The Spine Journal 19 (2018) 655–661

Clinical Study

Implementation of patient-reported outcome measures in appropriateness criteria of surgery for degenerative lumbar scoliosis

Eva Jacobs, MD[a], Sander M.J. van Kuijk, PhD[b], Johannes M.R. Merk, MSc[c], Mieke Vandewall-Peeters[a], Liesbeth M.C. Jütten-Brouwer, MSc[a], Lodewijk W. van Rhijn, MD, PhD[a], Paul C. Willems, MD, PhD[a,*]

[a] *Department of Orthopaedic Surgery, Research School CAPHRI, Maastricht University Medical Center, P. Debyelaan 25, 6229 HX Maastricht, the Netherlands*

[b] *Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Center, P. Debyelaan 25, 6202 AZ, Maastricht, the Netherlands*

[c] *Department of Finance, School of Business and Economics, Maastricht University, Tongersestraat 53, 6211 LM, Maastricht, the Netherlands*

Received 31 May 2018; revised 20 September 2018; accepted 21 September 2018

Abstract

BACKGROUND CONTEXT: Degenerative lumbar scoliosis (DLS) is an increasingly common spinal disorder of which current management is characterized by a substantial variety in treatment advice. To improve evidence-based clinical decision-making and increase uniformity and transparency of care, the Scoliosis Research Society established appropriateness criteria for surgery for DLS. In these criteria, however, the patient perspective was not formally incorporated. Since patient perspective is an increasingly important consideration in informed decision-making, embedding patient-reported outcome measures (PROMs) in the appropriateness criteria would allow for an objective and transparent patient-centered approach.

PURPOSE: To evaluate the extent that patient perspective is integrated into the appropriateness criteria of surgery for DLS.

STUDY DESIGN: Single center, retrospective, cohort study.

PATIENT SAMPLE: 150 patients with symptomatic degenerative lumbar scoliosis.

OUTCOME MEASURES: The association between appropriateness for surgery and various PROMs [Visual Analogue Scale for pain, Short Form 36 (SF-36), Pain Catastrophizing Scale (PCS), Hospital Anxiety Depression Scale (HADS), and Oswestry Disability Index (ODI)].

METHODS: Medical records of all patients with symptomatic DLS were reviewed and scored according to the appropriateness criteria. To assess the association between the appropriateness criteria and the validated PROMs, analysis of variance was used to test for differences in PROMS for each of the three categories resulting from the appropriateness criteria. To assess how well PROMs can discriminate between appropriate and inappropriate, we used a logistic regression analysis. Discriminative ability was subsequently determined by computing the area under the curve (AUC), resulting from the logistic regression analysis. Spearman rank analysis was used to establish a correlation pattern between the PROMs used and the appropriateness criteria.

RESULTS: There was a significant association between the appropriateness of surgery and the PROMs. The discriminative ability for appropriateness of surgery for PROMs as a group was strong (AUC of 0.83). However, when considered in isolation, the predictive power of any

FDA drug /drug status: Not applicable.

Author disclosures: ***EJ:*** Nothing to disclose. ***SMJVK:*** Nothing to disclose. ***JMRM:*** Nothing to disclose. ***MVP:*** Nothing to disclose. ***LMCJB:*** Nothing to disclose. ***LWVR:*** Nothing to disclose. ***PCW:*** Nothing to disclose.

The disclosure key can be found on the Table of Contents and at www.TheSpineJournalOnline.com.

* Corresponding author. Department of Orthopaedic Surgery, Research School CAPHRI, Maastricht University Medical Center, P. Debyelaan 25, 6229 HX Maastricht, the Netherlands.

E-mail address: p.willems@mumc.nl (P.C. Willems).

https://doi.org/10.1016/j.spinee.2018.09.012

individual PROMs was poor. The different categories of the appropriateness criteria significantly coincided with the PROMs used.
CONCLUSION: There is a statistically significant association between the appropriateness criteria of surgery for DLS and PROMs. Implementation of PROMs into the appropriateness criteria may lead to more transparent, quantifiable and uniform clinical decision making for DLS.

Keywords: Appropriateness criteria; Degenerative lumbar scoliosis; Patient reported outcome measures; PROMs; Surgical decision making

Introduction

The current management of de novo degenerative lumbar scoliosis (DLS) is characterized by a substantial variety of treatment options reflecting the absence of a clear, evidence-based, widespread stepwise approach [1]. Nonoperative treatment includes physical conditioning and exercise, pharmacological agents for pain control, use of orthotics, and pain interventions like epidural and facet injections [1–3]. Operative treatment can include instrumented stabilization and correction with posterior or anterior fusion, neurologic decompression, or a combination of these [1,3]. Satisfactory clinical results of operative interventions have been reported [4–6]. However, the overall rate of complications for such major invasive surgery in predominantly elderly patients can exceed 30% [4].

To improve evidence-based clinical decision-making as well as to increase uniformity, transparency, consistency and quality of care, the Scoliosis Research Society (SRS) assessed the appropriateness of surgery for DLS patients with specific clinical characteristics using the RAND/UCLA Appropriateness Method [1,7–9]. Appropriateness criteria are quantitative tools designed to guide decision-making for well-defined populations of patients based on a combination of the best available evidence and multidisciplinary expert opinion [1,7]. The aim of this method is to determine whether surgery for a particular patient is appropriate or not. When expected harms outweigh expected benefits, the procedure would be considered inappropriate [1].

Patient-reported outcome measures (PROMs) are now widely used in clinical practice to narrow the gap between the clinician's and patient's view of what constitutes a clinically meaningful improvement and to help tailor treatment plans to meet the patient's preferences and needs. PROMs quantify patients' perspectives on the severity of their symptoms, the impact of the disease on their functioning, and the degree to which it limits their health-related quality of life [10]. In the SRS appropriateness criteria for DLS, the patient perspective has not been formally incorporated [1]. The degree of symptoms in the appropriateness criteria is currently based on the patient's reported symptoms as they are perceived during outpatient assessment by the treating surgeon and is thereby susceptible for interpretation. Because patient perspective is an increasingly important consideration in informed decision-making, embedding PROMs into the appropriateness criteria for DLS would allow for a more quantifiable and transparent patient-centered approach in clinical decision-making.

The purpose of this study was to evaluate whether there is an association between the appropriateness criteria for degenerative lumbar scoliosis and the patient perspective as determined by validated health-related quality of life (HRQOL) instruments or PROMs.

Materials and Methods

For this retrospective study, the medical records of all consecutive DLS patients who had been treated at the multidisciplinary spine center of the Maastricht University Medical Centre (the Netherlands) between January 1, 2011 and December 31, 2017 were analyzed. Institutional review board (IRB) approval was obtained before study initiation (METC17-4-022). We included adult patients with de novo DLS with a Cobb angle of more than 10° in the coronal plane, available PROMs, and available standing two-direction full spine or lumbar spine plain radiographs [2,3]. Patients were excluded if they were suffering from idiopathic adolescent scoliosis or neuromuscular or congenital scoliosis, had undergone previous corrective spinal surgery and fusion, or complete data were unavailable.

For each patient, medical records and radiographs were reviewed and scored according to the SRS appropriateness criteria [1], which is comprised of seven clinical or radiographic characteristics: (1) severity of self-reported symptoms, (2) severity of central spinal or foraminal stenosis, (3) progression of the degree of curvature or certain other radiographic abnormalities, (4) presence of sagittal imbalance, (5) severity of risk factors for suboptimal outcomes, (6) degree of curvature, and (7) when applicable, the number of levels with at least moderate central spinal or foraminal stenosis. For the category "progression," plain radiographs at various time points were assessed (at least 3 months apart). However, in a few cases, radiographs were only available for one time point in which case "no progression" was scored (in accordance with the appropriateness criteria [1]). Based on the appropriateness criteria for each patient, it was determined whether or not surgery would have been inappropriate, appropriate, or necessary. To assess agreement between raters (interobserver reliability), a random sample of 26 patients was reviewed and

scored by two different raters (E.J. a resident orthopedic surgery and P.W. an orthopedic surgeon with 15 years of experience in spine surgery).

Clinical patient analysis was based on standard PROMs that were collected in standard care at the same day of radiographic acquisition and included the following instruments: visual analogue scale (VAS) for back and leg pain, Short Form 36 (SF-36), Pain Catastrophizing Scale (PCS), Hospital Anxiety Depression Scale (HADS), and, for the majority of patients, the Oswestry Disability Index (ODI). Although higher scores on the VAS, PCS, HADS, and ODI represent increasing pain and disability, the inverse holds true for the SF-36 instrument, in which higher scores reflect less disability or pain.

Statistical analysis

We had no prior hypotheses on the strength of the associations between PROMs and the appropriateness criteria. Therefore, no formal sample size calculation was performed based on the preferred power to detect such an association. We included all evaluable patients treated within the selected period of time and judged appropriateness of the size of our cohort using general recommendations for ANOVA and (logistic) regression (ie, 10 observations per variable for continuous outcomes, and 10 events per variable for dichotomous outcomes). With these in mind, 150 subjects would provide ample precision to reliably estimate the suggested associations.

Baseline characteristics of the patients were described using mean and standard deviation (SD) for continuous variables and as count and percentage for categorical variables. Subsequently, ordinal logistic regression analysis was used to determine which of the variables of the appropriateness criteria (ie, degree of symptoms and degrees of the curve) was most associated with the outcome (inappropriate, appropriate, or necessary).

To assess interobserver agreement on the scoring of medical records and radiographs, we computed overall agreement and the Cohen Kappa coefficient to correct for chance agreement.

To assess the association between the outcome of the appropriateness criteria and the patient perspective, analysis of variance (ANOVA) was used to test for differences in PROMs for each of the three categories resulting from the appropriateness criteria (ie, inappropriate, appropriate, and necessary). In case of severe non-normality, we used the Kruskal-Wallis test. Subsequently, the appropriateness criteria were dichotomized into two categories: appropriate (including both appropriate and necessary) and inappropriate. To assess how well PROMs can discriminate between appropriate and inappropriate, we used logistic regression analysis. Discriminative ability was subsequently determined by computing the area under the receiver operating characteristic curve, or AUC, resulting from the logistic regression analysis.

Correlations between the various PROMs and the separate variables contributing to the appropriateness criteria were estimated using the Spearman rank order correlation coefficient.

All analyses were performed using R version 3.5.1.

Results

Medical records of 195 patients were reviewed, and a total of 150 lumbar degenerative deformity patients met the inclusion criteria and were included for analysis. The study group consisted of 28 men and 122 women, and the average age of the study sample was 67.1 years (SD = 9.37, range 42 −89 years). The mean Cobb angle was 29.6° (SD = 11.3, range 11−64°). The distribution of patients according to the appropriateness criteria is shown in Table 1.

According to the appropriateness criteria, surgery was inappropriate for 60 patients (mean Cobb angle 22.4+8.7°), appropriate for 49 patients (mean Cobb angle 32+10°) and necessary for 41 patients (mean Cobb angle 37.3+10°). Ordinal logistic regression results showed that the appropriateness of surgery is most significantly dependent on the degree of symptoms, imbalance, and magnitude of curvature (Table 2).

The absolute interobserver agreement between the two raters of the medical records and radiographs to determine

Table 1
The distribution of patients according to the appropriateness criteria as defined by the Scoliosis Research Society (data shown as number of patients [percentage of total]) [4]. See http://links.lww.com/BRS/B79 for definitions

	Inappropriate	Appropriate	Necessary	Total
Degree of symptoms				
None	0 (0)	2 (1)	0 (0)	2 (1)
Mild	55 (37)	14 (9)	0 (0)	69 (46)
Moderate	5 (3)	33 (22)	35 (23)	73 (49)
Severe	0 (0)	0 (0)	6 (4)	6 (4)
Degree of stenosis				
None	32 (21)	17 (11)	6 (4)	55 (37)
Mild	26 (17)	15 (10)	14 (9)	55 (37)
Moderate	2 (1)	17 (11)	20 (13)	39 (26)
Severe	0 (0)	0 (0)	1 (1)	1 (1)
Progression				
No	60 (40)	48 (32)	41 (27)	149 (99)
Yes	0 (0)	1 (1)	0 (0)	1 (1)
Imbalance				
No	58 (39)	24 (16)	2 (1)	84 (56)
Yes	2 (1)	25 (17)	39 (26)	66 (44)
Risk factors				
None to mild	28 (19)	20 (13)	16 (11)	64 (43)
Moderate	28 (19)	26 (17)	23 (15)	77 (51)
Severe	4 (3)	3 (2)	2 (1)	9 (6)
Curvature				
Mild (10−19°)	27 (18)	5 (3)	0 (0)	32 (21)
Moderate (20−29°)	24 (16)	10 (7)	8 (5)	42 (28)
Moderate (30−39°)	7 (5)	25 (17)	17 (11)	49 (33)
Severe (≥40°)	2 (1)	9 (6)	16 (11)	27 (18)
Total	60 (40)	49 (33)	41 (27)	150 (100)

Table 2
Outcome of the ordinal logistic regression analysis, regressing the appropriateness of surgery against the various variables of the appropriateness criteria

	Coefficient	Standard error	T-value	p Value
Degree of symptoms	8.00	1.85	**4.33**	0.00*
Degree of stenosis	2.65	0.83	3.18	0.00*
Progression	6.36	13.79	0.46	0.65
Imbalance	9.66	2.04	**4.74**	0.00*
Risk factors	±3.37	1.12	−3.00	0.00*
Curvature	2.34	0.68	**3.44**	0.00*
Constant				
Inappropriate versus appropriate	43.31	16.48	2.63	0.01
Appropriate versus necessary	55.17	18.02	3.06	0.00

Bold values indicate appropriateness is most significantly dependent on the degree of symptoms, imbalance and magnitude of curvature.

the appropriateness category was 92.3%. Adjusted for chance-agreement, the Kappa coefficient was 0.88 on a scale from 0 to 1 ($p < .001$). The high Kappa score indicated that there was high consensus in the ratings given by the two raters.

Table 3 shows results of the ANOVA and in case of skewed variables Kruskal-Wallis statistics. It revealed a significant association between the appropriateness of surgery and virtually all PROMs (VAS, PCS, HADS, and ODI). Only three domains of the SF-36 (energy, role limitations physical and emotional) showed no significant association with the appropriateness of surgery.

The ability of PROMs to discriminate between those for whom surgery would be deemed inappropriate according to the criteria and those for whom surgery would be deemed appropriate or necessary was expressed as the AUC. The AUC for the combined VAS scores (ie, VAS right leg, left leg, and spine) was 0.69 (95% confidence interval [CI]: 0.60−0.77), for the SF-36 domains 0.70 (95% CI: 0.61−0.79), for PCS 0.67 (95% CI: 0.57−0.76), for the combined HADS domains 0.65 (95% CI: 0.56−0.74), and for the ODI 0.73 (95% CI: 0.61−0.85). For a model combining all PROMs, the AUC was 0.83 (95% CI: 0.71−0.95).

As demonstrated in Tables 4 to 6, the Spearman rank order correlation coefficients showed a statistically significant correlation between the degree of symptoms and each PROM ($p < .01$). Another statistically significant positive correlation was found between the VAS leg scores (left and right) and the degree of stenosis (ρ 0.56 [VAS left leg], ρ 0.60 [VAS right leg], $p < .01$). Moreover, the Spearman rank order correlation coefficients also demonstrated a statistically significant correlation between the appropriateness criteria and adverse health status score for all domains of the SF-36 score ($p < .01$) (Table 5) and a statistically significant positive correlation with the PCS score ($p < 0.01$), the HADS scores ($p < .01$), and the ODI questionnaires ($p < .01$) (Table 4).

Discussion

In the current study, a statistically significant association between the appropriateness of surgery for DLS and validated PROMs was found. Patients for whom surgery was

Table 3
Association between the outcome of the appropriateness criteria and the PROMs (data shown asmean / median [range])

	Inappropriate	Appropriate	Necessary	p Value for difference
VAS				
Right leg	27 / 10 [0−100]	35 / 20 [0−100]	45 / 50 [0−100]	**.01**
Left leg	20 / 0 [0−100]	31 / 20 [0−100]	45 / 50 [0−100]	**.00**
Spine	58 / 60 [0−100]	61 / 65 [0−100]	74 / 75 [30−100]	**.00**
SF-36				
Physical functioning*	44 / 45 [0−90]	36 / 30 [0−100]	24 / 20 [0−75]	**.00**
Social functioning*	60 / 63 [0−100]	60 / 63 [13−100]	42 / 50 [0−88]	**.00**
Role limitations (physical)*	27 / 0 [0−100]	23 / 0 [0−100]	15 / 0 [0−100]	.17
Role limitations (emotional)*	62 / 67 [0−100]	50 / 33 [0−100]	45 / 33 [0−100]	.09
Emotional well-being*	71 / 76 [4−92]	65 / 68 [12−96]	62 / 64 [16−100]	**.03**
Energy (fatigue)	49 / 50 [0−85]	47 / 50 [10−85]	43 / 45 [10−85]	.08
Pain	41 / 45 [0−80]	36 / 45 [0−88]	28 / 22 [0−100]	**.00**
General health	51 / 53 [10−90]	50 / 50 [5−80]	42 / 40 [0−75]	**.01**
Health change*	22 / 25 [0−75]	31 / 25 [0−50]	21 / 25 [0−50]	**.01**
PCS	19 / 17 [0−52]	25 / 25 [1−52]	27 / 26 [3−48]	**.00**
HADS				
Anxiety*	5 / 5 [0−19]	7 / 6 [1−16]	8 / 8 [0−17]	**.02**
Depression*	5 / 4 [0−17]	6 / 6 [0−18]	8 / 7 [1−17]	**.01**
ODI	33 / 32 [6−66]	41 / 44 [0−70]	51 / 51 [30−72]	**.00**

Bold values represent statistical significant differences between the outcome categories.

VAS, Visual Analogue Scale; SF-36, Short Form 36; PCS, Pain Catastrophizing Scale; HADS, Hospital Anxiety Depression Scale; ODI, Oswestry Disability Index; PROMs, patient-reported outcome measures.

* Because of ske wness of the data, the Kruskal-Wallis test was used.

Table 4
Spearman rank order correlation coefficient analysis for the various components of the appropriateness criteria and the different PROMs (analysis based on aggregate data)

	VAS spine	VAS left leg	VAS right leg	PCS	HADS anxiety	HADS depression	ODI
Symptoms	**.43***	**.21***	**.22***	**.34***	**.27***	**.32***	**.48***
Degree of stenosis	.11	**.56***	**.60***	**.24***	.15	**.19****	**.29***
Progression	−.09	.00	.02	−.07	−.11	−.08	NA
Imbalance	.13	.07	−.04	**.21****	.14	**.18****	**.35***
Risk factors	**.24***	.03	.03	**.20****	**.18****	**.27***	**.28****
Curvature	.07	.11	.02	.13	.10	.10	.18
Levels	**.18****	**.28***	**.36***	**.23****	.13	**.17****	.18

Bold values represent statistical significance.

VAS, Visual Analogue Scale; PCS, Pain Catastrophizing Scale; HADS, Hospital Anxiety Depression Scale; ODI, Oswestry Disability Index; PROMs, patient-reported outcome measures.

* $p \leq .01$.

** $p \leq .05$.

deemed necessary showed a higher degree of pain and disability, whereas the inverse holds true for patients for whom surgery would have been inappropriate. The appropriateness criteria for DLS as developed by the SRS represent a significant step towards evidence-based uniform treatment in spinal surgery [1]. However, as stated by Glassman et al. [7], it is important to recognize that the appropriateness criteria of surgery for DLS are a starting point, not an end point. From this statement one can infer that the current appropriateness criteria for DLS require further validation and refinement. One important limitation of the current criteria for DLS is the absence of PROMs as a validated, uniform, and transparent measure of patient perspective [1,7].

In this study, imbalance appeared to be the strongest determinant for appropriateness of surgery (Table 2, T-value 4.74). Imbalance was associated with a higher degree of pain and disability in patients suffering from DLS (Tables 4 and 5), which is in accordance with recent literature [11–14]. In a recent study, Daubs et al. [15], also using the RAND/UCLA Appropriateness Method, demonstrated that sagittal imbalance was a major factor affecting the appropriateness of surgery among patients with DLS.

A patient's health condition is often mainly determined by the subjective observation of the treating spinal surgeon and, as such, is highly susceptible to interpretation. PROMs, on the other hand, are defined as "Any report of the patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" [16]. By using individual patient PROMs, clinicians receive standardized information on a patient's individual health problem in order to identify or monitor symptoms and support shared decision-making. As shown in the current study, the discriminative ability for appropriateness of surgery for the PROMs as a group is strong (AUC 0.83 (95% CI: 0.71–0.95); however, when considered in isolation, the predictive power of the individual PROMs is poor (AUC between 0.65 and 073). For example, it feels contradictory that in the current study a high PCS and HADS score were positively associated with the appropriateness of surgery. In recent literature, it has been shown that preoperative depression is significantly associated with decreased improvement in quality of life after lumbar surgery [17]. Although there is a statistically significant association between the PCS or HADS and the category "risk factors" of the appropriateness criteria, this category apparently has little impact on the appropriateness

Table 5
Spearman rank order correlation coefficient analysis for the various components of the appropriateness criteria and the Short-From 36 (analysis based on aggregate data)

	Physical functioning	Social functioning	Role limitations (physical)	Role limitations (emotional)	Emotional well-being	Energy (fatigue)	Pain	General health	Health change
Degree of symptoms	**−.47***	**−.36***	**−.23***	**−.22***	**−.23***	**−.3***	**−.39***	**−.28***	−.20
Degree of stenosis	**−.21***	**−.24***	−.20	−.16	**−.23***	−.14	**−.37***	−.19	−.12
Progression	0.06	.13	.06	−.10	.06	.11	.11	.11	−.02
Imbalance	**−.22***	−.12	−.05	−.12	−.14	−.08	−.08	−.15	−.09
Risk factors	**−.31***	−.17	−.1	**−.21***	−.2	−.13	−.11	−.18	−.05
Curvature	.00	−.02	−.04	−.08	−.10	.02	−.02	−.05	−.09
Levels	**−.25***	−.2	−.17	−.10	−.17	−.17	**−.26***	−.14	−.04

Bold values represent statistical significance.

* $p \leq .01$.

Table 6
Association between the various subcategories as defined by the appropriateness criteria and their "matching" PROMs (data shown as mean / median [range])

Degree of symptoms	None	Mild	Moderate	Severe	p Value
VAS spine	20 / 20 [0–40]	53 / 50 [0–100]	73 / 75 [0–100]	79 / 78 [70–100]	**.00**
SF-36 phys. functioning	65 / 65 [50–80]	48 / 45 [0–100]	25 / 20 [0–75]	23 / 20 [5–50]	**.00**
ODI	NA*	31 / 29 [0–66]	48 / 50 [18–70]	54 / 60 [32–72]	**.00**
Degree of stenosis	*None*	*Mild*	*Moderate*	*Severe*	
VAS right leg	7 / 0 [0–70]	49 / 50 [0–100]	54 / 60 [0–100]	NA*	**.05**
VAS left leg	6 / 0 [0–90]	39 / 40 [0–100]	53 / 60 [0–100]	NA*	**.08**
Imbalance	*No*	*Yes*			
ODI	36 / 34 [6–70]	47 / 50 [0–72]			**.00**
Risk factors	*None to mild*	*Moderate*	*Severe*		
SF-36 role limit. (emotional)	61 / 67 [0–100]	52 / 67 [0–100]	7 / 0 [0–33]		**.00**
SF-36 emotional well-being	71 / 76 [24–100]	65 / 68 [4–96]	56 / 60 [28–72]		**.04**
PCS	21 / 21 [2–51]	24 / 24 [0–52]	34 / 33 [15–44]		**.03**
HADS anxiety	6 / 5 [0–17]	7 / 6 [0–19]	9 / 8 [5–12]		**.04**
HADS depression	6 / 5 [0–17]	7 / 6 [0–18]	10 / 8 [6–16]		**.00**

Bold values represent statistical significance ($p < 0.05$).

VAS, Visual Analogue Scale; SF-36, Short Form 36; PCS, Pain Catastrophizing Scale; HADS, Hospital Anxiety Depression Scale; ODI, Oswestry Disability Index; PROMs, patient-reported outcome measures.

* Too little observations.

of surgery. These findings support the use of preoperative PROMs (including a validated psychosocial questionnaire) as an additional tool for determination of appropriateness of surgery and for optimization of surgical outcome. In an effort to maximize the benefit of surgery, a trained multidisciplinary team may then address and treat anxiety or depression before proceeding with a surgical intervention.

Although the current study advocates the implementation of quantifiable, transparent PROMs into the appropriateness criteria, it should be further elucidated which PROMs can be used best and how they should be interpreted. Tables 4 and 5 show by means of a correlation analysis that the different categories of the appropriateness criteria significantly correlate with the various PROMs. The various PROMS all represent different components of the patient's health-related quality of life, and subsequently correspond to the different subcategories of the appropriateness criteria. As shown in Table 6, the VAS spine quantifies the amount of perceived back pain, the VAS leg defines the amount of perceived leg pain, the ODI questionnaire quantifies how DLS impacts functioning and health-related quality of life, and the HADS score demonstrates the level of depression and anxiety a patient experiences (subcategory risk factors). As an example, each PROM could be added to reinforce the corresponding appropriateness category. However, the exact thresholds of these individual PROMS to advise for or against surgery are yet to be determined. To determine such thresholds, and to determine which exact PROMs should be used, prospective studies with larger patient cohorts or analyses from spine registries will be necessary. Furthermore, patient preference with respect to surgery is not addressed by the PROMs or the appropriateness criteria itself. More insight into the patient preference could provide valuable information for clinical decision-making in patients suffering from DLS. Finally, as PROMs reflect the individually perceived burden due to a spinal disorder, their implementation into appropriateness criteria should account for cultural differences as well.

In conclusion, this study has revealed statistically significant association between validated PROMs and the appropriateness criteria for surgery in DLS. Implementation of PROMs into the appropriateness criteria could provide a more quantifiable, transparent, and uniform approach. Future studies with larger patient cohorts should further validate the incorporation of PROMs into the appropriateness criteria and optimize thresholds when to opt for surgery or not in DLS.

References

[1] Chen PG, Daubs MD, Berven S, Raaen LB, Anderson AT, Asch SM, et al. Surgery for degenerative lumbar scoliosis: the development of appropriateness criteria. Spine (Phila Pa 1976) 2016;41:910–8. doi: 10.1097/BRS.0000000000001392.

[2] Benner B, Ehni G. Degenerative lumbar scoliosis. Spine (Phila Pa 1976) 1979;4:548–52.

[3] Aebi M. The adult scoliosis. Eur Spine J 2005;14:925–48. doi: 10.1007/s00586-005-1053-9.

[4] Yamada K, Nakamae T, Shimbo T, Kanazawa T, Okuda T, Takata H, et al. Targeted therapy for low back pain in elderly degenerative lumbar scoliosis: a cohort study. Spine (Phila Pa 1976) 2016;41:872–9. doi: 10.1097/BRS.0000000000001524.

[5] Takahashi S, Delecrin J, Passuti N. Surgical treatment of idiopathic scoliosis in adults: an age-related analysis of outcome. Spine (Phila Pa 1976) 2002;27:1742–8.

[6] Shapiro GS, Taira G, Boachie-Adjei O. Results of surgical treatment of adult idiopathic scoliosis with low back pain and spinal stenosis: a study of long-term clinical radiographic outcomes. Spine (Phila Pa 1976) 2003;28:358–63. doi: 10.1097/01.BRS.0000048502.62793.0C.

[7] Glassman SD, Berven SH, Shaffrey CI, Mummaneni PV, Polly DW. Commentary: Appropriate use criteria for lumbar degenerative scoliosis: developing evidence-based guidance for complex treatment decisions. Neurosurgery 2017;80:E205–12. doi: 10.1093/neuros/nyw094.

[8] Lawson EH, Gibbons MM, Ingraham AM, Shekelle PG, Ko CY. Appropriateness criteria to assess variations in surgical procedure use

in the United States. Arch Surg 2011;146:1433–40. doi: 10.1001/archsurg.2011.581.

[9] Shekelle P. The appropriateness method. Med Decis Making 2004;24:228–31. doi: 10.1177/0272989X04264212.

[10] Falavigna A, Dozza DC, Teles AR, Wong CC, Barbagallo G, Brodke D, et al. Current status of worldwide use of patient-reported outcome measures (PROMs) in spine care. World Neurosurg 2017;108:328–35. doi: 10.1016/j.wneu.2017.09.002.

[11] Glassman SD, Bridwell K, Dimar JR, Horton W, Berven S, Schwab F. The impact of positive sagittal balance in adult spinal deformity. Spine (Phila Pa 1976) 2005;30:2024–9.

[12] Glassman SD, Berven S, Bridwell K, Horton W, Dimar JR. Correlation of radiographic parameters and clinical symptoms in adult scoliosis. Spine (Phila Pa 1976) 2005;30:682–8.

[13] Mac-Thiong JM, Transfeldt EE, Mehbod AA, Perra JH, Denis F, Garvey TA, et al. Can c7 plumbline and gravity line predict health related quality of life in adult scoliosis? Spine (Phila Pa 1976) 2009;34:E519–27. doi: 10.1097/BRS.0b013e3181a9c7ad.

[14] Schwab FJ, Blondel B, Bess S, Hostin R, Shaffrey CI, Smith JS, Boachie-Adjei O, et al. Radiographical spinopelvic parameters and disability in the setting of adult spinal deformity: a prospective multicenter analysis. Spine (Phila Pa 1976) 2013;38:E803–12. doi: 10.1097/BRS.0b013e318292b7b9.

[15] Daubs MD, Brara HS, Raaen LB, Chen PG, Anderson AT, Asch SM, et al. How does sagittal imbalance affect the appropriateness of surgical indications and selection of procedure in the treatment of degenerative scoliosis? Findings from the RAND AUC study. Spine J 2018;18:900–11. doi: 10.1016/j.spinee.2018.01.027.

[16] Santana MJ, Haverman L, Absolom K, Takeuchi E, Feeny D, Grootenhuis M, et al. Training clinicians in how to use patient-reported outcome measures in routine clinical practice. Qual Life Res 2015;24:1707–18. doi: 10.1007/s11136-014-0903-5.

[17] Miller JA, Derakhshan A, Lubelski D, Alvin MDA, McGirt MJ, Benzel EC, et al. The impact of preoperative depression on quality of life outcomes after lumbar surgery. Spine J 2015;15:58–64. doi: 10.1016/j.spinee.2014.06.020.