported complications: Did any complication arise as a consequence of your operation in our hospital (eg, problems with wound healing, paralysis, sensory disturbances)? Response categories: “yes,” “no.”

* Chi-square test was used.

[†] Student *t* test for independent samples was used.

[‡] p<.05 for differences between the study samples.

patients with degenerative lumbar spine disorders undergoing elective spine surgery were included, with 1-year (± 1.5 months; n=780) or 2-year (± 2 months; n=508) follow-up. The characteristics of the two study samples (mean age: 58.3 [standard deviation: 14.9] years) did not differ (Table 1).

Distribution of “treatment success” ratings

Table 2 shows the distribution of responses for the COMI SSWB item. At the 1-year and 2-year follow-ups, a comparable proportion of patients reported an acceptable symptom state (PASS 1): 43.3% and 43.9%, respectively. For the PASS 2, the proportions were 24.0% and 24.6%, respectively.

Area under the curve (AUC) and threshold for PASS

The ROC analyses revealed that the AUCs for each definition of PASS, for each subgroup, and for each follow-up assessment were >0.7 ($0.83 \leq \text{AUC} \leq 0.96$; Tables 3 and 4). The absolute ODI threshold for each definition of PASS was 22,

Table 2
Distribution of responses in relation to the COMI single item used to define the acceptability of the symptom state

	Total patient sample 1-year follow-up (n=780) n (%)	Total patient sample 2-year follow-up (n=508) n (%)
Acceptability of current state		
SSWB		
Very satisfied (1)	187 (23.9)	125 (24.6)
Somewhat satisfied (2)	151 (19.4)	98 (19.3)
Neither satisfied nor dissatisfied (3)	113 (14.5)	61 (12.0)
Somewhat dissatisfied (4)	152 (19.5)	78 (15.4)
Very dissatisfied (5)	177 (22.7)	146 (28.7)
Acceptable (1–2) : Unacceptable state (3–5) yes : no	338 (43.3) : 442 (56.7)	223 (43.9) : 285 (56.1)
Acceptable (1) : Unacceptable state (2–5) yes : no	187 (23.9) : 593 (76.1)	125 (24.6) : 383 (75.4)

COMI, Core Outcome Measures Index; SSWB, symptom-specific well-being.

Table 3
Results of ROC analyses

	AUC	95% CI	Absolute ODI threshold	Se %	Sp %	% Correctly classified
1-year follow-up (n=780)						
PASS 1	0.89	0.86–0.91	22	78.3	82.1	79.6
PASS 2	0.89	0.86–0.91	16	80.2	82.0	81.5
2-year follow-up (n=508)						
PASS 1	0.91	0.89–0.93	22	80.7	85.6	83.5
PASS 2	0.90	0.87–0.92	18	82.4	80.7	81.1

PASS 1, patient acceptable symptom state with COMI SSWB answers “very satisfied” and “somewhat satisfied”; PASS 2, patient acceptable symptom state with COMI SSWB answer “very satisfied”; Se, sensitivity; Sp, specificity; ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval; ODI, Oswestry Disability Index; COMI, Core Outcome Measures Index; SSWB, symptom-specific well-being.

irrespective of the time of follow-up (sensitivity [Se]: 78.3%, specificity [Sp]: 82.1%; and Se: 80.7%, Sp: 85.6%, with 79.6% and 83.5% correctly classified for the 1- and 2-year follow-ups, respectively; [Table 3](#) and [Fig. 2](#)).

Threshold for PASS

[Table 3](#) shows the absolute ODI threshold values for each definition of PASS and for each follow-up assessment. The ODI threshold for PASS was 22, irrespective of the time of follow-up (AUC: 0.89 [Se: 78.3%, Sp: 82.1%] and AUC: 0.91 [Se: 80.7%, Sp: 85.6], with 84.0% and 84.2% correctly classified for the 1- and 2-year follow-ups, respectively; [Fig. 2](#)).

Sensitivity analyses

Definition of success

For PASS 2, the ODI threshold was either 16 (1-year follow-up) or 18 (2-year follow-up) (respectively, AUC=0.91 [Se: 80.2%, Sp: 82.0%] and AUC=0.90 [Se: 82.4%, Sp: 80.4%], with 81.5% and 81.1% correctly classified; [Table 3](#)).

Subgroups

The absolute ODI-22 threshold (PASS 1) for the two follow-up assessments was robust ([Table 4A](#)). [Table 4A](#) shows that the ODI threshold was slightly higher in patients with previous surgery (ODI=23) than in those with no previous surgery (ODI=22). Similarly, the ODI for decompression patients was slightly higher (ODI=23) than for fusion patients (ODI=22).

At 2-year follow-up, the ODI threshold was 22 for all subgroups, except for those with perioperative surgical complications as documented by the surgeon (ODI=20) and those who self-reported no complications (ODI=24; [Table 4A](#)). For PASS 2, the ODI threshold values varied (14–18) depending on the subgroup under investigation and the follow-up used ([Table 4B](#)).

Discussion

In the present study, we identified the ODI (version 2.1a) score corresponding to a “patient acceptable symptom state”. The absolute cutoff ODI value (threshold) was generally estimated to be ≤ 22 and seemed to be robust for different subgroups of patients and at different follow-up assessments. This threshold is similar to the “normal” value as recently defined by van Hooff et al. [11].

Our finding of the ODI threshold of ≤ 22 is in line with other studies. Tonosu et al. calculated the cutoff values for the presence or absence of LBP in a random sample of people working at a Japanese Internet research company (n=1,200). The authors found a similar threshold of $ODI \geq 22$ in those who had LBP and disability [23]. In a large heterogeneous sample (n=774) of inpatients with spinal disorders (eg, acute and chronic LBP, herniation, stenosis, scoliosis), a discharge threshold of $ODI \leq 30$ was found [24]. To our knowledge, the present study is the first to report ODI threshold values based on PASS in patients with degenerative lumbar spine disorders who have undergone lumbar spine surgery.

To quantify the PASS for the ODI, we used different external criteria as the anchor for a successful outcome. When performing the sensitivity analyses with PASS 2 (“very satisfied”), we found lower ODI thresholds ($ODI \leq 16-19$), but with similar satisfactory areas under the curve ($AUC > 0.7$; [Table 3](#)). Overall, the results of the sensitivity analysis turned out to be robust; a more stringent definition of success yielded lower ODI threshold values, meaning that the anchor to perform the present study was well chosen.

As the use of absolute thresholds might be population-dependent, we performed different sensitivity analyses for subgroups of patients. In the main analyses for PASS (“very satisfied” and “somewhat satisfied”), the AUCs were satisfactory ($AUC > 0.7$) [22]. This means that for all subgroups of patients studied, the ODI discriminates between acceptable and unacceptable symptom states compared with the anchor used. The overall ODI threshold for each subgroup was ≤ 22 . Only for “previous surgery (yes)” and “surgical procedure (decompression; commonly performed for lumbar disc herniation and lumbar spinal stenosis)” a slightly different ODI threshold was found (each $ODI \leq 23$; [Table 4A](#)). However, these differences in ODI thresholds were too small to be clinically meaningful. To establish whether the ODI score for the PASS was time independent, we provided estimations for the ODI thresholds for two different commonly used times of follow-up. We found no meaningful differences between the 1-year and 2-year follow-up thresholds: The absolute values were similar and the 95% confidence intervals of the AUCs showed overlap ([Table 3](#)). This means that the ODI-22 threshold associated with a satisfied symptom state remains consistent.

Limitations

Some limitations of the present study should be mentioned. First, the EUROSPINE Spine Tango Registry contains

Table 4
 (A) Sensitivity analyses for PASS 1 for subgroups of patients and (B) sensitivity analyses for PASS 2 for subgroups of patients

	1-year follow-up (n=780)							2-year follow-up (n=508)								
	n	AUC	95% CI	Absolute ODI threshold	Se %	Sp %	Chi-square value	p-value	n	AUC	95% CI	Absolute ODI threshold	Se %	Sp %	Chi-square value	p-value
A																
Gender							2.97	.085							0.01	.996
Male	373	0.91	0.88–0.94	22	81.6	81.9			243	0.91	0.87–0.94	22	81.3	85.5		
Female	407	0.87	0.83–0.91	22	76.0	82.3			265	0.88	0.84–0.92	22	80.0	85.7		
Age group							0.07	.795							0.91	.340
<65	475	0.89	0.86–0.92	22	78.4	85.3			317	0.90	0.86–0.93	22	79.4	81.7		
≥65	305	0.88	0.84–0.92	22	80.6	77.3			191	0.92	0.88–0.96	22	82.7	81.8		
Previous surgery							0.01	.917							0.12	.726
Yes	180	0.88	0.84–0.93	23	71.2	81.0			122	0.90	0.83–0.96	22	81.3	87.5		
No	600	0.88	0.85–0.91	22	77.8	82.6			386	0.91	0.88–0.94	22	80.6	84.9		
Surgical procedure							9.02	.003*							1.31	.252
Decompression	487	0.91	0.89–0.94	23	82.2	84.6			335	0.92	0.89–0.95	22	83.6	87.2		
Fusion	293	0.83	0.78–0.88	22	72.3	77.9			173	0.88	0.84–0.93	22	80.7	82.2		
Complications—surgical							0.02	.887							2.18	.872
Yes	36	0.89	0.78–0.99	22	72.7	80.0			31	0.96	0.89–1.00	20	90.7	79.0		
No	736	0.88	0.86–0.91	22	76.2	82.0			472	0.91	0.88–0.93	22	79.9	85.9		
Complications—patient							1.27	.259							0.01	.934
Yes	192	0.91	0.86–0.95	22	75.9	87.6			133	0.90	0.86–0.95	22	84.5	84.0		
No	588	0.88	0.85–0.90	22	76.5	80.4			375	0.91	0.88–0.94	24	81.2	82.9		
B																
Gender							0.97	.325							0.05	.825
Male	355	0.91	0.87–0.94	15	81.4	84.3			243	0.89	0.85–0.94	16	82.5	83.3		
Female	435	0.91	0.87–0.94	18	80.2	80.1			265	0.90	0.86–0.94	18	82.3	82.3		
Age group							0.01	.908							0.01	.984
<65	475	0.89	0.86–0.92	18	82.4	80.9			317	0.90	0.86–0.94	16	82.3	85.2		
≥65	305	0.89	0.85–0.93	16	82.1	79.0			191	0.90	0.84–0.94	18	80.0	81.5		
Previous surgery							8.20	.004*							0.80	.374
Yes	180	0.95	0.91–0.98	16	85.7	84.7			122	0.86	0.77–0.95	16	75.0	89.0		
No	600	0.87	0.84–0.90	16	79.3	81.0			386	0.91	0.87–0.94	16	82.5	83.0		
Surgical procedure							0.95	.230							0.44	.507
Decompression	487	0.92	0.87–0.93	16	81.7	80.7			335	0.90	0.86–0.94	16	82.9	85.0		
Fusion	293	0.87	0.82–0.92	18	80.6	79.2			173	0.88	0.83–0.93	16	78.1	83.9		
Complications—surgical							0.00	.968							0.17	.676
Yes	36	0.88	0.73–1.00	18	71.4	72.4			31	0.96	0.89–1.00	12	71.4	87.5		
No	736	0.89	0.86–0.92	16	80.8	81.6			472	0.91	0.88–0.93	18	81.2	80.9		
Complications—patient							0.04	.851							0.02	.893
Yes	192	0.91	0.85–0.97	18	91.4	80.4			133	0.90	0.88–0.94	14	84.4	88.1		
No	588	0.91	0.88–0.94	18	84.6	84.9			375	0.91	0.86–0.95	18	78.5	80.5		

PASS 1, patient acceptable symptom state with COMI SSWB answers “very satisfied” and “somewhat satisfied”; PASS 2, patient acceptable symptom state with COMI SSWB answer “very satisfied”; Se, sensitivity; Sp, specificity; AUC, area under the curve; ODI, Oswestry Disability Index; CI, confidence interval; COMI, Core Outcome Measures Index; SSWB, symptom-specific well-being.

* p<.05 for comparison of the AUC between subgroups.

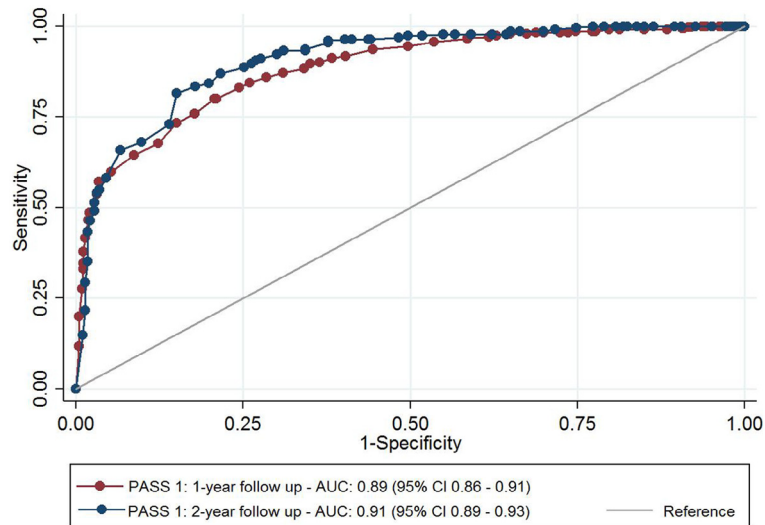


Fig. 2. ROC for patient acceptable symptom states (PASS 1) using 1- and 2-year follow-up ODI values. AUC, area under the curve; 95% CI, 95% confidence interval; PASS 1, Patient Acceptable Symptom State with “very satisfied” and “somewhat satisfied”; ODI, Oswestry Disability Index; ROC, receiver operating characteristic.

data from surgically treated patients in different centers and in various countries [16]. Participation in the registry is voluntary and the ODI is not a compulsory outcome measure for the majority of the centers, which instead use the COMI as standard. Therefore, selection bias might be introduced. The completion of the PROMs is dependent on patients' willingness to cooperate. In a previous registry study on Spine Tango data, it was suggested that the proportion of dural lesions in spinal stenosis was a good indicator of the completeness and the “honesty” of the data submitted [25]. The authors found comparable proportions to those seen in the Swedish SweSpine registry, which suggested credible or at least similarly honest reporting. Second, we performed several sensitivity analyses to test the robustness of the ODI-22 threshold. A previous study has shown that a higher preoperative pain level is associated with a higher acceptable pain level postoperatively [26]; as such, a sensitivity analysis including predefined subgroups with differing preoperative disability would have been interesting. However, the ODI is not a compulsory outcome measure in the registry and insufficient baseline data were available for such an analysis. Further research is needed to explore the influence of baseline ODI scores on the acceptable ODI score postoperatively. Moreover, we did not perform a specific subgroup analysis based on underlying pathologies of degenerative disorders of the lumbar spine. Although based on the results of a recently published study [26] a slightly lower threshold might be expected for lumbar disc herniation, the clinical relevance of such a difference is uncertain and might be arbitrary as this value could be regarded as a value seen in “normal” and healthy populations (ie, $ODI \leq 22$) [11]. Third, one of the participating centers used a version of the COMI that had not been formally translated and validated. However, we do not think this would have introduced any major bias, as we used only

the single “symptom-specific well-being” item of the COMI, and the original item [19,20] is simply and clearly stated and not likely to be subject to different interpretations. Fourth, to create the dataset for the current study, we started with 30,878 surgically treated patients. Because of the strict inclusion criteria, such as an ODI and COMI completed on the same day, a large amount of data were excluded, leaving us with 780 and 508 patients at the 1-year and 2-year follow-ups, respectively. Moreover, in the present study, we used the data of the last available follow-up on both ODI and COMI. By using this criterion, we might have lost the data of patients who completed the questionnaires at both time-points, as only the last assessment was included. However, both samples were large enough for analyses and had similar baseline characteristics, with the exception of the proportion that had previous surgery. Fifth, the COMI SSWB item was used as an external anchor to assess the achievement of an acceptable symptom state at follow-up. Although the COMI has been shown to be valid for patients with degenerative disorders of the lumbar spine undergoing surgery [19,20], the single SSWB item was not specifically tested for its unique psychometric properties. To evaluate the validity of the SSWB item as an anchor, we carried out a post-hoc analysis of the strength of the correlations between the postoperative ODI scores and the responses on the SSWB (Pearson r) for the 1- and 2-year follow-up assessments: the coefficients were $r=0.73$ and $r=0.77$, respectively. These correlations are interpreted as “strong” ($0.51 \leq r \leq 1.00$ [27,28]). In evaluating the validity of transition ratings, a strong correlation ($r \geq 0.50$) with change scores in a relevant health-related quality of life questionnaire suggests that the transition rating may be used as an anchor [29]. By analogy, we believe that, with correlation coefficients ≥ 0.50 , the use of the COMI SSWB item as a reference standard (anchor) to define the ODI threshold for

an acceptable symptom state is justified. Finally, for the main analysis, we used the responses “very satisfied” and “somewhat satisfied” on the COMI SSWB item to define an acceptable symptom state. This was supported by the findings of a previous study in which ROC analysis using the original dichotomized PASS (acceptance of the “current state” [yes or no]) as the external criterion revealed that “very satisfied,” “somewhat satisfied,” and “neither satisfied/nor dissatisfied” were considered as “acceptable” [30], together with our considerations as to what should be considered minimally acceptable as a response for a “state for the rest of one’s life.” We performed a sensitivity analysis with an even more stringent definition, which, as expected, resulted in lower ODI cutoff values. However, the decision as to which definition of “acceptable state” should be used remains somewhat arbitrary.

Conclusions

In the present study, we determined that an absolute ODI value of ≤ 22 best indicated a satisfactory symptom state in a large sample of patients from the Spine Tango Registry who had undergone surgery for degenerative disorders of the lumbar spine. As this threshold appeared to be fairly consistent across subpopulations, we suggest that the same common threshold could be used for all degenerative lumbar spine disorders when defining whether a patient has reached an acceptable state after spine interventions (ie, is a “responder”). As the concepts of PASS and change are complementary, in line with another group [31,32], we recommend using this threshold alongside the commonly used “change score” measure in reporting success. We suggest to report both measures and that the results should be expressed as the proportions of patients achieving both measures at follow-up assessment.

Acknowledgments

The authors would like to thank Johanna E. Vriezekolk for her critical and independent review of the analyses performed and the final version of the manuscript. The participants of the EUROSPINE Spine Tango Register are acknowledged for their continuous contribution that makes these studies possible, reflecting the daily practice of spine surgeons. The data of the following centers were used (in alphabetical order of country, city, hospital, and department): Department of Spinal Surgery in Royal Adelaide Hospital (Australia); Department of Orthopedic Surgery and Traumatology in the University Hospital of Cologne (Germany); Department of Special Spine Surgery in Leopoldina Hospital of Schweinfurt (Germany); Department of Orthopaedic Surgery in Tan Tock Seng Hospital Singapore (Singapore); Department of Orthopedic Surgery in the University Hospital of Ljubljana (Slovenia); Department of Neurosurgery in Bethesda Hospital of Basel (Switzerland); Department of Orthopedic Surgery in the Salem Hospital of Bern (Switzerland); Department of Spine Surgery in The Spine Center Thun (Switzerland); MEMdoc Hospital of

Thun-Simmental AG (Switzerland); Spine Unit of Nuffield Oxford Centre (United Kingdom); and Division of Spine Surgery in NYU Hospital for Joint Diseases of New York (United States).

References

- [1] Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386:743–800.
- [2] Poolman RW, Swiontkowski MF, Fairbank JC, Schemitsch EH, Sprague S, de Vet HC. Outcome instruments: rationale for their use. *J Bone Joint Surg Am* 2009;91(Suppl. 3):41–9.
- [3] Fairbank JC, Couper J, Davies JB, O’Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3.
- [4] Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine* 2000;25:2940–52.
- [5] PROQOLID. Patient-Reported Outcome and Quality of Life Instruments Database—Oswestry Disability Index (ODI) [Mapi Research Trust], 2014. Available at: <https://eprovide.mapi-trust.org/>. Accessed July 2016.
- [6] Clement RC, Welander A, Stowell C, Cha TD, Chen JL, Davies M, et al. A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta Orthop* 2015;86:523–33.
- [7] Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCI/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? *Ann Rheum Dis* 2007;66(Suppl. 3):iii40–1.
- [8] van der Roer N, Ostelo RW, Bekkering GE, van Tulder MW, de Vet HC. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine* 2006;31:578–82.
- [9] Kirwan JR. Minimum clinically important difference: the crock of gold at the end of the rainbow? *J Rheumatol* 2001;28:439–44.
- [10] de Vet HC, Foumani M, Scholten MA, Jacobs WC, Stiggelbout AM, Knol DL, et al. Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. *J Clin Epidemiol* 2015;68:518–24.
- [11] van Hooff ML, Spruit M, O’Dowd JK, van Lankveld W, Fairbank JC, van Limbeek J. Predictive factors for successful clinical outcome 1 year after an intensive combined physical and psychological programme for chronic low back pain. *Eur Spine J* 2014;23:102–12.
- [12] Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheum Dis* 2005;64:34–7.
- [13] Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, Bombardier C, et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. *Arthritis Care Res (Hoboken)* 2012;64:1699–707.
- [14] Dougados M. It’s good to feel better but it’s better to feel good. *J Rheumatol* 2005;32:1–2.
- [15] Roder C, Chavanne A, Mannion AF, Grob D, Aebi M. SSE Spine Tango—content, workflow, set-up. www.eurospine.org-Spine Tango. *Eur Spine J* 2005;14:920–4.
- [16] Melloh M, Staub L, Aghayev E, Zweig T, Barz T, Theis JC, et al. The international spine registry SPINE TANGO: status quo and first results. *Eur Spine J* 2008;17:1201–9.
- [17] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003;138:W1–12.

- [18] Mannion AF, Elfering A, Staerke R, Junge A, Grob D, Semmer NK, et al. Outcome assessment in low back pain: how low can you go? *Eur Spine J* 2005;14:1014–26.
- [19] Mannion AF, Porchet F, Kleinstuck FS, Lattig F, Jeszenszky D, et al. The quality of spine surgery from the patient's perspective. Part 1: the Core Outcome Measures Index in clinical practice. *Eur Spine J* 2009;18(Suppl. 3):367–73.
- [20] Mannion AF, Porchet F, Kleinstuck FS, Lattig F, Jeszenszky D, Bartanusz V, et al. The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. *Eur Spine J* 2009;18(Suppl. 3):374–9.
- [21] Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. *BMJ* 1994;309:188.
- [22] de Vet HC, Ostelo RW, Terwee CB, van der Roer N, Knol DL, Beckerman H, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res* 2007;16:131–42.
- [23] Tonosu J, Takeshita K, Hara N, Matsudaira K, Kato S, Masuda K, et al. The normative score and the cut-off value of the Oswestry Disability Index (ODI). *Eur Spine J* 2012;21:1596–602.
- [24] Park SW, Shin YS, Kim HJ, Lee JH, Shin JS, Ha IH. The dischargeable cut-off score of Oswestry Disability Index (ODI) in the inpatient care for low back pain with disability. *Eur Spine J* 2014;23:2090–6.
- [25] Munting E, Roder C, Sobottke R, Dietrich D, Aghayev E. 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* 2015;24:358–68.
- [26] Fekete TF, Haschtmann D, Kleinstuck FS, Porchet F, Jeszenszky D, Mannion AF. What level of pain are patients happy to live with after surgery for lumbar degenerative disorders? *Spine J* 2016;16:S12–18.
- [27] Cohen J. *Statistical power analysis for the behavioral sciences*. Second ed. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers; 1988.
- [28] Cohen J. A power primer. *Psychol Bull* 1992;112:155–9.
- [29] Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. *J Clin Epidemiol* 2002;55:900–8.
- [30] Impellizzeri FM, Mannion AF, Naal FD, Hersche O, Leunig M. The early outcome of surgical treatment for femoroacetabular impingement: success depends on how you measure it. *Osteoarthritis Cartilage* 2012;20:638–45.
- [31] Escobar A, Gonzalez M, Quintana JM, Vrotsou K, Bilbao A, Herrera-Espineira C, et al. Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement. Identification of cut-off values. *Osteoarthritis Cartilage* 2012;20:87–92.
- [32] Escobar A, Riddle DL. Concordance between important change and acceptable symptom state following knee arthroplasty: the role of baseline scores. *Osteoarthritis Cartilage* 2014;22:1107–10.