The Spine Journal 19 (2019) 95–103

Clinical Study

Responsiveness of the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire in lumbar surgery and its threshold for indicating clinically important differences

Takahito Fujimori, MD[a,b,*], Toshitada Miwa, MD[b], Takenori Oda, MD[b,c]

[a]*Department of Orthopedic Surgery, Japan Community Healthcare Organization, Osaka Hospital, 4-2-78 Fukushimaku, Fukushima, Osaka 553-0003, Japan*
[b]*Department of Orthopedic Surgery, Sumitomo Hospital, 5-3-20 Kitaku Nakanoshima, Osaka, Osaka 530-0005, Japan*
[c]*Department of Orthopedic Surgery, Osaka Minami Medical Center, 2-1 Kidohigashimachi, Kawachinagano, Osaka, Japan*

Received 22 December 2017; revised 11 April 2018; accepted 16 May 2018

Abstract

BACKGROUND CONTEXT: Introduced in 2007, the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire (JOABPEQ) has been widely used, but its psychometric properties have not been well studied.

PURPOSE: The objective of this study was to assess the responsiveness of the JOABPEQ in lumbar surgery and its threshold for indicating clinically important differences.

STUDY DESIGN: This is a prospective study.

PATIENT SAMPLE: Two hundred three consecutive patients underwent lumbar surgeries between July 2013 and November 2015 in a single hospital. Of the 203 patients, 181 patients who completed 1 year of follow-up were included.

OUTCOME MEASURES: Before and after surgery, the patients were asked to complete the questionnaire, including JOABPEQ, the 8-Item Short Form Health Survey (SF-8), and EuroQol-5D (EQ-5D). The participants were divided into five anchoring groups, ranging from "much better" to "much worse," according to reports from both physicians and patients.

MATERIALS AND METHODS: The responsiveness of measures was compared among five domains of the JOABPEQ ("low back pain," "walking ability," "lumbar function," "social function," and "mental health"), two domains of the SF-8 (the physical component summary [PCS] and the mental component summary [MCS]), and the EQ-5D. The responsiveness was assessed by the paired *t* test, the effect size, and the standardized response mean. The Spearman rank correlation coefficient and the receiver operating characteristic (ROC) curve were assessed using the five anchoring groups as external criteria. The clinically important differences, based on the ROC curve, were assessed.

RESULTS: Walking ability was most responsive, followed by low back pain and the PCS. The MCS was least responsive, followed by mental health and lumbar function. Social function and the EQ-5D had intermediate-level responsiveness. The substantial clinically important differences occurred at 20 points for low back pain and lumbar function, 23 points for walking ability, 14 points for social function, and 8 points for mental health.

CONCLUSIONS: The JOABPEQ domains are responsive measures in patients who undergo lumbar surgery. For physical function, the threshold for substantial clinically important differences was approximately 20 points for the JOABPEQ.

Keywords: Clinical important difference; Cutoff; Health-related outcome; JOABPEQ; Lumbar surgery; Patient-reported; Responsiveness; Scale; Threshold

FDA device/drug status: Not applicable.

Author disclosures: ***TF:*** Nothing to disclose. ***TM:*** Nothing to disclose. ***TO:*** Nothing to disclose.

No funds were received in support of this work.

There are no conflicts of interest associated with this study.

* Corresponding author. Department of Orthopedic Surgery, Japan Community Healthcare Organization, Osaka Hospital, 4-2-78 Fukushimaku, Fukushima, Osaka 553-0003, Japan. Tel.: +81 6 6441 5451; fax: +81 6 6441 8900.

E-mail address: takahito-f@hotmail.co.jp (T. Fujimori)

https://doi.org/10.1016/j.spinee.2018.05.013

Introduction

Lumbar spine disorders are common all over the world. With the promotion of an evidence-based approach to lumbar spine surgery, appropriate assessment methods of treatment effects become more important.

The Roland-Morris Disability Questionnaire (RMDQ) and the Oswestry Disability Index (ODI) are common disease-specific health-related patient-reported outcomes (HR-PROs). However, the RMDQ is limited to evaluation of back pain and may lack the ability to cover the multiple aspects of lumbar disorders. Additionally, the RMDQ has only two options for the questionnaire and patients may have difficulty in choosing the answer. As for the ODI, a low response rate of the sex question in some cultures has been controversial [1].

Considering these issues, the Japanese Orthopaedic Association created a new disease-specific HR-PRO named the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire (JOABPEQ) in 2007 [2–4]. Currently, English, Iranian [5], Arabic [6], Thai [7], Korean [8], and Turkish [9] versions are available. The JOABPEQ consists of 25 items divided into five domains: low back pain (pain-related disorder), lumbar function (lumbar spine dysfunction), walking ability (gait disturbance), social function (social life dysfunction), and mental health (psychological disorders). The questionnaire is set up so that scores in each domain range from 0 to 100, with higher scores indicating a better condition (Table 1). These multiple domains were developed with the intention of covering multiple aspects of lumbar spine disorders and are independent of each other. The weighting

Table 1
Japanese Orthopaedic Association Back Pain Evaluation Questionnaire

Question number	Question	Possible answers				
Q1-1	To alleviate low back pain, you often change your posture.	Yes		No		
Q1-2	Because of low back pain, you lie down more often than usual.	Yes		No		
Q1-3	Your lower back is almost always aching.	Yes		No		
Q1-4	Because of low back pain, you cannot sleep well.	Yes		No		
Q2-1	Because of low back pain, you sometimes ask someone to help you when you do something.	Yes		No		
Q2-2	Because of low back pain, you refrain from bending forward or kneeling down.	Yes		No		
Q2-3	Because of low back pain, you have difficulty in standing up from a chair.	Yes		No		
Q2-4	Because of low back pain, turning over in bed is difficult.	Yes		No		
Q2-5	Because of low back pain, you have difficulty putting on socks or stockings.	Yes		No		
Q2-6	Do you have difficulty in any one of the following motions: bending forward, kneeling, or stooping?	Great difficulty		Some difficulty	No	
Q3-1	Because of low back pain, you walk only short distances.	Yes		No		
Q3-2	Because of low back pain, you stay seated most of the day.	Yes		No		
Q3-3	Because of low back pain, you go upstairs more slowly than usual.	Yes		No		
Q3-4	Do you have difficulty in going upstairs?	Great difficulty		Some difficulty	No	
Q3-5	Do you have difficulty in walking more than 15 min?	Great difficulty		Some difficulty	No	
Q4-1	Because of low back pain, do you avoid any routine housework?	Yes		No		
Q4-2	Have you been unable to do your job or ordinary activities as well as you would like?	Always unable	Unable most of the time	Sometimes unable	Able most of the time	Always able
Q4-3	Has your work routine been hindered because of pain?	A great deal	Moderately	Slightly	Very little	Not at all
Q5-1	Because of low back pain, do you get irritated or get angry at other people more often than usual?	Yes		No		
Q5-2	How is your present health condition?	Poor	Fair	Good	Very good	Excellent
Q5-3	Have you been discouraged and depressed?	Always	Frequently	Sometimes	Rarely	Never
Q5-4	Do you feel exhausted?	Always	Frequently	Sometimes	Rarely	Never
Q5-5	Have you felt happy?	Never	Rarely	Sometimes	Almost always	Always
Q5-6	Do you think you are in decent health?	Not at all	Barely	Not very much	Fairly	Yes
Q5-7	Do you feel your health will get worse?	Very much so	A little bit	Sometimes	Not very much	Not at all

Note: Calculating formulas: low back pain: (Q1-1×20+Q1-3×20+Q1-7×20+Q1-11×10−70)×100/70; lumbar function: (Q2-1×10+Q2-2×10+Q2-3×20+Q2-4×10+Q2-5×30+Q2-6×20−100)×100/120; walking ability: (Q3-1×30+Q3-2×20+Q3-3×10+Q3-4×10+Q3-5×30−100)×100/140; social function: (Q3-5×4+Q4-1×2+Q4-2×6+Q4-3×10−22)×100/74; mental health: (Q5-1×3+Q5-2×4+Q5-3×6+Q5-4×6+Q5-5×3+Q5-6×3+Q5-7×3−28)×100/103.

coefficient is set for each question on the basis of the severity of symptoms. Reliability, internal consistency, and criteria validity were tested when the JOABPEQ was being developed [2–4].

The responsiveness of a scale is its ability to detect change over time in the construct to be measured [10] and is one of the important psychometric properties of the scale. If responsiveness is low, the change in score after an intervention may be small. In such a situation, the evaluator will have difficulty judging whether the intervention is an ineffective treatment or whether instead the scale is unfit for evaluation of the disease. However, to our knowledge, no previous study has assessed the responsiveness of the JOABPEQ.

The user's guide for the JOABPEQ specifies that if a score in a JOABPEQ domain increases more than 20 points after an intervention, the intervention can be regarded as effective [11]. This threshold is important for using the JOABPEQ. However, the guide is currently available only in Japanese, and the data have not been fully published [12].

We hypothesized that further verification of the psychometric properties of the JOABPEQ would encourage more widespread use of the scoring system. The purpose of the present study was, first, to assess the responsiveness of the JOABPEQ and to compare the findings with those of common types of HR-PRO, and then to determine the threshold of JOAPBEQ scores for indicating clinically important differences in patients' outcomes.

Materials and methods

Our study protocol was approved by our institutional review board. Between July 2013 and November 2015, two spine attending surgeons performed lumbar surgeries for 203 consecutive patients in a single hospital. The patients did not respond to conservative treatment, such as medication and epidural blocks, and were referred to the hospital.

Questionnaires

Before surgery, the patients were asked to complete questionnaires, including the JOABPEQ, the physical component summary (PCS) and the mental component summary (MCS) of the 8-Item Short Form Health Survey (SF-8), and the EuroQol-5D (EQ-5D). One year after surgery, the patients were asked to complete the same questionnaires. In addition, they were asked to respond to a question about symptom reduction caused by surgery. The question used a five-point global rating scale of responses: much better, better, neither better nor worse, worse, and much worse. A numerical value from 5 to 1 was attributed to each answer, with 5 representing "much better" and 1 representing "much worse." This global rating scale has high face validity and has been commonly used. However, some researchers questioned the reliability of such retrospective measures of change [13]. Additionally, from our experience, we were aware that some patients hesitated to answer such a direct question honestly because they were afraid of hurting their relationship with their surgeons [14]. To avoid this bias, the attending surgeons in our study were also asked to independently evaluate the improvement of the patients' condition using the same five-point rating scale. The final status was defined as the lower rating score from either the patient or the surgeon. For example, if a surgeon evaluated a patient's condition as worse but the patient gave an evaluation of "better," the patient was assigned to the "worse" group.

Responsiveness

We defined responsiveness as the ability to detect change over time in the construct being measured [10,15–17]. We used two approaches to examine responsiveness: internal and external [12,15,18,19]. Internal responsiveness is the ability of a measure to change over a particular time frame [15]. Internal responsiveness is based on the distribution of the scores. Internal responsiveness was assessed by the paired *t* test, the effect size, and the standardized response mean (SRM). With the paired *t* test, postoperative and preoperative scores were compared. The smaller the p-value, the better the responsiveness. Effect size was defined as the difference between preoperative and postoperative scores (ie, the change score) divided by the standard deviation of the preoperative score. The SRM was defined as the effect size divided by the standard deviation of the change score. The larger the effect size or the larger the SRM, the better the responsiveness. The internal responsiveness of measure depends on both the particular treatment and the particular outcomes used to determine treatment efficacy.

External responsiveness reflects the extent to which changes in a measure over a specified time frame relate to corresponding changes in a reference measure of health status. In contrast with internal responsiveness, external responsiveness is the relationship between change in the measure and change in the external standard [15]. In the present study, external responsiveness was assessed using the Spearman rank correlation coefficient and the receiver operating characteristic (ROC) curve, using the global rating scale as an external criterion. The higher the correlation coefficient and the larger the area under the curve, the better the responsiveness. The optimal cutoff value on the ROC curve was calculated as the best point that discriminated the "much better" and the "better" groups from the "neither better nor worse," the "worse," and the "much worse" groups.

For the JOABPEQ, we compared the sensitivity and the specificity at this ROC-based cutoff with those of the proposed 20-point threshold. A p-value of $<.05$ was considered statistically significant with the two-tailed test. SPSS Statistics (version 20; IBM, Armonk, NY, USA) and EZR (Saitama Medical Center, Jichi Medical University, Saitama, Japan) was used for statistical analysis. EZR is a graphical user interface for R (The R Foundation for Statistical Computing, Vienna, Austria) [20].

Table 2
Diagnosis and surgical procedures

		n	%
Diagnosis	Lumbar canal stenosis	65	36
	Degenerative spondylolisthesis	47	26
	Isthmic spondylolisthesis	9	5
	Foraminal stenosis	7	4
	Disc herniation	26	14
	Osteoporotic vertebral fracture	14	8
	Spinal tumor	8	4
	Spinal deformity	3	2
	Spondylitis	2	1
Operation	Fenestration (partial laminectomy)	60	33
	Posterior lumbar interbody fusion	70	39
	Discectomy	22	12
	Balloon kyphoplasty	7	4
	Posterior fusion	12	7
	Anterior fusion	4	2
	Anterior and posterior fusion	4	2
	Other	2	1

Table 3
Correspondence between patients' ratings and surgeons' ratings

	Surgeons' ratings				
Parameter	Much better	Better	Neither	Worse	Much worse
Patients' ratings					
Much better	51	65	6	1	1
Better	2	24	4	3	0
Neither	0	3	8	3	0
Worse	0	2	0	2	0
Much worse	0	0	3	3	0
Global rating scale as the gold standard	Much better	Better	Neither	Worse	Much worse
Number of patients	51	91	21	11	7
Percentage	28	50	12	6	4

Results

Patients' demographics

Of the 203 patients who underwent surgery, 181 patients (follow-up rate of 89%) who completed questionnaires both before and after surgery were included in our study. The mean age of the patients was 69±13 years. There were 65 patients with lumbar spinal stenosis, 47 with degenerative spondylolisthesis, 9 with isthmic spondylolisthesis, 7 with foraminal stenosis, 26 with lumbar disc herniation, 14 with osteoporotic vertebral fracture, 8 with a spinal tumor, 3 with spinal deformity, and 2 with spondylitis. Table 2 shows the details of surgical procedures.

Global rating scale

Table 3 shows the correspondence between patients' global ratings and surgeons' global ratings. Most patients (50%) were in the better group; the least amount (4%) were in the much worse group. Table 4 shows the mean change in score for each domain in each group.

Domain scores

Table 5 shows the mean scores for all domains for all patients. Fig. 1 shows the preoperative distribution of scores in each domain, and Fig. 2 shows the postoperative distributions.

Internal responsiveness

Paired t *test*

Scores for all domains were significantly better after surgery (Table 5). The smallest p-value was for "walking ability" scores of the JOABPEQ (7×10^{-29}), with the p-values for the other scores increasing in this order: the PCS of the SF-8, "low back

Table 4
Mean change in score before and after surgery in each global rating group

			JOABPEQ					SF-8		
Globalrating scale	n		Low back pain	Lumbar function	Walking ability	Social function	Mental health	PCS	MCS	EQ-5D
Much better	51	Mean	55	35	59	44	25	17	5	0.38
		Median	57	25	64	48	20	18	4	0.31
		SD	34	36	35	27	23	10	10	0.29
Better	91	Mean	44	26	42	25	15	11	4	0.23
		Median	43	25	47	23	13	11	6	0.20
		SD	33	32	34	26	19	10	9	0.23
Neither	21	Mean	9	7	14	15	3	7	3	0.14
		Median	0	9	0	13	0	10	3	0.06
		SD	41	28	31	29	18	12	10	0.22
Worse	11	Mean	−8	−10	−3	−13	3	0	1	0
		Median	0	−16	−2	−13	0	3	2	0
		SD	35	23	28	17	27	7	12	0
Much worse	7	Mean	−10	−26	−8	−11	−4	−2	−2	−0.13
		Median	−14	−25	−7	-15	0	−1	−2	−0.09
		SD	11	29	26	24	13	5	4	0.22

EQ-5D, EuroQol-5D; JOABPEQ, Japanese Orthopaedic Association Back Pain Evaluation Questionnaire; PCS, physical component summary; MCS, mental component summary; SD, standard deviation; SF-8, 8-Item Short Form Health Survey.

Note: Numeric values are mean, median, and SD.

Table 5
Scores and responsiveness in each domain in all patients

	JOABPEQ					SF-8		
Parameter	Low back pain	Lumbar function	Walking ability	Social function	Mental health	PCS	MCS	EQ-5D
Preoperative score (SD)	31 (29)	43 (33)	21 (23)	33 (22)	41 (21)	30 (8)	45 (9)	0.51 (0.23)
Postoperative score (SD)	68 (35)	65 (31)	60 (35)	59 (28)	56 (20)	41 (10)	49 (8)	0.74 (0.22)
Change in score (SD)	38 (39)	22 (36)	39 (39)	26 (31)	15 (22)	11 (11)	4 (9)	0.23 (0.27)
Internal responsiveness								
p-Value (paired *t* test)	2×10^{-27}	4×10^{-14}	7×10^{-29}	1×10^{-22}	1×10^{-16}	2×10^{-28}	5×10^{-8}	1×10^{-23}
Effective size	1.3	0.67	1.7	1.2	0.74	1.3	0.4	1.0
Standardized response mean	0.97	0.62	1.0	0.84	0.69	0.99	0.4	0.86
External responsiveness								
Spearman rank correlation coefficient with global rating scale	0.46	0.37	0.49	0.47	0.34	0.45	0.13	0.44
ROC analysis								
Area under the curve (95% confidence interval)	0.83 (0.75–0.90)	0.76 (0.68–0.83)	0.83 (0.75–0.89)	0.78 (0.69–0.86)	0.73 (0.64–0.81)	0.76 (0.67–0.84)	0.59 (0.49–0.7)	0.76 (0.67–0.85)
Cutoff	15–28	17–25	23–29	14	8	8.8	3.9	0.12
Sensitivity	0.82	0.55	0.71	0.73	0.69	0.68	0.59	0.75
Specificity	0.77	0.82	0.80	0.72	0.72	0.73	0.68	0.69
Sensitivity+specificity	1.59	1.37	1.51	1.44	1.41	1.41	1.26	1.45
20-point threshold								
Sensitivity	0.82	0.55	0.77	0.62	0.41	N/A	N/A	N/A
Specificity	0.77	0.82	0.69	0.74	0.80			
Sensitivity+specificity	1.59	1.37	1.46	1.36	1.20			

EQ-5D, EuroQol-5D; JOABPEQ, Japanese Orthopaedic Association Back Pain Evaluation Questionnaire; MCS, mental component summary; N/A, not applicable; PCS, physical component summary; ROC, receiver operating characteristics; SD, standard deviation; SF-8, 8-Item Short Form Health Survey.

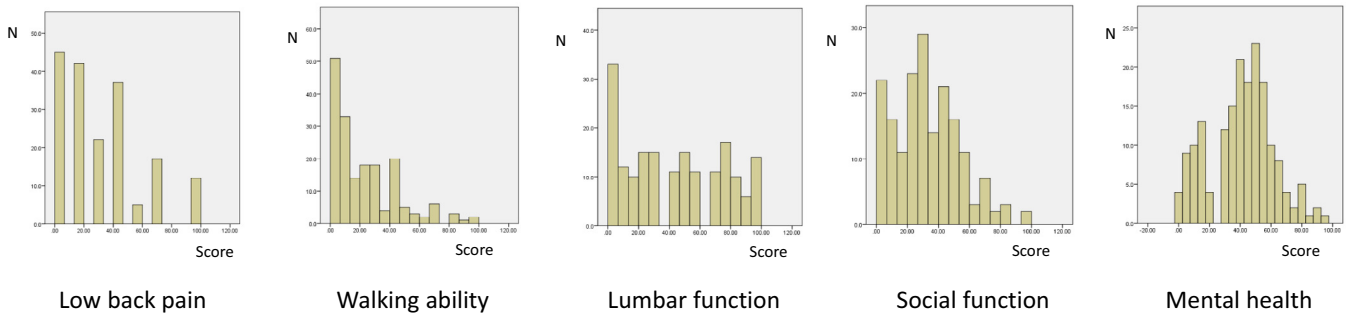

Fig. 1. Preoperative histogram of each domain of the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire. The horizontal line shows the score of the domain and the vertical line shows the number of patients. The "mental health" domain had a more normal distribution than the other domains.

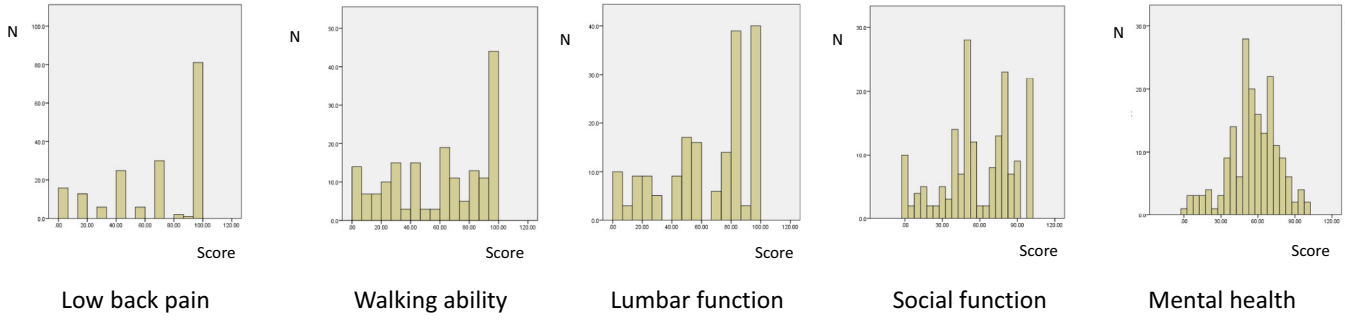

Fig. 2. Postoperative histogram of each domain of the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire. The horizontal line shows the score of the domain and the vertical line shows the number of patients. The "mental health" domain had a more normal distribution than the other domains.

pain" of the JOABPEQ, the EQ-5D, "social function" of the JOABPEQ, "mental health" of the JOABPEQ, "lumbar function" of the JOABPEQ, and the MCS of the SF-8 (5×10^{-8}).

Effect size

The largest effect size was for walking ability of the JOABPEQ (1.7), followed by the PCS of the SF-8, low back pain of the JOABPEQ, social function of the JOABPEQ, the EQ-5D, mental health of the JOABPEQ, lumbar function of the JOABPEQ, and the MCS of the SF-8 (0.4) (Table 5).

Standardized response mean

The largest SRM was for walking ability of the JOABPEQ (1.0), followed the SRMs of the PCS of the SF-8, low back pain of the JOABPEQ, the EQ-5D, social function of the JOABPEQ, mental health of the JOABPEQ, lumbar function of the JOABPEQ, and the MCS of the SF-8 (0.40) (Table 5).

External responsiveness

Correlation

Table 6 shows the correlation coefficients between the change score of each domain and the global rating scale. All HR-PRO domains were significantly correlated with the global rating scale ($p<.05$). The highest correlation was for walking ability of the JOABPEQ (0.49), followed by social function of the JOABPEQ, low back pain of the JOABPEQ, the PCS of the SF-8, the EQ-5D, lumbar function of the JOABPEQ, mental health of the JOABPEQ, and the MCS of the SF-8 (0.13).

Receiver operating characteristic curve

The largest area under the curve was found for low back pain of the JOABPEQ (0.83), followed by walking ability of the JOABPEQ, social function of the JOABPEQ, lumbar function of the JOABPEQ, the EQ-5D, the PCS of the SF-8, mental health of the JOABPEQ, and the MCS of the SF-8 (0.59) (Table 5).

The cutoff values for the JOABPEQ ranged from 15 to 28 points for low back pain, from 17 to 25 points for lumbar function, from 23 to 29 points for walking ability, 14 points for social function, and 8 for mental health. The cutoff values for the SF-8 were 8.8 for the PCS and 3.9 for the MCS. The cutoff value for the EQ-5D was 0.12 (Table 5).

As for the JOABPEQ's low back pain and lumbar function domains, the proposed 20-point threshold was in the range of ROC-based cutoff values. As for the domains walking ability, social function, and mental health, the sum of sensitivity and specificity at the ROC-based cutoff was slightly larger than that at 20 points (Table 5).

Discussion

We assessed the responsiveness of the JOABPEQ and its threshold for indicating clinically important differences in

Table 6
Spearman rank correlation between each domain

		JOABPEQ					SF-8		
	Global rating scale	Low back pain	Lumbar function	Walking ability	Social function	Mental health	PCS	MCS	EQ-5D
Global rating scale	1	0.46	0.37	0.49	0.47	0.34	0.45	0.13	0.44
JOABPEQ									
Low back pain		1.0	0.48	0.46	0.47	0.41	0.46	0.21	0.43
Lumbar function			1.0	0.41	0.42	0.49	0.38	0.24	0.53
Walking ability				1.0	0.71	0.43	0.60	0.21	0.56
Social function					1.0	0.51	0.66	0.3	0.56
Mental health						1.0	0.48	0.51	0.54
SF-8									
PCS							1.0	0.1	0.58
MCS								1.0	0.22
EQ-5D									1.0

EQ-5D, EuroQol-5D; JOABPEQ, Japanese Orthopaedic Association Back Pain Evaluation Questionnaire; MCS, mental component summary; PCS, physical component summary; SF-8, 8-Item Short Form Health Survey.

patients' outcomes by using well-accepted methods reported in the literature [18,19,21–23]. Generally, the JOABPEQ showed better responsiveness than the SF-8 or the EQ-5D, suggesting that the JOABPEQ is suitable for determining surgical efficacy in patients with lumbar spine disorders. The threshold for indicating clinically important differences of the JOABPEQ for low back pain, walking ability, lumbar function, and "social ability" was approximately 20 points. To the best of our knowledge, ours is the first study to assess the responsiveness of JOABPEQ domains, as well as of other HR-PROs. The JOABPEQ was a responsive HR-PRO for lumbar surgery. The most responsive domain was walking ability, and the second most responsive domain was low back pain.

We believe that these results reflect the results of surgery. From the perspective of functional recovery, the purpose of surgery for most of our patients was to decrease leg symptoms, such as radicular pain, motor weakness, or intermittent claudication caused by nerve compression. This belief was supported by the fact that the PCS of the SF-8 was as responsive as the walking ability domain of the JOABPEQ. As shown by the high correlation of the PCS with the walking ability and social function domains, the PCS is closely associated with ambulatory ability. Some researchers may prefer an HR-PRO that is specific for spinal stenosis, such as the Swiss Spinal Stenosis Questionnaire. However, our study demonstrated that the walking ability domain of the JOABPEQ was responsive enough for evaluating surgical outcome for patients undergoing lumbar surgery.

The threshold of JOABPEQ scores for indicating clinically important differences in patients' outcomes was approximately 20 points. The ROC-based cutoff for low back pain and lumbar function was 20 points. The cutoff was slightly higher than 20 in the walking ability domain and slightly lower than 20 in the social function domain. However, the cutoff of the mental health domain was only 8 points. In the mental health domain, the proportion of patients whose scores increased by more than 20 points was limited to 40%, even in the much better or the better groups (sensitivity of 0.4).

Although there have been few reports about the threshold for indicating clinically important differences in outcomes shown with the JOABPEQ, several reports have been published about such differences shown with the PCS and the EQ-5D. According to those reports, the threshold for determining clinically important differences for the PCS [24–26] ranges from 4.1 to 6.5, and the threshold for the EQ-5D [27,28] ranges from 0.03 to 0.54.

Compared with these numbers, the cutoff points for the PCS (8.8) and for the EQ-5D (0.12) in our study were slightly high. This was because we set a strict anchor that was based on the lower rating score from either the patient or the surgeon. These results suggest that the 20-point threshold in the JOABPEQ should be regarded as a high threshold and that it corresponds to substantial rather than minimal clinical differences. Recently, Kasai et al. [29] reported data on the threshold of the JOABPEQ for determining treatment effectiveness. According to Kasai et al.'s findings, the 25th percentile for change in scores in patients who rated their postoperative condition as "improved" was approximately 20 points. Kasai et al. concluded that the 20 points corresponded to the JOABPEQ's thresholds for indicating substantial clinical benefit.

Low back pain had the second highest responsiveness of the five JOABPEQ domains. The patients who underwent lumbar surgery often had low back pain as well as leg symptoms. Approximately 10% of our patients underwent surgery for low back pain (eg, osteoporotic vertebral fracture and spinal deformity).

Scores for the walking ability and the low back pain domains were directly related to patients' symptoms. The preoperative scores for those domains were lower than the preoperative scores for the other domains, and the difference between the preoperative and the postoperative scores for those two domains was also larger than the differences between the scores in the other domains.

In contrast, the responsiveness for the lumbar function and the mental health domains was lower than that for the other domains. We speculate that the reason lumbar function was less responsive is derived from the concept of the domain. The questions in the low back pain domain simply ask patients about the presence or the absence of pain. However, the questions in lumbar function ask patients about the function restricted by pain. Even though a patient might have less low back pain after surgery, that patient might still have some restrictions in some activities of daily life. In addition, the patient might have a limited range of motion of the lumbar spine because of spinal fusion or might need orthotic devices. Therefore, the lumbar function domain is less likely to be responsive than the walking ability or the low back pain domains.

The mental health domain had the lowest responsiveness of all the JOABPEQ domains. The preoperative scores for the mental health domain were higher than those for the other domains and were distributed more normally (Fig. 2). The postoperative change in scores for the mental health domain was smaller than those for the other domains. The MCS of the SF-8 showed a similar tendency. These results suggest that study participants were more affected by physical issues than by psychological issues. Although psychological aspects are important in treating chronic pain, the primary purpose of our surgery was to improve physical function. The mental health domain and the MCS may be useful for evaluating psychological aspects of spine disease that are related to issues such as chronic low back pain. However, these domains did not show good responsiveness in assessing surgically treated patients.

Our study had some limitations. First, an independent assessor was not adopted in the present study. There has been no consensus on the way to set the gold standard for HR-PRO [30]. Previous studies reported that surgeons tended to assess the result of surgery more optimistically than patients [30,31]. In the present study, this optimistic assessment could have been prevented by adopting the lower rating score from either the patient or the surgeon. However, this strict assessment may elevate the threshold of JOABPEQ. Second, our study included a heterogeneous population of lumbar spinal disorders. The responsiveness and the threshold could differ between the patients whose chief complaint was low back pain and the patients whose chief complaint was leg pain. If focusing only on the patients with osteoporotic compression fracture, the responsiveness of low back pain may be higher. Third, other types of disease-specific HR-PROs, such as the ODI, were not available in the present study. Comparison of the psychometric property between ODI and JOABPEQ will be considered in future tasks.

Despite these limitations, our findings are useful for designing a prospective study for lumbar surgery using the JOABPEQ. We recommend walking ability and low back pain as primary end points. These two domains have high responsiveness and are suitable for assessing the results of lumbar surgery. The 20-point threshold is a reasonable value for indicating substantial clinical differences for these domains. Because the mental health domain was less responsive, it would be difficult to detect a 20-point change with it for patients undergoing lumbar surgery. We recommend using a lower threshold, such as 10 points, for detecting clinically important differences in the mental health domain.

Conclusions

The JOABPEQ is a responsive assessment tool in patients who undergo lumbar surgery. Compared with more generic HR-PROs such as the SF-8 or EQ-5D, the JOABPEQ domains of walking ability and low back pain show greater responsiveness, whereas the lumbar function and the mental health domains show less responsiveness. The threshold for detecting substantial clinical differences for low back pain, walking ability, lumbar function, and social ability is 20 points.

Acknowledgments

The authors thank Ms. Aoki, Ms. Yamada, Ms. Kimura, Ms Kawaguchi, and Ms. Iwakiri for helping with the data collection. Medical editor Katharine O'Moore-Klopf, ELS (East Setauket, NY, USA), provided professional English-language editing of this article.

References

[1] Fujiwara A, Kobayashi N, Saiki K, Kitagawa T, Tamai K, Saotome K. Association of the Japanese Orthopaedic Association score with the Oswestry Disability Index, Roland-Morris Disability Questionnaire, and short-form 36. Spine 2003;28:1601–7.

[2] Fukui M, Chiba K, Kawakami M, et al. Japanese Orthopaedic Association Back Pain Evaluation Questionnaire. Part 3. Validity study and establishment of the measurement scale: Subcommittee on Low Back Pain and Cervical Myelopathy Evaluation of the Clinical Outcome Committee of the Japanese Orthopaedic Association, Japan. J Orthop Sci 2008;13:173–9.

[3] Fukui M, Chiba K, Kawakami M, et al. Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire (JOACMEQ): Part 2. Endorsement of the alternative item. J Orthop Sci 2007;12:241–8.

[4] Fukui M, Chiba K, Kawakami M, et al. JOA Back Pain Evaluation Questionnaire (JOABPEQ)/JOA Cervical Myelopathy Evaluation Questionnaire (JOACMEQ). The report on the development of revised versions. April 16, 2007. The Subcommittee of the Clinical Outcome Committee of the Japanese Orthopaedic Association on Low Back Pain and Cervical Myelopathy Evaluation. J Orthop Sci 2009;14:348–65.

[5] Azimi P, Mohammadi HR, Montazeri A. An outcome measure of functionality and pain in patients with lumbar disc herniation: a validation study of the Japanese Orthopedic Association (JOA) score. J Orthop Sci 2012;17:341–5.

[6] Alfayez SM, Bin Dous AN, Altowim AA, Alrabiei QA, Alsubaie BO, Awwad WM. The validity and reliability of the Arabic version of the Japanese Orthopedic Association Back Pain Evaluation Questionnaire: can we implement it in Saudi Arabia? J Orthop Sci 2017;22:618–21.

[7] Poosiripinyo T, Paholpak P, Jirarattanaphochai K, et al. The Japanese Orthopedic Association Back Pain Evaluation Questionnaire (JOABPEQ): a validation of the reliability of the Thai version. J Orthop Sci 2017;22:34–7.

[8] Jung KS, Jung JH, Jang SH, Bang HS, In TS. The reliability and validity of the Korean version of the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire. J Phys Ther Sci 2017;29:1250–3.

[9] Gunaydin G, Hazar Kanik Z, Karabicak GO, et al. Cross-cultural adaptation, reliability and validity of the Turkish version of the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire. J Orthop Sci 2016;21:295–8.
[10] Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol 1989;42:403–8.
[11] Japanese Orthopaedic Association. The Committee of Diagnosis and Evaluation Criteria of the Japanese Society for Spine Surgery and Related Research. The user's manual of JOABPEQ and JOACMEQ. Tokyo: Nankodo; 2012.
[12] Chien A, Lai DM, Cheng CH, Wang SF, Hsu WL, Wang JL. Responsiveness of the Chinese versions of the Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire and Neck Disability Index in postoperative patients with cervical spondylotic myelopathy. Spine 2015;40:1315–21.
[13] Kamper S. Global Rating of Change scales. Aust J Physiother 2009;55:289.
[14] Fujimori T, Okuda S, Iwasaki M, et al. Validity of the Japanese Orthopaedic Association scoring system based on patient-reported improvement after posterior lumbar interbody fusion. Spine J 2016;16:728–36.
[15] Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53:459–68.
[16] Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40:171–8.
[17] Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.
[18] Monticone M, Ambrosini E, Vernon H, et al. Responsiveness and minimal important changes for the Neck Disability Index and the Neck Pain Disability Scale in Italian subjects with chronic neck pain. Eur Spine J 2015;24:2821–7.
[19] Burgers PT, Poolman RW, Van Bakel TM, et al. Reliability, validity, and responsiveness of the Western Ontario and McMaster Universities Osteoarthritis Index for elderly patients with a femoral neck fracture. J Bone Joint Surg Am 2015;97:751–7.
[20] Kanda Y. Investigation of the freely available easy-to-use software "EZR" for medical statistics. Bone Marrow Transplant 2013;48:452–8.
[21] Chen P, Lin KC, Liing RJ, Wu CY, Chen CL, Chang KC. Validity, responsiveness, and minimal clinically important difference of EQ-5D-5L in stroke patients undergoing rehabilitation. Qual Life Res 2016;25:1585–96.
[22] Obradovic M, Lal A, Liedgens H. Validity and responsiveness of EuroQol-5 Dimension (EQ-5D) versus Short Form-6 Dimension (SF-6D) questionnaire in chronic pain. Health Qual Life Outcomes 2013;11:110.
[23] DeVine J, Norvell DC, Ecker E, et al. Evaluating the correlation and responsiveness of patient-reported pain with function and quality-of-life outcomes after spine surgery. Spine 2011;36(21 Suppl.):S69–74.
[24] Carreon LY, Glassman SD, Campbell MJ, Anderson PA. Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J 2010;10:469–74.
[25] Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study Questionnaire Short Form 36, and pain scales. Spine J 2008;8:968–74.
[26] Glassman SD, Copay AG, Berven SH, Polly DW, Subach BR, Carreon LY. Defining substantial clinical benefit following lumbar spine arthrodesis. J Bone Joint Surg Am 2008;90:1839–47.
[27] Coretti S, Ruggeri M, McNamee P. The minimum clinically important difference for EQ-5D index: a critical review. Expert Rev Pharmacoecon Outcomes Res 2014;14:221–33.
[28] Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 2005;14:1523–32.
[29] Kasai Y, Fukui M, Takahashi K, et al. Verification of the sensitivity of functional scores for treatment results—substantial clinical benefit thresholds for the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire (JOABPEQ). J Orthop Sci 2017;22:665–9.
[30] Lattig F, Grob D, Kleinstueck FS, et al. Ratings of global outcome at the first post-operative assessment after spinal surgery: how often do the surgeon and patient agree? Eur Spine J 2009;18(Suppl. 3):386–94.
[31] McGee MA, Howie DW, Ryan P, Moss JR, Holubowycz OT. Comparison of patient and doctor responses to a total hip arthroplasty clinical evaluation questionnaire. J Bone Joint Surg Am 2002;84-A:1745–52.