

*From the Japanese Orthopaedic Association*

**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

MITSURU FUKUI<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>,  
EJI 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, 35 Shinanomachi, Shinjuku, Tokyo 160-8582, 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, Saitama Medical School, Saitama, 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>Department of Orthopaedic Surgery, Hoshigaoka Koseinenkin Hospital, Osaka, Japan  
<sup>13</sup>National Hospital Organization, Osaka-Minami Medical Center, 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

**CONTENTS**

1. The aim of the revision and the story behind it - - - - - 349  
2. User's guide for the JOA Back Pain Evaluation Questionnaire (JOABPEQ) - - - - - 351  
3. JOA Back Pain Evaluation Questionnaire - - - - - 355  
4. User's guide for the JOA Cervical Myelopathy Evaluation Questionnaire (JOACMEQ) - - - - - 358  
5. JOA Cervical Myelopathy Evaluation Questionnaire - - - - - 362  
6. Q&As on the JOABPEQ and the JOACMEQ - - - - - 365

## The aim of the revision and the story behind it

With the trend toward “consumer first” — and this trend applies to the medical community as well — the initiative has been shifting from supplier (e.g., physicians) to recipients (e.g., patients). Naturally, systems and elements in the medical field must also change according to this trend. Among them is the issue of evaluation systems for treatment results of various diseases.

First, those who ultimately evaluate the treatment results should be patients (patient-based). From the viewpoint of patients, diseases are not just pathological matters; they are dysfunctions caused by the disease, disabilities due to such dysfunctions, psychological problems due to such disabilities, and problems encountered in the course of social life. The problems patients suffer have multidimensional aspects. Therefore, it is expected that treatment results should be evaluated multidimensionally (multidimensional evaluation).

Evidence-based medicine requires the assumption of various scientifically based criteria. Given this background, the Japanese Orthopaedic Association (JOA) established the Clinical Outcome Committee on September 3, 1999 and decided to revise various criteria for the evaluation of treatment results. It requested its associated academic societies to revise the criteria related to the focus of each society.

On June 8, 2000, in response to this request, the Japanese Society for Spine Surgery and Related Research (JSSR) and the Japanese Society of Lumbar Spine Disorders, jointly held the first working group meeting. Its aim was to revise the JOA scoring systems for cervical myelopathy and low back pain.

To establish new criteria based on the fundamental philosophy “patient-based, multidimensional, and scientific,” the members of the working group (individual names are listed elsewhere) devoted their endeavors and accomplished the revision described in the following section. We take this opportunity to acknowledge the members of JSSR, from whom we received enormous help in implementing various investigations.

### 1. First investigation: selection of items for the questionnaire<sup>1-3</sup>

For both cervical myelopathy and back pain, we selected candidate items that should be included in the new evaluation questionnaire from those used during conventional evaluations, those used in countries outside Japan, and those used to assess quality of life (QOL). We used this process to assemble a universal, international, multidimensional evaluation. For the evaluation of cervical myelopathy, we included items that evaluated cervical spine function that had not been included in the current system. The first draft was made

in the form of patient self-administered questionnaires, so the responses of the patients would not be biased by interactions with the treating physicians. For cervical myelopathy 216 healthy volunteers and 250 patients completed the draft questionnaire, and for back pain 216 healthy volunteers and 346 patients did so. Although the details of the analytical methods are not shown here, the items were further selected from among candidate items based on the following considerations: The items should be independent of each other; the items should not be of such ambiguity that they could cause response variations; and the questions were kept to a reasonable number. In addition, regarding the questions on QOL, we used items from the Japanese version of SF-36, the reliability of which has already been validated.

### 2. Second investigation: verification of the reproducibility of the selected items<sup>4,5</sup>

To verify that the selected items have reproducibility, we conducted second-step surveys of 304 patients for cervical myelopathy and 350 patients for back pain. The surveys were repeated twice at a 4-week interval for cervical myelopathy and at a 2-week interval for back pain. As a result, the reproducibility of the questionnaires was verified.

### 3. Third investigation: verification of the validity of the questionnaire and establishment of measurement scales<sup>6,7</sup>

Using the items fixed at the second investigation, we administered the questionnaires to patients with cervical myelopathy or lumbar diseases of different severity to ensure their accuracy and responsiveness: 369 patients for cervical myelopathy and 452 patients for back pain. The equations to calculate the scores for disease severity were established based on this third investigation.

### 4. Fourth investigation: verification of the sensitivity of the functional scores for treatment results

To assess whether the finalized questionnaire can reflect the change in patients' conditions after various treatments, we implemented the fourth investigation on 221 patients with cervical myelopathy and 313 patients with back pain (534 patients in total) who underwent surgery. The results indicated that the new evaluation questionnaire has high sensitivity for assessing treatment results.

We believe that the new evaluation questionnaire satisfies the requirements of the fundamental philosophy, “patient-based, multidimensional, and scientific.” However, even after undertaking the processes described above, some problems remain. Because QOL assessments are neither widely accepted nor have been used in Japan, we had to apply those used in other countries

with necessary changes; therefore, in some cases wording of the questions may be a little awkward when expressed in Japanese. For now, however, we do not have any choice but to wait for studies of this kind to prevail in Japan.

To fulfill the scientific integrity of the questionnaire, the equations to calculate the functional scores for the various domains are so complicated that it is difficult to calculate the scores manually. To solve this problem, we have developed an Excel file containing a spreadsheet that automatically calculates the functional scores after simply filling in the numbers in front of the answers for the items selected by the patients. The files can be downloaded from the web pages of the JOA and the JSSR.<sup>8,9</sup> We believe that most problems can be solved by using this file.

The most significant problem is how we can integrate the results obtained from the new questionnaire with those obtained using the conventional JOA scoring system. In essence, because the concept of this new evaluation instrument and its criteria are completely different from those of previous ones, the resultant data would not correspond to those of conventional evaluation criteria. However, as the approaches of clinical research appear to become more and more prospective, we believe that it is not necessary to stick to conventional evaluation criteria. For a while, however, in order to see long-term results retrospectively, we have to use

the conventional criteria. Thus, for the time being, it appears that we have no choice but to use both sets of criteria simultaneously.

## References

1. Fukui M, Chiba K, Kawakami M, et al. An outcome measure for patients with cervical myelopathy: Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire (JOACMEQ). Part 1. *J Orthop Sci* 2007;12:227–40.
2. 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.
3. Fukui M, Chiba K, Kawakami M, et al. JOA Back Pain Evaluation Questionnaire: initial report. *J Orthop Sci* 2007;12:443–50.
4. Fukui M, Chiba K, Kawakami M, et al. Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire: Part 3. Determination of reliability. *J Orthop Sci* 2007;12:321–6.
5. Fukui M, Chiba K, Kawakami M, et al. Japanese Orthopaedic Association Back Pain Evaluation Questionnaire. Part 2. Verification of its reliability. *J Orthop Sci* 2007;12:526–32.
6. Fukui M, Chiba K, Kawakami M, et al. Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire (JOACMEQ). Part 4. Establishment of equations for severity scores. *J Orthop Sci* 2008;13:25–31.
7. Fukui M, Chiba K, Kawasaki M, et al. Japanese Orthopaedic Association Back Pain Evaluation Questionnaire (JOABPEQ). Part 3. Validity study and establishment the measurement scale. *J Orthop Sci* 2008;13:173–9.
8. [http://www.joa.or.jp/english/english\\_frame.html](http://www.joa.or.jp/english/english_frame.html)
9. [http://www.jssr.gr.jp/jssr\\_web/html/e/index.html](http://www.jssr.gr.jp/jssr_web/html/e/index.html)

## User's guide for the JOA Back Pain Evaluation Questionnaire (JOABPEQ) (2007.3.8)

The first step is to have the patients themselves answer 25 questions. The examiner then calculates five functional scores for corresponding domains according to the provided calculating formulas. (See ref\*1, below.) Please use the free Excel file that is available on the website of the JOA and JSSR. The scores can then be calculated automatically.

- Because each functional score indicates an evaluation based on patient subjectivity, the symptoms that may appear similar to doctors' eyes might not necessarily result in the same score.
- The range of the score for each domain is from 0 to 100, with higher scores indicating a better condition.
- The five functional scores should be used independently. Adding all or some of the five scores makes no sense, so do not try to total the scores.
- As the five severity scores are not confirmed to follow a normal distribution (Table 1), nonparametric statistical tests should be used.

### At calculation

- The functional score should be calculated only if all questions for that particular domain are answered. For example, if the answer for question 1-1 is missing, the "Low Back Pain" score cannot be calculated, although all of the other 24 questions are answered. The other four severity scores, however, can be calculated).
  - Note that the answer for question 3-5 is required to calculate both "Walking ability" and "Social life function" scores.
- The functions for the domain for which the scores cannot be calculated are regarded as those that cannot be evaluated. The functional scores for the representative domains that can be evaluated, if not all five, can be used individually for statistical analyses for intergroup comparisons.
- The software in the Excel file is designed to calculate automatically the difference in the severity scores between two time points (e.g., pretreatment and posttreatment) for each person, so the effect of a specific treatment can be evaluated. However, if the results of only one time point are entered, the severity score for that single time point alone is displayed.

- Only the functional score for the domain, in which the answers for all questions are obtained, are displayed in the Excel software. If the difference in the functional scores of a domain between two time points increases by 20 points, that function is judged to be "effective."
  - If the patient did not give all answers for the questions in a domain:
    - The functional score of the domain that includes the unanswered question cannot be calculated, and this patient should be excluded from the score analysis of intergroup comparisons for that domain.
    - You can judge that a treatment is "effective" for a patient if:
      - 1) The patient give all answers for the questions necessary to calculate the functional score of a domain and an increase of  $\geq 20$  points is obtained for that score, *or*
      - 2) The functional score after treatment is  $>90$  points even if the answer for the unanswered questions was supposed to be the worst possible choice

The effectiveness of the treatment can be evaluated based only on the two above-mentioned conditions.

### At analysis

- How to express the functional scores appropriately in the analyses for a group of patients
 

*In the text:* median (minimum value – maximum value) or median (25 percentile value – 75 percentile value)

*In graphs:* box plot
- Comparison between different groups at one time point (e.g., pretreatment)
  - When you want to state that "there is no statistically significant difference between different groups at a certain time point (e.g., pretreatment)"
 

*For two groups:* Use the Mann-Whitney U-test (Wilcoxon rank-sum test can also be used).

*For more than two groups:* Use the Kruskal-Wallis test.

In both cases, if the null hypothesis is not rejected after setting the significance level to 20% or less,

you can state that “there is no significant difference between the groups.” (See ref \*3, below.)

- When you want to state that “there is a statistically significant difference between different groups at a certain time point (e.g., pretreatment)”

*For two groups:* Use the Mann-Whitney U-test (Wilcoxon rank-sum test can also be used).

*For more than two groups:* Use the Kruskal-Wallis test.

In both cases, if the null hypothesis is rejected against an appropriately set significance level, you can state that “there is a significant difference between the two groups.” As a post hoc test, a nonparametric multiple comparison test should be used.

- To evaluate a therapeutic effect of a treatment (e.g., performing surveys at two time points, such as pretreatment and posttreatment)

#### For individual patients

- 1) If the post-treatment score increases by  $\geq 20$  points compared to the pretreatment score
- 2) If the pretreatment score is  $< 90$  and the posttreatment score reaches  $\geq 90$  points

If either 1) or 2) is satisfied, the treatment is judged “effective.”

#### For a group

$(\text{no. of patients judged “effective”}) / [(\text{total no. of patients in the group}) - (\text{no. of patients whose pre- and posttreatment scores are both } \geq 90)]$

- Evaluating the difference in effectiveness of a treatment between different groups (one group, two groups, or more)

After excluding patients whose pretreatment and post-treatment scores are both  $\geq 90$  points from the analysis, use either of the following:

- 1) Calculate the effectiveness rate in each group and then perform tests of population proportion.
- 2) After calculating the differences of the scores between two time points in individual patients, perform parametric group comparison tests ( $t$ -test for two groups; one-way analysis of variance and multiple comparison post hoc tests for three groups or more) assuming that the scores in the groups follow normal distribution.
  - You can analyze the difference of scores for individual patients assuming that they follow a normal distribution. (See ref \*2, below.)
  - It is misleading to decide that there is a significant difference in the therapeutic effect between two groups when no significant difference is detected before treatment but is detected after treatment by a group comparison because the correspondence of the data between pretreatment and posttreatment is not considered.

**References**

**\*1:**

Low back pain

$$(Q1-1 \times 20 + Q1-2 \times 20 + Q1-3 \times 20 + Q1-4 \times 10 - 70) \times 100 \div 70$$

Lumbar function

$$(Q2-1 \times 10 + Q2-2 \times 10 + Q2-3 \times 20 + Q2-4 \times 10 + Q2-5 \times 30 + Q2-6 \times 20 - 100) \times 100 \div 120$$

Walking ability

$$(Q3-1 \times 30 + Q3-2 \times 20 + Q3-3 \times 10 + Q3-4 \times 10 + Q3-5 \times 30 - 100) \times 100 \div 140$$

Social life function

$$(Q3-5 \times 4 + Q4-1 \times 2 + Q4-2 \times 6 + Q4-3 \times 10 - 22) \times 100 \div 74$$

Mental health

$$(Q5-1 \times 3 + Q5-2 \times 4 + Q5-3 \times 6 + Q5-4 \times 6 + Q5-5 \times 3 + Q5-6 \times 3 + Q5-7 \times 3 - 28) \times 100 \div 103$$

**\*2:**

**Table 1.** The normality parameter of functional scores

	The 3rd survey			The 1st test of the 4th survey			The difference between the 1st and 2nd tests of the 4th survey		
	n	Skewness	Kurtosis	n	Skewness	Kurtosis	n	Skewness	Kurtosis
Low back pain	447	0.39	-0.97	244	0.85	-0.31	222	-0.37	-0.49
Lumbar function	438	0.08	-1.05	241	0.13	-0.97	216	-0.10	-0.24
Walking ability	435	0.34	-1.00	241	0.70	-0.60	216	-0.16	-0.51
Social life function	446	0.28	-0.39	243	0.32	-0.35	220	-0.02	-0.29
Mental health	445	-0.12	0.04	244	-0.02	-0.41	223	0.20	0.21

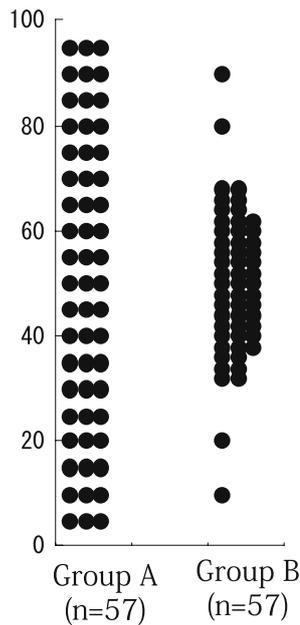
Threshold value beyond which normality is negated ( $\alpha = 0.05$ )

n	Skewness	Kurtosis
200	0.339	0.679
250	0.304	0.607
300	0.277	0.554
350	0.257	0.513
400	0.240	0.480
450	0.226	0.453

**\*3:**

- In general, “no significant difference” does not necessarily mean “equal.” It may be because the test simply lacks sufficient statistical power to detect the difference. To avoid such problems, when the result “no significant difference” is obtained even after lowering the significance level you may state that “there is no significant difference between the groups” and assume that the groups are equal and then proceed to further analyses.
- As any of these tests compare the center position of distribution and because of the issue of variation in the statistical power, the results do not necessarily

guarantee the equality of the compared groups. For example, in cases such as in Fig. 1, as the centers of the distribution curves in both groups A and B are approximately 50, and even though the distributions are evidently different, a significant difference in the median would not be detected between the two groups. In this case, select 30–70 samples that would have almost the same distribution for both groups and repeat the analysis. If you obtain results similar to those for the whole population, it will ensure your decision and you may be able to avoid errors arising from the difference in distribution in the two groups.

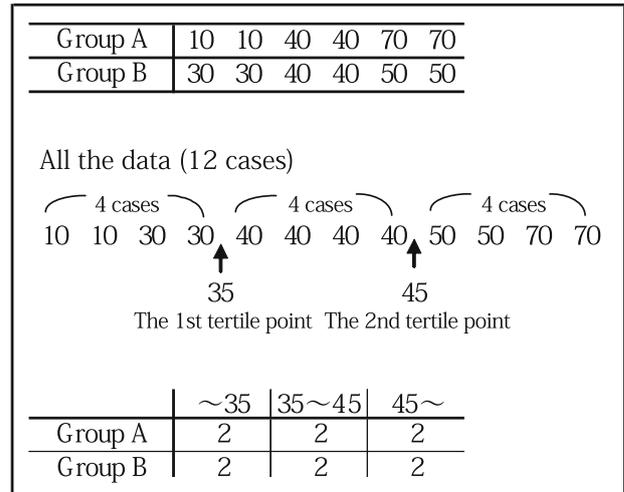


**Fig. 1.** An example with difference distributions

It is not always necessary to disclose results of the analysis conducted with such selected cases in an article. Even if the situation is not as extreme as in Fig. 1, “the analyses with selected cases” may be necessary for the following occasions.

1) Divide the cases into five classes by using quintile points if the total number of cases is approximately the group number  $\times 50$  or more, and into three classes by using tertile points if the total number is approximately the group number  $\times 30$ . (See Example 1, below.)

**Example 1.** Classification using tertile points



- 2) If there is a class with fewer than five cases → perform “analysis with selected cases.”
- 3) If there is no class with fewer than 5 cases → if the result of the  $\chi^2$  test shows a significant difference, also perform “analysis with selected cases.”

## JOA Back Pain Evaluation Questionnaire

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

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

- 1) Yes                      2) No

**Q1-2 Because of the low back pain, you lie down more often than usual.**

- 1) Yes                      2) No

**Q1-3 Your lower back is almost always aching.**

- 1) Yes                      2) No

**Q1-4 Because of the low back pain, you cannot sleep well.** (If you take sleeping pills because of the pain, select "No.")

- 1) No                        2) Yes

**Q2-1 Because of the low back pain, you sometimes ask someone to help you when you do something.**

- 1) Yes                      2) No

**Q2-2 Because of the low back pain, you refrain from bending forward or kneeling down.**

- 1) Yes                      2) No

**Q2-3 Because of the low back pain, you have difficulty standing up from a chair.**

- 1) Yes                      2) No

**Q2-4 Because of the low back pain, turning over in bed is difficult.**

- 1) Yes                      2) No

**Q2-5 Because of the low back pain, you have difficulty putting on socks or stockings.**

- 1) Yes                      2) No

**Q2-6 Do you have difficulty with 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

**Q3-1 Because of the low back pain, you walk only short distances.**

- 1) Yes                      2) No

**Q3-2 Because of the low back pain, you stay seated most of the day.**

- 1) Yes                      2) No

**Q3-3 Because of the low back pain, you go up the stairs more slowly than usual.**

- 1) Yes                      2) No

**Q3-4 Do you have difficulty going up the stairs?**

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

**Q3-5 Do you have difficulty walking more than 15 minutes?**

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

**Q4-1 Because of the low back pain, you do not do any routine housework these days.**

- 1) No
- 2) Yes

**Q4-2 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.

**Q4-3 Has your work routine been hindered because of the pain?**

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

**Q5-1 Because of the low back pain, you get irritated or get angry at other persons more often than usual.**

- 1) Yes
- 2) No

**Q5-2 How is your present health condition?**

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

**Q5-3 Have you been discouraged and depressed?**

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

**Q5-4 Do you feel exhausted?**

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

**Q5-5 Have you felt happy?**

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

**Q5-6 Do you think you are in decent health?**

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

**Q5-7 Do you feel your health will get worse?**

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

On a scale of 0 to 10, regarding 0 as “no pain/numbness at all” and 10 as “the most intense pain/numbness imaginable,” mark a point between 0 and 10 on the lines below to show the degree of your pain or numbness when your symptom was at its worst during the last week.

	0		10
Degree of low back pain	<hr/>		
Degree of pains in buttocks and lower limb(s)	<hr/>		
Degree of numbness in buttocks and lower limb(s)	<hr/>		

## User's guide for the JOA Cervical Myelopathy Evaluation Questionnaire (JOACMEQ) (2007.3.8)

First, have the patients answer 24 questions by themselves, and then you calculate five functional scores for corresponding domains according to the provided calculating formulas. (See ref\*1.) Please use the Excel file that is available on the websites of JOA and JSSR for free. The scores can then be calculated automatically.

- Because each functional score indicates an evaluation based on patient subjectivity, the symptoms that may appear similar to doctors' eyes might not necessarily result in the same score.
- The range of each functional score is from 0 to 100, with higher scores indicating better condition.
- Five functional scores should be used independently. Adding all or some of the five scores makes no sense, so do not try to total the scores.
- As the five functional scores are not confirmed to follow normal distribution (Table 2), nonparametric statistical tests should be used.

### At calculation

- The functional scores should be calculated only if all the questions for the domain are answered (e.g. If the answer for the question 1-1 is missing, the cervical spine function score cannot be calculated, although all other 23 questions are answered; the other 4 severity scores, however, can be calculated).
  - The answer for question 1-4 is required to calculate both "cervical spine function" and "upper extremity function" scores.
  - The answer for question 3-1 is required to calculate both "upper extremity function" and "lower extremity function" scores.
- The functions of the domain, the score for which cannot be calculated, are regarded as those that cannot be evaluated. The functional scores for the respective domains that could be evaluated, if not all five, can be used individually for statistical analyses for intergroup comparisons.
- The software in the Excel file is designed to calculate automatically the difference in the functional scores between two time points (e.g., pretreatment and posttreatment) for each person, so the effect of a specific treatment can be evaluated. However, if the results of only one time point are entered, the functional score for that single time point alone is displayed.

- Only the functional score, in which the answers for all questions are obtained, is displayed in the Excel software. If the difference in the functional scores of a domain between two time points increase by 20 points, that function can be judged to be "effective."
  - If the patient did not give all answers for the questions in a domain:
    - The functional score of the domain cannot be calculated, and this patient should be excluded from the intergroup comparisons for that domain.
    - You can judge that a treatment is "effective" for a patient if:
      - 1) The patient give all answers for the questions necessary to calculate the functional score of a domain and an increase of  $\geq 20$  points is obtained for that score, *or*
      - 2) The functional score after treatment is  $>90$  points even if the answer for the unanswered questions was supposed to be the worst possible choice.

The effectiveness of the treatment can be evaluated based only on the two above-mentioned conditions.

### At analysis

- How to express the functional scores appropriately in the analyses for a group of patients
 

*In the text:* median (minimum value – maximum value) or median (25th percentile value – 75th percentile value)

*In graphs:* box plot
- Comparison between different groups at one time point (e.g., pretreatment)
  - When you want to state that "there is no statically significant difference between different groups at a certain time point (e.g., pretreatment)"
 

*For two groups:* Use the Mann-Whitney U-test (Wilcoxon rank-sum test can also be used).

*For more than two groups:* Use the Kruskal-Wallis test.

In both cases, if the null hypothesis is not rejected after setting the significance level to  $\leq 20\%$ , you can state that "there is no significant difference between the groups."
  - When you want to state that "there is a statistically significant difference between different groups at a certain time point (e.g., pretreatment)"

*For two groups:* Use the Mann-Whitney U-test (Wilcoxon rank-sum test can also be used).

*For more than two groups:* Use the Kruskal-Wallis test.

In both cases, if the null hypothesis is rejected against an appropriately set significance level, you can state that “there is a significant difference between the two groups.” As a post hoc test, a nonparametric multiple comparison test should be used.

- To evaluate a therapeutic effect of a treatment (e.g., performing surveys at two time points such as pretreatment and posttreatment)

#### For individual patients

- 1) If the posttreatment score increases by  $\geq 20$  points compared to the pretreatment score
- 2) If the pretreatment score is  $< 90$ , and the posttreatment score reaches  $\geq 90$  points

If either 1) or 2) is satisfied, the treatment is judged “effective.”

#### For a group

$(\text{no. of patients judged “effective”}) / [(\text{total no. of patients in the group}) - (\text{no. of patients whose pre- and posttreatment scores are both } \geq 90)]$

- Evaluating the difference in effectiveness of a treatment between different groups (one group, two groups, or more)

After excluding patients whose pretreatment and posttreatment scores are both  $\geq 90$  points from the analysis, use either of the following:

- 1) Calculate the effectiveness rate in each group and then perform tests of population proportion.
- 2) After calculating the differences of the scores between two time points in individual patients, perform nonparametric group comparison tests (either the Mann-Whitney U-test or the Wilcoxon rank-sum test for two groups or the Kruskal-Wallis test and multiple comparison post hoc tests for three groups or more).
  - It is not confirmed whether the difference of scores for individual patients also follows normal distribution. (ref \*2)
  - It is misleading to decide that there is a significant difference in the therapeutic effect between two groups when no significant difference is detected before treatment but is detected after treatment by a group comparison because the correspondence of the data between pretreatment and posttreatment is not considered.

References

**\*1:**

Cervical spine function

$$(Q1-1 \times 20 + Q1-2 \times 10 + Q1-3 \times 15 + Q1-4 \times 5 - 50)$$

Upper extremity function

$$(Q1-4 \times 5 + Q2-1 \times 10 + Q2-2 \times 15 + Q2-3 \times 5 + Q3-1 \times 5 - 40) \times 100 \div 95$$

Lower extremity function

$$(Q3-1 \times 10 + Q3-2 \times 10 + Q3-3 \times 15 + Q3-4 \times 5 + Q3-5 \times 5 - 45) \times 100 \div 110$$

Bladder function

$$(Q4-1 \times 10 + Q4-2 \times 5 + Q4-3 \times 10 + Q4-4 \times 5 - 30) \times 100 \div 80$$

Quality of life

$$(Q5-1 \times 3 + Q5-2 \times 2 + Q5-3 \times 2 + Q5-4 \times 5 + Q5-5 \times 4 + Q5-6 \times 3 + Q5-7 \times 2 + Q5-8 \times 3 - 24) \times 100 \div 96$$

**\*2:**

**Table 1.** The normality parameter of functional scores

	The 3rd survey			The 1st test of the 4th survey			The difference between the 1st and 2nd tests of the 4th survey		
	n	Skewness	Kurtosis	n	Skewness	Kurtosis	n	Skewness	Kurtosis
Cervical spine function	360	-0.48	-0.91	178	-0.37	-1.06	158	-0.06	1.79
Upper extremity function	364	-0.87	0.19	179	-0.60	-0.57	163	0.26	1.15
Lower extremity function	357	-0.35	-0.84	181	-0.23	-1.12	167	0.56	0.22
Bladder function	360	-1.04	1.01	181	-0.69	-0.29	165	0.64	1.29
Quality of life	352	-0.04	-0.11	178	-0.04	-0.54	160	0.57	0.30

Threshold value beyond which normality is negated ( $\alpha = 0.05$ )

n	Skewness	Kurtosis
200	0.339	0.679
250	0.304	0.607
300	0.277	0.554
350	0.257	0.513
400	0.240	0.480
450	0.226	0.453

**\*3:**

- In general, “no significant difference” does not necessarily mean “equal.” This may be because the test simply lacks sufficient statistical power to detect the difference. To avoid such problems, when the result “no significant difference” is obtained even after lowering the significance level, you may state that “there is no significant difference between the groups” and assume that the groups are equal and then proceed to further analyses.

- As any of these tests compare the center position of distribution and because of the issue of a variation in the statistical power, the results do not necessarily guarantee the equality of the compared groups. (For example, in cases such as that in Fig. 1, as the centers of the distribution curves in both groups A and B are approximately 50 and even though the distributions are evidently different, a significant difference in the median would not be detected between the two groups.) In this case, select 30–70 samples that would



## JOA Cervical Myelopathy Evaluation Questionnaire

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

**Q1-1 While in the sitting position, can you look up at the ceiling by tilting your head upward?**

- 1) Impossible
- 2) Possible to some degree (with some effort)
- 3) Possible without difficulty

**Q1-2 Can you drink a glass of water without stopping despite the neck symptoms?**

- 1) Impossible
- 2) Possible to some degree
- 3) Possible without difficulty

**Q1-3 While in the sitting position, can you turn your head toward the person who is seated to the side but behind you and speak to that person while looking at his/her face?**

- 1) Impossible
- 2) Possible to some degree
- 3) Possible without difficulty

**Q1-4 Can you look at your feet when you go down the stairs?**

- 1) Impossible
- 2) Possible to some degree
- 3) Possible without difficulty

**Q2-1 Can you fasten the front buttons of your blouse or shirt with both hands?**

- 1) Impossible
- 2) Possible if I spend time
- 3) Possible without difficulty

**Q2-2 Can you eat a meal with your dominant hand using a spoon or a fork?**

- 1) Impossible
- 2) Possible if I spend time
- 3) Possible without difficulty

**Q2-3 Can you raise your arm? (answer for the weaker side)**

- 1) Impossible
- 2) Possible up to shoulder level
- 3) Possible although the elbow and/or wrist is a little flexed
- 4) I can raise it straight upward

**Q3-1 Can you walk on a flat surface?**

- 1) Impossible
- 2) Possible but slowly even with support
- 3) Possible only with the support of a handrail, a cane, or a walker
- 4) Possible but slowly without any support
- 5) Possible without difficulty

**Q3-2 Can you stand on either leg without the support of your hand? (Do you need to support yourself?)**

- 1) Impossible with either leg
- 2) Possible on either leg for more than 10 seconds
- 3) Possible on both legs individually for more than 10 seconds

**Q3-3 Do you have difficulty going up stairs?**

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

**Q3-4 Do you have difficulty with 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

**Q3-5 Do you have difficulty walking more than 15 minutes?**

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

**Q4-1 Do you have urinary incontinence?**

- 1) Always
- 2) Frequently
- 3) When retaining urine over a period of more than 2 hours
- 4) When sneezing or straining
- 5) No

**Q4-2 How often do you go to the bathroom at night?**

- 1) Three times or more
- 2) Once or twice
- 3) Rarely

**Q4-3 Do you have a feeling of residual urine in your bladder after voiding?**

- 1) Most of the time
- 2) Sometimes
- 3) Rarely

**Q4-4 Can you initiate (start) your urine stream immediately when you want to void?**

- 1) Usually not
- 2) Sometimes
- 3) Most of the time

**Q5-1 How is your present health condition?**

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

**Q5-2 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.

**Q5-3 Has your work routine been hindered because of the pain?**

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

**Q5-4 Have you been discouraged and depressed?**

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

**Q5-5 Do you feel exhausted?**

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

**Q5-6 Have you felt happy?**

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

**Q5-7 Do you think you are in decent health?**

- 1) Not at all (my health is very poor)
- 2) Barely (my health is poor)
- 3) Not very much (my health is average)

- 4) Fairly (my health is better than average)
- 5) Yes (I am healthy)

**Q5-8 Do you feel your health will get worse?**

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

On a scale of 0 to 10, regarding 0 as “no pain (numbness) at all” and 10 as “the most intense pain (numbness) imaginable,” mark a point between 0 and 10 on the lines below to show the degree of your pain or numbness when your symptom was at its worst during the last week.

If you feel pain or stiffness in your neck or shoulders, mark the degree.

0 10

---

If you feel tightness in your chest, mark the degree.

0 10

---

If you feel pain or numbness in your arms or hands, mark the degree. (If there is pain in both limbs, judge the worse of the two.)

0 10

---

If you feel pain or numbness from chest to toe, mark the degree

0 10

---

## Q&As on the JOABPEQ and the JOACMEQ

**Q. Is it not necessary to include items for objective findings examined by the treating physicians?**

A. The aim of both the JOABPEQ and the JOACMEQ is to make the evaluation from the patient's standpoint. It is not possible to assess all aspects of patient's condition with these questionnaires. Therefore, it is advisable to include additional evaluations from physicians' points of view at the same time.

**Q. The English seems awkward.**

A. Both questionnaires were first translated into English by a Professor of English language and then checked and revised by several English-native MDs and finalized by the Subcommittee on Low Back Pain and Cervical Myelopathy Evaluation. We are planning to introduce the JOABPEQ and the JOACMEQ to various foreign countries, although validation of the English versions has not been done and will be necessary.

**Q. There are five functional scores. Should the statistical analyses be performed separately on each score?**

A. Yes. Adding the five scores to obtain a total score does not make sense for the JOABPEQ or the JOACMEQ.

**Q. Calculation of the severity scores is cumbersome.**

A. A program that can easily calculate the scores is available from the website. In addition, the development of a program in which a patient can answer the question by simply touching the answers displayed on a touch screen, with the scores being calculated automatically, is almost completed. It should soon be available for download from the Japanese Orthopaedic Association website.

**Q. Visual analogue scale scores are included. Are VAS scores related to the functional scores?**

A. VAS has nothing to do with the functional scores. They are included for supplemental use.

**Q. What happens if a patient gives no answer to a question?**

A. The functional score in the domain that contains the unanswered question cannot be calculated for this patient.

**Q. How were the coefficients in the equations to calculate the functional scores determined?**

A. A series of articles describing the details of the revision processes of the JOABPEQ and the JOACMEQ have been or will be published in the *Journal of Orthopaedic Science*.

**Q. If a patient has difficulty filling in the questionnaire because of a physical or mental reason, can others complete it instead?**

A. You can complete the questionnaire instead of the patient by explaining the questions directly to him or her on the spot. You can also assist the patient in using the touch screen. However, it is not advisable to make the inquiry on telephone.

**Q. Are the functional scores interchangeable with the conventional JOA scores?**

A. No, they are not interchangeable with the JOA scores. Whereas conventional JOA scores were made to evaluate patients' conditions from the physicians' standpoint, the JOABPEQ and the JOACMEQ are designed to make the evaluation from the patient's standpoint.

**Q. The Japanese wordings of the questions are awkward. Can these be improved?**

A. The questionnaires were validated through various investigations and statistical analyses while also examining the Japanese wording closely. Changing the wording may affect the reliability and reproducibility of the JOABPEQ and the JOACMEQ; therefore, we have no thought of changing the Japanese wording or modifying the questions.

**Q. There are no items to evaluate the lower limb symptoms in the JOABPEQ.**

A. The JOABPEQ is intended to evaluate common symptom "back pain" and is not aimed at evaluating the condition of a patient with a specific disease. For evaluating lower limb symptoms, please use other evaluation methods as needed.

**Q. It seems that there are other questions that should be included in the JOABPEQ and JOACMEQ.**

A. The questions in both JOABPEQ and JOACMEQ were selected carefully from many preexisting questionnaires using rigorous statistical analyses, thereby producing the functional scores of multidimensional evaluations of the patient's overall status. The selection and validation processes of the questions are described in detail in the listed articles.

**Q. Why are questions 1–4 ("Can you look at your feet when you go down the stairs?") and 3-1 ("Can you walk on a flat surface?") included in "upper limb function" in the JOACMEQ?**

A. The 25 questions in the BPEQ and the 24 questions in the CMEQ were selected through strict theoretical selection processes followed by rigorous statistical analysis. After several surveys, these questions were divided into five factors based on "factor analysis," which is a purely statistical methodology. We then chose a title for each factor that tends to reflect the context of the various questions. In other words, the names of the factors were assigned for the sake of convenience. As a result, the context of some of the questions does not entirely correspond to the titles under which they appear. For example, question 1–4 is included in both "upper extremity function" and "cervical spine function," and question 3-1 is included in both "upper extremity function" and "lower limb function."