

## Reliability, validity, and responsiveness of the Japanese version of the patient-rated elbow evaluation

Tadamasa Hanyu · Mikihiro Watanabe · Takashi Masatomi · Keiichiro Nishida · Teruhiko Nakagawa · Yasumasa Nishiura · Hiroyuki Ohi

Received: 31 January 2013 / Accepted: 16 April 2013 / Published online: 11 June 2013  
© The Japanese Orthopaedic Association 2013

### Abstract

**Background** The patient-rated elbow evaluation (PREE) is a joint-specific, self-administered questionnaire consisting of a pain scale (PREE-P) and a functional scale (PREE-F), the latter consisting of specific function (PREE-SF) and usual function (PREE-UF). The purpose of this study was to cross-culturally adapt the PREE into Japanese (PREE-J) and to test its reliability, validity, and responsiveness.

**Methods** A consecutive series of 74 patients with elbow disorder completed the PREE-J, the Japanese version of the disabilities of the arm, shoulder, and hand (DASH–JSSH) questionnaire, and the official Japanese version of the 36-Item Short-Form Health Survey (SF-36). Of the 74 patients, 53 were reassessed for test–retest reliability 1 or 2 weeks later. Reliability was investigated in terms of reproducibility and internal consistency. The validity of the PREE-J was examined by factor analysis, and correlation coefficients were obtained using the PREE-J, DASH–JSSH,

and SF-36. Responsiveness was examined by calculating the standardized response mean (SRM) and effect size after elbow surgery in 53 patients.

**Results** Cronbach’s  $\alpha$  coefficients for PREE-P, PREE-F, and PREE were 0.92, 0.97, and 0.97, respectively, and the corresponding intraclass correlation coefficients were 0.92, 0.93, and 0.94, respectively. Unidimensionality of PREE-P and PREE-F was confirmed by factor analysis. The coefficients of correlation between PREE-P and PREE-F or DASH–JSSH were 0.81 and 0.74, respectively; that between PREE-F and DASH–JSSH was 0.86, and those between DASH–JSSH and PREE-SF or PREE-UF were 0.85 and 0.82, respectively. Moderate correlation was observed in “physical functioning” for SF-36 and PREE-F ( $r = -0.69$ ) or PREE ( $r = -0.68$ ). The SRMs/effect sizes of PREE-P (1.31/1.32) or PREE (1.28/1.12) were more responsive than the DASH–JSSH (0.99/0.85), “bodily pain” ( $-1.15/-1.43$ ), and “physical functioning” ( $-0.70/-0.44$ ) in SF-36.

**Conclusion** The PREE-J represents a reliable, valid, and responsive instrument and has evaluation capacities equivalent to those of the original PREE.

For the Clinical Outcomes Committee of the Japanese Orthopaedic Association and the Functional Evaluation Committee of the Japan Elbow Society.

T. Hanyu (✉)  
Department of Orthopaedic Surgery, Nagaoka Red Cross Hospital, 2-297-1 Senshu, Nagaoka, Niigata 940-2085, Japan  
e-mail: ra-hanyu@nagaoka.jrc.or.jp

M. Watanabe  
Department of Orthopaedic Surgery, Nihon Koukan Hospital, Kawasaki, Japan

T. Masatomi  
Osaka Yukioka Hospital Hand Center, Osaka, Japan

K. Nishida  
Department of Orthopaedic Surgery, Graduate School of Medicine, Okayama University, Okayama, Japan

T. Nakagawa  
Department of Orthopaedic Surgery, The Fraternity Memorial Hospital, Tokyo, Japan

Y. Nishiura  
Department of Orthopaedic Surgery, University of Tsukuba, Tsukuba, Japan

H. Ohi  
Hand Surgery and Microsurgery Center, Seirei Hamamatsu General Hospital, Hamamatsu, Japan

## Introduction

Traditional methods of reporting clinical outcomes include data on complication and mortality rates, and these data should be an integral part of any outcome study. Patient self-administered questionnaires acquire data on a patient's experience of pain, functional disability, and general health status. The collection of data by patient reports removes the possibility of observer bias, as the questionnaires are self-administered by the patient at each assessment. These questionnaires do not replace traditional measures of clinical endpoints, but will be additions to data collection [1]. Patient questionnaires developed so far have focused on general health and quality of life, such as the 36-Item Short-Form Health Survey (SF-36) [2] or EuroQol (EQ-5D) [3]. Several measures for the evaluation of upper extremity function have also been developed [4–8]. Some of them are disease-specific [5], joint-specific [4] or region-specific measures [6–8]. Especially for the wrist and hand region, the most commonly used outcome measures described in the literature are the disabilities of the arm, shoulder, and hand (DASH) questionnaire [6] and the patient-rated wrist evaluation (PRWE) questionnaire [5]. The DASH score is the best instrument for evaluating patients with disorders involving multiple joints of the upper limb. In the Japanese Society for Surgery of the Hand (JSSH), the DASH questionnaire has been subjected to cross-cultural adaptation, and a Japanese version of the DASH–JSSH [9] has been developed. The PRWE questionnaire has also undergone cross-cultural adaptation, and a Japanese version of the PRWE (PRWE-J) has been developed by the Impairment Evaluation Committee of the JSSH. It has been reported that the PRWE-J has evaluation capacities equivalent to those of the original PRWE [10], and it is now used for patients with wrist and hand disorders.

MacDermid has reported its reliability and validity for patients with elbow pathology [11], and a German version of the PREE is available [12].

We, the Functional Evaluation Committee of the Japan Elbow Society, have completed cross-cultural adaptation and development of the Japanese version of the PREE (PREE-J). The purpose of the present study was to test the reliability, validity, and responsiveness of the PREE-J, with a view to making it available for use in Japan.

## Materials and methods

Our aim was to adapt the PREE culturally for Japanese patients, as the Functional Evaluation Committee of the JSSH has done already for the PRWE-J.

## Adaptation process

Similarly to other transcultural adaptation projects, we followed a cycle of forward/backward translation [13]. The English version of the PREE [11] was translated into Japanese by two translators, one with and one without a medical background, whose first language was Japanese. Then, the two forward translations were synthesized into a single one after a review and discussion by the committee. Parts of the PREE which were also found in the PRWE were translated identically, since the PRWE-J had already been published and it was important to be consistent. This Japanese version was translated back into English by two other translators whose first language was English. After we had compared those two back-translations with the original PREE, we developed a temporary version of the PREE-J. After pilot testing, the final PREE-J version was then evaluated for reliability, validity, and responsiveness.

## PREE questionnaire

The PREE questionnaire contains two subscales: a pain scale (PREE-P) and a functional scale (PREE-F). The PREE-P consists of five items that have multiple choice responses, which are scored from 0 point (no pain) to 10 points (worst possible). The pain score (0 = best to 50 = worst) is calculated as the sum of the scores for the five individual items (4 items for intensity, 1 item for frequency).

The PREE-F consists of a specific functional scale (PREE-SF) and a usual functional scale (PREE-UF). The answers are rated from 0 point (no difficulty) to 10 points (unable to perform at all). The PREE-SF has eleven specific elbow functional activities and is calculated as the total sum of all eleven items (0 = best to 110 = worst). The PREE-UF has four usual elbow functional activities and is calculated as the total sum of all four items (0 = best to 40 = worst). The overall score for PREE-F (0 = best to 50 = worst) is calculated as the sum of PREE-SF and PREE-UF divided by three. The total PREE score is the sum of PREE-P and PREE-F, greater pain and disability being indicated by a higher score (0 = best, 100 = worst).

## Patients and setting

The study was conducted on a consecutive series of 74 patients with elbow disorders seen on an out-patient or inpatient basis at six departments of orthopedic surgery in Japan (Table 1).

Informed consent was obtained from each patient to participate in this study. Minors had parental consent

**Table 1** Descriptive summary of subjects

Parameter	Entry	Test– retest	After surgery
No. of subjects: 74			
Sex(M/F): 39/35			
Age (years), mean $\pm$ SD: 46.7 $\pm$ 20.7 (13–82)			
Affected side (right/left): 55/19			
Diagnosis			
Rheumatoid arthritis	35	30	17
Osteoarthritis	8	6	6
Osteochondritis dissecans	6	6	5
Medial ligament injury	5	1	5
Posterior impingement	2	0	2
Lateral epicondylitis	5	4	5
Cubital tunnel syndrome	4	1	4
Post-trauma contracture	3	3	3
Loose bodies in elbow joint	2	1	2
Olecranon fracture	2	0	2
Synovial osteochondromatosis	1	0	1
Rupture of biceps tendon	1	1	1
Total	74	53	53

instead. They answered the PREE-J questionnaire, the DASH-JSSH questionnaire, and the official Japanese version of the SF-36 (version 2.0). The data collected from the 74 patients were used as baseline values. Among the 74 patients, 53 were readministered the PREE-J questionnaire for test–retest reliability 1 or 2 weeks later. Fifty-three patients who underwent surgery conducted by six elbow surgeons answered the PREE-J, the DASH-JSSH, and SF-36 questionnaires twice: preoperatively and 3 months after surgery.

#### Assessment of reliability, validity, and responsiveness

Reliability was investigated by assessment of reproducibility and internal consistency based on the test–retest method. The following analyses were conducted for examination of validity. Factor analysis (principal axis factoring) was conducted to examine the construct validity and unidimensionality of the PREE-P and PREE-F. Completeness of the item responses for the PREE-J was also examined.

Coefficients of correlation between the PREE-P or PREE-F and the DASH-JSSH were obtained. The following hypotheses were examined to investigate concurrent validity: (1) the PREE-P would exhibit moderate association with DASH-JSSH; (2) the PREE-F (SF and UF) would exhibit the strongest association with DASH-JSSH.

Coefficients of correlation between the PREE-P or PREE-F and the SF-36 were also obtained. The following hypotheses were examined to investigate concurrent validity: (1) the PREE-P would exhibit the strongest association with “bodily pain” (SF36-BP) among the SF-36 subscales; (2) the PREE-F (SF and UF) would exhibit the strongest association with “physical functioning” (SF36-PF) or “role-physical” (SF36-RP). Those three subscales of the SF-36 were chosen because the correlation between the DASH-JSSH and the three SF-36 subscales was more than moderate [7, 10].

The responsiveness of the PREE-J, DASH-JSSH, and SF-36 was examined by calculating the standardized response mean (SRM) (mean change/SD) [14] and effect size (mean change/SD of baseline value) [15] after elbow surgery in 53 patients.

The protocol of this study was reviewed and approved by the institutional review board of Nagaoka Red Cross Hospital prior to implementation.

#### Statistical analysis

Distribution of the PREE-J, DASH-JSSH, and SF-36, and ages of the subjects were assessed. The interval measurements (age, PREE-J except PREE-SF, DASH-JSSH, all subscales of the SF-36 except “physical functioning”, “social functioning”, and “role-emotional”) were normally distributed; the other interval measurements (PREE-SF, “physical functioning”, “social functioning”, and “role-emotional” of the SF-36) were not normally distributed. Cronbach’s  $\alpha$  was then used to assess the internal consistency of the PREE-J (P and F). The instrument test–retest reliability of the PREE-J (P and F) was assessed with the intraclass correlation coefficient (ICC). All coefficients of correlation among the PREE-J (P and F), DASH-JSSH, and SF-36 were calculated using Spearman’s correlation (a nonparametric test) because some subscales of the SF-36 were not normally distributed. Changes in measurements after elbow surgery were assessed with a parametric test (paired *t* test). All statistical analyses were conducted using the Statistical Package for the Social Sciences (SPSS) version 19.0J software package. Differences at  $P < 0.05$  were considered to be significant.

## Results

#### Completeness of item responses

No patients had difficulty completing the PREE-J questionnaire. Of the 74 patients, 3 did not answer one or two items (two failed to answer 2 items and 1 failed to answer one item). The items that they failed to respond to were as

**Table 2** Scores for PREE, DASH, and SF-36

Instrument	No.	Mean	SD	Median	Minimum	Maximum
PREE-P	73	22.4	12.7	24.0	0 <sup>a</sup>	50 <sup>b</sup>
PREE-SF	72	36.6	32.1	26.5	0 <sup>a</sup>	110 <sup>b</sup>
PREE-UF	72	15.5	12.4	13.0	0 <sup>a</sup>	40 <sup>b</sup>
PREE-F	72	17.4	14.4	13.3	0 <sup>a</sup>	50 <sup>b</sup>
PREE	71	39.8	26.0	34.7	0.3	100 <sup>b</sup>
DASH-JSSH	74	34.8	23.8	31.8	0 <sup>a</sup>	80.8
SF36-PF_N	74	35.8	19.2	39.8	-14.4	57.8
SF36-RP_N	74	36.8	15.9	39.1	2.6	55.7
SF36-BP_N	74	37.7	9.4	35.4	21.5	61.7
SF36-GH_N	74	48.3	12.8	48.5	19.1	69.8
SF36-VT_N	74	49.4	11.1	49.8	27.4	69.1
SF36-SF_N	74	49.9	11.5	50.6	11.9	57.0
SF36-RE_N	74	45.7	13.2	51.9	6.1	56.1
SF36-MH_N	74	49.7	9.9	51.8	22.3	65.2

*PREE-P* pain scale of the Japanese version of PREE, *-SF* specific functional scale, *-UF* usual functional scale, *-F* functional scale, *DASH-JSSH* disability/symptom scale of the Japanese version of the disabilities of the arm, shoulder, and hand (DASH) questionnaire, *SF-36-PF\_N* standardized physical functioning subscale of the 36-Item Short-Form Health Survey (SF-36), *-RP\_N* standardized role-physical subscale, *-BP\_N* standardized bodily pain subscale, *-GH\_N* standardized general health subscale, *-VT\_N* standardized vitality subscale, *-SF\_N* standardized social functioning subscale, *-RE\_N* standardized role-emotional subscale, *-MH\_N* standardized mental health subscale

<sup>a</sup> Maximum health status scores

<sup>b</sup> Minimum health status scores

follows: item 5 of the PREE-P (unanswered by one patient), item 7 of the PREE-SF and item 4 of the PREE-UF (unanswered by one patient), and item 10 of the PREE-SF and item 4 of the PREE-UF (unanswered by one patient).

The mean, median, standard deviation, and range of the PREE-J, DASH-JSSH, and SF-36 are shown in Table 2. Two, 7, 7, and 3 patients had a minimum disability score of zero (ceiling) on the PREE-P, PREE-SF, PREE-UF, and PREE-F, respectively. One patient each had a maximum disability score (floor) on the PREE-P, PREE-SF, PREE-UF, PREE-F, and PREE, respectively. One patient had a maximum disability score on the PREE.

**Reliability**

Internal consistency was assessed using Cronbach’s  $\alpha$  coefficient (Table 3). The  $\alpha$  coefficient for the five items in the PREE-P was 0.92 ( $n = 73$ ); when calculated for each of the five items by eliminating each of them one by one, the range was 0.88–0.93. The  $\alpha$  coefficient for the eleven items in the PREE-SF was 0.96 ( $n = 72$ ); after eliminating

**Table 3** Internal consistency of PREE-J

Instrument scale	No.	Cronbach’s $\alpha$	Cronbach’s $\alpha$ range
PREE-P	73	0.92	0.88–0.93
PREE-SF	72	0.96	0.95–0.96
PREE-UF	72	0.92	0.88–0.92
PREE-F	72	0.97	0.97–0.97
PREE	71	0.97	0.97–0.97

**Table 4** Intraclass correlation coefficient of PREE-J

Instrument scale	All cases		RA cases	
	No.	ICC	No.	ICC
PREE-P	52	0.92	29	0.90
PREE-SF	51	0.93	29	0.86
PREE-UF	51	0.92	29	0.85
PREE-F	51	0.93	29	0.86
PREE	50	0.94	28	0.90

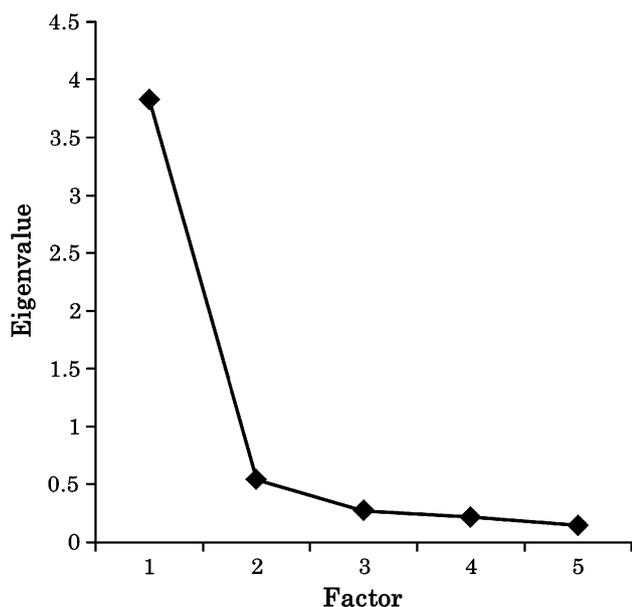
*ICC* intraclass correlation coefficient

each item one by one, the range was 0.95–0.96. The  $\alpha$  coefficient for the four items in the PREE-UF was 0.92 ( $n = 72$ ); after eliminating each item one by one, the range was 0.88–0.92. The  $\alpha$  coefficient for the fifteen items in the PREE-F was 0.97 ( $n = 72$ ); after eliminating each item one by one, the range was 0.97–0.97. The  $\alpha$  coefficient for the twenty items in the PREE was 0.97 ( $n = 71$ ); after eliminating each item one by one, the range was 0.97–0.97. Among all of the above, no items were found to change the internal consistency substantially.

Instrument test–retest reliability was assessed with the intraclass correlation (ICC) (Table 4). With regard to test–retest reliability, 52 of the 53 patients assessed had no missing items, and the mean period between the first and second tests was  $12 \pm 3$  days (range 5–19 days). The ICC for the PREE-P, PREE-SF, PREE-UF, PREE-F, and PREE in all cases ( $n = 53$ ) was 0.92, 0.93, 0.92, 0.93, and 0.94, respectively. The ICC for the PREE-P, PREE-SF, PREE-UF, PREE-F, and PREE in RA patients ( $n = 30$ ) was 0.90, 0.86, 0.85, 0.86, and 0.90, respectively. All ICCs for the PREE-J subscales and total scale indicated sufficient reproducibility.

**Validity**

Factor analysis (principal axis factoring) was conducted to confirm the unidimensionality of the PREE-P and PREE-F. The first factor of the PREE-P had an eigenvalue (amount of variation in the total sample attributable to that factor) of 3.67, which accounted for 71 % of the total variance of the PREE-P scores (Fig. 1). The unidimensionality of the



**Fig. 1** Solid plot of the patient-rated elbow evaluation, pain scale (PREE-P) factors

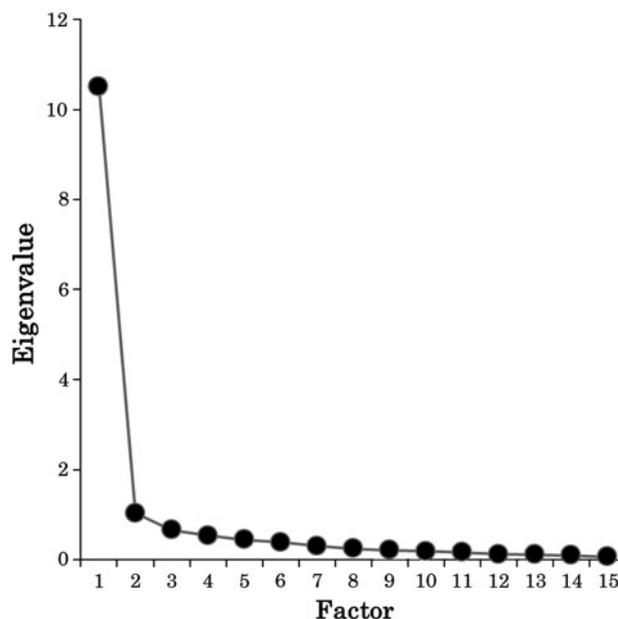
**Table 5** Component matrix of factor analysis for PREE-P

Item	Component
Pain-1	0.79
Pain-2	0.71
Pain-3	0.90
Pain-4	0.94
Pain-5	0.87

PREE-P was found to be strong as a result of the low eigenvalue of the second factor (0.54) (Fig. 1). Assessment of the first factor loading for each item showed that all items had a loading of 0.4 or higher (Table 5).

The first factor of the PREE-F had an eigenvalue of 10.50, which accounted for 70 % of the total variance of the PREE-F scores (Fig. 2). The second factor of the PREE-F had an eigenvalue of 1.03, which accounted for 7 % of the total variance of the PREE-F scores and 77 % of the cumulative of PREE-F scores (Fig. 2). The third factor of the PREE-F had an eigenvalue of 0.65, which accounted for 4 % of the total variance of the PREE-F scores. Assessment of the first factor loading for each item showed that all items had a loading of 0.4 or higher (Table 6).

Factor analysis indicated two factors in the PREE-F, suggesting a bidimensional structure. When looking at the second factor loading for each item, all items pertaining to usual function had plus values. Although many items of specific function had minus values, three items related to specific function had high plus values (Table 6). The bidimensionality of the PREE-F was not separated clearly into PREE-SF and PREE-UF.



**Fig. 2** Solid plot of the PREE-F (where  $F$  represents the functional scale) factors

**Table 6** Component matrix of factor analysis for PREE-F

Item	Component	
	1	2
SF-1	0.89	-0.32
SF-2	0.84	-0.38
SF-3	0.87	0.29
SF-4	0.83	-0.11
SF-5	0.89	0.24
SF-6	0.72	-0.12
SF-7	0.83	0.08
SF-8	0.84	-0.28
SF-9	0.80	-0.18
SF-10	0.78	0.08
SF-11	0.78	0.23
UF-1	0.90	0.17
UF-2	0.84	0.23
UF-3	0.72	0.21
UF-4	0.79	0.30

*SF* specific functional scale of PREE-J, *UF* usual functional scale of PREE-J

The coefficients of correlation between the PREE-P and PREE-F or DASH-JSSH were 0.81 and 0.74, respectively (Table 7) ( $P < 0.01$ ). These results indicated moderate correlations between the PREE-P and the PREE-F and between the PREE-P and the DASH-JSSH. The coefficient of correlation between the PREE-F and DASH-JSSH was 0.86 (Table 7) ( $P < 0.01$ ), which indicated a strong

**Table 7** Correlation of PREE, DASH, and SF-36

Instrument scale	No.	Spearman's correlation					
		PREE-P	PREE-SF	PREE-UF	PREE-F	PREE	DASH
PREE-P	73						
PREE-SF	71	0.80**					
PREE-UF	71	0.73**	0.85**				
PREE-F	71	0.81**	0.98**	0.92**			
PREE	71	0.94**	0.94**	0.87**	0.95**		
DASH	73	0.74**	0.85**	0.82**	0.86**	0.84**	
SF36-PF_N	73	-0.59**	-0.69**	-0.66**	-0.69**	-0.68**	-0.77**
SF36-RP_N	73	-0.46**	-0.48**	-0.56**	-0.53**	-0.52**	-0.54**
SF36-BP_N	73	-0.65**	-0.58**	-0.60**	-0.62**	-0.66**	-0.58**
SF36-GH_N	73	-0.47**	-0.55**	-0.48**	-0.55**	-0.51**	-0.52**
SF36-VT_N	73	-0.48**	-0.48**	-0.44**	-0.48**	-0.50**	-0.53**
SF36-SF_N	73	-0.38**	-0.43**	-0.27*	-0.39**	-0.39**	-0.40**
SF36-RE_N	73	-0.63**	-0.62**	-0.52**	-0.61**	-0.64**	-0.61**
SF36-MH_N	73	-0.52**	-0.47**	-0.41**	-0.48**	-0.49**	-0.54**

\*  $P < 0.05$ , \*\*  $P < 0.01$

correlation between them. The coefficients of correlation between the DASH-JSSH and PREE-SF or PREE-UF were 0.85 and 0.82, respectively (Table 7) ( $P < 0.01$ ). These results demonstrated strong correlations between the PREE-SF and DASH-JSSH as well as between the PREE-UF and DASH-JSSH, thus supporting the proposed hypotheses (Table 7).

The correlations between the PREE-P score and the subscales of the SF-36 scale ranged from -0.38 to -0.65 (Table 7). A moderate correlation was observed between the SF36-BP and PREE-P ( $r = -0.65$ ). These results supported the proposed hypotheses. The correlations between the PREE-P and SF36-PF or SF36-RP were somewhat weak, thus the proposed hypotheses were not supported.

The correlations between the PREE-SF score and the subscales of the SF-36 scale ranged from -0.43 to -0.69 (Table 7). A moderate correlation was observed between the SF36-PF and PREE-SF ( $r = -0.69$ ), PREE-F ( $r = -0.69$ ), or PREE ( $r = -0.68$ ). These results supported the proposed hypotheses. The correlations between the PREE-SF and SF36-BP or SF36-RP were somewhat weak, and thus the hypotheses were not supported.

The correlations between the PREE-UF score and the subscales of the SF-36 scale ranged from -0.27 to -0.66 (Table 7). A moderate correlation was observed between the PREE-UF and SF36-PF ( $r = -0.66$ ) or SF36-BP ( $r = -0.60$ ), and thus supported the proposed hypotheses.

Multiple regression analysis of PREE adjusted by age and sex was conducted because a significant difference ( $P < 0.001$ ) in mean age ( $\pm$ SD) was found between men ( $36 \pm 19$  years) and women ( $59 \pm 15$  years). The PREE

**Table 8** Multiple regression of PREE

Parameter	B	SE B	$\beta$	P	95 % CI of B	
					Lower	Upper
Constant	43.20	22.17		0.06	-1.56	88.00
Age	0.18	0.14	0.14	0.22	0.41	1.02
Sex	-8.80	6.26	-0.16	0.17	-21.44	3.84
DASH	0.71	0.15	0.65	0	0.41	1.02
SF36_PF_N	0.33	0.24	0.22	0.19	-0.16	0.82
SF36_RP_N	0.08	0.16	0.05	0.61	-0.24	0.41
SF36_BP_N	-0.73	0.30	-0.25	0.02	-1.34	-0.12
SF36_GH_N	-0.19	0.26	-0.09	0.46	-0.71	0.33
SF36_VT_N	0.63	0.37	0.27	0.09	-0.10	1.35
SF36_SF_N	-0.04	0.24	-0.02	0.86	-0.54	0.46
SF36_RE_N	-0.43	0.26	-0.21	0.11	-0.91	0.11
SF36_MH_N	-0.34	0.42	-0.13	0.42	-1.18	0.50

$R^2 = 0.80$  ( $n = 53$ )

B unstandardized coefficient, SE B standard error of B

scores were explained by the DASH scores and SF36-BP (Table 8).

**Responsiveness**

Fifty-three patients who underwent elbow surgery completed the PREE, DASH-JSSH, and SF-36 at 3 months (mean  $\pm$  SD:  $98 \pm 23$  days) after surgery. The mean age of those patients was  $42 \pm 20$  years (range 13–77 years). There were 32 men and 21 women. The calculated SRMs and effect sizes of PREE-P, PREE-SF, PREE-UF, PREE-F, PREE, DASH-JSSH, and SF36-BP ( $n = 53$ ) were 1.31/

**Table 9** Standardized response means and effect size of PREE, DASH, and SF-36

Instrument scale	Total			Rheumatoid arthritis		
	No.	SRM	Effect size	No.	SRM	Effect size
PREE-P	53***	1.31	1.32	17***	2.31	2.06
PREE-SF	53***	0.94	0.81	17***	1.34	1.32
PREE-UF	53***	1.03	0.91	17***	1.41	1.29
PREE-F	53***	1.02	0.86	17***	1.44	1.35
PREE	53***	1.28	1.12	17***	2.21	1.75
DASH	53***	0.99	0.85	17***	1.12	0.94
SF36-PF_N	53***	-0.70	-0.44	17*	-0.56	-0.46
SF36-RP_N	53***	-0.68	-0.71	17	-0.38	-0.41
SF36-BP_N	53***	-1.15	-1.43	17***	-1.12	-1.29
SF36-GH_N	53***	-0.49	-0.32	17	-0.51	-0.32
SF36-VT_N	53*	-0.35	-0.37	17	-0.06	-0.07
SF36-SF_N	53**	-0.39	-0.37	17	-0.45	-0.49
SF36-RE_N	53*	-0.36	-0.31	17	-0.35	-0.40
SF36-MH_N	53	-0.25	-0.27	17	-0.12	-0.17

SRM standardized response means

\* Significant difference between the preoperative and postoperative mean value ( $P < 0.05$ )

\*\* Significant difference between the preoperative and postoperative mean value ( $P < 0.01$ )

\*\*\* Significant difference between the preoperative and postoperative mean value ( $P < 0.001$ )

1.32, 0.94/0.81, 1.03/0.91, 1.02/0.86, 1.28/1.12, 0.99/0.85, and -1.15/-1.43, respectively (Table 9). There were significant differences between the mean values of preoperative and postoperative PREE-P, PREE-SF, PREE-UF, PREE-F, PREE, DASH-JSSH, and all subscales of the SF-36 except for the “mental health” subscale ( $n = 53$ ).

Among those patients who underwent surgery, 17 of 53 who had rheumatoid arthritis underwent total elbow arthroplasty. The calculated SRMs and effect sizes of PREE-P, PREE-SF, PREE-UF, PREE-F, PREE, DASH-JSSH, and SF36-BP were 2.31/2.06, 1.34/1.32, 1.41/1.29, 1.44/1.35, 2.21/1.75, 1.12/0.94, and -1.12/-1.29, respectively (Table 9). There were significant differences between the mean values of preoperative and postoperative PREE-P, PREE-SF, PREE-UF, PREE-F, PREE, DASH-JSSH, SF36-PF, and SF36-BP ( $n = 17$ ).

## Discussion

We had previously developed a Japanese self-administered questionnaire based on an English version of the Western Ontario and McMaster Universities osteoarthritis index (WOMAC) to measure subjective function and pain status of patients undergoing total knee arthroplasty [16]. With

the cross-cultural translation and adaptation of the original PREE, we produced a joint-specific instrument for subjective assessment of elbow pain and function in Japanese-speaking patients. The purpose of this study was to examine the psychometric qualities of the PREE-J by assessing its psychometric standards in the areas of reliability, validity, and responsiveness.

Internal consistency was assessed using Cronbach's  $\alpha$  coefficient (Table 3). The  $\alpha$  coefficients for pain, function, and total scores in PREE-J (0.92/0.97/0.97) were equivalent to those of the original version (data not shown) and the German version (-/-/0.96) [12].

Instrument test-retest reliability was assessed with the ICC (Table 4). The ICCs of the pain, function, and total scores in PREE-J (0.92/0.93/0.94) were equivalent to those of the original version (0.88/0.89/0.95) [11], and the German version (0.73/0.82/0.80) [12]. This indicated that the ICCs for the PREE-J subscales and total scale had sufficient reproducibility. Even in patients with RA who showed a diurnal variation and day-to-day variation, ICCs had good scores (0.90/0.86/0.90) (Table 4).

The process of validation of the PREE-J questionnaire has shown that it has validity similar to those of the German version and the original PREE. The strong correlations between the PREE-J and DASH-JSSH supported this validity (Table 7). Although the correlations between the PREE-J and SF-36 were weak, bodily pain and physical functioning assessed by SF-36 showed moderate correlation with the German version as well as the original PREE. DASH-JSSH had higher correlations with physical functioning assessed by SF-36 than that assessed by PREE-J. This is thought to be because the DASH covers a broader region of the whole body than the PREE. These results demonstrated that the PREE-J measures only one area of health-related quality of life.

The pain scales of PREE-J exhibited high unidimensionality (Table 5; Fig. 1), and there was no low item-scale correlation. The loading of this scale was very high. These results indicated that the pain scale of PREE-J has a high validation quality.

The functional scale of PREE-J exhibited bidimensionality (Table 6; Fig. 2), and there was no low item-scale correlation. However, the two factors could not be clearly separated into specific function and usual function in the PREE-J. Imaeda et al. [10] reported that the two factors in the PRWE-J were clearly separable into specific function (6 items) and usual function (4 items). Therefore, we would like to consider changing some of the items related to special function in the PREE-J.

Cohen's rule-of-thumb for interpreting the “effect size index” (a value of 0.2 is small, 0.5 moderate, and  $\geq 0.8$  large) can be applied to the SRM [14]. Imaeda et al. [10] reported that the SRMs/effect sizes of PRWE-P, PRWE-F

or PRWE were excellent: 1.7/2.2, 1.2/1.3, and 1.6/1.9, respectively. In the present study, the responsiveness (SRM/ES) of the pain scale and functional scale of the PREE-J for the patients overall was excellent 3 months after surgery, and larger than that of the DASH-JSSH.

The Kinemax Outcome Group reported that measurement of the SRM showed the Knee Society knee score to be more responsive (SRM 2.2) than the WOMAC (SRM 2.0 for pain and 1.4 for function) and the SF-36 (SRM 1.0 for bodily pain and 1.1 for physical functioning). The Knee Society function score was the least responsive measure (SRM 0.8). The WOMAC and SF-36 have high internal consistency and are more responsive measures of the outcomes of total knee arthroplasty [17]. In this study, the responsiveness of the pain and functional scale of the PREE-J for patients undergoing total elbow arthroplasty were equivalent to the results 3 months after surgery.

## Conclusions

We conclude that the Japanese version of PREE (PREE-J) has evaluation capacities comparable to those of the original PREE. We expect that use of this scale in Japan for self-assessment by patients receiving treatment will make a meaningful contribution to improving the outcomes of patients with elbow problems.

**Conflict of interest** None of the authors received any type of support, benefits, or funding from any commercial party related directly to the subject of this article.

## References

- Lingard EA, Hashimoto H, Sledge CB. Development of outcome research for total joint arthroplasty. *J Orthop Sci.* 2000;5:175–7.
- Fukuhara S, Bito S, Green J, Hirota A, Kurokawa K. Translation, adaptation, and validation of the SF-36 health survey for use in Japan. *J Clin Epidemiol.* 1998;51:1037–44.
- Group EuroQol. EuroQol — a new facility for the measurement of health-related quality of life. *Health Policy.* 1990;16:199–208.
- Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Care Res.* 1991;4:143–9.
- Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, Katz JN, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am.* 1993;75:1585–92.
- Hudak PL, Amadio PC, Bombardier C. Upper Extremity Collaborative Group (UECO). Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder, and hand). *Am J Ind Med.* 1996;29:602–8.
- MacDermid JC. Development of a scale for patient rating of wrist pain and disability. *J Hand Ther.* 1996;9:178–83.
- Chung KC, Pillsbury MS, Walter MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg.* 1998;23:575–87.
- Imaeda T, Toh S, Nakao Y, Nashida J, Hirata H, Ijichi M, Kohri C, Nagano A, Impairment Evaluation Committee, Japanese Society for Surgery of the Hand. Validation of the Japanese Society for Surgery of the Hand Version of the Disability of the Arm, Shoulder, and Hand (DASH-JSSH) questionnaire. *J Orthop Sci.* 2005;10:353–9.
- Imaeda T, Uchiyama S, Wada T, Okinaga S, Sawaizumi T, Omokawa S, Momose T, Moritomo H, Gotani H, Abe Y, Nishida J, Kanaya F, Clinical Outcomes Committee of the Japanese Orthopaedic Association and the Functional Evaluation Committee of the Japanese Society for Surgery of the Hand. Reliability, validity, responsiveness of the Japanese version of the patient-rated wrist evaluation. *J Orthop Sci.* 2010;15:509–17.
- MacDermid JC. Outcome evaluation in patients with elbow pathology: issues in instrument development and evaluation. *J Hand Ther.* 2001;14:105–14.
- John M, Angst F, Pap G, Junge A, Mannion AF. Cross-cultural adaptation, reliability and validity of the patient rated elbow evaluation (PREE) for German-speaking patients. *Clin Exp Rheumatol.* 2007;25:195–205.
- Guillemin F. Cross-cultural adaptation and validation of health status measures. *Scand J Rheumatol.* 1995;24:61–3.
- Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care.* 1990;28:632–42.
- Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care.* 1989;27:S178–89.
- Hashimoto H, Hanyu T, Sledge CB, Lingard EA. Validation of a Japanese patient-derived outcome scale for assessing total knee arthroplasty: comparison with WOMAC osteoarthritis index. *J Orthop Sci.* 2003;8:288–93.
- Lingard EA, Katz JN, John Wright R, Wright EA EA, Sledge CB, the Kinemax Outcomes Group. Validity and responsiveness of the knee society clinical rating system in comparison with the SF-36 and WOMAC. *J Bone Joint Surg Am.* 2001;83:1856–64.