Minimum Clinically Important Difference of the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire for Patients with Lumbar Spine Disease Undergoing Posterior Surgery by Generation

Minori Kato[1)], Hidetomi Terai[1)], Takashi Namikawa[2)], Akira Matsumura[2)], Masatoshi Hoshino[2)], Hiromitsu Toyoda[1)], Akinobu Suzuki[1)], Shinji Takahashi[1)], Koji Tamai[1)], Yuta Sawada[1)], Masayoshi Iwamae[1)], Yuki Okamura[1)], Yuto Kobayashi[1)] and Hiroaki Nakamura[1)]

1) Department of Orthopaedics, Osaka Metropolitan University Graduate School of Medicine, Osaka, Japan
2) Department of Orthopaedics, Osaka City General Hospital, Osaka, Japan

Abstract:

Introduction: Few studies have assessed the minimum clinically important difference (MCID) of each Japanese Orthopaedic Association Back Pain Evaluation Questionnaire (JOABPEQ) domain. This study assessed MCIDs of JOABPEQ in patients with lumbar spine disease by generation.

Methods: We evaluated the JOABPEQ score of 805 consecutive patients with lumbar spine disease undergoing posterior surgery preoperatively and 1 year postoperatively. MCIDs of each JOABPEQ domain were determined using anchor- and distribution-based methods according to age. A question based on the concept of a health transition item was used as the anchor for the MCID decision.

Results: Overall, MCIDs of the JOABPEQ were 28.6 and 27.3 points for pain-related disorder and gait disturbance, respectively. The MCID for the lumbar spine dysfunction domain did not reach 0.6 over the area under the curve. Regarding the differences among generations, MCIDs of pain-related disorder and gait disturbance domains differed slightly between the elderly and middle-aged. The psychological disorder domain did not reflect clinically meaningful changes in the elderly. MCIDs of the social life disturbance domain decreased with age.

Conclusions: Focusing on achieving the ideal responsiveness of patient-reported outcomes across generations, MCIDs of the pain-related disorder and gait disturbance domains may be valuable for patients, regardless of age, when adopting the JOABPEQ for patients with lumbar spine disease undergoing surgery. This study only evaluated cases that underwent posterior lumbar surgery. Future research will necessitate conducting surveys concerning the outcomes of various treatments for lumbar spine disease.

Keywords:

Lumbar spine, Outcome measures, Minimum clinically important difference, Japanese Orthopaedic Association Back Pain Evaluation, Patient-reported outcomes

Spine Surg Relat Res 2024; 8(5): 518-527
dx.doi.org/10.22603/ssrr.2023-0293

Introduction

Recently, there has been an increased focus on evaluating evidence-based medicine in the medical field. Clinicians are using validated outcome measures increasingly, and other investigators have attempted to assess the effects of both operative and nonoperative therapeutic modalities. In the analysis of treatment outcomes, the concept of minimum clinically important difference (MCID) is spreading and affirming efficacy with the most negligible difference in health-related quality of life[1-8]. These values are crucial for clinical decision-making, economic evaluation, and sample size estimation in clinical research. Two general approaches have been used to determine the MCID. The anchor-based

Corresponding author: Hidetomi Terai, terai@omu.ac.jp
Received: November 30, 2023, Accepted: January 26, 2024, Advance Publication: April 3, 2024

method uses an external criterion called anchor, which must correlate with the health status instrument in the study to determine what patients consider an essential improvement. The distribution-based method detects the degree to which the change between the baseline and different time points exceeds what would be expected from chance alone, helping to confirm measurement errors. Each approach calculates MCID based on the following definitions: (1) the average change in responders[4], (2) the minimum detectable change approach that defines MCID as the minor change above measurement error[5], (3) the difference in the average change score between responders and nonresponders[6], and (4) the receiver operating characteristic curve (ROC) analysis, which defines MCID as the change with the highest sensitivity and specificity in positive responders[7].

Lumbar spine diseases, such as lumbar disc herniation, lumbar canal stenosis, degenerative lumbar scoliosis, and degenerative spondylolisthesis, result in low back pain, leg pain, neurogenic claudication, and excretory disorders. These symptoms negatively affect daily life and time off work more than other medical condition[9]. If conservative treatment for lumbar spine disease is not practical, surgery is performed. However, despite reasonable surgical outcomes, low patient satisfaction has been reported, which results from the difficulty of directly interpreting the meaning or clinical importance of numerical scores obtained from patient-reported outcomes (PROs)[10-12]. The Japanese Orthopaedic Association (JOA) Back Pain Evaluation Questionnaire (JOABPEQ) is an accurate outcome measurement following the treatment of lumbar spine disease, created by the JOA in 2007[13-15]. The questionnaire score on each domain is effective at >20 points after treatment. However, the threshold has not been retested since the JOABPEQ has spread worldwide. The cutoff value of the JOABPEQ was established on the basis of the concept of substantial clinical benefit and discussions of the JOA[15,16]. However, the effectiveness of PROs has often been determined by the MCID calculated using anchor- and distribution-based methods. To date, only some reports have focused on determining MCIDs for each JOABPEQ domain in patients undergoing surgery for lumbar spine disease[17,18]. Additionally, with aging becoming a global concern, it is debatable whether older adults should be assessed using the same thresholds of PROs as younger and middle-aged people[16,19,20]. This issue highlights the need to consider the differential status associated with aging, which potentially affects the applicability and interpretation of PROs in different age groups. However, the differences in baseline and change threshold of the JOABPEQ scores by generation, including the elderly population, have not been reported. This study aimed to calculate the MCID of the JOABPEQ in patients with lumbar spine disease who underwent surgery and to assess the JOABPEQ results in patients treated with posterior surgery by generation.

Materials and Methods

The Institutional Review Board of Osaka City General Hospital approved the study protocol (No. 1912112). Written informed consent was obtained from the patients to publish current research and any accompanying images.

Patients

We retrospectively studied 805 consecutive patients treated at a single institution between 2009 and 2019 who received posterior surgical treatment for lumbar spine disease. The clinical indication for surgical treatment was pain or numbness, or both, inducing intermittent claudication, mainly from canal and foraminal stenoses. Patients with trauma, tumors, or adult spinal deformity were excluded. Initially, 1,589 candidates were identified. However, the final cohort comprised 805 cases (a follow-up rate of 51.4%), selected on the basis of availability of preoperative and 1-year postoperative data, including general health status and the JOABPEQ. For reoperation cases, this study included those for which data were collected 1 year after the initial operation.

We performed microsurgical bilateral decompression via a unilateral approach preserving posterior central elements as decompression surgery[21] and posterior or transforaminal lumbar interbody fusion as fusion surgery[22]. We decided on the surgery based on the imaging findings; decompression surgery was indicated for patients without disc instability and severe scoliosis. Fusion surgery was performed for patients with degenerative lumbar scoliosis with Cobb's angle ≥20°, degenerative spondylosis with Meyerding grade ≥II, or posterior disc opening ≥5° during anterior flexion of the affected intervertebral level. Patients were divided into three groups based on age. The middle-aged, young-elderly, and old-elderly groups were <65, 65-74, and ≥75 years, respectively. The preoperative physical condition of all patients was evaluated using the American Society of Anesthesiologists (ASA) physical status classification system[23].

Questionnaires

Clinical outcomes were evaluated using the JOA score and the recovery ratio of the JOA score using Hirabayashi's calculation[24,25]. The PRO scale was assessed using the visual analog scale (VAS), including low back and leg pain (VAS of 0-100 mm)[26] and the JOABPEQ[13]. The scores of all questionnaires were calculated preoperatively and 1 year postoperatively.

Responsiveness

The standardized response mean (SRM) was adopted as an indicator of responsiveness for JOABPEQ domains. The SRM is a type of effect size used to assess the magnitude of the effect of a particular intervention on a group. It indicates the degree of change before and after the intervention, which is calculated by dividing the mean change over time by the standard deviation (SD) of that change. SRM values

Table 1. Patient Demographics.

	Total	Middle-aged group	Young-elderly group	Old-elderly group	p
Number of cases (%)	805	226 (28.1)	316 (39.3)	263 (32.7)	
Age (years)	69.5±10.6	55.8±8.8	70.1±2.9	79.5±2.8	<0.01*
Gender [number (%)]					
Women	434 (53.9)	114 (14.2)	165 (20.5)	155 (19.3)	N.S.**
Men	371 (46.1)	112 (13.9)	151 (18.8)	108 (13.4)	
BMI (kg/m^2)	24.5±3.8	25.2±4.0	24.5±3.9	23.7±3.2	N.S.*
Operative time (min)	173±85.5	170±85.5	185±88.9	162±79.3	N.S.*
Blood loss (mL)	145±218	129±195	162±226	137±225	N.S.*
Surgery [number (%)]					N.S.**
Decompression	468 (58.1)	130 (16.2)	178 (22.1)	160 (19.9)	
Fusion	337 (41.9)	69 (8.6)	138 (17.1)	103 (12.8)	
ASA-PS [number (%)]					<0.01**
I	128 (15.9)	75 (9.3) †	42 (5.2)	11 (1.4) †	
II	600 (74.5)	135 (16.8) †	251 (31.2) †	214 (26.6) †	
III	77 (9.6)	16 (2.0)	23 (2.9)	38 (4.7) †	

BMI, body mass index; ASA-PS, American Society of Anesthesiologists Physical Status Classification System; ±, standard deviation; N.S., not significant

*, analyzed using the Tukey–Kramer test among the three groups.

**, analyzed using Pearson's chi-squared test for parameters on a categorical scale among the three groups.

†p<0.05, analyzed using residual analysis.

of 0.2, 0.5, and ≥0.8 were defined as small, medium, and large responsiveness, respectively[27].

Calculating the MCID

We adopted anchor- and distribution-based methods to determine the MCID as previously recommended. A question based on the concept of health transition item (HTI) was used as the anchor for the MCID decision[28,29]. The patients were asked to rate their general condition 1 year postoperatively as "much better," "somewhat better," "about the same," "somewhat worse," or "much worse" compared with their preoperative general condition. On the basis of the results of these questionnaires, they were allocated an external assessment rating reaching the anchors. Patients who rated their state as recovered, "much better," or "somewhat better" were classified as improved, and those who responded to "about the same," "somewhat worse," or "much worse" were classified as nonimproved. On the basis of these anchors, we evaluated the MCID calculated using ROC analysis as the anchor-based method. This value was used to identify the MCID of PROs as the value with equal sensitivity and specificity in the improved case[7,30,31]. The discriminative ability of the model was assessed using the area under the ROC (AUC). An AUC score of 0.50-0.60 was classified as failed. The distribution-based method calculated half (0.5) the SD of the preoperative baseline scores. The distribution-based method provides an objective, statistical perspective in determining the MCID. This approach reduces the reliance on subjective interpretation. Changes in the target subject's data are considered within the measurement error range for PROs and not clinically significant when they are lower than the MCID determined by the distribution method. When using the anchor-based method, the MCID should be higher than the distribution-based value to provide statistical reliability beyond the measurement error range[1-3]. Statistical analyses were conducted using SPSS version 23 (IBM Corp., Armonk, NY, USA). The t-test, Mann-Whitney test, and Tukey-Kramer test were used to compare continuous data. Pearson's chi-squared test and residual analysis were used to determine the significance of the study parameters on a categorical scale[32]; p-values of <0.05 were considered statistically significant.

Results

A total of 805 patients (434 women and 371 men) were enrolled in this study. The average age was 69.5 (range 37-89) years. ASA III was significantly more common in the old-elderly group than in the other groups (Table 1). Overall, the JOA score changed from 12.5 points preoperatively to 22.5 points 1 year postoperatively. The recovery ratio of the JOA score was 60.7% 1 year postoperatively. The decreasing VAS score for low back and leg pain was 24.0 and 40.3 mm, respectively (Table 2). Focusing on the difference among generations, the improvement in the JOA and VAS scores for leg pain in the old-elderly group was significantly lower than that in the middle-aged and young-elderly groups. There was no significant difference in the improvement of the VAS score for low back pain among the three groups (Fig. 1).

Regarding the increasing JOABEPQ score 1 year postoperatively, each JOABPEQ domain (pain-related disorder, gait disturbance, and social life disturbance) was >20 points (Table 3). Fig. 2 shows the differences in the increasing JOABEPQ scores among the three groups. The increase in the score of the JOABPEQ domains (pain-related disorder, lum-

Table 2. Outcome Measurement (JOA Score and VAS) and the Difference Baseline and 1-year Assessments.

	Total			Middle-aged group			Young-elderly group			Old-elderly group		
	Preop.	Postop.	Δ	Preop.	Postop.	Δ	Preop.	Postop.	Δ	Preop.	Postop.	Δ
JOA score (point)	12.5±5.1	22.5±5.7	10.0*	13.2±5.0	24.3±5.2	11.0*	12.6±5.1	22.7±5.1	10.2*	11.9±5.1	20.7±5.7	8.9*
VAS (mm)												
Low back pain	53.9±28.3	29.9±26.3	−24.0*	54.8±25.8	29.5±25.3	−25.2*	53.3±29.4	28.4±26.2	−25.0*	53.6±29.0	31.8±27.1	−22.0*
Leg pain	67.2±26.2	26.9±28.0	−40.3*	66.3±26.5	22.3±26.2	−44.0*	68.9±26.0	26.8±28.1	−42.2*	65.8±26.0	31.0±28.7	−35.0*

JOA, Japanese Orthopaedic Association; VAS, visual analog scale; Preop., preoperative data; Postop., postoperative data; Δ, postoperative data−preoperative data; ±, standard deviation

*p<0.05, analyzed the difference between preoperative and postoperative data.

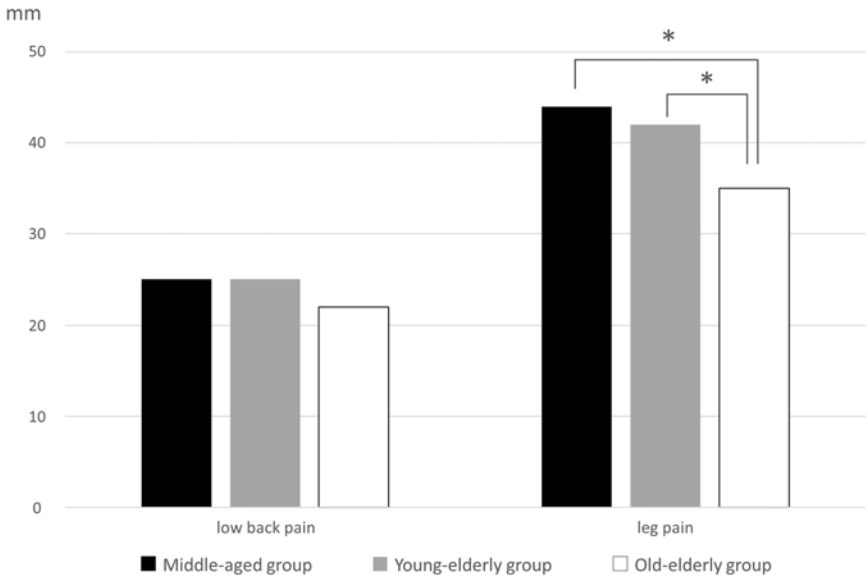

Figure 1. Difference between baseline and 1-year assessments (VAS). *p<0.05, analyzed using the Tukey–Kramer test. VAS, visual analog scale

Table 3. Outcome Measurement (JOABPEQ) and the Difference between Baseline and 1-year Assessments.

	Total			Middle-aged group			Young-elderly group			Old-elderly group		
	Preop.	Postop.	Δ	Preop.	Postop.	Δ	Preop.	Postop.	Δ	Preop.	Postop.	Δ
JOABPEQ (point)												
Pain-related disorder	39.7±33.7	67.8±35.6	28.0*	36.7±33.0	69.2±36.8	32.6*	40.9±33.1	70.8±33.3	30.0*	41.1±34.7	62.9±36.7	21.8*
Lumbar spine dysfunction	52.8±30.0	68.8±29.3	16.0*	53.9±30.0	74.7±28.1	20.8*	55.0±29.6	71.0±28.2	15.9*	49.2±30.3	61.2±29.9	12.0*
Gait disturbance	24.7±24.4	64.4±32.3	39.6*	28.0±25.3	75.9±31.2	47.9*	23.0±23.7	66.0±31.2	43.0*	24.0±24.3	52.6±30.6	28.6*
Social life disturbance	33.7±21.3	60.3±25.7	26.5*	34.3±20.8	67.8±26.1	33.5*	33.5±21.0	60.9±24.4	27.4*	33.6±22.2	53.1±25.0	19.6*
Psychological disorder	42.1±19.1	55.1±19.6	13.0*	44.3±18.3	57.9±20.1	13.6*	42.4±19.1	56.0±19.3	13.7*	40.0±19.4	51.6±19.0	11.6*

JOABPEQ, Japanese Orthopaedic Association Back Pain Evaluation Questionnaire; Preop., preoperative data; Postop., postoperative data; Δ, postoperative data–preoperative data; ±, standard deviation

*p<0.05, analyzed the difference between preoperative and postoperative data.

bar spine dysfunction, gait disturbance, and social life disturbance) in the old-elderly group was significantly lower than that in the middle-aged group.

Table 4 shows the responsiveness of JOABPEQ. Overall, the SRM values for gait and social life disturbances were 11.6 and 0.92, respectively, indicating a large responsiveness. The SRM value for lumbar spine dysfunction was 0.48, indicating a small responsiveness. In the middle-aged group, all responsiveness for each domain was medium or above. In the old-elderly group, the SRM value for gait disturbance was 0.92, indicating a large responsiveness.

Table 5 shows the rating groups based on the HTI ques-

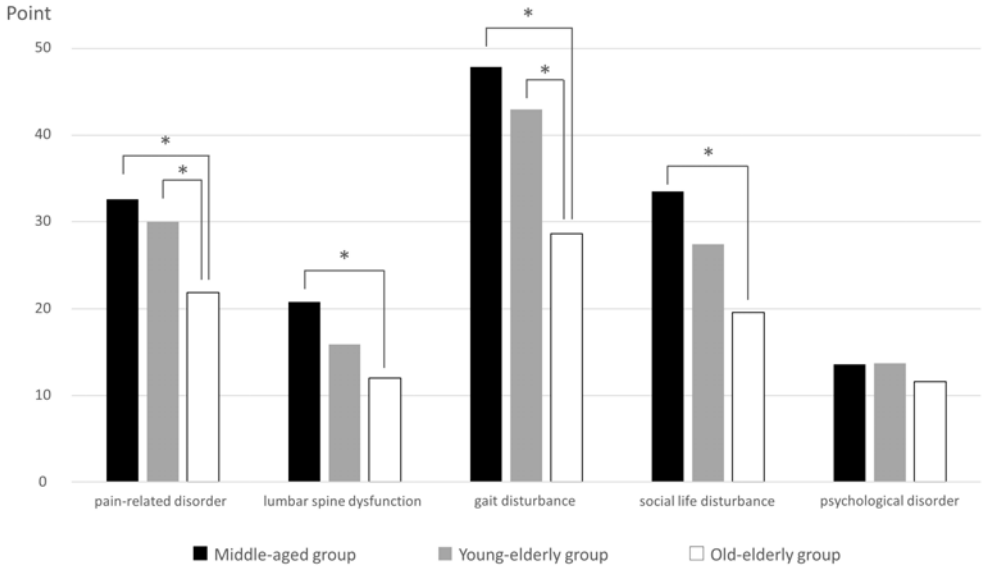

Figure 2. Difference between baseline and 1-year assessments (JOABPEQ). *p<0.05, analyzed using the Tukey–Kramer test. JOABPEQ, Japanese Orthopaedic Association Back Pain Evaluation Questionnaire

Table 4. Responsiveness of the JOABPEQ.

	Total		Middle-aged group		Young-elderly group		Old-elderly group	
	SRM	Responsiveness	SRM	Responsiveness	SRM	Responsiveness	SRM	Responsiveness
JOABPEQ								
Pain-related disorder	0.71	Medium	0.79	Medium	0.78	Medium	0.55	Medium
Lumbar spine dysfunction	0.48	Small	0.58	Medium	0.48	Small	0.39	Small
Gait disturbance	1.16	Large	1.38	Large	1.27	Large	0.92	Large
Social life disturbance	0.92	Large	1.10	Large	0.99	Large	0.71	Medium
Psychological disorder	0.67	Medium	0.68	Medium	0.72	Medium	0.60	Medium

JOABPEQ, Japanese Orthopaedic Association Back Pain Evaluation Questionnaire; SRM, standardized response mean
The SRM values of 0.2, 0.5, and ≥0.8 were defined as small, medium, and large responsiveness, respectively.

Table 5. Distribution of Ratings by Response to the Anchor Question.

	Total	Middle-aged group	Young-elderly group	Old-elderly group	p
Number of cases [number (%)]					
Improved case	466 (57.9)	156 (19.4) **	186 (23.0)	124 (15.5) **	<0.05*
Non-improved case	339 (42.1)	70 (8.7) **	130 (16.2)	139 (17.3) **	
Response to the anchor question [number (%)]					
Much better	223 (27.7)	91 (11.3) **	92 (11.4)	40 (5.0) **	<0.05*
Somewhat better	243 (30.2)	65 (8.1)	94 (11.7)	84 (10.4)	
About the same	192 (23.9)	38 (4.7) **	82 (10.2)	72 (8.9)	
Somewhat worse	113 (14.0)	29 (3.6)	34 (4.2) **	50 (6.2) **	
Much worse	34 (4.2)	3 (0.4) **	14 (1.7)	17 (2.1) **	

*, analyzed using Pearson's chi-squared test for parameters on a categorical scale among the three groups.
**p<0.05, analyzed using residual analysis.

tionnaire answers and the likelihood of improved or nonimproved cases. In the middle-aged, young-elderly, and old-elderly groups, 156, 186, and 124 cases were classified as improved cases, respectively. The proportion of improved cases was significantly lower with increasing age. Fig. 3 shows the increase in the JOABPEQ scores for each domain according to the anchor based on the answers. Overall, all domains showed significant differences between patients reporting "much better" and patients reporting "about the same," except for the lumbar spine dysfunction domain. In

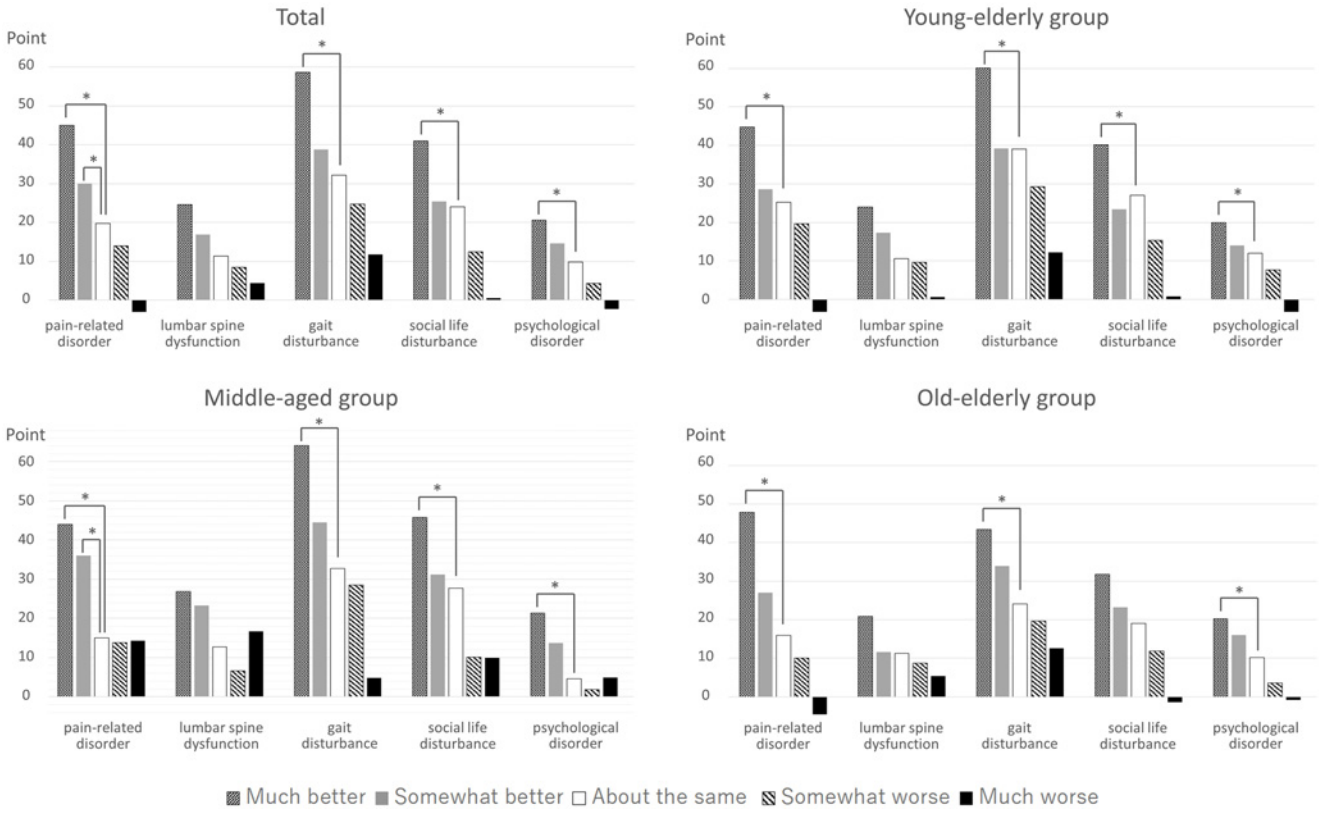

Figure 3. Difference of JOABPEQ between baseline and 1-year assessments by response to the anchor question. *p<0.05, analyzed using the Tukey–Kramer test. JOABPEQ, Japanese Orthopaedic Association Back Pain Evaluation Questionnaire

all three groups, there was no significant difference in the lumbar spine dysfunction domain among all patients anchored by ratings. In the social life disturbance domain, a significant difference was observed in the middle-aged and young-elderly groups between patients reporting "much better" and those reporting "about the same." In the old-elderly group, however, there was no significant difference in the social life disturbance domain among all patients anchored by ratings.

Table 6 shows the 0.5 SD values of the preoperative baseline scores and MCIDs for each JOABPEQ domain. Overall, MCIDs of the JOABPEQ using ROC analysis were 28.6, 16.7, 27.3, 18.9, and 12.6 points for pain-related disorder, lumbar spine dysfunction, gait disturbance, social life disturbance, and psychological disorder domains, respectively. The MCID for lumbar spine dysfunction did not reach an AUC score of 0.6; thus, it was classified as a failure. All MCIDs were >0.5 SD. Regarding the differences among generations, MCID of social life disturbance decreased with increasing age. Comparing 0.5 SD values, MCIDs of lumbar spine dysfunction, social life disturbance, and psychological disorder domains in the old-elderly group and the psychological disorder domain in the young-elderly group were <0.5 SD. These results indicate that these thresholds are inappropriate as MCIDs for elderly people according to the concept of MCID determination using anchor- and distribution-based methods.

Discussion

The JOA score has been used worldwide to assess low back pain and lumbar spine disease. However, the JOA score has some limitations, such as no evaluation of health conditions, no psychological or social activity evaluation, and a lack of patient subjectivity. To overcome these limitations, the JOABPEQ was developed for patient-based evaluation by the JOA because it comprises five domains, and multifaceted evaluation is possible. The JOABPEQ had the strongest correlation with patient satisfaction following surgery for lumbar spine disease compared with the other PROs[13,14]. Kasai et al., in the original study regarding the JOA, concluded that a correlation was noted between patients' self-rating and acquired JOABPEQ score, indicating that 20 acquired points can be interpreted as substantial clinical benefit thresholds for the JOABPEQ; however, as they noted, the study had some limitations, including the small sample size, and they only investigated data within 200 days after receiving treatment, which varied inconsistently[15]. Despite the worldwide spread of the JOABPEQ, the threshold for substantial clinical benefit has not been reexamined on the basis of the MCID concept. Few studies have focused on determining MCIDs of different JOABPEQ domains in patients who underwent lumbar surgery. Azimi et al. and Ogura et al. only evaluated posterior lumbar decompression cases with lumbar spine disease, and included <200 cases, which is not a large sample size[17,18]. Arima et al. fo-

Table 6. MCID of the JOABPEQ.

	0.5 SD	MCID (point)	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)
Total						
Pain-related disorder	16.8	28.6	0.66	0.62	0.67	0.56
Lumbar spine dysfunction	15.0	16.7	0.59	0.57	0.58	0.57
Gait disturbance	12.2	27.3	0.68	0.67	0.77	0.53
Social life disturbance	10.7	18.9	0.66	0.65	0.72	0.55
Psychological disorder	9.5	12.6	0.67	0.63	0.61	0.68
Middle-aged group						
Pain-related disorder	16.5	28.3	0.67	0.64	0.73	0.51
Lumbar spine dysfunction	14.9	8.3	0.61	0.57	0.73	0.48
Gait disturbance	12.7	29.6	0.70	0.69	0.84	0.46
Social life disturbance	10.4	24.3	0.69	0.68	0.7	0.63
Psychological disorder	9.2	12.6	0.72	0.68	0.64	0.75
Young-elderly group						
Pain-related disorder	16.6	26.6	0.61	0.61	0.68	0.49
Lumbar spine dysfunction	14.8	16.7	0.60	0.58	0.57	0.6
Gait disturbance	11.8	27.2	0.63	0.66	0.82	0.47
Social life disturbance	10.5	16.2	0.64	0.63	0.77	0.5
Psychological disorder	9.6	9.5	0.65	0.63	0.67	0.59
Old-elderly group						
Pain-related disorder	17.4	27.2	0.67	0.62	0.64	0.61
Lumbar spine dysfunction	15.1	8.3	0.57	0.57	0.58	0.57
Gait disturbance	12.2	21.4	0.65	0.62	0.76	0.48
Social life disturbance	11.1	8.1	0.63	0.61	0.75	0.49
Psychological disorder	9.7	8.7	0.66	0.61	0.67	0.59

MCID, minimum clinically important difference; JOABPEQ, Japanese Orthopaedic Association Back Pain Evaluation Questionnaire; AUC, area under the receiver operating characteristic curve; 0.5 SD, half the standard deviation
MCID was calculated using the receiver operating characteristic curve analysis.

cused only on cases of spinal deformity in adults[19].

This study aimed to determine the MCID of patients who underwent posterior lumbar spine surgery. Using both anchor- and distribution-based methods, this study also included an analysis of test responsiveness based on the SRM. This approach provided a comprehensive and precise assessment of the efficacy of the questionnaire in clinical settings. This study has the following strengths: (1) the number of patients (n=805) was the largest, (2) the follow-up period was homogeneous and long (1 year), and (3) a wide patient age range was assessed. In this study, for the total patient population, MCIDs for each domain were 28.6, 16.7, 27.3, 18.9, and 12.6 points for pain-related disorder, lumbar spine dysfunction, gait disturbance, social life disturbance, and psychological disorder, respectively. The MCID of lumbar spine dysfunction dmain did not reach an AUC score of 0.6, and the degree of change in the lumbar spine dysfunction domain was small responsiveness evaluated using SRM. The efficacy of the questionnaire in the lumbar spine dysfunction domain may be inappropriate. This failure is likely because we indicated surgery mainly for patients with leg pain or numbness, or both, inducing intermittent claudication. The lumbar spine dysfunction domain is a specific outcome measurement for low back pain. Our selection bias may have influenced the small responsiveness of the lumbar spine dysfunction domain to lumbar spine disease. Another reason is that the preoperative score of lumbar spine dysfunction was high (52.8 points), which can lead to small responsiveness. In surveys assessing PROs, high initial scores can make additional point improvements difficult, often resulting in reduced responsiveness.

This study demonstrated the MCID value by generation. To the best of our knowledge, no study has evaluated the JOABPEQ responsiveness for patients with lumbar spine disease by generation. Some studies have reported that the threshold value of PROs in the elderly population differs[16,19,20]. In this study, the improvement in each JOABPEQ domain decreased with age; therefore, it is worth reconsidering whether it is appropriate to evaluate middle-aged and elderly people on the same scale. On the basis of the distribution-based approach, MCIDs of lumbar spine dysfunction, social life disturbance, and psychological disorder domains in the old-elderly group and the psychological disorder domain in the young-elderly group were <0.5 SD values. MCIDs of social life disturbance and psychological disorder domains decreased with increasing age. These results imply that these thresholds for elderly people are inappropriate. In the psychological disorder domain, the MCID was 8.7 and 9.5 points in the old-elderly and young-elderly groups, respectively. Patients' mental conditions influenced by physical status alone might be limited because their mental state may also be affected by social and financial situ-

ations. Some studies have reported that the MCID for the psychological disorder domain tends to be lower than that of other domains[17,18]. The psychological disorder domain may not reflect clinically meaningful changes. MCIDs of the social life disturbance domain decreased with age. Responsiveness in the social life domain is essential for middle-aged individuals. This responsiveness may be due to more frequent social interactions in middle-aged people than in elderly people[19]. Middle-aged patients engage in more social activity than elderly patients, requiring more remarkable improvement after surgery. Meanwhile, regarding MCIDs of pain-related disorder and gait disturbance, there was a slight difference between the middle-aged and elderly groups. Even in the elderly groups, the change in these domains demonstrated medium or large responsiveness, as assessed using the SRM. The plausible responsiveness of these domains (pain-related disorder and gait disturbance) may be considered an appropriate outcome measure for lumbar spine disease. Severe pain and poor walking ability negatively affect daily life regardless of age. Especially for the elderly generation, the baseline physical condition is not good; therefore, slight deterioration adversely affects daily life. Pain and walking ability significantly influence patient satisfaction following clinical improvements in lumbar spine function[33,34]. In focusing on the responsiveness of PROs[35,36], MCIDs for pain-related disorder and gait disturbance domains may be more appropriate indices across generations. Additionally, emphasizing the MCID of the gait disturbance domain may be particularly beneficial for patients who have undergone lumbar spine surgery.

Our study has several limitations. First, it was necessary to verify whether the method for determining the MCID was appropriate. There is no consensus regarding the optimal method, and various MCIDs have been proposed for clinical use. In this study, no MCIDs had an AUC of ≥0.7, which is considered acceptable. However, these results do not invalidate the JOABPEQ because it has been evaluated and used worldwide. This study investigated which JOABPEQ domains are more useful across generations, using anchor- and distribution-based methods and responsiveness for domains. We consider the comparison using AUC to be valid. There have been some reports in the past using a cutoff in the 0.6 range[37,38]. However, it may be necessary to reexamine the validity of JOABPEQ and improve the performance of JOABPEQ. Second, this study excluded patients treated with conservative therapies; only patients who underwent posterior surgery were included. Therefore, MCIDs used in this study do not apply to conservatively treated cases. In terms of surgical cases, the number of cases indicated for decompression and fusion surgeries was almost equivalent. The treatment in this study is considered representative of most surgical variations, despite only the posterior approach being included. Future research will necessitate conducting surveys concerning the outcomes of various treatments for lumbar spine disease. Third, our study's data, sourced from one facility, may not accurately represent general populations, potentially leading to selection bias. This limitation affects the generalizability of our findings. Future research encompassing multiple institutions is recommended to ensure more diverse and representative data collection. In the future, surveys will need to be conducted to overcome these limitations and improve the performance of JOABPEQ.

Conclusions

MCIDs of the JOABPEQ using ROC analysis were 28.6 and 27.3 points for pain-related disorder and gait disturbance, respectively. The MCID for the lumbar spine dysfunction domain did not reach 0.6 over the AUC score, indicating that the value failed. MCIDs of pain-related disorder and gait disturbance domains differed slightly between the elderly and middle-afged generations. Focusing on achieving the ideal responsiveness of PROs across generations, it might be appropriate to emphasize MCIDs of the pain-related disorder and gait disturbance domains. These domains could be considered valuable for patients, regardless of age, when adopting the JOABPEQ for patients with lumbar spine disease undergoing posterior surgery. This study only evaluated cases that underwent posterior lumbar surgery. Future research will necessitate conducting surveys concerning the outcomes of various treatments for lumbar spine disease.

Conflicts of Interest: The authors declare that there are no relevant conflicts of interest.

Sources of Funding: None.

Author Contributions: M.K., H.T., H.N., T.N., A.M., and M.H. designed the study; H.T. and A.S. analyzed the data; S.T., K.T., Y.S., and M.I. performed statistical analysis; and M.K., Y.O., and Y.K. wrote the manuscript.

Ethical Approval: The Institutional Review Board of Osaka City General Hospital approved the study protocol (IRB No. 1912112).

Informed Consent: Informed consent for publication was obtained from all study participants.

References

1. Copay AG, Subach BR, Glassman SD, et al. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J. 2007;7(5):541-6.
2. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61(2):102-9.
3. Asher AL, Kerezoudis P, Mummaneni PV, et al. Defining the minimum clinically important difference for grade I degenerative lumbar spondylolisthesis: insights from the Quality Outcomes Database. Neurosurg Focus. 2018;44(1):E2.
4. Gunaydin G, Citaker S, Meray J, et al. Reliability, validity, and cross-cultural adaptation of the Turkish version of the

Bournemouth questionnaire. Spine. 2016;41(21):E1292-7.
5. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003;56(5):395-407.
6. Parker SL, Adogwa O, Paul AR, et al. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. J Neurosurg Spine. 2011;14(5):598-604.
7. Carreon LY, Glassman SD, Campbell MJ, et al. 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(6):469-74.
8. Glassman SD, Copay AG, Berven SH, et al. Defining substantial clinical benefit following lumbar spine arthrodesis. J Bone Joint Surg Am. 2008;90(9):1839-47.
9. AlEissa SI, Tamai K, Konbaz F, et al. SPINE20 A global advocacy group promoting evidence-based spine care of value. Eur Spine J. 2021;30(8):2091-101.
10. Wang YC, Hart DL, Stratford PW, et al. Baseline dependency of minimal clinically important improvement. Phys Ther. 2011;91(5):675-88.
11. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol. 1997;50(8):869-79.
12. Porchet F, Lattig F, Grob D, et al. Comparison of patient and surgeon ratings of outcome 12 months after spine surgery: presented at the 2009 Joint Spine Section Meeting. J Neurosurg Spine. 2010;12(5):447-55.
13. Fukui M, Chiba K, Kawakami M, et al. Japanese Orthopaedic Association Back Pain Evaluation Questionnaire. Part 2. Verification of its reliability: the Subcommittee on Low Back Pain and Cervical Myelopathy Evaluation of the Clinical Outcome Committee of the Japanese Orthopaedic Association. J Orthop Sci. 2007;12(6):526-32.
14. 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(3):348-65.
15. 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(4):665-69.
16. Wada E, Fukui M, Takahashi K, et al. Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire (JOACMEQ): part 5. Determination of responsiveness. J Orthop Sci. 2019;24(1):57-61.
17. Azimi P, Yazdanian T, Benzel EC. Determination of minimally clinically important differences for JOABPEQ measure after discectomy in patients with lumbar disc herniation. J Spine Surg. 2018;4(1):102-8.
18. Ogura Y, Ogura K, Kobayashi Y, et al. Minimally clinically important differences for the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire (JOABPEQ) following decompression surgery for lumbar spinal stenosis. J Clin Neurosci. 2019;69:93-6.
19. Arima H, Carreon LY, Glassman SD, et al. Age variation in the minimum clinically important difference in SRS-22r after surgical treatment for adult spinal deformity - A single institution analysis in Japan. J Orthop Sci. 2018;23(1):20-5.
20. Tanaka N, Konno S, Takeshita K, et al. An outcome measure for patients with cervical myelopathy: the Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire (JOACMEQ): an average score of healthy volunteers. J Orthop Sci. 2014;19(1):33-48.
21. Kato M, Konishi S, Matsumura A, et al. Clinical characteristics of intraspinal facet cysts following microsurgical bilateral decompression via a unilateral approach for treatment of degenerative lumbar disease. Eur Spine J. 2013;22(8):1750-7.
22. Taneichi H, Suda K, Kajino T, et al. Unilateral transforaminal lumbar interbody fusion and bilateral anterior-column fixation with two Brantigan I/F cages per level: clinical outcomes during a minimum 2-year follow-up period. J Neurosurg Spine. 2006;4(3):198-205.
23. Hurwitz EE, Simon M, Vinta SR, et al. Adding examples to the ASA-physical status classification improves correct assignment to patients. Anesthesiology. 2017;126(4):614-22.
24. Yone K, Sakou T, Kawauchi Y, et al. Indication of fusion for lumbar spinal stenosis in elderly patients and its significance. Spine. 1996;21(2):242-8.
25. Hirabayashi K, Miyakawa J, Satomi K, et al. Operative results and postoperative progression of ossification among patients with ossification of cervical posterior longitudinal ligament. Spine. 1981;6(4):354-64.
26. Kersten P, White PJ, Tennant A. Is the pain visual analogue scale linear and responsive to change? An exploration using Rasch analysis. PLoS One. 2014;9(6):e99485.
27. Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000;53(5):459-68.
28. Hägg O, Fritzell P, Nordwall A; Swedish Lumbar Spine Study Group. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J. 2003;12(1):12-20.
29. Copay AG, Glassman SD, Subach BR, et al. 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(6):968-74.
30. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. Phys Ther. 1998;78(11):1197-207.
31. Stratford PW, Binkley JM, Riddle DL, et al. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. Phys Ther. 1998;78(11):1186-96.
32. Haberman, SJ. The analysis of residuals in cross-classified tables. Biometrics. 1973;29(1):205-20.
33. Yao M, Yang L, Zhu B, et al. Simplified Chinese Version of the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire: agreement, responsiveness, and minimal important change for patients with chronic low back pain. Spine. 2018;43(20):1438-45.
34. Fujimori T, Miwa T, Oda T. Responsiveness of the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire in lumbar surgery and its threshold for indicating clinically important differences. Spine J. 2019;19(1):95-103.
35. Daltroy LH, Cats-Baril WL, Katz JN, et al. The North American spine society lumbar spine outcome assessment instrument: reliability and validity tests. Spine. 1996;21(6):741-9.
36. Hart RA, Gundle KR, Pro SL, et al. Lumbar Stiffness Disability

Index: pilot testing of consistency, reliability, and validity. Spine J. 2013;13(2):157-61.

37. Matsukura Y, Egawa S, Inose H, et al. Preoperative symptom duration influences neurological recovery and patient-reported outcome measures after surgical treatment of cervical ossification of the posterior longitudinal ligament. Spine. 2023;48(18):1259-65.

38. Pitamberwale A, Mahmood T, Ansari AK, et al. Biochemical parameters as prognostic markers in severely Ill COVID-19 patients. Cureus. 2022;14(8):e28594.