

## 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

MITSURU FUKU<sup>1</sup>, KAZUHIRO CHIBA<sup>2</sup>, MAMORU KAWAKAMI<sup>3</sup>, SHINICHI KIKUCHI<sup>4</sup>, SHINICHI KONNO<sup>4</sup>, MASABUMI MIYAMOTO<sup>5</sup>, ATSUSHI SEICHI<sup>6</sup>, TADASHI SHIMAMURA<sup>7</sup>, OSAMU SHIRADO<sup>8</sup>, TOSHIHIKO TAGUCHI<sup>9</sup>, KAZUHISA TAKAHASHI<sup>10</sup>, KATSUSHI TAKESHITA<sup>6</sup>, TOSHIKAZU TANI<sup>11</sup>, YOSHIAKI TOYAMA<sup>2</sup>, KAZUO YONENOBU<sup>12</sup>, EIJI WADA<sup>13</sup>, TAKASHI TANAKA<sup>14</sup>, and YOSHIO HIROTA<sup>15</sup>

<sup>1</sup>Laboratory of Statistics, Osaka City University Faculty of Medicine, Osaka, Japan

<sup>2</sup>Department of Orthopaedic Surgery, Keio University, Tokyo, Japan

<sup>3</sup>Department of Orthopaedic Surgery, Wakayama Medical University, Wakayama, Japan

<sup>4</sup>Department of Orthopaedic Surgery, School of Medicine, Fukushima Medical University, Fukushima, Japan

<sup>5</sup>Department of Orthopedic Surgery, Nippon Medical School, Tokyo, Japan

<sup>6</sup>Department of Orthopaedic Surgery, the University of Tokyo, Tokyo, Japan

<sup>7</sup>Department of Orthopaedic Surgery, Iwate Medical University School of Medicine, Morioka, Japan

<sup>8</sup>Department of Orthopaedic Surgery, Faculty of Medicine, Saitama Medical University, Moroyama, Saitama 350-0495, Japan

<sup>9</sup>Department of Orthopedic Surgery, Yamaguchi University School of Medicine, Yamaguchi, Japan

<sup>10</sup>Department of Orthopedic Surgery, Graduate School of Medicine, Chiba University, Chiba, Japan

<sup>11</sup>Department of Orthopaedics, Kochi Medical School, Kochi, Japan

<sup>12</sup>National Hospital Organization Osaka-Minami Medical Center, Kawachinagano, Japan

<sup>13</sup>Department of Orthopaedic Surgery, Hoshigaoka Koseinenkin Hospital, Osaka, Japan

<sup>14</sup>Department of Internal Medicine, Houai Hospital, Osaka, Japan

<sup>15</sup>Department of Public Health, Osaka City University Faculty of Medicine, Osaka, Japan

### Abstract

**Background.** The project to develop a new Japanese Orthopaedic Association (JOA) score rating system for low back disorders, the JOA Back Pain Evaluation Questionnaire (JOABPEQ), is currently in progress. Part 1 of the study selected 25 “candidate” items for use on the JOABPEQ. The purpose of this current Part 2 of the study was to verify the reliability of the questionnaire.

**Methods.** A total of 161 patients with low-back disorders of any type participated in the study. Each patient was interviewed twice at an interval of 2 weeks using the same questionnaire. The reliability of the questionnaire was evaluated by determining the extension of the kappa and weighted kappa coefficients.

**Results.** Both kappa and weighted kappa were more than 0.50 for all but one item, which was 0.48. The lower 95% confidence interval exceeded 0.4 in all but two items, which was 0.39. This implied that the test–retest reliability of JOABPEQ was acceptable as a measure of outcome.

**Conclusions.** The tentative questionnaire of the JOABPEQ with 25 items was confirmed to be reliable enough to describe the quality of life of patients who suffer low back disorders.

*Offprint requests to:* O. Shirado

Received: December 26, 2006 / Accepted: July 3, 2007

This report was composed by the Subcommittee on Low Back Pain and Cervical Myelopathy Evaluation of the Clinical Outcome Committee of the Japanese Orthopaedic Association

### Introduction

Measurement of the outcome is critical for any decision-making and results of evaluations in all medical circumstances. This should be applicable when managing patients who have lumbar spine-related problems. The Japan Orthopaedic Association (JOA) developed and published a specific instrument to measure outcomes for patients with low back problems in 1986.<sup>1</sup> It was called the JOA score rating system for low back pain, with a full score being 29 points. Since then, the instrument has been widely utilized to evaluate the functional results of many types of intervention for patients with such problems. It has been referred to not only in articles by Japanese investigators<sup>2</sup> but also in those by non-Japanese-speaking investigators.<sup>3,4</sup> One of the major criticisms of this specific instrument, however, is that it is not a patient-oriented measurement but a physician-based one. It is now widely accepted that a patient’s perspective is essential for making medical decisions and for evaluating the results of interventions.<sup>5</sup> Based on the current needs for measuring outcome, the JOA was urged to revise its original score rating system and to develop a new one. In 2002, a Subcommittee on Evaluation of Back Pain and Cervical Myelopathy was organized in the Clinical Outcome Committee of JOA,

and work began on revising the original JOA scoring system.

This revision process consisted of four steps: Parts 1 to 4. As described in the previous literature concerning Part 1, the original JOA scoring system was revised and a new scoring system (the JOA Back Pain Evaluation Questionnaire — JOABPEQ) was developed.<sup>6</sup> The key points of this revision were to make the original JOA score more patient-oriented. For the survey in the Part

1 study, we first created a preliminary questionnaire consisting of 60 items. The questionnaire was a self-administered, disease-specific measure that was created with reference to the Japanese editions of the short form health survey with 36 questions (SF-36)<sup>7</sup> and the Roland-Morris Disability Questionnaire (RDQ)<sup>8</sup> to assess health-related quality of life. From the survey, a total of 25 items were selected for tentative use on a draft of the JOABPEQ (Table 1).

**Table 1.** Items ( $n = 25$ ) selected for the draft of the JOABPEQ evaluated in this study

*With regard to your health condition during the last week, please circle the item number of the answer for the following questions that best applies. If your condition varies depending on the day or time, circle the item number when your condition is at its worst.*

Q1-1. To alleviate low back pain, you often change your posture.

- 1) Yes
- 2) No

Q1-2. Because of low back pain, you do not do any routine housework these days.

- 1) No
- 2) Yes

Q1-3. Because of low back pain, you lie down more often than usual.

- 1) Yes
- 2) No

Q1-4. Because of low back pain, you sometimes ask someone to help you when you do something.

- 1) Yes
- 2) No

Q1-5. Because of low back pain, you refrain from bending forward or kneeling down.

- 1) Yes
- 2) No

Q1-6. Because of low back pain, you have difficulty standing up from a chair.

- 1) Yes
- 2) No

Q1-7. Your lower back aches most of the time.

- 1) Yes
- 2) No

Q1-8. Because of low back pain, turning over in bed is difficult.

- 1) Yes
- 2) No

Q1-9. Because of low back pain, you have difficulty putting on socks or stockings.

- 1) Yes
- 2) No

Q1-10. Because of low back pain, you walk only short distances.

- 1) Yes
- 2) No

Q1-11. Because of low back pain, you cannot sleep well. (If you take sleeping pills because of the pain, select “No.”)

- 1) No
- 2) Yes

Q1-12. Because of low back pain, you stay seated most of the day.

- 1) Yes
- 2) No

Q1-13. Because of low back pain, you become irritated or angry at other persons more often than usual.

- 1) Yes
- 2) No

Q1-14. Because of low back pain, you go up stairs more slowly than usual.

- 1) Yes
- 2) No

**Table 1.** *Continued*

---

Q2-1. How is your present health condition?

- 1) Excellent
- 2) Very good
- 3) Good
- 4) Fair
- 5) Poor

Q2-2. Do you have difficulty in climbing stairs?

- 1) I have great difficulty.
- 2) I have some difficulty.
- 3) I have no difficulty.

Q2-3. Do you have difficulty in any one of the following motions: bending forward, kneeling, or stooping?

- 1) I have great difficulty.
- 2) I have some difficulty.
- 3) I have no difficulty.

Q2-4. Do you have difficulty walking more than 15 minutes?

- 1) I have great difficulty.
- 2) I have some difficulty.
- 3) I have no difficulty.

Q2-5. Have you been unable to do your work or ordinary activities as well as you would like?

- 1) I have not been able to do them at all.
- 2) I have been unable to do them most of the time.
- 3) I have sometimes been unable to do them.
- 4) I have been able to do them most of the time.
- 5) I have always been able to do them.

Q2-6. Has your work routine been hindered because of the pain?

- 1) Greatly
- 2) Moderately
- 3) Slightly (somewhat)
- 4) Little (minimally)
- 5) Not at all

Q2-7. Have you been discouraged or depressed?

- 1) Always
- 2) Frequently
- 3) Sometimes
- 4) Rarely
- 5) Never

Q2-8. Do you feel exhausted?

- 1) Always
- 2) Frequently
- 3) Sometimes
- 4) Rarely
- 5) Never

Q2-9. Do you feel happy?

- 1) Always
- 2) Almost always
- 3) Sometimes
- 4) Rarely
- 5) Never

Q2-10. Do you think you are in reasonable health?

- 1) Yes (I am healthy.)
- 2) Fairly (my health is better than average)
- 3) Not (very much)/particularly (my health is average)
- 4) Barely (my health is poor)
- 5) Not at all (my health is very poor)

Q2-11. Do you feel your health will get worse?

- 1) Very much so
  - 2) A little at a time
  - 3) Sometimes yes and sometimes no
  - 4) Not very much
  - 5) Not at all
-

The purpose of the Part 2 study in this project was to evaluate the reliability of the 25 items selected for the draft JOABPEQ; for this, test–retest reliability was ascertained.

## Materials and methods

### Recruitment of patients

Altogether, 460 of the 829 Japanese board-certified spine surgeons were randomly selected, and each was asked to recruit two patients to evaluate the JOABPEQ between January and June 2004. The recruited patients were scheduled to reply to the questionnaire twice at a 2-week interval. Patient criteria were as follows: (1) patients could be any age of either sex; (2) patients had any lumbar spine disorder and were currently visiting an outpatient clinic; (3) the severity of the symptoms was expected to be at the same level between the two interviews. Exclusion criteria were the presence of: (1) other musculoskeletal diseases requiring medical treatment; (2) psychiatric disease (e.g., dementia), potentially leading to inappropriate answers; (3) a postoperative condition; (4) having participated in previous surveys of the related study.

### Testing the questionnaire

Each patient was asked to complete the same questionnaire twice at an interval of 2 weeks ( $\pm 3$  days). The attending surgeon filled out the patient information on the diagnosis and the presence or absence of concomitant diseases, followed by judging the severity of symptoms using a three-step rating scale (mild, moderate, severe). Symptom severity was determined subjectively by the attending surgeon, who was asked not to select a similar patient solely on the basis of severity. Patients who had the same level of severity as judged by all surgeons were then selected and analyzed to verify the reliability of the questionnaire.

This study was approved by the Ethics Committee of the Japanese Society for Spine Surgery and Related Research. Informed consent was obtained from each patient.

The reliability of the questionnaire was evaluated by determining the extension of the kappa coefficients. The weighted kappa coefficient was calculated in the items with three choices or more. The kappa and weighted kappa coefficients were calculated based on a formula using Microsoft Office Excel 2003. Kappa and weighted kappa coefficients of 0.4 or above were judged to be reliable.<sup>9</sup> The 95% confidence intervals (95% CI) were calculated for all reliability coefficients using the bootstrap method.

## Results

### Patient characteristics

A total of 350 patients participated in this study and completed the questionnaire twice following the project's plan. However, 135 patients were excluded because the severity of their symptoms had changed between the two interviews or they violated the interval period. Of the remaining 215 patients, 54 were ineligible because of other musculoskeletal diseases, such as knee and hip osteoarthritis. As a result, a total of 161 patients were available for the analysis in this study: 86 men and 75 women with a mean age of 57.7 years (SD 16.3 years). The clinical diagnosis included degenerative lumbar canal stenosis in 49 patients, lumbar disc herniation in 44, spondylolisthesis in 20, spondylosis in 16, degenerative disc disease in 13, mechanical low back pain in 11, and miscellaneous in 8. The patients' age varied from their twenties to their eighties, and symptom severity varied from mild to severe (Table 2). Neurological and physical status was evaluated for each patient using the current JOA score rating system and finger–floor distance (Table 3). Neurological deficits varied from mild to severe, and trunk flexibility varied among the subjects as well.

### Face validity

Face validity was checked in terms of the completion rate for filling out the questionnaire. The distribution of

**Table 2.** Distribution of age and severity of symptoms in the patients analyzed ( $n = 161$ )

Age (years)	No. of patients, by severity of symptoms			
	Mild	Moderate	Severe	Total
Men				
20–	2	2	3	7
30–	4	4	1	9
40–	2	2	0	4
50–	7	9	1	17
60–	13	8	3	24
70–	6	18	0	24
80–	0	0	1	1
Total	34	43	9	86
Women				
20–	2	3	0	5
30–	4	5	0	9
40–	6	1	1	8
50–	9	4	1	14
60–	7	11	1	19
70–	8	9	1	18
80–	1	1	0	2
Total	37	34	4	75
Total no.	71	77	13	161

**Table 3.** Current Japanese Orthopaedic Association score rating system and finger to floor distance for the patients analyzed ( $n = 161$ )

Parameter	No.
SLR test	
Normal	124
30°–70°	35
<30°	2
Motor function	
Normal	113
Slight weakness (MMT: good)	38
Severe weakness (MMT: less than good)	10
Sensory function	
Normal	80
Slight disturbance	59
Severe disturbance	22
Bladder function	
Normal	147
Mild dysuria	12
Severe dysuria	2
Finger to floor distance (cm)	
~–15	1
–14––5	17
–4––4	41
5~14	40
15~24	32
25~34	9
35~44	7
45~54	7
55~64	4
65~74	1
Immeasurable	2
Total no.	161

SLR, straight leg raising; MMT, manual muscle testing

the answers for all question items was then checked to ensure that there were no biased answers. Items remaining unanswered accounted for less than 5% in the first test, and there was no skewed distribution, such as “floor and ceiling” effects (Table 4).

### Reliability

The test–retest reliability was confirmed by calculating the kappa and weighted kappa coefficients for each item (Tables 5A, 5B). Both kappa and weighted kappa were more than 0.50 in all items, except in one item with 0.48. The lower 95% CI exceeded 0.4 in all items, except in two items with 0.39. This implied that the test–retest reliability of JOABPEQ was acceptable as a measurement of outcome.

### Discussion

Measurement of the outcome is generally divided into two categories: generic and disease-specific measures.<sup>5,10</sup>

SF-36 has been commonly used as representative of a measurement of generic health status.<sup>5,7,10</sup> The RDQ and the Oswestry Disability Index are widely used as disease-specific measurements for back pain.<sup>8,11</sup> The JOA score rating system for low back pain, developed in 1986, was also a disease-specific measuring instrument for back disorders and injuries and has been widely utilized in clinical research and the decision-making process in Japan. However, this is not a patient-based outcome measure reliable enough to describe the objective status of the function and quality of life (QOL) of patients with low-back disorders. There has, to date, been insufficient psychometric analysis to confirm the validity and reliability of this JOA score rating system.

The project for developing the new questionnaire, JOABPEQ, was initiated to create a self-administered, disease-specific method for measuring low back pain. This instrument should include functions of the lumbar spine as well as health-related QOL. The reliability of the questionnaire that includes the 25 suggested items was evaluated using psychometric analysis as Part 2 of this project. Kappa and weighted kappa coefficient were utilized to verify the test–retest reliability.<sup>12,13</sup>

In terms of external validity, biased data were inevitable because one criterion that was included was that the severity of the symptoms was expected to be at the same level between the two interviews. However, there was no bias on the choices of answer to each question. This implies that test–retest reliability was acceptable even if the subjects had symptoms of different severity. The older the patients were, the worse was the interpretation of each question. There were small numbers of patients of younger generations, such as those in their thirties and forties, in this study. Thus, the reliability would not deteriorate even if the number of young people were to increase.

In terms of English expression, there is a possibility of ambiguity in questions 1-2 and 1-11, where double negatives (two “no’s” in the answer) may be confusing. It is necessary to reconsider and revise the English expression so it is more easily understood by native English-language users. The number of choices for the answer in all questions varied from two to five, which is also a point to be reconsidered in the future.

The current study demonstrated that the 25 items had enough reliability to describe the QOL in patients suffering low back disorders. However, further studies are needed to complete the project, including a factor analysis to determine the underlying cluster of the questionnaire items, a formula for calculating the severity score, and confirmation of the responsiveness to the questionnaire.

**Table 4.** Reproducibility of each item ( $n = 161$ )

Item	Choices for answer					No answer
	1	2	3	4	5	
Q1-1	117 72.7%	43 26.7%				1 0.6%
Q1-2	32 19.9%	127 78.9%				2 1.2%
Q1-3	76 47.2%	83 51.6%				2 1.2%
Q1-4	42 26.1%	119 73.9%				
Q1-5	77 47.8%	84 52.2%				
Q1-6	31 19.3%	130 80.7%				
Q1-7	68 42.2%	93 57.8%				
Q1-8	65 40.4%	95 59.0%				1 0.6%
Q1-9	72 44.7%	89 55.3%				
Q1-10	87 54.0%	74 46.0%				
Q1-11	35 21.7%	122 75.8%				4 2.5%
Q1-12	41 25.5%	119 73.9%				1 0.6%
Q1-13	36 22.4%	125 77.6%				
Q1-14	115 71.4%	45 28.0%				1 0.6%
Q2-1		18 11.2%	59 36.6%	66 41.0%	16 9.9%	2 1.2%
Q2-2	15 9.3%	92 57.1%	52 32.3%			2 1.2%
Q2-3	25 15.5%	93 57.8%	38 23.6%			5 3.1%
Q2-4	35 21.7%	70 43.5%	55 34.2%			1 0.6%
Q2-5	15 9.3%	12 7.5%	85 52.8%	35 21.7%	13 8.1%	1 0.6%
Q2-6	13 8.1%	36 22.4%	66 41.0%	32 19.9%	11 6.8%	3 1.9%
Q2-7	12 7.5%	8 5.0%	79 49.1%	39 24.2%	22 13.7%	1 0.6%
Q2-8	8 5.0%	27 16.8%	88 54.7%	24 14.9%	12 7.5%	2 1.2%
Q2-9	8 5.0%	42 26.1%	74 46.0%	31 19.3%	5 3.1%	1 0.6%
Q2-10	13 8.1%	59 36.6%	42 26.1%	34 21.1%	12 7.5%	1 0.6%
Q2-11	17 10.6%	48 29.8%	56 34.8%	28 17.4%	11 6.8%	1 0.6%

**Table 5A.** Kappa coefficient with 95% CI for items Q1-1 to Q1-14

Item	$\kappa$	95% CI
Q1-1	0.69	0.60–0.77
Q1-2	0.62	0.51–0.73
Q1-3	0.67	0.60–0.75
Q1-4	0.65	0.56–0.75
Q1-5	0.48	0.39–0.57
Q1-6	0.55	0.43–0.66
Q1-7	0.65	0.57–0.73
Q1-8	0.55	0.47–0.64
Q1-9	0.71	0.64–0.78
Q1-10	0.63	0.55–0.72
Q1-11	0.50	0.39–0.61
Q1-12	0.56	0.46–0.65
Q1-13	0.65	0.55–0.74
Q1-14	0.72	0.64–0.80

CI, confidence interval

**Table 5B.** Weighted kappa coefficient with 95% CI for items Q2-1 to Q2-11

Item	Weighted $\kappa$	95% CI
Q2-1	0.51	0.43–0.57
Q2-2	0.61	0.52–0.68
Q2-3	0.57	0.49–0.64
Q2-4	0.73	0.68–0.78
Q2-5	0.54	0.47–0.60
Q2-6	0.61	0.55–0.67
Q2-7	0.53	0.46–0.59
Q2-8	0.55	0.48–0.61
Q2-9	0.54	0.46–0.60
Q2-10	0.54	0.47–0.61
Q2-11	0.53	0.46–0.60

## Conclusions

The tentative JOABPEQ with 25 items was confirmed to be reliable enough to describe the QOL of patients suffering low back disorders.

## References

- Izumida S, Inoue S. Assessment of treatment for low back pain. *J Jpn Orthop Assoc* 1986;60:391–4 (in Japanese).
- Takahashi K, Kitahara H, Yamagata M, Murakami M, Takata K, Miyamoto K, et al. Long-term results of anterior interbody fusion for treatment of degenerative spondylolisthesis. *Spine* 1990;15:1211–5.
- Tiusanen H, Hurri H, Seitsalo S, Osterman K, Harju R. Functional and clinical results after anterior interbody lumbar fusion. *Eur Spine J* 1996;5:288–92.
- Jeong GK, Bendo JA. Lumbar intervertebral disc cyst as a cause of radiculopathy. *Spine J* 2003;3:242–6.
- Kopec JA. Measuring functional outcomes in persons with back pain: a review of back-specific questionnaires. *Spine* 2000;25:3110–4.
- Fukui M, Chiba K, Kawakami M, Kikuchi S, Konno S, Miyamoto M, et al. JOA Back Pain Evaluation Questionnaire: initial report. *J Orthop Sci* 2007;12:443–50.
- Fukuhara S, Bito S, Green J, Hsiao A, Kurokawa K. Translation, adaptation, and validation of the SF-36 health survey for use in Japan. *J Clin Epidemiol* 1998;51:1037–44.
- Suzukamo Y, Fukuhara S, Kikuchi S, Konno S, Roland M, Iwamoto Y, et al. Validation of the Japanese version of the Roland-Morris Disability Questionnaire. *J Orthop Sci* 2003;8:543–8.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 2000;25:3100–3.
- Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine* 2000;25:3115–24.
- Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology* 1996;7:199–202.
- Hernandez-Cruz B, Cardiel MH. Intra-observer reliability of commonly used outcome measures in rheumatoid arthritis. *Clin Exp Rheumatol* 1998;16:459–62.