



EUROSPINE 2017 FULL PAPER AWARD: Time to remove our rose-tinted spectacles: a candid appraisal of the relative success of surgery in over 4500 patients with degenerative disorders of the lumbar spine, hip or knee

Anne F. Mannion¹ · Franco M. Impellizzeri¹ · Michael Leunig² · Dezső Jeszenszy³ · Hans-Jürgen Becker³ · Daniel Haschtmann³ · Stefan Preiss² · Tamas F. Fekete³

Received: 11 August 2017 / Accepted: 6 January 2018 / Published online: 19 February 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Purpose Studies comparing the outcome of spine surgery with that of large-joint replacement report equivocal findings. The patient-reported outcome measures (PROMs) used in such studies are typically generic and may not be sufficiently sensitive to the successes/failures of treatment. This study compared different indices of “success” in patients undergoing surgery for degenerative disorders of the lumbar spine, hip, or knee, using a validated, multidimensional, and joint-specific PROM.

Methods Preoperatively and 12 months postoperatively, 4594 patients (3937 lumbar spine, 368 hip, 269 knee) undergoing first-time surgery completed a PROM that included the Core Outcome Measures Index (COMI) for the affected joint. The latter comprises a set of single items on pain, function, symptom-specific well-being, quality of life, and disability—all in relation to the specified joint problem. Other single-item ratings of treatment success were made 12 months postoperatively.

Results In multiple regression analyses, controlling for confounders, the mean improvement in COMI at 12 months was greatest for the hip patients and lowest for those with degenerative spinal deformity (= the statistical reference group) ($p < 0.05$). Compared with spinal deformity, the odds of achieving “success” were: higher for hip (OR 4.6; 95% CI 2.5–8.5) and knee (OR 4.0; 95% CI 2.1–7.7) (no difference between spine subgroups) for “satisfaction with care”; higher for hip (OR 16.9; 95% CI 7.3–39.6), knee (OR 6.3; 95% CI 3.4–11.6), degenerative spondylolisthesis (OR 1.6; 95% CI 1.2–2.2), and herniated disc (OR 1.7; 95% CI 1.2–2.4) for “global treatment outcome”; and higher for hip (OR 13.8; 95% CI 8.8–21.6), knee (OR 5.3; 95% CI 3.6–7.8), degenerative spondylolisthesis (OR 1.6; 95% CI 1.3–2.1), and herniated disc (1.5; 95% CI 1.1–2.0) for “patient-acceptable symptom state”. Patient-rated complications were the greatest in degenerative spinal deformity (29%) and the lowest in hip (18%).

Conclusions The current study is the largest of its kind and the first to use a common, but joint-specific instrument to report patient-reported outcomes after surgery for degenerative disorders of the spine, hip, or knee. The findings provide a sobering account of the significantly poorer outcomes after spine surgery compared with large-joint replacement. Further work is required to hone the indications and patient selection criteria for spine surgery. The data should be used to lobby research

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00586-018-5469-4>) contains supplementary material, which is available to authorized users.

✉ Anne F. Mannion
anne.mannion@yahoo.com

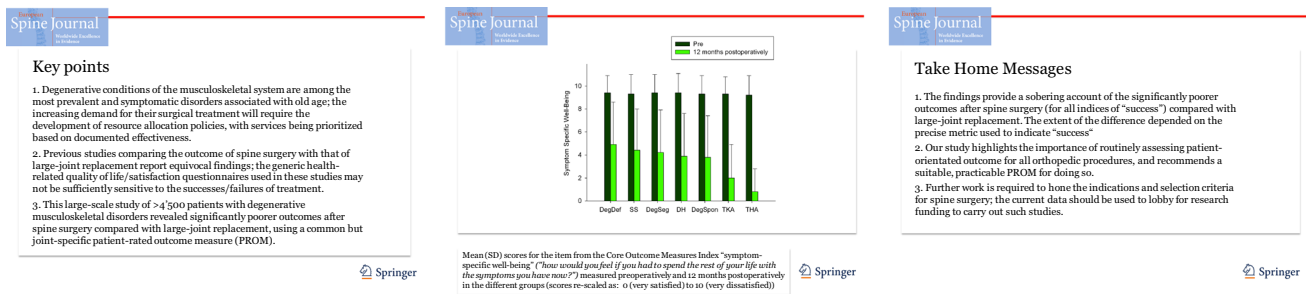
¹ Department of Teaching, Research and Development, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland

² Department of Hip and Knee Surgery, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland

³ Department of Spine Surgery and Neurosurgery, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland

funding-bodies, governmental agencies, industry, and charitable foundations to invest more in spine research/registries, in the hope of ultimately improving spine outcomes.

Graphical abstract These slides can be retrieved under Electronic Supplementary Material.



Keywords Spine · Hip · Knee · Surgery · Outcomes

Introduction

The most recent Global Burden of Disease study reveals that the single greatest cause of years lost with disability (YLD)—causing more than 146 million YLDs in 2013, an increase of 61% since 1990—is chronic low back pain [1]. Osteoarthritis was in 13th place. Population aging is clearly one of the reasons behind these statistics (<http://www.unfpa.org/ageing>). Degenerative conditions of the musculoskeletal system are among the most prevalent and symptomatic disorders associated with middle and old age [2] and the increasing number of aged people will hence be paralleled by an increase in the number suffering from degenerative joint diseases. Analyses suggest that patients of the 2020s will be more demanding of treatment and less willing to live with their symptoms than our current elderly [2]. This will result in a greater demand for elective surgery and a consequent need for resource allocation policies, with services being prioritized based on documented effectiveness. In turn, this requires that we embrace a more evaluative culture, with surgical procedures being systematically registered and scrutinized regarding their respective patient-oriented outcomes, using comparative effectiveness research (CER) methodology [3].

Anecdotally, lumbar spine surgery is believed to deliver poorer patient outcomes than total hip (THR) and knee (TKR) replacement surgery; however, the formal studies hitherto conducted in small, select groups of spine patients do not always substantiate this [4–12]. Notably, these studies have predominantly employed generic health-related quality of life (HRQL) outcome measures, such as the Short Form 36 (SF-36), which do not always adequately reflect the benefit of specific types of surgery [9, 13] and are less well able to discriminate between the successes and failures

of treatment [14, 15]. It is well known that the proportion of patients that can be considered a success after treatment depends very much on how success is defined, in terms of both the specific metric employed and the cutoff values applied [16, 17]. Although the spine and the large joints of the lower extremity are not comparable from a (patho) anatomical/physiological or biomechanical point of view, disorders in these regions impact the same “core domains” of importance to the patient (pain, function, quality of life, etc.), which allows them to be compared, given an appropriate set of questions that tap these domains. “Success” can be measured by the achievement of a “minimal clinically important change (MCIC) score” on the given outcome instrument, but this measure is not without its drawbacks. Firstly, it is influenced by baseline scores [18], and secondly, although indicating “improvement”, it does not necessarily tell us whether the patient is doing well in the end. Recent studies suggest that the patient’s achievement (or not) of an “acceptable symptom state” (PASS) may offer a more rigorous measure of success and better tease out differences between treatments [19]. Finally, enquiry as to the patient’s perspective on complications arising after surgery may provide a hitherto poorly investigated, but extremely important aspect of patient outcome [20].

The aim of the present study was to compare the outcomes after surgery in a large number of patients with different degenerative disorders of the lumbar spine, hip, or knee, using a brief patient-reported outcome measure (PROM) that includes the “Core Outcome Measures Index” (COMI). The latter is a set of single items on pain, function, symptom-specific well-being, quality of life, and disability, all in relation to the specified joint problem. It was developed and validated [21–23] following initial expert group recommendations for assessing outcomes in back patients [24], and then further validated for use in other joints/locations, including the hip and knee [25–27] and in many languages <http://www.eurospine.org/forms.htm>. The PROM also includes single

items on satisfaction with care, global treatment outcome, reoperations, and complications, all rated from the patient's perspective (as per the "Patient Self-Assessment form" of the EUROSPINE Spine Tango Registry). This parsimonious set hence provides a comprehensive and candid reflection of surgical outcome, in all the major areas that matter most to patients undergoing elective surgery.

Materials and methods

We carried out an analysis of PROM data, collected prospectively from patients operated in our orthopedic hospital between 2005 and 2014 (Spine) or in 2014 (hip and knee) and stored within our in-house Surgical Outcomes registry. The spine data were collected using the framework of the EUROSPINE Spine Tango Registry [28] <http://www.eurospine.org/spine-tango.htm>. We included patients with a good understanding of one of the languages in which the PROM was available [28] undergoing first-time surgery for the given joint/region of the spine (and unilaterally, for THR and TKR), and having reached at least 1 year postoperative. We included spine patients who had lumbar degenerative disorder as the main pathology (with "pain relief" indicated as one of the goals of surgery), further categorized as herniated disc, spinal stenosis without spondylolisthesis, degenerative spondylolisthesis, degenerative deformity, or degenerative disc/segment disease, according to the Spine Tango diagnostic groups algorithm (http://www.eurospine.org/cm_data/def_of_degen_patho.pdf). Patients underwent decompression and/or instrumented fusion as deemed appropriate by the treating surgeon. For the hip and knee groups, we took all patients undergoing total joint replacement for primary osteoarthritis. For the hip, direct anterior or posterior approaches were used, which are both abductor sparing; for the knee, classic approaches with or without tuberosity osteotomies were used. We had no exclusion criteria. We obtained ethics committee approval for the re-use of routinely collected data.

Questionnaire

The PROM we used is shown in Table 1. Preoperatively, it comprised the Core Outcome Measures Index (COMI) [21–23], a validated, multidimensional index (scored 0–10) containing one question for each domain, formulated in relation to the affected joint/region. At 12 months' follow-up, in addition to the COMI, the PROM included a range of single-item measures of treatment success (Table 1).

Patients completed the questionnaire at home, having received it from the research department by post, so the information given was free of care provider influence.

Statistical analysis

We present descriptive data as mean \pm standard deviation (SD) or percentages. We used crosstabs/contingency analyses with Chi squared to assess the significance of differences between group distributions for various nominal variables.

We applied multiple linear regression analysis to investigate the influence of group (coded as dummy variables) on the COMI score at 12 months' follow-up (FU) while controlling for potential confounders [age, sex, BMI (in categories corresponding to those of the Spine Tango surgery form; < 20, 20–25, 26–30, 31–35, > 35), smoking status (no, yes), and American Society of Anesthesiologists Physical Status Score (ASA 1–5)].

We used multivariable logistic regression analysis to evaluate the influence of pathology/group on different indices of "success", controlling for the aforementioned confounders. The indices included: achievement of MCIC (reduction of ≥ 2.2 points between the preoperative and postoperative COMI score [29]); achievement of a good global treatment outcome (operation helped/helped a lot); satisfaction with care (satisfied/very satisfied); achievement of an acceptable symptom state (satisfied/very satisfied to spend rest of life with current symptoms).

We set an alpha level for the analyses of 0.05, and used SPSS version 22 (IBM Corp., USA, 2013) and StatView 5.0 software (StatView, SAS Inc., Berkeley, CA) for the analyses.

Results

Final study group

A total of 4594 patients (3937 spine (with different pathologies), 368 hip and 269 knee) fulfilled the study inclusion criteria (Fig. 1; Table 2). Women predominated in the degenerative spondylolisthesis, degenerative deformity, degenerative segment, and knee groups, while men predominated in the stenosis, herniated disc, and hip groups. Patients were significantly younger and had lower ASA grades in the herniated disc and degenerative segment groups. A greater proportion of patients were overweight or obese ($BMI > 25 \text{ kg m}^{-2}$) in the knee and spinal stenosis groups. There were more smokers in the herniated disc and degenerative segment groups, and fewer in the hip and knee groups.

Group differences in baseline COMI scores

Preoperatively, the hip group showed slightly but significantly lower scores (i.e., a better status) for the COMI sum

Table 1 Items and response options for the PROM used in the study

Items	Response options
COMI [21–23] (items 1–6) completed at baseline and 12 months postoperatively	
(1a) Pain symptoms How severe was your back (<i>hip/knee</i>) pain in the last week?	Graphic rating scale from 0 (no pain) to 10 (worst pain I can imagine)
(1b) Pain symptoms (spine only) How severe was your leg pain (<i>sciatica</i>)/ buttock pain in the last week?	Graphic rating scale from 0 (no pain) to 10 (worst pain I can imagine)
(2) Function During the past week , how much did your back (<i>hip/knee</i>) problem interfere with your normal work (including both work outside the home and housework)?	5-point scale (1) not at all; (2) a little bit; (3) moderately; (4) quite a bit; (5) extremely
(3) Symptom-specific well-being ^a If you had to spend the rest of your life with the symptoms you have right now , how would you feel about it?	5-point scale: (1) very satisfied; (2) somewhat satisfied; (3) neither satisfied nor dissatisfied; (4) somewhat dissatisfied; (5) very dissatisfied
(4) General quality of life: Please reflect on the last week . How would you rate your quality of life?	5-point scale: (1) very good; (2) good; (3) moderate; (4) bad; (5) very bad
(5) Social disability During the past 4 weeks how many days did you cut down on the things you usually do (work, housework, school, recreational activities) because of your back (<i>hip/knee</i>) problem?	5-point scale: (1) none; (2) between 1 and 7 days; (3) between 8 and 14 days; (4) between 15 and 21 days; (5) more than 21 days
(6) Work disability During the past 4 weeks how many days did your back (<i>hip/knee</i>) problem keep you from going to work (job, school, housework)?	5-point scale: (1) none; (2) between 1 and 7 days; (3) between 8 and 14 days; (4) between 15 and 21 days; (5) more than 21 days
Patient-reported indices of success measured 12 months postoperatively	
(7a) Did any complications arise as a consequence of your operation 1 year ago (e.g., problems with wound healing, paralysis, sensory disturbances (<i>hip/knee: e.g., wound/bone healing, injury to nerves/blood vessels, infection, etc.</i>)? [36]	(1) No; (2) yes If yes, please describe these... (<i>free text</i>)
(7b) How bothersome were these complications?	5-point scale: (1) not at all bothersome; (2) slightly bothersome; (3) moderately bothersome; (4) very bothersome; (5) extremely bothersome
(8) Since the operation 1 year ago, have you had any further operation(s) on your lumbar spine (back) (<i>hip/knee</i>)?	(1) No; (2) yes, but at a different level of spine (<i>yes but on the other hip/knee</i>); (3) yes, at the same level of the spine/same segment (<i>same hip/knee</i>)
(9) Over the course of treatment for your back (<i>hip/knee</i>) problem, how satisfied were you with your overall medical care in our hospital? [24]	5-point scale: (1) very satisfied; (2) somewhat satisfied; (3) neither satisfied nor dissatisfied; (4) somewhat dissatisfied; (5) very dissatisfied
(10) Overall, how much did the operation that you had 1 year ago help your back (<i>hip/knee</i>) problem? [29]	5-point scale: (1) helped a lot; (2) helped; (3) helped only a little; (4) didn't help; (5) made things worse

The main wording is for the spine patients, and that for the hip/knee patients is given in italics in parentheses

SCORING: questions 1–6 comprise the Core Outcome Measures Index, scored as follows: the higher of the two pain scores (items 1a and 1b) is taken as the “pain” score (for hip/knee, simply take the pain score). The average of the two disability items (5 and 6) forms the “disability” score. In patients who are unable to conceive of anything that they do (housework, paid work, etc.) as being considered “work”, this item is not scored and the response to social disability represents the “disability” score. The five domain scores for “pain” (the higher of items 1a and 1b for spine patients), function (item 2), symptom-specific well-being (item 3), general quality of life (item 4), and “disability” (average of items 5 and 6) are then averaged to give a COMI score from 0 to 10. The five response options of the adjectival and Likert scales are scored as 0, 2.5, 5, 7.5, and 10, respectively

The 5-point scales for questions 3, 7b, 9, and 10 are dichotomized (top two categories vs bottom three categories) for some further analyses

^aThe “symptom-specific well-being” item is also used to determine the “patient-acceptable symptom state” (PASS), dichotomized as per above into “acceptable” (very satisfied/somewhat satisfied) and “not acceptable” (neither satisfied nor dissatisfied, somewhat dissatisfied, very dissatisfied)

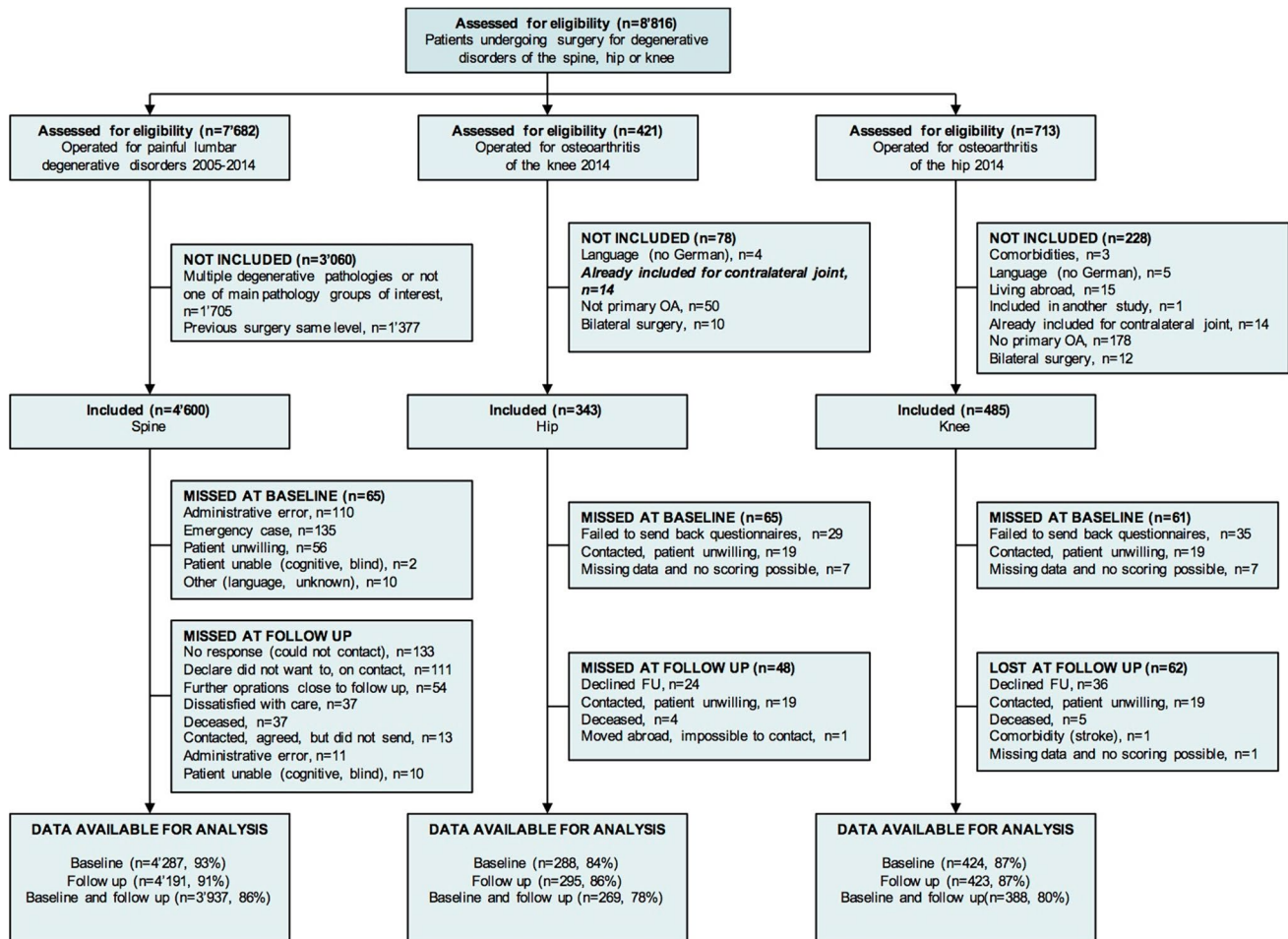


Fig. 1 Flowchart showing the flow of patients through the study

score and most of the COMI sub-domains than all other groups (Table 3). The knee group showed significantly lower scores than almost all the spine subgroups. The only COMI item for which scores did not differ significantly between the groups was “symptom-specific well-being”.

Group differences in the improvement in COMI from preoperatively to 12 months postoperatively

From preoperatively to 12 months’ postoperatively, there was a significant reduction in the COMI score (i.e., improvement in status) in each of the groups (Table 4). However, compared with the hip group, the magnitude of the improvement was significantly less in all other groups (ranging from 0.6 points less (for TKA) to 2.7 points less (for degenerative deformity), after adjusting for confounders; Table 3). A similar pattern was observed for all the COMI sub-domain scores (detailed results not shown; see example for symptom-specific well-being item in Fig. 2).

Group differences in dichotomized ratings of success at 12 months postoperatively

The achievement of the different indices of “success” are shown in Table 5.

“Satisfaction with medical care” gave the highest proportion of successful outcomes all round, at 96% for the hip and knee groups and between 83 and 90% for the spine subgroups. In adjusted analyses, the odds of being satisfied were approximately four to five times greater ($p < 0.0001$) for the hip and knee patients compared with the statistical reference group, spine degenerative deformity (with no difference between the spine subgroups).

A good “global treatment outcome” was reported by 95–98% hip/knee patients and 73% (stenosis) to 84% (herniated disc) of the spine patients; the figures were approximately 5–10% lower in each group for the respective proportions of patients achieving the MCIC for the COMI (Table 4). Adjusted analyses showed that, compared with spine degenerative deformity patients, the odds of a good

Table 2 Demographic data

	Cont. coef.	Whole group	Lumbar spine					Hip	Knee
			DegDef	DegSeg	DegSpon	HD	SS		
Number		4594	410	272	853	1099	1303	388	269
Age [years (SD)]		63 (15)	69 (11)	52 (14)#	68 (11)	49 (14)#	70 (10)	68 (10)	69 (9)
Sex	0.22								
Men		2131 (46%)	133 (32%)*	107 (39%)	247 (29%)*	613 (56%)*	722 (55%)*	202 (52%)	107 (40%)
Women		2463 (54%)	277 (68%)	165 (61%)	606 (71%)	486 (44%)	581 (45%)	168 (48%)	162 (60%)
ASA ^b	0.47								
I		1068 (23%)	38 (9%)**	107 (40%)**	98 (12%)**	644 (59%)**	138 (11%)**	33 (9%)**	10 (4%)**
II		2355 (52%)	217 (53%)	127 (48%)	505 (59%)	389 (36%)**	695 (54%)	254 (66%)	168 (63%)
III		1123 (25%)	149 (37%)	32 (12%)	248 (29%)	57 (5%)**	450 (35%)**	96 (25%)	91 (34%)
IV		18 (0.4%)	3 (1%)	1 (0%)	0 (0%)	0 (0%)	11 (1%)	3 (1%)	0 (0%)
BMI category ^c	0.20								
Normal (< 25)		1837 (45%)	179 (49%)	130 (54%)	351 (47%)	548 (56%)**	384 (34%)**	174 (45%)	71 (26%)*
Overweight (26–30)		1555 (38%)	125 (34%)	78 (32%)	259 (35%)	343 (35%)	488 (43%)	141 (37%)	121 (45%)
Obese (> 30)		734 (18%)	60 (17%)	35 (14%)	140 (19%)	97 (10%)**	254 (23%)	71 (18%)	77 (29%)
Smoking status ^a	0.20								
Yes		855 (22%)	61 (17%)	87 (37%)**	131 (19%)	295 (31%)**	207 (19%)	48 (12%)**	26 (10%)**

DegDef degenerative deformity, DegSeg degenerative segment, DegSpon degenerative spondylolisthesis, HD herniated disc, SS spinal stenosis without spondylolisthesis, Cont. coef. contingency coefficient

^a13% missing cases

^b1% missing cases

^c10% missing cases

#*p* < 0.001 compared to all other groups

p* < 0.05 and *p* < 0.001 indicate higher or lower counts than expected

Table 3 Preoperative COMI scores and sub-domain scores

COMI items	Lumbar spine					Hip (6)	Knee (7)
	DegDef (1)	DegSeg (2)	DegSpon (3)	HD (4)	SS (5)		
Pain	7.4 (1.9) ^{4,6,7}	7.3 (1.9) ^{6,7}	7.3 (1.9) ^{6,7}	7.0 (2.2) ^{1,6,7}	7.1 (2.1) ^{6,7}	6.0 (2.0) ^{1-5,7}	6.6 (1.9) ¹⁻⁶
Function	7.6 (2.0) ⁵⁻⁷	7.6 (2.2) ^{6,7}	7.3 (2.3) ^{6,7}	7.6 (2.4) ⁵⁻⁷	7.2 (2.4) ^{1,4,6,7}	5.5 (2.4) ^{1-5,7}	6.4 (2.3) ¹⁻⁶
Symptom-specific well-being	9.4 (1.5)	9.4 (1.6)	9.3 (1.6)	9.4 (1.7)	9.3 (1.7)	9.2 (1.7)	9.3 (1.5)
Quality of life	7.6 (1.9) ^{6,7}	7.4 (2.1) ^{6,7}	7.3 (2.0) ^{6,7}	7.5 (2.1) ^{6,7}	7.3 (2.0) ^{6,7}	5.8 (2.2) ¹⁻⁵	6.0 (2.3) ¹⁻⁵
Disability ^a	7.0 (3.3) ^{6,7}	6.8 (3.3) ^{6,7}	6.4 (3.5) ^{6,7}	7.1 (3.2) ⁵⁻⁷	6.6 (3.5) ^{4,6,7}	4.6 (3.4) ¹⁻⁵	5.1 (3.5) ¹⁻⁵
COMI index	7.8 (1.6) ^{6,7}	7.7 (1.7) ^{6,7}	7.5 (1.8) ^{6,7}	7.7 (1.7) ⁵⁻⁷	7.5 (1.8) ^{4,6,7}	6.2 (1.9) ^{1-5,7}	6.7 (1.8) ¹⁻⁶

Superscript numbers indicate the group differing at a *p* level < 0.0026 (Bonferroni adjustment)

DegDef degenerative deformity; DegSeg, degenerative segment; DegSpon, degenerative spondylolisthesis; HD, herniated disc; SS, spinal stenosis without spondylolisthesis

^aDisability is calculated from the average of the social and work disability items. In patients who are unable to conceive of anything that they do (housework, paid work, etc.) as being considered “work”, the work item is not scored and the response to social disability represents the “disability” score

global treatment outcome and achievement of MCIC were significantly higher (odds ratios 1.6–16.9; Table 5) for hip, knee, degenerative spondylolisthesis, and herniated disc patients.

The proportion of patients achieving the patient-acceptable symptom state (PASS; item 3 in Table 1) was the outcome that varied most widely between the groups, from as low as 44% for the spine degenerative deformity

Table 4 Group differences in the improvement in COMI from preoperatively to 12 months postoperatively

Groups	Preoperatively Unadjusted mean (SD)	Postoperatively* Unadjusted mean (SD)	Adjusted ^a difference in change score compared with hip group ^b (95% CI)	<i>p</i> value
Hip	6.2 (1.9)	0.8 (1.4)	Reference group	
Knee	6.7 (1.8)	1.7 (1.9)	0.6 (0.2–1.0)	0.006
Lumbar HD	7.7 (1.7)	3.1 (2.7)	1.9 (1.6–2.2)	< 0.001
Lumbar DegSpon	7.5 (1.8)	3.3 (2.7)	2.0 (1.7–2.3)	< 0.001
Lumbar DegSeg	7.7 (1.7)	3.6 (2.9)	2.4 (1.9–2.8)	< 0.001
Lumbar SS	7.5 (1.8)	3.8 (2.8)	2.5 (2.2–2.8)	< 0.001
Lumbar DegDef	7.8 (1.6)	4.1 (2.8)	2.7 (2.4–3.1)	< 0.001

DegDef degenerative deformity, *DegSeg* degenerative segment, *DegSpon* degenerative spondylolisthesis, *HD* herniated disc, *SS* spinal stenosis without spondylolisthesis

^aAdjusted for gender, age, body mass index, ASA, smoking status, and preoperative COMI scores

^bDifference compared to the change score of the hip group; e.g., the knee group improved 0.6 points less than the hip group

*Postoperative COMI scores are all significantly lower ($p < 0.001$) than pre scores

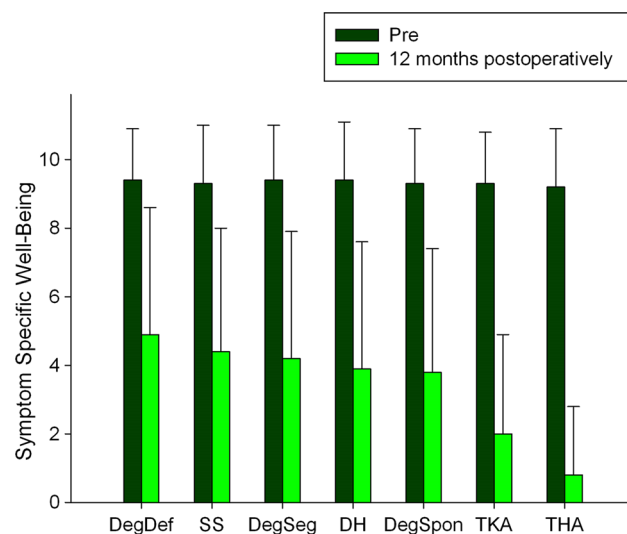


Fig. 2 Mean (SD) scores for “symptom-specific well-being” measured preoperatively and 12 months postoperatively in the different groups (scores re-scaled as 0–10)

group to 93% for the hip group. Again, adjusted analyses showed that, compared with degenerative deformity patients, the odds of achieving PASS were significantly higher for hip and knee patients (odds ratios, 13.8 and 5.3, respectively; Table 4), and for herniated disc and degenerative spondylolisthesis patients (odds ratios 1.4 and 1.6, respectively; Table 5).

Group differences in patient-reported complications and reoperations

Overall, 25% of patients reported that some type of complication had arisen as a consequence of their operation 1 year earlier. The highest rate (29%) was seen in patients with degenerative deformity and the lowest (18%) in the hip patients (Table 6). Approximately half of the hip and knee patients reported that the complications were at least “moderately bothersome”; the corresponding figure was around 70% for patients with degenerative deformity, degenerative segment or herniated disc, and 80% for those with degenerative spondylolisthesis and spinal stenosis (Table 6). The two most common patient-reported complications in spine, hip, and knee patients alike were sensory disturbances (20–23% of all complications reported) and continuing/new pain (23–42% of all complications) (Table 7). Problems with wound healing were more commonly reported by hip and knee patients (11–16% of complications) than spine patients (7%).

The presence of a complication had a significant negative association with satisfaction and with achievement of a good global treatment outcome, MCIC and PASS (each $p < 0.0001$). Patient-reported reoperation rates did not differ significantly between the groups.

Discussion

When PROMs are used to serially evaluate the course of change in large numbers of patients, brief instruments ease administration, reduce respondent burden, increase completion rates, and lower costs. These factors are critical to the adoption of systematic evaluation in routine practice [30]. Using a very brief, multidimensional instrument to cover

Table 5 Proportion of patients perceiving a successful surgery according to different criteria

Groups	Satisfaction with care		Good GTO		Achieving MCIC		Achieving PASS	
	%	OR (95% CI)	%	OR (95% CI)	%	OR (95% CI)	%	OR (95% CI)
Lumbar DegDef	84	Reference group	75	Reference group	67	Reference group	44*	Reference group
Lumbar SS	83	0.97 (0.70–1.35)	73	0.93 (0.70–1.25)	67	1.17 (0.89–1.53)	48**	1.25 (0.96–1.61)
Lumbar DegSeg	88	1.44 (0.96–2.16)	81	1.20 (0.78–1.83)	74	1.31 (0.89–1.95)	53	1.44 (1.00–2.06)
Lumbar DegSpon	88	1.37 (0.95–1.98)	83	1.58 (1.15–2.18)	73	1.55 (1.16–2.08)	56	1.63 (1.25–2.13)
Lumbar HD	90	1.44 (0.96–2.16)	84	1.69 (1.19–2.40)	79	1.79 (1.30–2.46)	55	1.48 (1.11–1.97)
Knee	96	4.04 (2.10–7.74)	95	6.25 (3.37–11.59)	90	6.93(4.34–11.08)	81**	5.32 (3.63–7.80)
Hip	96	4.62 (2.50–8.52)	98**	16.9 (7.25–39.61)	93**	11.62 (7.21–18.72)	93**	13.79 (8.80–21.59)
Study group average	88		81		75		57	
Contingency coeff.	0.22		0.20		0.19		0.26	

Responses 1 or 2 to the satisfaction, GTO, and PASS, and a COMI change score above the MCIC indicated successful surgery

DegDef degenerative deformity, *DegSeg* degenerative segment, *DegSpon* degenerative spondylolisthesis, *HD* herniated disc, *SS* spinal stenosis without spondylolisthesis, *GTO* global treatment outcome, *MCIC* minimal clinically important change score for COMI, *PASS* patient-acceptable symptom state

* $p < 0.05$ and ** $p < 0.001$ indicate larger or lower counts than expected

Table 6 Patient-rated complications and reoperations within the first postoperative year

	Cont. coeff.	Study group average	Lumbar					Hip	Knee
			DegDef	DegSeg	DegSpon	HD	SS		
Reoperation < 1 year	0.07								
Reoperated same joint/segment		116 (3%)	14 (3%)	4 (2%)	14 (2%)	40 (4%)	40 (3%)	2 (0%)	2 (1%)
Presence of complication(s)	0.07								
Patient-rated complication		1158 (25%)	120 (29%)	66 (24%)	231 (27%)	260 (24%)	351 (27%)	69 (18%)	61 (23%)
Odds ratio for “no complications” (95% CI) ^a			Ref. group	1.24 (0.83–1.84)	1.05 (0.78–1.40)	1.46 (1.06–2.00)	1.08 (0.82–1.42)	1.65 (1.15–2.37)	1.33 (0.91–1.94)
Bothersomeness of complications	0.30								
Not at all bothersome		21 (2%)	2 (2%)	0 (0%)	3 (2%)	2 (1%)	2 (1%)	9 (13%)*	3 (5%)
Slightly bothersome		255 (26%)	28 (26%)	15 (31%)	38 (19%)	62 (28%)	56 (20%)	27 (39%)	29 (47%)
Moderately bothersome		286 (29%)	27 (26%)	13 (27%)	65 (32%)	66 (30%)	88 (31%)	13 (19%)	14 (23%)
Very bothersome		276 (28%)	38 (36%)	9 (18%)	64 (32%)	53 (24%)	95 (33%)	8 (12%)	9 (15%)
Extremely bothersome		155 (16%)	11 (10%)	12 (25%)	32 (16%)	36 (16%)	45 (16%)	12 (17%)	7 (11%)

DegDef degenerative deformity, *DegSeg* degenerative segment, *DegSpon* degenerative spondylolisthesis, *HD* herniated disc, *SS* spinal stenosis without spondylolisthesis

* $p < 0.05$; indicates larger or lower counts than expected

^aWhen the 95% CI does not include 1 the results are significant ($p < 0.05$)

all the core domains of importance to patients—including some novel and sensitive indices that are not included in existing joint-specific or generic instruments—we showed that the extent to which THR proved superior to TKR and

Table 7 Description of main patient-rated complications

Complications category ^a	Lumbar	Hip	Knee
Sensory	428 (42%)	14 (20%)	17 (28%)
Pain	234 (23%)	16 (23%)	11 (18%)
Motor	102 (10%)	1 (1%)	0 (0%)
Pain and sensory	75 (7%)	2 (3%)	2 (3%)
Problems with wound healing	68 (7%)	11 (16%)	7 (11%)
Motor, sensory	45 (4%)	0 (0%)	0 (0%)
Infection	18 (2%)	8 (11%)	1 (2%)
Internal medicine	13 (1%)	6 (9%)	0 (0%)
Implant failure	3 (0%)	3 (4%)	2 (3%)
Stiffness	1 (0%)	1 (1%)	3 (5%)
Inflammation	0 (0%)	0 (0%)	11 (18%)
Other	41 (4%)	8 (11%)	7 (11%)

^aPatient-rated complications categorized as described in Grob et al. [36]

spine surgery was highly sensitive to the method used to categorize success. This may explain some of the previous discrepancies in the literature based mainly on generic quality of life measures [4–12] or satisfaction [4]. Our PROM contained questions relevant to any patient with a painful musculoskeletal disorder, but formulated in reference to the particular joint problem. In this sense, it represented an instrument that was both joint specific (and hence highly responsive) and yet generic (and hence comparable between different musculoskeletal conditions). The open nature of the questions makes them similar to those in “patient-individualized” questionnaires, known to be especially responsive [31]. For example, the function item enquires about “your normal” work/housework, as opposed to listing specific activities (that the patient may or may not normally do). This serves to ensure that the items’ content is always relevant to the patient, making them more likely to be “shifting” items (i.e., susceptible to change) with effective treatment. From a practical perspective, a “one size fits all” instrument simplifies data collection and comparative analyses, making it useful not only in research but also in hospital-wide quality control and benchmarking activities. These are the reasons underlying the decision of EUROSPINE, the Spine Society of Europe, to adopt the PROM used in the present study for its Spine Tango Spine Surgery Registry, for all spinal pathologies alike.

THR is considered to be one of the most successful orthopedic procedures available today [5], and the results of our study also substantiated this. It was in top place for all indices, including satisfaction with care, improvement, current state, patient-rated complications, and repeat surgery. However, the extent to which it distinguished itself from the other treatments clearly depended on the precise metric used. Scores on “satisfaction with care” showed the least

difference among the groups. The odds of being satisfied were the same for hip and knee patients, and were approximately fourfold those for the spine patients; however, all groups showed respectable figures, with greater than 83% patients being satisfied. Satisfaction with care, which is influenced by the patient–provider relationship and concerns treatment delivery, typically yields higher proportions of success than constructs focused on therapeutic improvement [23]. It can be an important concern in quality improvement initiatives—e.g., to document that patients are not disgruntled, when popular but ineffective treatment or unnecessary imaging is denied [24]—but its external validity/generalizability may be limited. The effectiveness of a procedure can be measured as either “the extent of improvement” (doing better) or the “actual state” (doing well) following treatment. Our indices of improvement teased apart further differences between the groups, with > 95% of hip and knee patients reporting a good global treatment outcome, compared with 73–84% spine patients. A difference of this size, if it were seen between two treatments for the same condition, would yield a highly relevant and clinically significant “number needed to treat” of as few as five. Compared with the hip group, the improvement in COMI score 12 months postoperatively was significantly lower for all other groups, by 0.6 points for knee and by 2–3 points for spine patients, the latter again being a clinically relevant difference [21].

Of all the indices, the PASS was the index that revealed the lowest rates of success for all pathologies and the greatest differences between the pathologies, with only about half of the patients in the spine group achieving an acceptable symptom state, compared with 81% of the knee and 93% of the hip patients. This highlights the fact that even large and statistically significant improvements in outcome scores do not necessarily mean that an acceptable state is reached in the end. This is perhaps our most poignant take-home message. Compared with the stringent measures used in the present study—measures that capture the impact of the precise symptoms surgery aims to relieve—it seems that the quality of life instruments (SF-36, EQ 5D) used in previous studies may have been too generic and too insensitive, resulting in their painting an overly rosy picture of what spine surgery can achieve. That these findings are not peculiar to our own hospital, but are instead typical of those reported on a wider geographical basis, is shown by the similarity of the results reported for other spine units in the EUROSPINE Spine Tango Registry [28, 32].

The success of surgery seemed to diminish in line with the increasing complexity of the “motion segment” (hip, knee, spine). Multisegmental spine pathology (present in 50% of our spine patients) might serve to increase the complexity again, as might previous surgery at a different spinal level (12% patients). Problems with the hip and knee are often unilateral and may be relieved by resting

the joint, whereas disorders of the centrally located spine may result in limitations that are more difficult to cope with. The common involvement of the neurologic system in spine patients is a confounding factor that is generally not an issue in hip or knee arthritis. Baseline status was significantly better in THR patients, as has been reported before [33], suggesting the threshold for surgery was lower or the impact of the problem was less severe; this may provide for better chances of full recovery postoperatively. The indications are usually much clearer for large-joint replacement than for most spine surgery. In spine, it is generally the case that the greater the concordance between symptoms, multimodal imaging (X-ray, MRI), and the rationale for the planned procedure, the better is the outcome. There are clear subsets of spine patients that benefit more from a given surgery—e.g., herniated disc patients with greater leg pain than back pain undergoing decompression [34] and discogenic pain patients with a distinct pain pattern undergoing fusion [35]—and we need to better identify such subgroups.

In summary, there are numerous reasons why outcomes are, and can be expected to be, significantly worse for spine surgery than for THR or TKR. However, if this is not exposed by the use of sensitive and stringent measures, and we instead elect to believe the generic quality of life data that suggest comparable or even superior results for spine surgery [4, 5, 8], then we will fail to seek and attract the necessary investment in research to improve the situation. Our study shows the importance of routinely assessing patient-orientated outcome for all orthopedic procedures, and recommends a suitable, practicable instrument for doing so. The findings should provide the evidence required to lobby research funding bodies, governmental agencies, industry, and charitable foundations to make greater investments in spine registries and spine research, in the hope of ultimately improving spine patient-rated outcomes.

Acknowledgements We thank all the patients of the Schulthess Klinik who have contributed data to our in-house registry. We thank François Porchet, Frank Kleinstück, and all present and past surgeons working at the Schulthess Klinik Spine Center for their compliance with the surgical forms of the Spine Tango Registry. We thank Dave O’Riordan, Gordana Balaban, Julian Amacker, Kirsten Clift, Sara Preziosa, Stéphanie Dosch, Riccardo Curatolo, Selina Nauer, Danica Mauz, Vanessa Wellhauer, and Myrta Viloz for the administration of the surgical forms and patient-rated outcome measures in our registry and for the preparation of the data files for analysis.

Funding The study was approved by the local ethics committee.

Compliance with ethical standards

Conflict of interest The authors have no conflict of interest concerning the materials or methods used in this study or the findings described in this paper.

References

- Vos T (2015) 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* 386:743–800. [https://doi.org/10.1016/s0140-6736\(15\)60692-4](https://doi.org/10.1016/s0140-6736(15)60692-4)
- Buckwalter JA, Heckman JD, Petrie DP, AOA (2003) An AOA critical issue: aging of the North American population: new challenges for orthopaedics. *J Bone Jt Surg Am* 85-A:748–758
- Issues and Options for an Expanded Federal Role (2017) Congress of the United States, Congressional Budget Office Publication Number 2975. <https://www.cbo.gov/sites/default/files/110th-congress-2007-2008/reports/12-18-comparativeeffectiveness.pdf>. Accessed 10 March 2017
- Lim JB, Chou AC, Yeo W, Lo NN, Chia SL, Chin PL, Tay DK, Yeo SJ (2015) Comparison of patient quality of life scores and satisfaction after common orthopedic surgical interventions. *Eur J Orthop Surg Traumatol* 25:1007–1012. <https://doi.org/10.1007/s00590-015-1635-0>
- Juul O, Sigmundsson FG, Ovesen O, Andersen MO, Ernst C, Thomsen K (2006) No difference in health-related quality of life in hip osteoarthritis compared to degenerative lumbar instability at pre- and 1-year postoperatively: a prospective study of 101 patients. *Acta Orthop* 77:748–754. <https://doi.org/10.1080/17453670610012935>
- Rampersaud YR, Ravi B, Lewis SJ, Stas V, Barron R, Davey R, Mahomed N (2007) Assessment of health-related quality of life after surgical treatment of focal symptomatic spinal stenosis compared with osteoarthritis of the hip or knee. *Spine J* 8:296–304. <https://doi.org/10.1016/j.spinee.2007.05.003>
- Rampersaud YR, Wai EK, Fisher CG, Yee AJ, Dvorak MF, Finkelstein JA, Gandhi R, Abraham EP, Lewis SJ, Alexander DI, Oxner WM, Davey JR, Mahomed N (2011) Postoperative improvement in health-related quality of life: a national comparison of surgical treatment for focal (one- to two-level) lumbar spinal stenosis compared with total joint arthroplasty for osteoarthritis. *Spine J* 11:1033–1041. <https://doi.org/10.1016/j.spinee.2011.10.011>
- Mokhtar SA, McCombe PF, Williamson OD, Morgan MK, White GJ, Sears WR (2010) Health-related quality of life: a comparison of outcomes after lumbar fusion for degenerative spondylolisthesis with large joint replacement surgery and population norms. *Spine J* 10(4):306–312. <https://doi.org/10.1016/j.spinee.2010.01.018>
- Hozack WJ, Rothman RH, Albert TJ, Balderston RA, Eng K (1997) Relationship of total hip arthroplasty outcomes to other orthopaedic procedures. *Clin Orthop Relat Res* 344:88–93
- Hansson T, Hansson E, Malchau H (2008) Utility of spine surgery: a comparison of common elective orthopaedic surgical procedures. *Spine (Phila Pa 1976)* 33:2819–2830. <https://doi.org/10.1097/brs.0b013e31818e2914>
- Rampersaud YR, Lewis SJ, Davey JR, Gandhi R, Mahomed NN (2014) Comparative outcomes and cost-utility after surgical treatment of focal lumbar spinal stenosis compared with osteoarthritis of the hip or knee—part 1: long-term change in health-related quality of life. *Spine J* 14:234–243. <https://doi.org/10.1016/j.spinee.2013.12.010>
- Tso P, Walker K, Mahomed N, Coyte PC, Rampersaud YR (2012) Comparison of lifetime incremental cost:utility ratios of surgery relative to failed medical management for the treatment of hip, knee and spine osteoarthritis modelled using 2-year postsurgical values. *Can J Surg* 55:181–190. <https://doi.org/10.1503/cjs.033910>
- Dawson J, Fitzpatrick R, Murray D, Carr A (1996) The problem of ‘noise’ in monitoring patient-based outcomes: generic, disease-specific and site-specific instruments for total hip replacement. *J Health Serv Res Policy* 1:224–231

14. Suarez-Almazor ME, Kendall C, Johnson JA, Skeith K, Vincent D (2000) Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. *Rheumatology (Oxford)* 39:783–790
15. Fanuele JC, Birkmeyer NJ, Abdu WA, Tosteson TD, Weinstein JN (2000) The impact of spinal problems on the health status of patients: have we underestimated the effect? *Spine* 25:1509–1514
16. Copay AG, Martin MM, Subach BR, Carreon LY, Glassman SD, Schuler TC, Berven S (2010) Assessment of spine surgery outcomes: inconsistency of change amongst outcome measurements. *Spine J* 10:291–296
17. Howe J, Frymoyer JW (1985) The effects of questionnaire design on the determination of end results in lumbar spinal surgery. *Spine* 10:804–805
18. Judge A, Arden NK, Kiran A, Price A, Javaid MK, Beard D, Murray D, Field RE (2012) Interpretation of patient-reported outcomes for hip and knee replacement surgery: identification of thresholds associated with satisfaction with surgery. *J Bone Jt Surg Br* 94:412–418. <https://doi.org/10.1302/0301-620x.94b3.27425>
19. Fekete TF, Haschtmann D, Kleinstueck FS, Porchet F, Jeszenszky D, Mannion AF (2016) What level of pain are patients happy to live with after surgery for lumbar degenerative disorders? *Spine J* 16:S12–S18. <https://doi.org/10.1016/j.spinee.2016.01.180>
20. Stauff MP, Cheng I (2013) Complications: a critical component of patient outcome. *Spine J* 13:625–627. <https://doi.org/10.1016/j.spinee.2013.03.008>
21. Mannion AF, Elfving A, Staerkle R, Junge A, Grob D, Semmer NK, Jacobshagen N, Dvorak J, Boos N (2005) Outcome assessment in low back pain: how low can you go? *Eur Spine J* 14:1014–1026
22. Ferrer M, Pellise F, Escudero O, Alvarez L, Pont A, Alonso J, Deyo R (2006) Validation of a minimum outcome core set in the evaluation of patients with back pain. *Spine* 31:1372–1379 (**discussion 1380**)
23. Mannion AF, Porchet F, Kleinstueck F, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective: part 1. The Core Outcome Measures Index (COMI) in clinical practice. *Eur Spine J* 18:367–373
24. Deyo RA, Battié M, Beurskens AJHM, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G (1998) Outcome measures for low back pain research. A proposal for standardized use. *Spine* 23:2003–2013
25. Impellizzeri FM, Mannion AF, Naal FD, Leunig M (2013) A Core Outcome Measures Index (COMI) for patients undergoing hip arthroplasty. *J Arthroplasty* 28:1681–1686. <https://doi.org/10.1016/j.arth.2013.01.014>
26. Staerkle RF, Villiger P (2011) Simple questionnaire for assessing core outcomes in inguinal hernia repair. *Br J Surg* 98:148–155
27. Impellizzeri FM, Leunig M, Preiss S, Gugli T, Mannion AF (2016) The use of the Core Outcome Measures Index (COMI) in patients undergoing total knee replacement. *Knee*. <https://doi.org/10.1016/j.knee.2016.11.016>
28. EUROSPINE (2016) <http://www.eurospine.org/spine-tango.htm>. Accessed 7 April 2016
29. Mannion AF, Porchet F, Kleinstueck FS, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) 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* 18:374–379
30. Falavigna A, Dozza DC, Teles AR, Wong CC, Barbagallo G, Brodke D, Al-Mutair A, Ghogawala Z, Riew KD (2017) Current Status of Worldwide Use of Patient-Reported Outcome Measures (PROMs) in Spine Care. *World Neurosurg* 108:328–335. <https://doi.org/10.1016/j.wneu.2017.09.002>
31. Donnelly C, Carswell A (2002) Individualized outcome measures: a review of the literature. *Can J Occup Ther* 69:84–94
32. van Hooff ML, Mannion AF, Staub LP, Ostelo RW, Fairbank JC (2016) 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. *Spine J* 16:1221–1230. <https://doi.org/10.1016/j.spinee.2016.06.010>
33. Becerra Fontal JA, Bago Granell J, Garre Olmo J, Roig Busquets R, Peris Prats F, Villanueva Leal C (2013) Evaluation of health-related quality of life in patients candidate for spine and other musculoskeletal surgery. *Eur Spine J* 22:1002–1009. <https://doi.org/10.1007/s00586-012-2617-0>
34. Kleinstueck FS, Fekete T, Jeszenszky D, Mannion AF, Grob D, Lattig F, Mutter U, Porchet F (2011) The outcome of decompression surgery for lumbar herniated disc is influenced by the level of concomitant preoperative low back pain. *Eur Spine J* 20:1166–1173
35. Nystrom B, Weber H, Schillberg B, Taube A (2017) Symptoms and signs possibly indicating segmental, discogenic pain. A fusion study with 18 years of follow-up. *Scand J Pain* 16:213–220. <https://doi.org/10.1016/j.sjpain.2016.10.007>
36. Grob D, Mannion AF (2009) The patient's perspective on complications after spine surgery. *Eur Spine J* 18:380–385