Cross-cultural adaptation, reliability and validity of the Patient Rated Elbow Evaluation (PREE) for German-speaking patients

M. John[1,2], F. Angst[2,3], G. Pap[1,2], A. Junge[2], A.F. Mannion[2]

[1]Department of Orthopaedics, Otto-von-Guericke University, Magdeburg, Germany; [2]Schulthess Klinik, Zurich, Switzerland; [3]Rehaclinic Zurzach, Switzerland.

Abstract

Objective

Patient-orientated questionnaires are important instruments for the assessment of outcome in the clinical environment and in musculoskeletal research. The objective of this study was to cross-culturally adapt the Patient Rated Elbow Evaluation (PREE) into German (PREE-G) and to test its reliability, validity and psychometric properties.

Methods

The PREE was cross-culturally adapted for the German language, according to established guidelines. Fifty-six patients who had undergone elbow arthroplasty for osteoarthritis or chronic polyarthritis, on average 11 years previously, were assessed using the PREE-G, the Short Form 36 (SF-36), the Disabilities of Arm, Shoulder, Hand (DASH) and the modified American Shoulder and Elbow Surgeons (mASES) clinical evaluation.

Results

The test-retest reliability (intraclass correlation coefficient) of the PREE-G was 0.80, and the internal consistency 0.96. The PREE-G correlated with the DASH ($r = 0.73$) and the physical component summary of the SF-36 ($r = 0.57$) but not with the mental component summary ($r = -0.02$). The PREE-G correlated moderately with certain clinical findings (mASES) ($r = 0.36$-0.54; $p < 0.01$).

Conclusion

The PREE-G represents a reliable and valid instrument to evaluate subjective outcome in German speaking patients with elbow pathology.

Key words

PREE, elbow, questionnaire, self-assessment, DASH, SF-36, mASES, cross-cultural adaptation, validity, reliability.

Michael John, MD; Felix Angst, MD, MPH; Géza Pap, MD; Astrid Junge, PhD; Anne F. Mannion, PhD.

This study was supported by the Georg and Bertha Schwyzer-Winiker Foundation, Zürich, Switzerland.

Please address correspondence and reprint requests to: Michael John, MD, Weberstraße 1, 39112 Magdeburg, Germany.
E-mail: michajohn100@web.de

Received on May 15, 2006; accepted in revised form on September 28, 2006.

Introduction

The assessment of outcome after medical treatment is of major importance in determining the quality and effectiveness of the intervention administered (1). With regards to the musculoskeletal system, the patient's self-rated pain reduction and/or range of motion (ROM) or other 'clinical' surrogates are often the only measures that are used to supplement the standard functional or structural indices of success (2). For therapeutic, health-economic, and ethical reasons, the patient's own rating of treatment outcome is becoming increasingly important, especially within the framework of evidence-based medicine (3, 4). Patient-orientated monitoring of treatment effects usually encompasses documentation of the patient's health-related quality of life and ability to carry out activities of daily living (ADL), each of which is important in relation to the patient's reintegration into the social and occupational environment (2).

With the help of appropriate measurement instruments, the patient's subjective functional capacity or disability can also be evaluated at the organ-specific or symptom-specific level (2). Hence, in addition to the generic instruments used for measuring quality of life (e.g. the Short-Form 36 (SF-36) (5), various joint/region-specific questionnaires have also been introduced. Compared with other regions of the body, a relatively large number of self-rating questionnaires are available for the upper extremities. The Patient Rated Elbow Evaluation [PREE (6)], the Patient-Rated Forearm Evaluation Questionnaire [PRFEQ (7)], the modified American Shoulder and Elbow Surgeons Score (mASES (8)) and the Mayo Elbow Performance Index [MEPI (9, 10, 11)] are all available in the English language only; the Disabilities of the Arm, Shoulder and Hand (DASH) has been validated in both German and English (12, 13).

A quality management programme, initiated in our orthopaedic hospital in 2001, revealed the need for a standardised questionnaire for the self-assessment of elbow function in German-speaking patients. The DASH questionnaire was deemed to be too unspecific, as it considers the whole upper extremity as an independent functional unit, assessing the symptoms and functional capacity of the whole arm including the shoulder and hand (14, 15). The PRFEQ has only been validated in relation to the specific diagnosis of tennis elbow (7). The MEPI is a widely used instrument for the global assessment of the elbow joint, but it incorporates both subjective and objective aspects of function; its use as a purely patient-orientated self-assessment questionnaire is thus limited. In some studies the MEPI has been used as a self-rating questionnaire (11, 16, 17) and in other studies, as a clinical evaluation tool (9, 11, 18). All items of the MEPI are covered in the DASH and/or the PREE. The numerical rating scale, and the more detailed questioning in relation to the same ADLs enquired about in the MEPI, confer certain advantages upon the PREE and the mASES (6, 18). A prospective, comparative study of scoring systems found that the DASH and the mASES performed a better assessment of pain and function than the MEPI (18).

For these reasons, the PREE and the patient-orientated part of the mASES were considered to be the questionnaires of choice for cross-cultural adaptation into German (the results for the mASES will be discussed in a separate article). The consistent use of the same, established instruments is becoming increasingly important for the successful execution of multi-centre, cross-national studies, and for allowing comparison/meta-analyses of the results of different studies carried out in different countries; as such, it was not considered prudent to develop an entirely new questionnaire.

The aim of the present study was to carry out a cross-cultural adaptation of the English version of the patient-orientated questionnaire, PREE, according to currently accepted guidelines, for use with German-speaking patients. The psychometric properties of the adapted instrument (reliability and validity) were investigated in a population of patients from the German-speaking part of Switzerland.

Materials and methods

The Patient Rated Elbow Evaluation (PREE) questionnaire

The PREE is a short, patient-orientated, self-assessment questionnaire that was developed from the Patient Rated Wrist Evaluation (PRWE) (19). On average, 3 minutes are required to complete all 20 items of the PREE. A visual analog scale (VAS) from 0 = best rating to 10 = worst rating for each item enquires about elbow-related pain (4 items for intensity, 1 item for frequency) and functional limitations of the elbow in relation to ADL (11 items for specific activities and 4 items for usual activities). The unweighted means of the items are used to determine a pain score (0 = best to 50 = worst) and a function score (0 = best to 150 = worst; division of this score by three rescales it to a range from 0 = best to 50 = worst). The total score weights pain and disability equally by simple addition, with greater pain and disability being indicated by a higher score (0 = best, 100 = worst). The validity and reliability of the English version of the PREE was reported by MacDermid (6), although no corresponding information was given regarding the total number of 'missing' answers allowed. In the present study, an individual's scores were included in the group cross-sectional analyses as long as 65% of the individual questions for each scale or sub-scale had been completed [i.e. a minimum of 3/5 items on the pain scale and 10/15 items on the function scale (3)]. However, for the test-retest reliability analysis, only questionnaires in which a minimum of 85% items had been answered (4/5 pain, 13/15 function) were analysed.

Translation and cross-cultural adaptation

The translation and cross-cultural adaptation of the original English questionnaire into German was carried out in accordance with published recommendations and guidelines (20, 21). This process consists of several steps, each of which is documented with a written report.

a) Forward-translation and synthesis

Two native German speakers (T1; T2), with different educational and job profiles, carried out independent translations of the questionnaire from English to German. T1 (informed translator) was familiar with the concepts being examined and the clinical content of the questionnaires. T2 (uninformed translator) was a non-clinician, who was familiar with the general construction of questionnaires but not with the concepts being examined in the PREE [the "naive translator" (21)].

The two translations were compared with one another and with the original English version and then, after reaching consensus in relation to any discrepancies, synthesised to form one common translation, T-12.

b) Back-translation

Two native English speakers (one American and one Briton) with German as their second-language (BT1; BT2) carried out independent back-translations of the version (T-12) into English. Both translators fulfilled the requirements of bilingualism (22) and had no prior knowledge of the instrument. A third bilingual person (native English, German as second language) compared the two back-translations with each other and with the original-questionnaire and highlighted any conceptual errors or gross inconsistencies in the content of the translated versions, in preparation for the expert committee meeting.

c) Expert Committee

An expert committee was formed consisting of the forward and back-translators, 1 orthopaedic surgeon, 1 psychologist/methodologist, 1 rheumatologist and 1 bilingual clinical research scientist. The group examined the translations, the back-translations, and the notes made in carrying out/comparing the translations, and consolidated these to produce a "pre-final" version of the German PREE-G. The task of this expert committee was to assure semantic and idiomatic equivalence and experiential and conceptual equivalence between the German and the original English version of the questionnaire. For all parts of the questionnaire (instructions, items and response options) consensus was always found between the members of the committee.

d) Test of the pre-final version

A heterogeneous group of 30 people (patients, visitors and employees of the hospital) were given the pre-final version of the PREE-G questionnaire to complete. They were probed regarding their understanding of each question and the chosen response and about the questionnaire in general (layout, wording, ambiguities, ease of understanding, etc.). All the findings were evaluated by the work-group before the final German version of the PREE was produced and subject to further psychometric testing.

Psychometric properties of the German version of the PREE (PREE-G)

Patient group

Between March and October 2003, all patients from our orthopaedic hospital who had received an elbow prosthesis (GSB, Sulzer Medica, Baar, Switzerland), on average 11 years previously (between 1984 and 1996), were identified.

From 192 potential study participants, 71 (37%) patients had died, 11 (6%) could not be traced, 31 (16%) declined participation in the study and 11 (14%) were not able to visit the clinic due to severe illness and/or the long distance between home and the clinic. Hence, 68 (86%) patients were available for examination.

A clinical assessment, including x-ray and ultrasound examinations, preceded the completion of a set of various self-assessment questionnaires. The latter included the SF-36, the DASH and the new German version of the PREE (PREE-G). The findings from the clinical investigation of the elbow were used to complete the clinical mASES (modified American Shoulder and Elbow Score). Joint range of motion for the mASES was assessed with a goniometer, and grip strength was measured using a hand-grip dynamometer (JAMAR®).

Test-retest reliability of the PREE-G

In order to assess the test-retest reliability of the PREE-G, all patients were

required to complete the questionnaire again, 3 to 4 days after the initial assessment (questionnaire sent and returned by post).

Statistical analysis

The test-retest reliability of the PREE-G (pain, function, total score) was determined by calculation of the intraclass correlation coefficient (ICC) (23). Cronbach's alpha was used to determine the internal consistency of the questionnaire (24, 25).

The distribution of individual responses within the sub-scales/total score and the presence of any floor or ceiling effects were examined with stem and leaf plots. These allow visualization of the discriminative power of the questionnaire in relation to individual limitations/pain levels.

The criterion/construct (convergent) validity of the PREE-G was determined by analysing the relationship between its scores and the scores from: the Physical Component Summary (PCS) of the generic SF-36; all the physical sub-scores of the PCS and the DASH (an upper extremity specific questionnaire).

The construct validity of the PREE-G was further investigated by examining the relationship between the PREE-G scores and those derived from the clinical joint assessment (mASES).

All the statistical analyses were carried out using SPSS 11.0 for Windows® (SPSS Inc., Chicago, IL, USA). Statistical significance was accepted at the 5% level, but the precise p values for each analysis are given in the corresponding text/tables.

Results

Cross-cultural adaptation of the PREE

Overall, the translation and back-translation of the PREE questionnaire proceeded without any major difficulties. In the back-translation of the first German version, there were slight discrepancies with the original English version for 4 questions. The first question in the function sub-section – "comb my hair" – was back-translated as "combing your hair", which was a direct translation of the German expression that had been used. Nonetheless, in accordance with the usual German formulation of such phrases, the committee agreed to continue to use the phrase „Ihre Haare kämmen" ("combing your hair"). The function question "use my arm to rise from a chair" caused problems during translation into German, on account of the two possible ways of translating the word chair ("Stuhl" or "Sessel"). In our first German version of the PREE, the word "Sessel" was used for "chair", but this was then back-translated into English as "armchair". Although most chairs do have arm-rests, it was decided that the word "armchair" referred to a specific type of "easychair". Thus, "chair" was translated with the word "Stuhl", which refers to the general kind of chairs that one finds in waiting rooms, offices, cafes, restaurants, etc; the arm would then be used to assist with standing-up by leaning on some external object (e.g. the table top, edge of the chair, crutches, walking stick). Question 18, "amount of difficulty with household work (cleaning, maintenance)", was at first translated as "Haushaltarbeit (Saubermachen, Aufräumen)" which resulted in the back-translation "housework (cleaning, tidying)". However, "tidying" was not considered to have the same meaning as "maintenance" (and it was also feared that with "cleaning, tidying" a number of men may have ticked this question "not applicable"!) and so an additional example "doing small repairs" was added to the German version; no better translation of "maintenance" could be found. Further, for "cleaning", the word "Saubermachen" was replaced with "Putzen", which is the more commonly used term in Switzerland. This question was thus changed to "Haushaltarbeit (Putzen, Aufräumen, kleine Reparaturen)". In the original English version of the PREE, question 5 in the 'function' section was formulated with imperial units of weight ("carry a 10 lb object with my arm at my side"); this was converted to the nearest whole kilogram in forming the German version (= 5 kg). It was considered that the slightly greater weight of 5 kg compared with 10 lb (approx 0.5 kg) could be neglected.

Pre-testing of the PREE-G

The testing of the pre-final version showed that the questionnaire was by and large understandable in its content and easy to complete. Two patients considered the function question "use a telephone" to be too inexact, as various differing actions were being enquired about (lifting up the receiver, dialling the number (with either touch-buttons or a rotating dial), holding the phone at the ear for a certain length of time). A further person wrote alongside two of the pain questions ("pain when lifting a heavy object" and "pain when doing a task with repeated elbow movement") that they would have no pain because they consciously avoided doing these activities. Neither of these issues concerned the cross-cultural adaptation of the questionnaire and thus did not result in any further changes to our German version.

Completion rate for the PREE-G

The data from 56 patients with German as their mother tongue (51 Swiss German, 5 German) could be evaluated. 5 patients did not fill out the PREE-G questionnaire sufficiently completely (more than 7 missing answers, which exceeded the 65% completeness criterion) and 7 patients were of a non-German speaking nationality (1 Rumanian, 2 Spanish, 4 Italian). Thus, these twelve questionnaires were excluded. In the final study group (N=56) there were 37 women and 19 men and their mean age was 63.7 (SD=11.4) years. The indications for elbow arthroplasty were primary or posttraumatic osteoarthritis in 17 and joint destruction secondary to chronic polyarthritis in 39 cases.

The test-retest reliability analysis of the PREE-G was carried out on the data from 46 patients (30 women, 16 men) who returned the second questionnaire within 14 days; 10 patients failed to return a second questionnaire or the questionnaire was incompletely filled out (more than 3 missing answers, which exceeded the 85 % completeness criterion).

Analysis of the distribution of individual questions answered in the 56 questionnaires revealed no tendency for any particular question to be consistently missed out: of the 20 individual questions in the PREE-G, 12 were answered

Table I. Intraclass correlation coefficients (ICC) for the test-retest reliability of the PREE-G.

Pain scale		Function scale			
Question:	ICC	Question:	ICC	Question:	ICC
1	0.76	6	0.79	14	0.76
2	0.67	7	0.68	15	0.83
3	0.56	8	0.69	16	0.74
4	0.68	9	0.48	17	0.71
5	0.56	10	0.78	18	0.66
		11	0.67	19	0.61
		12	0.64	20	0.65
		13	0.72		
Pain sum scale:	0.73			Function sum scale:	0.82
(95 % - C.I.)		n = 46		Total score: 0.80	

by every single patient; 7 were not answered by up to 2 patients; and 1 question was not answered by 3 patients.

Test-retest reliability of the PREE-G
The test-retest reliability was calculated for all the individual questions, for the separate pain and function sub-scales, and for the PREE-G total score. The intraclass correlation coefficient (ICC) was between 0.61 and 0.83 for 17 of the individual questions, for the two sub-scales, and for the total score. The highest reliability was shown for the function questions 6, 10 and 15 (ICC = 0.78 - 0.83). The test-retest agreement for the answers to questions 3, 5 and 9 was low to moderate, with ICCs ranging from 0.48 to 0.56 (see Table I). The correlation coefficients for pairs of questions within a given scale, and for individual questions with their corresponding sub-scale score or whole scale score are shown in Table II.

Internal reliability (internal consistency) of the PREE-G
The Cronbach's alpha coefficients were 0.93 for the pain sub-scale, 0.95 for the function sub-scale and 0.96 for the total PREE-G score.

Distribution of answers (floor and ceiling effects) for the PREE-G
The distribution of the individual values for the PREE-G sub-scales (pain and function) and for the total PREE-G score are shown as stem and leaf plots in Table III. In each case, the stem runs from 0 to 10, corresponding to the rating scale 0 = best, 10 = worst. The stem represents the first digit of each individual's score and the leaf the subsequent first digit after the decimal place (e.g. for the function scores 8.7 and 8.9: the score of 8.7 is positioned under the stem value of 8 with a leaf value of 7; 8.9 is under the stem of 8 with a leaf value of 9). Asymmetrical distributions were observed for each sub-scale and also for the total score, with a greater number of scores being registered at the beginning of the rating scale (i.e. the distribution was skewed to the left/positively skewed). The best possible score (i.e. no pain/no dysfunction at all; score = 0) was shown for 8 patients for the pain scale, 3 patients for the function scale and 3 patients for the total score.

Construct/concurrent validity of the PREE-G
The pain sub-scale, the function sub-scale and the total score from the PREE-G each showed significant correlations with the DASH score ($p < 0.001$; Table IV). The PREE-G pain and function sub-scales also showed significant correlations with the sum-scores of the DASH's symptom questions (questions 24-29) and function questions (question 1-21) respectively (Table IV).
There was no significant correlation between the PREE-G scores and the Mental Component Summary score of the SF-36. The PREE-G scores showed no, or only weak correlations with the scores from the following SF-36 dimensions: vitality, social functioning, role emotional, mental health and general health ($r < 0.34$, $p < 0.05$).
The correlations between the PREE-G and the physical health dimensions of the SF-36 (bodily pain, role physical and physical functioning) were moderate but higher ($r = 0.27$ - 0.66) and statistically significant ($p < 0.05$ or $p < 0.01$; Table IV). PREE-G pain correlated weakly but significantly with the Physical Component Summary of the SF-36 ($r = 0.32$, $p < 0.05$) whilst PREE-G function and PREE-G total score showed moderate significant correlations with the PCS ($r = 0.57 – 0.67$, $p < 0.001$; Table IV).
The correlations between the PREE-G and the scores from the clinical assessment (clinical mASES) are shown in Table IV. PREE-G sub-scale

Table II. Psychometric characteristics of the PREE-G questionnaire: inter-item and item-scale correlation coefficients.

n = 56	PREE - Pain (0 – 50 points)	PREE - Function (0 – 50 points)	PREE - Totalscore (0 – 100 points)
Mean value (standard deviation)	14.8 (13.2)	16.7 (12.8)	31.5 (23.8)
Pearson correlation coefficients for pairs of individual questions	0.62 – 0.88 (questions 1 – 5)	0.37 – 0.8 (questions 6 – 20) 0.27 – 0.32# (for question 12 vs. 19, 20)	
Pearson correlation coefficients for individual questions and their corresponding scales	0.81 – 0.91	0.69 – 0.84	0.79 – 0.88 (questions 1 – 5) 0.61 – 0.81 (questions 6 – 20)

#$p < 0.05$, elsewhere: $p < 0.01$.

Table III. Stem and leaf plots of the scores from the PREE-G sub-scales (pain and function) and for the total PREE scores.

PREE-G Pain	N = 56	PREE-G Function	N = 56	PREE-G Total score	N = 56
10	0	10	0	10	0
9	0	9	0	9	0
8 0	1	8 79	2	8 36	2
7 268	3	7 5	1	7	0
6 048	3	6 34599	5	6 0367	4
5 066888	6	5 1347	4	5 1577899	7
4 028	3	4 0034579	7	4 0149	4
3 000068	6	3 8999	4	3 00011668	8
2 24444468	8	2 123345566789	12	2 1234455567	10
1 02222446	8	1 11557999	8	1 023667788	9
0 000000002444666888	18	0 0001123334558	13	0 000223444788	12

scores and the PREE-G total score each showed similar values for the strength of their correlation with the clinical mASES "signs and symptoms" scores ($r = 0.48$ - 0.54, $p < 0.001$). The PREE-G total score showed a rather low but significant correlation with the dimension "elbow strength" of the mASES ($r = 0.40$, $p < 0.001$); the sub-scale PREE-G function showed a somewhat higher, significant correlation with the mASES "hand-grip strength" ($r = 0.48$, $p < 0.001$).

Discussion

With the cross-cultural translation and adaptation of the PREE we have produced a joint-specific instrument for the subjective assessment of elbow pain and function in German-speaking patients. The questionnaire is quick and easy for both the patient to complete and the investigator to evaluate. As far as we are aware, this version of the PREE (PREE-G) represents the only elbow-specific "self-assessment" instrument available in the German language (the DASH is available in German but this is not elbow-specific). The translation and adaptation of the PREE was carried out in relation to the language and cultural conditions that prevail in Switzerland and was tested on a Swiss German group of patients. The successful use of this same version in other German-speaking lands (Austria and Germany) is not automatically guaranteed, despite the cultural and linguistic similarities of these neighbouring countries. According to the guide-

Table IV. Relationship between the PREE-G scores and various other (region-specific and generic) self-assessment instruments and clinical scores. (Spearman's rank - correlation).

PREE-G	Pain		Function		Total	
	r	p	r	p	r	p
SF-36 Physical functioning	0.27	0.034	0.64	< 0.001	0.50	< 0.001
SF-36 Role physical	0.27	0.035	0.46	< 0.001	0.39	0.002
SF-36 Bodily pain	0.49	< 0.001	0.66	< 0.001	0.66	< 0.001
SF-36 General Health	0.20	0.129	0.32	0.013	0.31	0.017
SF-36 Vitality	0.31	0.016	0.29	0.026	0.32	0.013
SF-36 Social functioning	0.31	0.014	0.31	0.014	0.34	0.007
SF-36 Role emotional	0.16	0.221	0.26	0.049	0.22	0.091
SF-36 Mental health	0.32	0.014	0.16	0.224	0.26	0.044
SF-36 PCS	0.32	0.017	0.67	< 0.001	0.57	< 0.001
SF-36 MCS	0.11	0.435	-0.12	0.388	-0.02	0.890
DASH Symptoms	0.61	< 0.001	0.72	< 0.001	0.73	< 0.001
DASH Function	0.32	0.12	0.83	< 0.001	0.65	< 0.001
DASH	0.45	< 0.001	0.87	< 0.001	0.73	< 0.001
cmASES Motion	0.15	0.200	0.03	0.836	0.06	0.596
cmASES Stability	-0.03	0.798	0.13	0.334	0.08	0.541
cmASES Strength	0.36	0.002	0.38	0.001	0.40	< 0.001
cmASES Grip strength (kg)	0.04*	0.737	0.48*	< 0.001	0.29*	0.014
cmASES Signs & Symptoms	0.48	< 0.001	0.51	< 0.001	0.54	< 0.001
cmASES	0.04	0.747	0.35	0.006	0.24	0.076

Pearson correlation*
(The scores of PREE, DASH and cmASES were transformed into a scale from 0 = worse to 100 = best.)

lines of Guillemin *et al*. (example 3: same language, different country, different cultures), a further adaptation of the questionnaire may still be necessary before it can be implemented in these other countries (21). Nonetheless, as there are no grammatical or semantic differences in the use of the written language amongst the German-speaking countries/regions in Europe, we do not anticipate that any major changes to the current version would be required. The language and cultural differences between these lands/regions exist mainly in the spoken language (different dialects). Further, in putting together this German version of the PREE (PREE-G), wherever there was more than one way of translating a particular English phrase, we paid special attention to choosing words that were just as commonly used in Germany as in Switzerland (two of the people in the expert committee were of German nationality). Thus, with the only prerequisite for the use of our German PREE being an adequate understanding of written German, we believe that the present version can most likely be used without difficulty in other German-speaking European countries. However, this awaits verification by our colleagues in Germany and Austria.

The process of translating and back-translating the English PREE was carried out in strict adherence to the "AAOS Guidelines for the Cross-cultural Adaptation of Self-report Instruments" (20, 21). These guidelines form the basis for producing a reliable and valid adaptation of the questionnaire that shows a high degree of agreement with the original version. In the present study, with the successful pre-testing of the translated version, the PREE-G was also examined for its face/content validity (20, 21). Due to the relationship between the English and German languages, the translation of the questionnaire could be carried out largely with literal equivalence. Only few questions required discussion or modification by the committee as a result of significant deviations from the original highlighted by the forward and back translations. A consensus was always found within the committee regarding any questionable wording. The testing of the pre-final version of the instrument revealed only few additional suggestions or comments by the test persons, and none of these resulted in any major changes to the questionnaire.

The German version of the PREE showed overall good reliability, especially for the sum-scales and the total score (ICC, 0.73 - 0.82). The intraclass correlation coefficient (ICC) was also acceptable for 17 of the 20 individual questions (ICC = 0.61 – 0.83). In general, the reliability was better for the questions in the function scale than those in the pain scale. The test-retest reliability for the individual items in the original English version of the PREE was somewhat higher, with ICCs ranging from 0.6 to 0.88 (6). In the latter study, the items in the pain sub-scale showed a much higher reliability than in the present study. There are two possible reasons for this. Firstly, the dimension "pain" is most likely subject to greater day-to-day variability than are many aspects of function e.g. joint range of motion. This is particularly so in relation to the typical pain pattern associated with polyarthritis, which was the indication for surgery in the majority of patients (n = 39, 70%) in the present study. (McDermid did not give any specific details about the patients involved in his study other than that they "were from the waiting area of their hand clinic and had a variety of elbow pathologies".) Secondly, whilst our patients were required to return their second questionnaire within 14 days of the first, in McDermid's study the test-retest period was 2-7 days. Naturally, the longer time-interval that we used would be expected to lead to a greater variability in scores, and even more so in relation to the pain ratings, for the aforementioned reason.

The Cronbach's alpha for the total PREE-G score, which indicates the internal consistency of the questionnaire, was very high (0.96). In the clinical situation, where the value of a scale for an individual is of interest, an alpha coefficient of at least 0.9 is recommended (24). The PREE-G would therefore clearly satisfy this condition. An even higher Cronbach's alpha value is not necessarily desirable, as this would indicate that a number of items are measuring the same characteristic (12).

In the present study, the group's responses to the PREE-G were by no means normally distributed. On a scale adjusted to 0 - 10, approximately 60% of the patients rated their elbow joint function and pain between 0 and 2, i.e. the distribution was skewed to the left (positively skewed), indicating that these patients had next to no problems with their elbow joint.

The ceiling effects (i.e. score = 0) for the pain and function scales were 14.3% and 5.3% respectively. This would mean that for these few patients it would be difficult to document any further improvement in their condition using the PREE-G score (3).

In order to assess the criterion/construct validity of the PREE-G, the correlation between the PREE-G scores and those of the self-rating questionnaires DASH and the mental and physical component summary scales of the SF-36 was examined. The whole scores from the PREE-G and the DASH showed a high correlation (r = 0.73). The symptom score of the DASH (which includes 3 pain questions) correlated moderately with the PREE-G pain sub-scale (r = 0.61). The highest correlations were found between the scores from the DASH (total score and function scale score) and the scores from the function sub-scale of the PREE-G (r = 0.87 and 0.83 respectively). The correlations between the PREE-G (total, pain sub-scale, function sub-scale) and the DASH (global score) differed somewhat from those reported by MacDermid (total score, present study r = 0.73 vs MacDermid r = 0.85; pain sub-scale r = 0.45 vs r = 0.71; function sub-scale, r = 0.87 vs r = 0.78) (6). Whilst the DASH includes questions about symptoms and functional limitations of the whole arm (including the shoulder and hand) and the associated psychosocial problems that result from the joint problem, the PREE focuses purely on pain and functional limitations of the elbow. Hence, although some correlation is expected between the two questionnaires, it is not totally

surprising that the relationship is not consistently high. The difference between studies may reflect differences in the degree of problems in the other two joints (shoulder and hand) experienced by the respective patient groups examined. Indeed, the lack of a perfect correlation between the two questionnaires emphasises the need for a specific questionnaire for the elbow joint alone.

In contrast to the findings of MacDermid, in the present study the subscales and total scores from the PREE-G correlated with the physical but not with the mental component summary scores of the SF-36 (6). Further, and as expected, there was no or only a minimal correlation between the PREE-G and the various dimensions of the mental scales (vitality, social functioning, role emotional, mental health) ($r = 0.26 – 0.32$, $p < 0.05$). Psychosocial problems as a result of the joint problem, as enquired about with the mental summary scales of the SF-36, are scarcely touched upon in the PREE, which primarily focuses on symptoms and functional limitations in everyday activities. The pain questions and questions 19 and 20 (difficulties at work (occupation/everyday activities) and during everyday activities) could to a certain extent be influenced by psychosocial effects of the disorder, and this may have accounted for the slight correlations observed between the PREE-G and some of the mental component summary dimensions. However, the PREE (PREE-G) and the mental summary scales of the SF-36 basically measure different constructs and so a high correlation between the scores is not to be expected. Indeed, this is to be viewed as a positive finding in relation to the validity testing of the PREE-G. Confirming this finding, Offenbaecher *et al*. reported a very low correlation between the DASH and the psychosocial sub-scales (social functioning, role emotional, mental health) of the SF-36 ($r = -0.19$ and -0.26), when validating the German version of the DASH questionnaire (which we used in the present study) (13). The validation study for the English version of the PREE revealed similarly low correlation coefficients for the relationship between the PREE sub-scales or total score and the mental component summary of the SF-36 ($r = -0.12$ to -0.30) (6).

Moderate correlations were found between the PREE-G (for total score and function sub-scale) and the physical dimensions of the SF-36 (i.e. bodily pain, physical functioning) ($r = 0.50$ to 0.66). Similar correlations were found by MacDermid (PREE total score – PCS (SF-36): present study $r = 0.57$ vs MacDermid $r = -0.56$; PREE pain scale - PCS (SF-36): $r = 0.32$ vs $r = -0.49$; PREE function scale - PCS (SF-36): $r = 0.67$ vs $r = -0.52$) (6). The degree of correlation between the DASH and the different sub-scales of the SF-36 reported by Offenbaecher *et al*. was similar to that between PREE-G and the SF-36 sub-scales found in the present study ($r = -0.58$ to -0.79, Table IV) (13). (The scores of the PREE were not adjusted in the original publication of MacDermid, which accounts for the negative correlation coefficients.)

The relationship between the pain ratings of both instruments (PREE-G pain sub-scale versus bodily pain in the SF-36) was only moderate and not as high as expected ($r = 0.49$). This may be partly the result of the relatively low number of pure pain questions (only 2 items), directed at specific regions or joints, in the SF-36. The SF-36 represents a comprehensive, but unspecific measurement instrument for assessing health-related quality of life, in which items concerned with the functioning/limitations of the lower extremities tend to dominate (2, 5). The correlations reported between the PREE-G and the SF-36 are therefore to be considered adequate.

The construct validity of the PREE was determined by examining the relationship between the PREE-G scores and the data collected during the clinical assessment (mASES assessment). The pain sub-scale of the PREE-G correlated moderately with the mASES "clinical signs and symptoms" ($r = 0.48$), which is dominated by 12 pain items (about tenderness, and pain upon movement in both the joint and the surrounding muscles) together with items about deformity, muscle atrophy and neurological deficit. The function sub-scale of the PREE-G showed a similarly moderate correlation with the mASES "grip-strength" ($r = 0.48$) and "clinical signs and symptoms" ($r = 0.51$). The PREE-G total score correlated moderately with the mASES "clinical signs and symptoms" ($r = 0.54$) and "elbow strength" (in pronation, supination, flexion, extension) ($r = 0.4$). The discrepancy between the results of clinical assessment and those of self-assessment questionnaires is a well-known problem in musculoskeletal research and is not only specific to the elbow joint (3, 4, 18). In general, the results of the clinical examination tend to be better than those of the self-rated assessment (3). There was no re-lationship between the results of the PREE-G and the clinically-determined mobility of the elbow joint (range of motion in pronation, supination, flexion, extension). Many studies have previously shown that limitations in joint range of motion, which often arise after joint arthroplasty, have no effect on patient satisfaction (9, 11, 17, 26).

The present study did not seek to examine the sensitivity of the PREE-G to differentiate between different disorders of the elbow joint or between "healthy" and "injured" joints; these will be examined in our future studies.

The present study has served to provide a valid and reliable version of the PREE for use in the German-speaking region of Switzerland (and, with possibly slight modifications, in other German-speaking areas of Europe). In doing so, it extends the list of currently available self-assessment questionnaires available in the German language. The cross-cultural adaptation of the PREE has addressed the need for standardised measurement instruments for the assessment of quality of life and treatment effectiveness within the framework of quality management in musculoskeletal medicine.

Acknowledgments

We thank Roberta Schefer for the management of the patients, the questionnaires, and the database.

References

1. BOERS M, BROOKS P, SIMON LS, STRAND V, TUGWELL P: OMERACT: An international initiative to improve outcome measurement in rheumatology. *Clin Exp Rheumatol* 2005; 23: 10-13.
2. PINCUS T, SOKKA T: Complexities in the quantitative assessment of patients with rheumatic diseases in clinical trials and clinical care. *Clin Exp Rheumatol* 2005; 23(Suppl. 39): S1-S9.
3. ANGST F, PAP G, MANNION AF *et al*.:Comprehensive assessment of clinical out-come and quality of life after total shoulder arthroplasty. Usefulness and validity of subjective outcome measurement. *Arthitis Rheum* 2004; 51: 819-28.
4. ANGST F, STUCKI G, AESCHLIMAN A: Quality of life assessment in osteoarthritis. *Expert Rev. Pharmacoeconomics Outcome Res* 2003; 3: 623-36.
5. WARE JE, SNOW KK, KOSINSKI M, GANDEK B: SF-36 Health survey: *Manual and interpretation guide*. 2nd ed. Lincoln (RI): QualityMetric Incorporated 2000.
6. MACDERMID JC: Outcome evaluation in patients with elbow pathology: Issues in instrument development and evaluation. *J Hand Ther* 2001; 14: 105-14.
7. OVEREND TJ, WUORI-FEARN JL, KRAMER JF, MACDERMID JC: Reliability of a Patient-rated Forearm Evaluation Questionaire for patients with lateral epicondylitis. *J Hand Ther* 1999; 12: 31-7.
8. KING GJ, RICHARDS RR, ZUCKERMANN JD *et al.:* Standardized method for assessment of elbow function. Research Commitee, American Shoulder and Elbow Surgeons. *J Shoulder Elbow Surg* 1999; 8: 351-4.
9. MORREY F, ADAMS RA: Semiconstrained arthroplasty for the treatment of rheumatoid arthritis of the elbow. *J Bone Joint Surg (Am)* 1992; 74A: 479-90.
10. RODGERS WB, KHARRAZI FD, WATERS PM, KENNEDY JG, MCKEE MD, LHOWE DW: The use of osseous suture anchors in the treatment of severe, complicated elbow dislocations. *Am J Orthop* 1996; 25: 794-8.
11. COBB TK, MORREY BF: Total elbow arthroplasty as primary treatment for distal humeral fractures in elderly patients. *J Bone Joint Surg (AM)* 1997; 79A: 826-32.
12. MCCONNELL S, BEATON DE, BOMBARDIER C: The DASH outcome measure: A user's manual. Toronto, Ontario, Institute for Work & Health, 1999: 95.
13. OFFENBAECHER M, EWERT T, SANGHA O, STUCKI G: Validation of a German version of the disabilities of arm, shoulder, and hand questionnaire (DASH-G). *J Rheumatol* 2002; 29: 401-2.
14. HUDAK PL, AMADIO PC, BOMBARDIER C: Development of an upper extremity outcome measure: The DASH (Disabilities of the Arm, Shoulder, and Hand). *Am J Indust Med* 1996; 29: 602-8.
15. NAVSARIKAR A, GLADMANN DD, HUSTED JA, COOK RJ: Validity assessment of the Disabilities of Arm, Shoulder, and Hand Questionaire (DASH) for patients with psoriatic arthritis. *J Rheumatol* 1999; 26: 2191-4.
16. BAILEY CS, MACDERMID JC, PATTERSON SD, KING GJ: Outcome of plate fixation of olecranon fractures. *Orthop Trauma* 2001; 15: 542-8.
17. HILDEBRAND KA, PATTERSON SD, REGAN WD, MACDERMID JC, KING GJW: Functional outcome of semiconstrained total elbow arthroplasty. *J Bone Joint Surg (Am)* 2000; 82A: 1379-86.
18. TURCHIN DC, BEATON DE, RICHARDS RR: Validity of observer–based aggregate scoring systems as descriptors of elbow pain, function, and disability. *J Bone Joint Surg (Am)* 1998; 80A: 154-62.
19. MACDERMID JC: Development of a scale for patient rating of wrist pain and disability. *J Hand Ther* 1996; 9: 178-83.
20. BEATON DE, BOMBARDIER C, GUILLEMIN F, FERRAZ MB: Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000; 25: 3186-91.
21. GUILLEMIN F, BOMBARDIER C, BEATON D: Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46: 1417-32.
22. DEYO RA: Pitfalls in measuring the health status of Mexican Americans: comparative validity of the English and Spanish sickness impact profile. *Am J Public Health* 1984; 74: 569-73.
23. BÜHL A, ZÖFEL P: SPSS 11: *Einführung in die moderne Datenanalyse unter Windows*. München/Germany: Pearson Studium; 2002.
24. BLAND JM, ALTMAN DG: Statistics notes: Cronbach's alpha. *BMJ* 1997; 314: 572.
25. CRONBACH LJ: Coefficient Alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297-334.
26. SIMMEN BR, GSCHWEND N: Elbow arthroplasty – Cinderella of arthroplasty: reconstruction of the elbow with the GSB III prosthesis. *In*: BAUMGARTNER H, DVORAK J, GROB D, MUNZINGER U, SIMMEN BR (Eds): *Rheumatoid arthritis: current trends in diagnostics, conservative treatment, and surgical reconstruction*. Stuttgart, Thieme 1995: 198-221.

Appendix I: Patient Rated Elbow Evaluation [MacDermid (5)]

The questions below will help us understand the amount of difficulty you have had with your elbow in the past week. You will be describing your **average** elbow symptoms **over the last week** on a scale from 0 to 10. If you did not experience an activity in the past week, please estimate your difficulty.

1. PAIN

Rate the average amount of pain in your elbow over the past week by circling the number that best describes your pain on a scale from 0 to 10. A zero (0) means that you did not have any pain, and a ten (10) means that you had the worst pain you have ever experienced.

Sample scale:	0 1 2 3 4 5 6 7 8 9 10	
	No pain	Worst ever
RATE YOUR PAIN:		
When it is at its worst	0 1 2 3 4 5 6 7 8 9 10	
At rest	0 1 2 3 4 5 6 7 8 9 10	
When lifting a heavy object	0 1 2 3 4 5 6 7 8 9 10	
When doing a task with repeated elbow movement	0 1 2 3 4 5 6 7 8 9 10	
	Never	Always
How often do you have pain?	0 1 2 3 4 5 6 7 8 9 10	

2. FUNCTION

A. Specific Activities

Rate the **amount of difficulty** you experienced performing each of the items listed below, over the past week, by circling the number that best describes your difficulty on a scale of 0 to 10. A **zero (0)** means you did not experience any difficulty, and a **ten (10)** means it was so difficult you were unable to do it at all.

Sample scale:	0 1 2 3 4 5 6 7 8 9 10	
	No Difficulty	Unable to Do
Comb my hair	0 1 2 3 4 5 6 7 8 9 10	
Eat with a fork or spoon	0 1 2 3 4 5 6 7 8 9 10	
Pull a heavy object	0 1 2 3 4 5 6 7 8 9 10	
Use my arm to rise from a chair	0 1 2 3 4 5 6 7 8 9 10	
Carry a 10 lb object with my arm at my side	0 1 2 3 4 5 6 7 8 9 10	
Throw a small object, such as a tennis ball	0 1 2 3 4 5 6 7 8 9 10	
Use a telephone	0 1 2 3 4 5 6 7 8 9 10	
Do up buttons on the front of my shirt	0 1 2 3 4 5 6 7 8 9 10	
Wash my opposite armpit	0 1 2 3 4 5 6 7 8 9 10	
Tie my shoe	0 1 2 3 4 5 6 7 8 9 10	
Turn the doorknob and open a door	0 1 2 3 4 5 6 7 8 9 10	

B. Usual Activities

Rate the **amount of difficulty** you experienced performing your **usual** activities in each of the areas listed below, over the past week, by circling the number that best describes your difficulty on a scale of 0 to 10. By „usual activities“ we mean the acticvities that you performed before you started having a problem with your elbow. A **zero (0)** means you did not experience any difficulty, and a **ten (10)** means it was so difficult you were unable to do any of your usual activities.

1.	Personal care activities (dressing, washing)	0 1 2 3 4 5 6 7 8 9 10
2.	Household work (cleaning, maintenance)	0 1 2 3 4 5 6 7 8 9 10
3.	Work (your job or everyday work)	0 1 2 3 4 5 6 7 8 9 10
4.	Recreational activities	0 1 2 3 4 5 6 7 8 9 10

Comments:

Appendix II: German version of the PREE – Score [Patient Rated Elbow Evaluation (PREE-G)]

Die unten aufgeführten Fragen sollten uns helfen, das Ausmass der Schwierigkeiten, die Sie wegen Ihres Ellbogens in der letzten Woche hatten, zu verstehen. Sie werden gebeten, auf einer Skala von 0 – 10 anzugeben, wie stark Ihre Ellbogenbeschwerden bei den folgenden Tätigkeiten in der letzten Woche durchschnittlich waren. Wenn Sie eine (oder mehrere) der Tätigkeiten in der letzten Woche nicht ausgeführt haben, SCHÄTZEN Sie bitte das Ausmass der Schwierigkeiten ein.

1. SCHMERZEN

Geben Sie bitte die durchschnittliche Stärke der Schmerzen im Ellbogen in der letzten Woche an, indem Sie auf einer Skala von 0 bis 10 die Zahl mit einem Bleistift ausfüllen, die Ihre Schmerzen am besten beschreibt. Null (0) bedeutet, dass Sie keinerlei Schmerzen hatten und Zehn (10) bedeutet, dass Sie die schlimmsten Schmerzen hatten, die Sie jemals erlebt haben.

	0 1 2 3 4 5 6 7 8 9 10	
	keinerlei Schmerzen	schlimmste Schmerzen
Bitte geben Sie die Stärke der Schmerzen an:		
Wenn sie am stärksten sind	0 1 2 3 4 5 6 7 8 9 10	
In Ruhe	0 1 2 3 4 5 6 7 8 9 10	
Beim Heben eines schweren Gegenstandes	0 1 2 3 4 5 6 7 8 9 10	
Bei Tätigkeiten mit wiederholter Bewegung des Ellbogens	0 1 2 3 4 5 6 7 8 9 10	
	Nie	Immer
Wie häufig haben Sie Schmerzen	0 1 2 3 4 5 6 7 8 9 10	

2. FUNKTION

A. Funktionsfähigkeit – Bestimmte Tätigkeiten
Bitte geben Sie an, wie viele Schwierigkeiten Sie bei den unten aufgeführten Tätigkeiten in der letzten Woche hatten. Bitte füllen Sie mit einem Bleistift die Zahl von 0 bis 10 aus, die das Ausmass Ihrer Schwierigkeiten am besten beschreibt. Null (0) bedeutet, dass Sie keinerlei Schwierigkeiten hatten und Zehn (10) bedeutet, dass die Schwierigkeiten so gross waren, dass Sie die Tätigkeit nicht ausführen konnten.

	0 1 2 3 4 5 6 7 8 9 10	
	keinerlei Schwierigkeiten	nicht möglich
Ihre Haare kämmen	0 1 2 3 4 5 6 7 8 9 10	
Mit Gabel oder Löffel essen	0 1 2 3 4 5 6 7 8 9 10	
Einen schweren Gegenstand ziehen	0 1 2 3 4 5 6 7 8 9 10	
Den Arm benutzen, um von einem Stuhl aufzustehen	0 1 2 3 4 5 6 7 8 9 10	
Mit dem Arm einen 5 kg schweren Gegenstand seitlich tragen	0 1 2 3 4 5 6 7 8 9 10	
Einen kleinen Gegenstand werfen, wie z.B. einen Tennisball	0 1 2 3 4 5 6 7 8 9 10	
Ein Telefon benutzen	0 1 2 3 4 5 6 7 8 9 10	
Ein Hemd / eine Bluse vorne zuknöpfen	0 1 2 3 4 5 6 7 8 9 10	
Die Achselhöhle auf der Gegenseite waschen	0 1 2 3 4 5 6 7 8 9 10	
Ihre Schuhe zubinden	0 1 2 3 4 5 6 7 8 9 10	
Einen Türknauf drehen und eine Tür öffnen	0 1 2 3 4 5 6 7 8 9 10	

B. Gewöhnliche Tätigkeiten
Bitte geben Sie an, wie viele Schwierigkeiten Sie in der letzten Woche bei Ihren gewöhnlichen Tätigkeiten in jedem der unten genannten Bereiche hatten. Bitte füllen Sie mit einem Bleistift die Zahl von 0 bis 10 aus, die das Ausmass Ihrer Schwierigkeiten am besten beschreibt. Unter „gewöhnliche Tätigkeiten“ verstehen wir die Aktivitäten, die Sie ausführten, bevor die Probleme mit Ihrem Ellbogengelenk begannen. Null (0) bedeutet, dass Sie keinerlei Schwierigkeiten hatten und Zehn (10) bedeutet, dass Sie so grosse Schwierigkeiten hatten, dass Sie keine dieser gewöhnlichen Tätigkeit ausführen konnten.

1.	Persönliche Körperpflege (Anziehen, Waschen)	0 1 2 3 4 5 6 7 8 9 10
2.	Hausarbeit (Putzen, Aufräumen, kleine Reparaturen)	0 1 2 3 4 5 6 7 8 9 10
3.	Arbeit (Beruf oder Alltagstätigkeiten)	0 1 2 3 4 5 6 7 8 9 10
4.	Freizeitaktivitäten	0 1 2 3 4 5 6 7 8 9 10

Bemerkungen: