THE JOURNAL OF BONE & JOINT SURGERY (BRITISH)

Assessing the outcome of disorders of the hand

IS THE PATIENT EVALUATION MEASURE RELIABLE, VALID, RESPONSIVE AND WITHOUT BIAS?

J. J. Dias, B. Bhowal, C. J. Wildin, J. R. Thompson

From Leicester Royal Infirmary, Leicester, England

The different attributes of the Patient Evaluation Measure (PEM) questionnaire were investigated in 80 patients with a fracture of the scaphoid. Assessments were made at 2, 8, 12, 26 and 52 weeks. Reliability was assessed by measurement of the internal consistency of the different questions in 275 completed PEM forms.

Cronbach's alpha, which needs to lie between 0.7 and 0.9, was 0.9 for the PEM. Pain, tenderness, swelling, wrist movement and grip strength correlated with the PEM score confirming the validity of the assessment. Changes in the different variables between visits correlated significantly with changes in the PEM score; its effect size and standardised response mean were comparable to those of grip strength and movement, confirming the responsiveness of this questionnaire. Gender, dominance and the side injured did not influence the scores. Older patients had a poorer outcome as assessed by the score which appeared to be a true effect and not age bias. Our study confirmed that the PEM is a reliable, valid and responsive instrument in assessing outcomes of disorders of the hand.

J Bone Joint Surg [Br] 2001;83-B:235-40.
Received 14 December 1999; Accepted after revision 9 August 2000

There are many patient-completed questionnaires for the assessment of outcomes in the hand which need evaluation to establish that they are robust, relevant and responsive. Such evaluation is critical in choosing a measurement of outcome which can be used during treatment and be employed confidently by post after discharge from hospital.

J. J. Dias, MD, Consultant Orthopaedic Surgeon
B. Bhowal, FRCS, Consultant Orthopaedic Surgeon
C. J. Wildin, FRCS, Specialist Orthopaedic Registrar
Department of Orthopaedics
J. R. Thompson, PhD, Senior Lecturer in Epidemiology
Department of Ophthalmology
Leicester Royal Infirmary, Infirmary Square, Leicester LE1 5WW, UK.

Correspondence should be sent to Mr J. J. Dias at The White Hall, Brook Lane, Billesdon, Leicestershire LE7 9AB, UK.

0301-620X/01/210838 $2.00

The Patient Evaluation Measure (PEM) developed in the UK in 1995[1] has a simple layout with questions asked in a visual analogue form. Patients are asked to read and comprehend the question alone and not the description of each interval answer. The form is uncluttered and easy to understand, complete, enter into a database and to analyse. A computerised optical mark reader can be used to minimise errors in entering the data. However, the reliability, validity, responsiveness and bias have not been determined. Our study, independent of the originators of the form, investigated these attributes for disorders of the wrist.

Patients and Methods

Between October 1996 and July 1999 we studied 80 patients with acute fractures of the scaphoid at 2, 8, 12, 26 and 52 weeks after the injury. There were nine women and 71 men. Their mean age was 29.7 years (SD ± 10.1; range 16 to 61). Thirteen were left-handed and 66 were right-handed. One was ambidextrous. These patients formed part of a study on scaphoid fractures. Patients with pre-existing disorders of the wrist or hand were excluded. At each visit information on the type, site, severity and frequency of pain and swelling were collected. The hand was examined for the site and severity of tenderness. The ranges of wrist movements were measured in a standard manner using a goniometer. The sum of flexion, extension and radial and ulnar deviation was expressed as a proportion of the sum of these movements on the opposite, unaffected side.

The grip strength of both hands was measured using a single calibrated Jamar dynamometer (Preston, Jackson, Michigan). That of the affected side was then expressed as a percentage of the grip strength measured in the opposite hand on that visit. This technique of measuring grip strength should overcome the known variations in its assessment.

The PEM (Fig. 1) was given to the patient in the clinic at each visit. The answers to the questions are presented as a seven-interval visual analogue scale for each item. The patients attributed values for each of the five items in the first part of the questionnaire, which assessed their view of that particular consultation, and the third part which gave a general view on their treatment and of the condition of the hand. The second part of the PEM, the Hand Health

Part one - treatment

Please put a circle around the number that is closest to the way you feel about how things have been for you. There are no right or wrong answers.

1. Throughout my treatment I have seen the same doctor:
   1 2 3 4 5 6 7
   Every time — Not at all

2. When the doctor saw me, he or she knew about my case:
   1 2 3 4 5 6 7
   Very well — Not at all

3. When I was with the doctor, he or she gave me the chance to talk:
   1 2 3 4 5 6 7
   As much as I wanted — Not at all

4. When I did talk to the doctor, he or she listened and understood me:
   1 2 3 4 5 6 7
   Very much — Not at all

5. I was given information about my treatment and progress:
   1 2 3 4 5 6 7
   All that I wanted — Not at all

Part two - how is your hand now

Hand health profile

1. The feeling in my hand is now:
   1 2 3 4 5 6 7
   Normal — Absent

2. When my hand is cold and/or damp, the pain is now:
   1 2 3 4 5 6 7
   Non-existent — Unbearable

3. Most of the time, the pain in my hand is now:
   1 2 3 4 5 6 7
   Non-existent — Unbearable

4. The duration my pain is present is:
   1 2 3 4 5 6 7
   Never — All the time

5. When I try to use my hand for fiddly things, it is now:
   1 2 3 4 5 6 7
   Skilful — Clumsy

(Part two contd)

6. Generally, when I move my hand it is:
   1 2 3 4 5 6 7
   Flexible — Stiff

7. The grip in my hand is now:
   1 2 3 4 5 6 7
   Strong — Weak

8. For everyday activities, my hand is now:
   1 2 3 4 5 6 7
   No problem — Useless

9. For my work, my hand is now:
   1 2 3 4 5 6 7
   No problem — Useless

10. When I look at the appearance of my hand now, I feel:
    1 2 3 4 5 6 7
    Unconcerned — Embarrassed & self-conscious

11. Generally, when I think about my hand I feel:
    1 2 3 4 5 6 7
    Unconcerned — Very upset

Part three - overall assessment

1. Generally, my treatment at the hospital has been:
   1 2 3 4 5 6 7
   Very satisfactory — Very unsatisfactory

2. Generally, my hand is now:
   1 2 3 4 5 6 7
   Very satisfactory — Very unsatisfactory

3. Bearing in mind my original injury or condition, I feel my hand is now:
   1 2 3 4 5 6 7
   Better than I expected — Worse than I expected

Fig. 1

The PEM Questionnaire.

Questionnaire, has ten questions investigating different attributes of hand health and function. A question on the duration of pain (number 4) was added to the original version of the questionnaire.

This was first given to the patient when followed up after two weeks. At each subsequent visit the patients were told what their answer had been at the previous visit for each question and they were then asked to indicate on the visual analogue scale the direction and magnitude of change for each item in the time period between assessments. This

ensured that each was assessed by the patient as a transition question. The PEM score was then calculated by summing the values for each item in parts two and three and expressing it as a percentage of the maximum possible score. The first part, which assessed the patient's view of the consultation, was excluded.

We collected complete unmodified data on these variables from 275 patient visits. They were used to assess reliability, validity and bias.

Reliability. This was assessed by measuring the internal consistency of the 11 items in the Hand Health Questionnaire part of the PEM and three items in the overall assessment. We measured the correlation of each item with each of the others and generated Cronbach's alpha.[2] The questionnaire was considered to be internally consistent if Cronbach's alpha was between 0.7 and 0.9. Internal consistency investigates how well different items in the scale measure different aspects of a single assessment, namely the health of the hand in this study. If the internal consistency is too high, >0.9, then the items are measuring the same aspect twice.

Validity. The validity of the PEM was assessed by determining how well responses to each item of the questionnaire, and the total score, correlated with pain, tenderness, swelling, wrist movement expressed as a percentage of the movement of the opposite wrist, and grip strength expressed as a percentage of that in the opposite hand. The correlations were calculated by combining data on repeated assessments at different points in the treatment of the same patients. Because of the repeated measures, conventional standard errors were too small and therefore the correlations have been given without confidence intervals. Thus, different subjective and objective parameters of hand function were used to assess whether the PEM was measuring what it was designed to accomplish.

Responsiveness. The ability of the PEM to detect clinically important change, that is its responsiveness, was assessed.[3] The change between visits for each patient in pain, tenderness, swelling, wrist movement, grip strength and the PEM score was assessed. Changes in the PEM score were correlated with changes in each of the other variables.[4] In addition, responsiveness was measured by the effect size[5] and by assessing the standardised response mean.[6]

Bias. Item bias[7] was investigated by measuring the variance of each item of the PEM and the full score for gender, hand dominance, the side injured and injuries to the dominant hand as opposed to those in the non-dominant hand. Age bias in the responses was studied by correlating age with the PEM score. The effects of subject factors such as age and gender were tested in a repeated-measures analysis of variance, which incorporated a fixed effect for the visit and a random effect for the subject.

Results

Reliability. Table I shows the correlation between each item of part two (Hand Health Profile) and part three of the PEM scale and the remaining items in this part of the questionnaire. This correlation was between the 14 items in 80 subjects, some on five occasions. There is overall correlation across visits and for each visit. Cronbach's alpha for each visit was 0.874, 0.905, 0.939, 0.912, 0.911 for weeks 2, 8, 12, 26 and 52, respectively. These figures tend to be slightly lower than the overall value (0.932) because the latter reflects the extra correlation due to changes in the group mean score over time. An individual Cronbach's alpha measures the consistency of the items at a single point in time and the overall Cronbach's alpha the consistency across time.

The correlations of the use of the hand for fiddly tasks, hand grip and overall satisfaction were high with the rest of the scale at 0.83, 0.84 and 0.84, respectively. The correlations of hospital treatment, appearance and feeling were particularly low at 0.34, 0.43 and 0.36, respectively. The internal consistency was assessed by calculating Cronbach's alpha from these correlations and found to be 0.9. The degree of contribution of each item was studied by measuring Cronbach's alpha excluding that particular item.

Table I. Reliability analysis of the PEM expressed as a percentage of the maximum possible score using Cronbach's alpha (14 items, n = 275)

	Mean	Variance	Alpha	Correlation*
PEM	*26.8*	170	0.932	
Deleting single scale items				
Feeling	*27.4*	189	0.934	0.36
Pain in cold	*27.0*	181	0.931	0.51
Pain severity	*27.0*	176	0.926	0.72
Pain duration	*26.8*	175	0.926	0.73
Use	*26.5*	160	0.921	0.83
Move	*26.4*	160	0.923	0.79
Grip	*26.2*	155	0.921	0.84
Activities	*26.6*	161	0.919	0.90
Work	*26.3*	157	0.926	0.75
Appearance	*27.2*	182	0.933	0.43
Think	*26.9*	170	0.926	0.70
Treatment	*27.6*	193	0.935	0.34
Satisfaction	*26.9*	169	0.922	0.84
Better/worse	*26.9*	172	0.925	0.74

* correlation between the deleted item and the PEM based on the remaining 13 items

This is demonstrated in column 4 of Table I. If an item is contributing a disproportionate amount to the internal consistency then deleting that item will cause a fall in Cronbach's alpha from the scale value of 0.9. If the item is not influencing the internal consistency, the Cronbach's alpha will rise. As Cronbach's alpha remained close to the scale value, our interpretation was that each item contributed to the overall picture of the health of the hand and that the scale was reliable.

Validity. A fundamental requirement of any questionnaire is that it measures what it is meant to. This can be investigated in different ways, but perhaps the most measurable is its validity. This assesses the different items on the scale against a number of other measures such as disorder-specific questionnaires or objective evaluation of different aspects of the disorder. Table II shows that the PEM was highly valid since it correlated with other commonly used assessments of the function of the wrist and the hand. Pain, tenderness, swelling, range of movement and grip strength all correlated with individual items and the total score of the PEM.

Responsiveness. A questionnaire which is used for repeated assessments of abnormality of the hand must be able to reflect the change in the underlying disorder by parameters such as pain, tenderness, stiffness and weakness which are commonly used for assessment. Of the various statistical measures to assess responsiveness, the easiest to understand is the correlation between the variation in the questionnaire to changes in pain, tenderness, swelling, movement and strength.[4] Table III shows that severity of pain and grip strength correlate significantly ($p < 0.001$) with changes in the PEM score between the second and eighth weeks, and that the severity of tenderness, range of movement and grip strength correlated significantly ($p < 0.03$) between weeks 8 and 52. The effect size, calculated by dividing the mean of change in the PEM score between visits by the standard deviation of the baseline PEM score, was 1.12. This needed

Table II. Correlation of the 14-item PEM score with other clinical assessments to assess construct validity

	Clinical assessments				
	Pain severity	**Tenderness severity**	**Swelling severity**	**Range of movement***	**Grip strength***
Number	274	256	265	266	266
Correlation	0.56	0.64	0.55	-0.69	-0.76
PEM items					
Feeling	0.21	0.18	0.10	-0.27	-0.34
Pain in cold	0.54	0.35	0.20	-0.33	-0.40
Pain severity	0.60	0.50	0.36	-0.55	-0.53
Pain duration	0.54	0.50	0.33	-0.48	-0.52
Use	0.47	0.61	0.54	-0.61	-0.68
Move	0.48	0.60	0.62	-0.68	-0.66
Grip	0.45	0.61	0.58	-0.69	-0.77
Activities	0.50	0.63	0.56	-0.65	-0.73
Work	0.41	0.59	0.54	-0.61	-0.63
Appearance	0.14	0.28	0.19	-0.31	-0.37
Think	0.35	0.35	0.30	-0.41	-0.51
Treatment	0.09	0.19	0.18	-0.22	-0.23
Satisfaction	0.43	0.50	0.39	-0.52	-0.61
Better/worse	0.42	0.49	0.36	-0.45	-0.53

* expressed as a percentage of the value for the opposite hand

Table III. Responsiveness of the PEM assessed as the effect size, standardised response mean and correlation of change in the PEM score *versus* other assessments

	Number	**Effect size**	**Standardised response mean**	**Correlation of change with change in PEM**	**p value**
Between weeks 2 and 8					
PEM	46	-1.12	-1.46		
Pain severity	58	-0.35	-0.27	**0.51**	**0.0003**
Tenderness severity	36	-0.94	-0.82	0.27	0.17
Swelling severity	45	-1.09	-1.14	0.11	0.54
Range of movement*	47	1.27	1.15	0.04	0.82
Grip strength*	47	1.63	1.30	**-0.50**	**0.001**
Between weeks 8 and 52					
PEM	35	-1.10	-1.47		
Pain severity	39	-0.61	-0.59	0.18	0.31
Tenderness severity	40	-0.64	-0.66	**0.49**	**0.003**
Swelling severity	40	-0.85	-0.85	0.19	0.28
Range of movement*	40	1.93	2.45	**-0.41**	**0.02**
Grip strength*	40	1.16	1.32	**-0.37**	**0.03**

* expressed as a percentage of the value for the opposite hand.
Significant correlations ($p < 0.05$) are shown in bold

Table IV. An assessment of bias based on the association between PEM and various subject factors

Subject factors		Regression coefficient	Standard error	95% confidence limits		p value
				Lower	Upper	
Gender	Female	5.41	3.35	-1.16	11.98	0.11
Dominance	Left	-2.22	2.88	-7.84	3.44	0.12
	Ambidextrous	17.89	9.56	-0.85	36.63	
Side	Left	1.61	2.30	-2.90	6.12	0.49
Age	per year	0.28	0.11	0.06	0.50	0.01

to be greater than 0.8 to be valuable. This effect size matched that of grip strength, range of movement and severity of swelling and tenderness between assessments at weeks 2 and 8 and between weeks 8 and 52. The standardised response mean, calculated by dividing the mean of change in the PEM score between visits at weeks 2 and 8 and between weeks 8 and 52 by the standard deviation of the change in PEM score, was 1.46 and 1.47, respectively. These assessments established that the PEM is a highly responsive scale.

Bias. Table IV shows the relationship between the PEM score and assumed independent variables such as age, gender, hand dominance and the side of injury, to determine whether there is item bias. The regression coefficients were obtained separately from models which also adjusted for the subject through a random term and the time of the visit through a fixed effect. There was no item bias for gender, dominance and side injured. The correlation with age was significant at 0.01, suggesting that older patients had a slightly lower PEM score. Instead of an age bias this could suggest a poorer outcome in older individuals, which was confirmed by a regression analysis of age and grip strength (regression coefficient, -0.51; standard error, 0.16; p value = 0.002).

Discussion

Assessing the outcome of treatment is the only way to monitor the effectiveness of the intervention and the quality of care. The task is difficult. Several techniques have evolved to look at outcomes of injury or intervention for disorders of the hand. Most are specific to a single anatomical region or a single disorder. Some have been validated and, although these, such as the Jebsen test, have been shown to be robust, their routine use is uncommon. Other tests such as the SF36 or the Nottingham Health Profile are so general in assessing the patient's 'whole' state that it is difficult to even obtain an indication of what aspect should be addressed in order to improve outcome.

We believe that a patient-completed questionnaire is the most efficient way of collecting information on outcome for routine use. Specific outcome measures may be required for audit and comparative studies, but cannot be used routinely in busy hand clinics. The forms, shown in Figure 1, provide a method of assessing outcome without the need for the patient to attend an outpatient clinic. This has the benefit of giving clinicians valuable information about the treatment or progress of a disorder, without the cost and time implications of repeated follow-up visits. An added benefit is that more clinic time could be available to devote to those disorders which do need careful monitoring. Patient-based outcome measures may also help to identify those patients who need formal review by doctors or by therapists.

While there are a number of questionnaires which assess outcomes of only one disorder (e.g., the Levine questionnaire for carpal tunnel syndrome[8]), or for one region of the hand (e.g., the wrist score[9]), there are very few that provide an overall view. In the USA two forms have been developed recently, the Disabilities of the Arm, Shoulder and Hand (DASH)[10] assessment and the Michigan Hand Index (MHI).[11] These were found by the originators to be relevant and robust. They are being used increasingly around the world. In the UK the PEM has been used. Each of these forms has positive attributes and drawbacks, the investigation of which was not the purpose of our study. In order to be used routinely such assessments must be valid, robust and relevant. They must also be simple to complete and have the ability to indicate in which way the outcomes are suboptimal. Our study tried to identify the various characteristics of the PEM form alone.

We established that the items in the PEM are internally highly consistent. Each item on the scale contributed to building a reasonably complete picture of hand function. The questions addressed different facets of the use and function of the hand. Three questions, in particular, assessed the impact of the disorder on the patient. These are (i) 'When I look at the appearance of my hand now, I feel: unconcerned..... embarrassed and self-conscious'; (ii) 'Generally, when I think about my hand I feel: unconcerned..... Very upset'; and (iii) 'Generally, my hand is now: very satisfactory..... very unsatisfactory'. These attributes are not usually addressed in objective assessments of outcome, which tend to concentrate mostly on pain, tenderness, range of movement and pinch and grip strength. The face validity of the PEM questionnaire therefore appears sound.

The style of the questionnaire is simple. It uses short clear questions with a visual analogue scale with seven intervals to determine the magnitude of each answer. This provides an optimal interval scale[12] which is less prone to error than the continuous analogue measurement or a four- or five-interval Likert scale. A five-interval Likert scale of responses has been used extensively in the DASH and MHI questionnaires.

The only question in the Hand Health Profile section (part two) of the PEM to trouble the patient was the first: 'The feeling in my hand is normal.... completely absent.' In common English the word 'feeling' is more than just 'sensation of touch' and includes 'feeling of stiffness', 'feeling swollen' and 'feeling not right'. It is possible that changing the text of the question to, 'The feeling of touch in my hand....' may make this question more targeted to the assessment of sensation. If the disorder is known, keeping the general common language sense may be of equal value.

The picture of the validity of a questionnaire is built up from a number of disparate assessments carried out by different groups providing converging evidence rather than on a single study. The statement that a questionnaire has been 'validated' must therefore be seen in the context of 'for which disorder' and 'against what other measure'. The correlations between each item of the scale and the whole score with other objective assessments of range of movement, tenderness and grip strength confirmed that the PEM measures what it is designed to measure and confirmed the form's validity, at least for disorders of the wrist.

This study also established that the PEM questionnaire is highly responsive to change in the commonly assessed parameters of hand function such as pain, tenderness and swelling. It is similarly responsive to changes in movement and grip strength.

It is probable that the method of administration determined this responsiveness. After the initial visit, on subsequent assessment the patients were told their previous answer and were asked to indicate the direction (worse, no change, better) and the magnitude of any change since their last evaluation. In effect each question was assessed as a transition question. This overcame a number of extraneous influences in the completion of the form such as a chance entry a single interval away from the previous mark, mood, memory or the effects of medication. It should be possible to indicate the previous response on the questionnaire so that the patients could provide information of a change rather than selecting an absolute value on each occasion. Based on this study we recommend that if the PEM were to be used repeatedly the patient should know their previous answer.

The sensitivity and specificity[13] of this measure in detecting change could not be assessed confidently because of the very small numbers of observations where no change was detected in this study. There could also be an error in assuming that if patients had no change in their pain and movement they were completely stable clinically. For example, the quality of rate and rhythm of movement may have altered without any change in the range. The aspect of specificity requires further investigation in a population with a larger proportion of clinically stable individuals.

There is a threat to outcome assessment from systematic bias when using a questionnaire.[14] Item bias can occur when independent variables such as gender and age affect responses to questions. This study confirmed that no bias existed for the PEM with regard to gender, age, hand dominance or side injured.

In order to maintain the comparability of outcomes we strongly recommend that this questionnaire should be used by different units without modification. Any suggestions for modification should be forwarded to the originator of the form who could then propose, validate and institute a new version. The interval between modifications should be sufficiently long, probably five years, to avoid a variety of different versions creeping into use. Alternatively, modifications should only change the number of items studied while keeping the construct of each item unchanged. This would make any new version backwardly compatible.

As a result of our study we conclude that the PEM is an internally consistent form, which is highly valid and responsive. Gender, age, handedness and side injured did not cause bias in the responses. When this form is being administered repeatedly the patient should probably be informed of their previous answer to provide a better measure of change.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

References

1. **Macey AC, Burke FD.** Outcomes of hand surgery. *J Hand Surg* 1995;20-B:841-55.
2. **Cronbach L.** Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:287-334.
3. **Guyatt GH, Walter S, Normal GR.** Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171-8.
4. **Meenan RF, Anderson JJ, Kazis LE, et al.** Outcome assessment in clinical trials: evidence for the sensitivity of a health status measure. *Arthritis Rheum* 1984;27:1344-52.
5. **Kazis LE, Callahan LF, Meenan RF, Pincus T.** Health status reports in the care of patients with rheumatoid arthritis. *J Clin Epidemiol* 1990;43:1243-53.
6. **Liang MH, Fossel AH, Larson MG.** Comparisons of five health status instruments for orthopaedic evaluation. *Med Care* 1990;28:632-42.
7. **Groenwold M, Bjorner JB, Klee MC, Kreiner S.** Test for item bias in a quality of life questionnaire. *J Clin Epidemiol* 1995;48:805-16.
8. **Levine DW, Simmons BP, Koris MJ, 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-A:1585-91.
9. **Green DP, O'Brien ET.** Open reduction of carpal dislocations: indications and operative techniques *J Hand Surg* 1978;3:250-65.
10. **Hudak PL, Amadio PC, Bombardier C.** Development of an upper extremity outcome measure: the DASH (Disabilities of the arm, shoulder and hand: the upper extremity collaborative group). *Am J Ind Med* 1996;29:602-8.
11. **Chung KC, Pillsbury MS, Walters MR, Hayward RA.** Reliability and validity testing of the Michigan Hand Outcomes questionnaire. *J Hand Surg [Am]* 1998;23:575-87.
12. **Avis NE, Smith KW.** Conceptual and methodological issues in selecting and developing quality of life measures. In: Fitzpatrick R, ed. *Advances in medical sociology.* London: JAI Press Inc, 1994:255-80.
13. **Deyo RA, Inni TS.** Toward clinical applications of health status measures: sensitivity of scales to clinically important changes. *Health Serv Res* 1984;19:275-89.
14. **Boucet C, Guillemin F, Braincon S.** Nonspecific effects in longitudinal studies: impact on quality of life measures. *J Clin Epidemiol* 1996;49:15-20.