Test-retest reliability of the Short-Form McGill Pain Questionnaire

GRAFTON, Kate, FOSTER, Nadine and WRIGHT, Christine

Title Page

Title:

Test-retest reliability of the Short-Form McGill Pain Questionnaire: Assessment of intraclass correlation coefficients and limits of agreement in patients with osteoarthritis.

Authors:

Kate V Grafton Grad.Dip.Phys. MSc MCSP MMACP – Senior Lecturer, School of Health and Social Care, Sheffield Hallam University, Sheffield S10 2BP, UK.

Nadine E Foster BSc(Hons) DPhil MCSP – Lecturer, Department of Physiotherapy Studies, Keele University, Staffordshire ST5 5BG, UK

Christine C Wright BSc(Hons) FSS, C.Math MIMA, - Principal Lecturer in Research Methods, School of Health and Social Sciences, Coventry University, Coventry CV1 5FB, UK

Author for correspondence:

Kate V Grafton Grad.Dip.Phys. MSc MCSP MMACP - School of Health and Social Care, Sheffield Hallam University, Collegiate Crescent Campus, Sheffield S10 2BP, UK. Tel: +44 (0) 114 2252498 Fax:+44 (0) 114 2252430 E-mail: k.grafton@shu.ac.uk

Abstract

Objectives

No published study has previously evaluated the test-retest reliability of the Short-Form McGill Pain Questionnaire (SFMPQ), yet it is increasingly being used as a measure of pain. This study evaluates the test-retest reliability in patients with osteoarthritis.

Methods

A prospective, observational cohort study was undertaken using serial evaluation of 57 subjects at two time-points. A sample of patients awaiting primary hip or knee joint replacement surgery, were recruited in clinic or via post. SFMPQs were delivered by post 5 days apart, and a supplementary questionnaire was completed on the second occasion to ensure that the subjects' pain report had remained stable.

Results

The intra-class correlation coefficient (ICC) was used as an estimate of reliability. For the total, sensory, affective and average pain scores high ICCs were demonstrated (.96, .95, .88 and .89, respectively). The current pain component demonstrated a lower ICC of .75. The coefficient of repeatability (CoR) was calculated as an estimation of the minimum metrically detectable change. The CoR for the total, sensory, affective, average, and current pain components were 5.2, 4.5, 2.8, 1.4cm and 1.4, respectively.

Discussion

Problems of adequate completion of the SFMPQ were highlighted in this sample and supervision via telephone contact was required. Patients recruited in clinic, who had practised completing the SFMPQ, demonstrated fewer errors than those

recruited by post. The SFMPQ was demonstrated to be a highly reliable measure of pain. These results should not be generalised to a more elderly population, as increasing age was correlated with greater variability of the sensory component scores.

1. Introduction

Pain is prevalent, costly and the most common reason for which people seek healthcare [1-5]. The measurement of pain is complex, but considered important in facilitating a diagnosis and as a measure of treatment effectiveness [6]. Pain measurement methods must adhere to key issues of reliability, validity and responsiveness to change [7, 8].

It is often considered that simple pain rating scales are inappropriate to evaluate what is acknowledged to be a multidimensional experience [9]. A multidimensional measure may provide a more comprehensive estimate of patients' pain experiences, for which they are seeking treatment and upon which clinical decisions are based. The most commonly used multidimensional pain measure is the McGill Pain Questionnaire (MPQ) [10, 11].

The validity of the MPQ has been generally supported by an abundance of studies [12-23], many of which have employed factor analysis to determine independent groupings of items for specific patient populations [24]. There is some evidence for the test-retest stability of the MPQ, principally from a study on 65 patients with chronic low back pain prior to receiving chiropractic treatment [25]. However, this study used Pearson's correlation coefficient as a measure of reliability, which is inappropriate, as it measures linear association rather than agreement [26, 27]. Also, there is no statement regarding the stability of the patients' pain over the days between completing the questionnaire.

Two studies have reported the consistency of choice of descriptions as ranging from 66% to 80% agreement [19, 28], when the MPQ was completed at different time-points by the same patients.

Whilst being well validated, the MPQ takes 15 to 20 minutes to complete [10] and some patients have difficulty with the complexity of the vocabulary used [29]. The Short-Form McGill Pain Questionnaire (SFMPQ) was developed to address these issues [30], and is increasingly being used as an outcome measure in both research and clinical practice but very few studies evaluate its psychometric utility [31-36].

The SFMPQ was developed from the most frequently used descriptors on the MPQ and assesses the sensory and affective dimensions of pain. Each descriptor is ranked as none, mild, moderate or severe [30]. A visual analogue scale (VAS) for pain intensity and a verbal descriptor scale (VDS) for present pain, are included. The SFMPQ is frequently used as a self-report questionnaire, but there are no standardised patient instructions published. It is multidimensional, evaluating dimensions homologous with those on the MPQ, so supporting its content validity [30]. There are no published studies evaluating its factor structure. It's construction has been derived logically from the MPQ [30], which has itself demonstrated construct validity. Neither of these theoretical aspects of validity has been extensively investigated by published studies.

Two studies have demonstrated that the concurrent criterion validity of the SFMPQ with the MPQ is good [30, 37]. As the SFMPQ is a subset of the MPQ, a good correlation between the scores would be expected. It is suggested that when two measures designed to evaluate the same construct are tested for association, they will always demonstrate a statistically significant association [38].

A search of databases from 1987 (the year of the SFMPQ's initial publication) revealed only one published study that had evaluated aspects of the SFMPQ's

test-retest reliability [39]. The study assessed the intercorrelation of the SFMPQ and five other pain assessment instruments but demonstrated poor research design, quality of data and statistical methods. From a sample of 31 patients with chronic low back pain, the responses of 17 patients were not included in the final analysis due to insufficient data. The patients were assessed initially in a hospital outpatient clinic and then weekly for four or five weeks. Whilst no medication or physical treatment was administered, general and specific ergonomic advice was given, potentially affecting the stability of the patients' symptoms over the evaluation period. The median coefficient of variation was calculated as the estimation of measurement error, but its use is not recommended for assessing reliability [38, 40]. The study results did not support the test-retest reliability of the SFMPQ, with the coefficients of variation ranging from 19% to 69%.

The SFMPQ is reported to be sensitive to clinical improvement in a variety of populations [30-32, 37]; these studies only imply responsiveness and do not specifically evaluate responsiveness to change. Without evidence of the stability of the SFMPQ when no change in pain has occurred, any reported change in SFMPQ score must be interpreted with caution. The change in SFMPQ score may reflect the measurement error of the SFMPQ and not a change in the pain being measured. Thus, evidence of the SFMPQ's test-retest reliability is a precondition to evidence of it's responsiveness to change.

No studies have been published that adequately demonstrate the reliability, responsiveness or validity of the SFMPQ. The primary aim of this study was to investigate the test-retest reliability of the SFMPQ, as a first step towards supporting the questionnaire's use as an outcome measure in research and clinical practice.

2. Materials and Methods

2.1 Setting and subjects

This prospective, observational cohort study was undertaken with consecutive subjects who attended the outpatient orthopaedic clinic at a large teaching hospital in the North of England. The subjects were selected from a sample who were on the waiting list for primary hip or knee joint replacement surgery for 'pain-dominant' osteoarthritis (OA), as they formed a suitable, accessible target population. Recruitment occurred either in clinic or via post after their clinic attendance. Written informed consent was obtained. Subjects were excluded from the study if they were unable to read or understand English, or could not indicate their pain description by marking the appropriate box on the SFMPQ. Ethical approval for the study was gained from the local research ethics committee. Over four months, 80 consecutive subjects were referred from the orthopaedic clinics for inclusion into the study; three were subsequently excluded as two had rheumatoid arthritis and one was partially sighted (Figure 1). Of the 77 eligible subjects, 71 (a 92.2% response rate) completed both SFMPQs. Subjects' responses were excluded from the analysis if they were unable to adequately complete the SFMPQ (n=5), or they reported a change in their pain on the supplementary questionnaire, so reflecting a pain status that was not stable (n=9). It is important to only include those subjects who judge that their pain is unchanged, as the variable being measured (pain) must be unchanged in order to evaluate the stability of the SFMPQ as a pain measure [41-43]. Subjects who reported a change in their pain also reported a change in their health or physical function. Table 1 summarises the number of errors in completion and telephone clarifications required by the sample. The respondents who omitted the VAS for

average pain but adequately completed the sensory, affective and current pain sections were included in the analysis for the completed sections. There were no statistically significant differences (using t-tests and chi-square analysis with p<0.05) between the demographics of the final sample of 57 subjects included in the statistical analysis and the initial 71 respondents. The mean age was 64.8 years (range 36 to 81, SD 10.4); there were 21 (36.8%) males and 36 (63.2%) females; 41 (71.9%) were awaiting hip arthroplasty and 16 (28.1%) knee arthroplasty. Further statistical exploration of the demographic characteristics indicated that there were no statistically significant differences between subjects recruited in clinic and via post, or between subjects reporting unstable and stable pain. However, an independent t-test demonstrated a statistically significant difference in the mean ages of subjects with problems completing the SFMPQ and those without completion problems (t=2.4, df=69, p=0.021). The mean age of subjects with completion problems was 68.3 years (SD 10.3), compared with 62.5 years (SD 10.2) for subjects without completion problems.

2.2 *Pilot study*

A pilot study with 23 subjects was undertaken to optimise the recruitment method, SFMPQ completion, and to establish the estimations upon which the sample size calculations were based. The modifications made arising from the pilot were:

- subjects recruited in clinic completed a practice SFMPQ to correct any initial completion errors

- the primary researcher carried out telephone clarification of any ambiguous responses on returned SFMPQs

- any reported change in the subject's pain, indicated by the supplementary questions was clarified by telephone (to ensure that subjects were referring

to the five day period between completing the SFMPQs and not to a
general, more long-term change).

The predicted reliability level (ρ1) of the SFMPQ component and total scores was
estimated as 0.9 from the pilot, the minimal acceptable reliability level (ρ0) for the
SFMPQ was set at 0.8 (based upon the limited consensus in the literature) [44], β
was set at 80% and α at 0.05. Using tables provided by Walter et al.'s [45]
functional approximation method, a sample of 46 subjects was identified as
required for the main study.

2.3 Procedure

Subjects completed the SFMPQ (Figure 2) at two time points, ten days after their
clinic attendance (test 1) and a further five days later (test 2). At the second time-
point, subjects also answered four supplementary questions which asked about
any change in health, physical function, pain or medication since completion of
the first questionnaire. Non-responders were telephoned to remind them to return
the questionnaire. To minimise bias or errors in the data handling, an independent
observer, blinded to the study's aim, checked all SFMPQ scoring and a random
20% of the data entry.

2.4 Data analysis

The SFMPQ was scored as recommended by Melzack [30] and therefore, was
regarded as interval level data. It is acknowledged that there is some evidence to
suggest that the assumption that the data from the rank descriptors is continuous
is incorrect, and that the category items do not exhibit the assumed homogeneity
of spacing [46]. However, the authors felt it necessary to initially establish the
reliability of the questionnaire as devised and scored by the developers. There is
a lack of consistency in the literature regarding choice of reliability estimates. The

9

intraclass correlation coefficient (ICC) is a preferred method of estimating

reliability as it relates the size of the error in repeated measurements to the

variation of interest [47, 48]. A disadvantage of the ICC is that it is a 'unit-less'

value, giving no indication of the actual measurement range or biases, and so it is

difficult to interpret clinically [40, 44]. The standard error of measurement (SEM) is

the standard deviation of the measurement error and is easier to interpret

clinically, as it is expressed in the units of measure [44]. Bland and Altman [38]

advocate the use of scatter diagrams and limits of agreement. These methods

identify biases in the scores and provide an estimate of a range of error that must

be interpreted in the context of the variance in the individual outcome measures

[38, 49]. Bland [49] also recommends the coefficient of repeatability (CoR) for

reliability estimation involving repeated measures. The CoR may be defined as

the value below which the absolute difference between test-retest scores may be

expected to lie with 95% probability [50]. It reflects the measurement error and

represents the clinical minimum detectable change in the unit of measurement.

The current consensus is that, independently, each method has its weakness but

by combining the methods, a more complete estimation of reliability may be

achieved [26, 40, 44].

Inferential statistics have not been used to explore the internal consistency of the

item selection, as although the SFMPQ uses a summated rating scale for the item

selection, it is not expected that all questions would score similarly. Indeed, the

SFMPQ may theoretically be used to discriminate between diagnostic pain traits

[30].

Data were analysed using the Statistical Package for the Social Sciences (SPSS,

Woking, Surrey, UK, Version 9). Plots of the difference (mean and absolute)

between tests 1 and 2 against the mean of tests 1 and 2 were constructed [38, 49].

The reliability of each component score ('sensory', 'affective', 'average' and 'current' pain) and the total (sensory and affective) score of the SFMPQ (Figure 3) were estimated using the ICC (1,1), SEM and CoR equations shown in Figure 4. Pearson's correlation was carried out to test for any association between the variables. Two-tailed significance was set at 0.05.

3. Results

Values of the difference in SFMPQ scores over the two time-points were normally distributed and demonstrated homoscedasticity. Table 2 summarises the mean score for each component of the SFMPQ and the total score across the two time points. It also shows the mean and absolute differences, and 95% confidence intervals (CI) for the mean differences between the two scores for each component and the total score. The 95% CI's for the mean difference in component scores at the two time points include zero, indicating no significant bias towards subjects scoring higher or lower at time point two. Also, the small mean absolute differences indicate a small magnitude of variation between the scores at the two time points.

Scatter diagrams of the absolute difference against the mean, and of the mean difference against the mean, for each pair of scores (Figure 5) demonstrated no evidence of systematic bias between the magnitude of the differences and the magnitude of the component and total SFMPQ scores. Furthermore, no bias was evident for subjects scoring higher or lower on the SFMPQ at the second time-point [38, 49].

The ICC values were high for the sensory, affective, average and total pain components (.95, .88, .89 and .96, respectively) with narrow CIs, indicating precise estimation of the reliability coefficient (Table 3). The 'current' pain dimension demonstrated a lower ICC of .75 and wider 95% CI, indicating less precision in this estimated coefficient.

The SEM for the total score was 1.87, the mean total score was 18.9. The sensory, affective, average and current pain components all demonstrated the expected smaller SEMs; 1.64, 1.01, 0.52 and 0.51, respectively. The CoR values demonstrate a change of at least 5.18 must be evident in the total score (4.54, 2.80, 1.44cm and 1.41 for the sensory, affective, average pain and current pain scores, respectively) if the change in pain is to be interpreted as a clinical change. The CoR reflects the precision of the SFMPQ and is considered as part of the overall reliability. Together, the ICC values and the CoR support the precision and reliability of the SFMPQ.

Pearson's correlation coefficient demonstrated that the association between age and absolute difference between SFMPQ scores was statistically significant for the sensory component only (r= 0.31, p=0.045); indicating that the older the subject, the greater the difference between the sensory pain scores across the two time-points.

4. Discussion

4.1 Statistical Analysis

Statistical theory provides no clear guidelines for acceptable reliability, with many authors ascribing different interpretations of 'acceptable' ICC values [43, 44, 48]. The high ICCs (.88 to .96) obtained for the SFMPQ total score, and the sensory,

affective and average pain dimensions would suggest that the SFMPQ is a highly reliable, multidimensional measure of pain in this population. The SEM helps to place the ICC in the context of the data from which it was derived, with smaller SEMs indicating greater reliability [44]. However, the CoR is more clinically useful by identifying the difference between the test-retest scores with confidence 0.95 [44, 49, 50]. In the total score, an ICC of .96 reflected a variation in the test-retest scores of 5.2, and the sensory component's ICC of .95 reflected a variation in the test-retest scores of 4.5. Hence, for a subject's recorded change in SFMPQ to be detected as a clinical change it must be greater than 5.2 for the total score, or 4.5 for the sensory component. Any change in score that is less than these values reflects the measurement error of the SFMPQ and is not attributable to a clinically meaningful change.

4.2 Current pain component score

A lower ICC (.75) was obtained for the current pain dimension. The pre-determined, minimally acceptable ICC value was set at 0.8, and so the values of the ICC (.75) and 95% CI (.61 to .84) do not support the reliability of this component of the SFMPQ for use in this population of OA patients. The low ICC implies either a lack of stability of the current pain being evaluated, or a lack of stability of the measure. In this study, the current pain component evaluated the pain experience at a point in time, whereas, the other components of the SFMPQ evaluated the pain experienced over the previous 30 days. Other studies have demonstrated that patients' point estimation of current pain is less reliable than the recall of usual or average pain [51]. Therefore, this seems to suggest that it is the stability of the measure that is unreliable, rather than the stability of the current pain.

Alternatively, the current pain report might be expected to differ at the two time-points, as factors such as medication and functional activities prior to completing the questionnaire were not controlled. Such pain variations do not reflect a change in the overall stability of the pain, but are a characteristic of the mechanical nature of the pain associated with OA. This may be considered to be a characteristic of the current research design and not a true reflection on the stability of this component of the measurement tool. Deyo et al [41] consider that such variability may be related to regression to the mean. Further research is required to more fully investigate each assertion.

The low ICC of .75 reflects a clinical minimum detectable change of 1.4 in the current pain dimension. Thus, a recorded change of two points on the current pain score may be interpreted as a clinical change in pain, which may be clinically acceptable for some applications.

4.3 Completion Problems

It is evident from the data in Table 1 that the SFMPQ as presented to the sample was associated with completion problems, despite the modifications that were made following the pilot study, in which 57% (n=12) demonstrated completion problems. Many of the completion errors were minor and simple clarification of ambiguous responses ensured that the data could be used in all but 7% (n=5) of the sample. Such a high number of completion errors may reflect poor face validity of the SFMPQ, but if so, a low response rate would also be expected. This was not the case, as a 92% response rate was achieved. The SFMPQ response format may have resulted in the completion problems, but the same format (discrete tick box responses) was adopted for the supplementary questionnaire without similar occurrences.

Lack of familiarity with the descriptors offered, or the absence of words that the individuals would use to describe their pain, may also have affected the completion of the SFMPQ. A study evaluating choice of descriptors by patients with rheumatoid arthritis in Manchester, UK, demonstrated a disparity between those offered by the MPQ and those identified by the patients [52]. Papageorgiou's North of England population is perhaps more comparable to the study population than the Canadian population, upon whom the SFMPQ's and MPQ's development were based, so supporting the supposition that unfamiliar descriptors on the SFMPQ may affect its completion. Verbal feedback obtained, whilst clarifying the responses, over the telephone identified that some descriptors were unfamiliar and, also, identified that the written instructions lacked clarity.

The SFMPQ written instructions were brief, and the difference in completion error rates between the recruitment methods supports the suggestion that the instructions lacked clarity. Subjects recruited by post received only the SFMPQ with the integral instructions; 75% (n=15) made errors on completion. Whereas, the subjects recruited in clinic also received a verbal explanation and completed a practice questionnaire; 27% (n=14) made errors on completion. Although the verbal explanation and practice certainly assisted completion, several patients still had problems.

Increasing age has been associated with problems in self-report questionnaires [53]. This is evidenced in this study by the statistically significant difference in the age of the subjects with and without completion problems, and by the correlation of age and greater absolute difference in the sensory component scores. Both findings indicate that increasing age is associated with inferior utility of the SFMPQ.

The cause of the completion problems is likely to be a combination of all the sources of error identified. To ensure the clinical utility of the SFMPQ as a self-report instrument, the rate of completion errors needs to be reduced substantially, otherwise it seems inappropriate for this measure to be used as a self-report measure in large scale pragmatic trials.

4.4 Bias Investigations

The visual estimation and statistical analysis of bias, indicated that the method of recruitment did not affect the reliability of adequately completed questionnaires. Whilst the postal method of recruitment resulted in more SFMPQ completion errors, it did not affect the variability in the SFMPQ responses.

The statistically significant association between increasing age and a greater difference between the sensory pain scores across the two time-points, suggests that age may affect the reliability of the SFMPQ. This finding has implications for the generalizability of the results, particularly for more elderly populations.

4.5 Study flaws and areas for further research

A key flaw in this study may be considered to be the lack of control regarding the SFMPQ completion. There was no certainty that the subject, and not a relative or friend, completed the SFMPQ, at both or either time-points. There was no artificial control of environmental factors, time of day, previous activity levels or location, all of which may affect pain recall and the SFMPQ completion [54]. This lack of 'control' however, reflects the way in which research utilising self-report questionnaires is undertaken.

Having established the reliability and the clinical minimal detectable change of the SFMPQ, a principal area for further research would be to establish the responsiveness of the SFMPQ to changing pain in the same research population.

Further work is required to establish SFMPQ reliability in different populations, and the validity of the pain descriptors for a British population.

4.6 Conclusion

The results of this study are important, as no other published work has satisfactorily established the test-retest reliability of the SFMPQ. The SFMPQ total score, sensory, affective and average pain components all demonstrated excellent reliability, when the subjects were followed-up and questionnaire responses clarified. The current pain component's reliability was not supported by the ICC and the 95% CI, but may still be clinically useful as it's clinical minimum detectable change was 1.4 on a 6 point descriptor scale. Problems were identified in ensuring subjects could adequately complete the SFMPQ. It is suggested that further development of the instructions may assist the completion of the SFMPQ. These results apply only to the population from which the sample was drawn, patients with OA awaiting primary hip or knee joint replacement surgery. Any generalization of the results must be undertaken with caution, especially with regard to more elderly populations and where telephone follow-up is not available.

There were no conflicts of interest represented.

80 consecutive subjects referred

58 subjects contacted in clinic

22 subjects contacted by post, after a clinic attendance

2 refused consent

2 returned only 1 SFMPQ

3 were excluded:
1 was partially sighted,
2 had rheumatoid arthritis

2 failed to respond to letters

51 subjects returned both SFMPQ's

20 subjects returned both SFMPQ's

71 subjects in sample

14 subjects excluded:
5 made non-clarifiable errors on SFMPQ completion
9 had unstable, reported pain status

57 subjects in final analysis
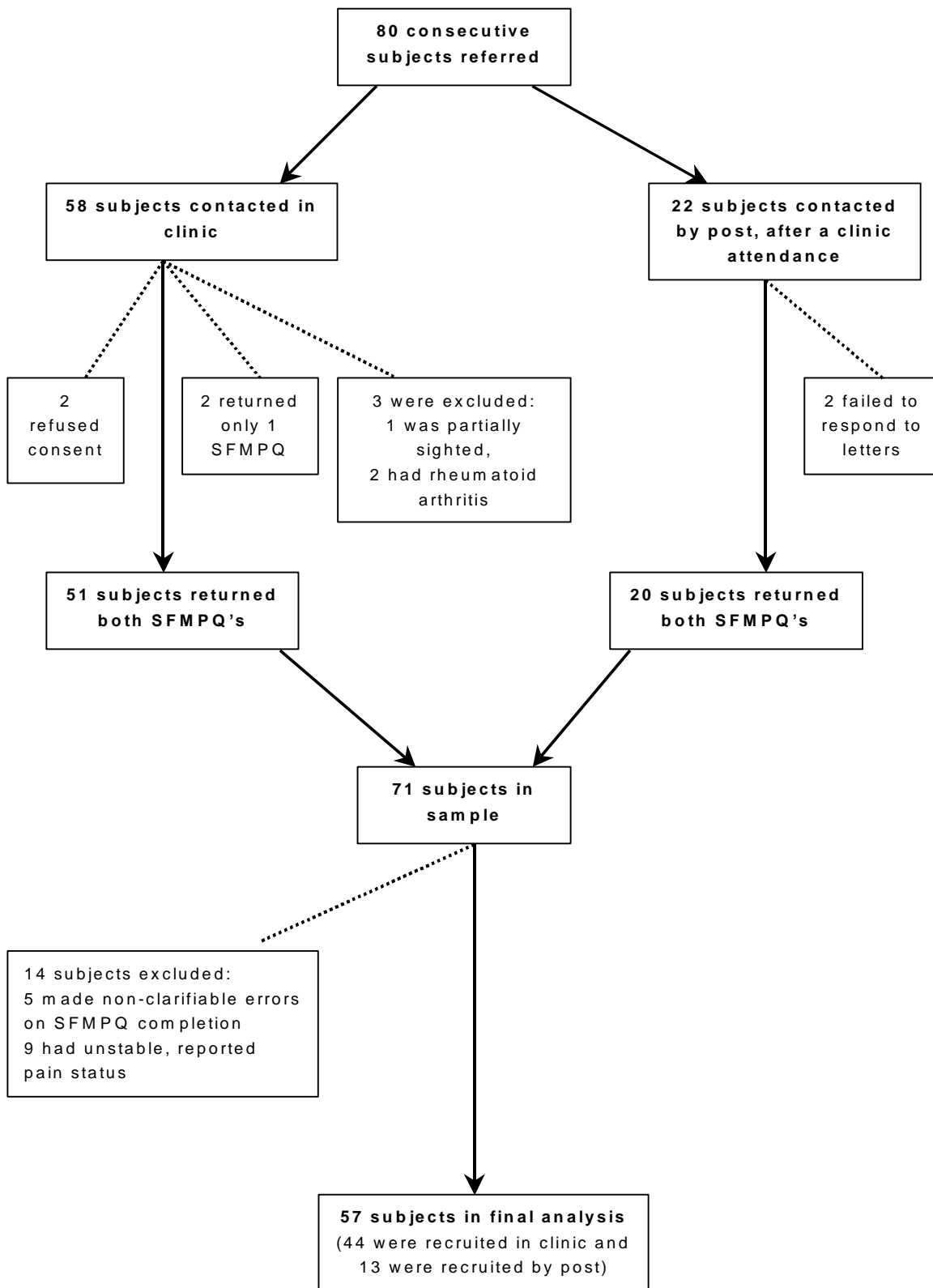(44 were recruited in clinic and 13 were recruited by post)

Figure 1: Flow chart summarizing subject recruitment and loss to follow-up

Date completed: Patient Hospital Number:

Short Form McGill Pain Questionnaire

Please indicate in the boxes below the type of pain(s), if any, that you have experienced in the previous 30 days in relation to your injury/disability.

	None	Mild	Moderate	Severe
Throbbing	☐	☐	☐	☐
Shooting	☐	☐	☐	☐
Stabbing	☐	☐	☐	☐
Sharp	☐	☐	☐	☐
Cramping	☐	☐	☐	☐
Gnawing	☐	☐	☐	☐
Hot burning	☐	☐	☐	☐
Aching	☐	☐	☐	☐
Heavy	☐	☐	☐	☐
Tender	☐	☐	☐	☐
Splitting	☐	☐	☐	☐
Tiring\Exhausting	☐	☐	☐	☐
Sickening	☐	☐	☐	☐
Fearful	☐	☐	☐	☐
Punishing\Cruel	☐	☐	☐	☐

MARK ON THIS LINE THE **AVERAGE INTENSITY** OF THE PAIN YOU SUFFER

NO PAIN ⌊_____⌋ WORST
POSSIBLE PAIN

CURRENT PAIN

i.e. the pain you are experiencing at this present moment

Please tick **one** box only

No pain	☐	Distressing	☐
Mild	☐	Horrible	☐
Discomforting	☐	Excruciating	☐

Melzack 1984

Figure 2: Short-Form McGill Pain Questionnaire [30]

- Allocation of a rank score to each pain descriptor (0 - none; 1 - mild; 2 - moderate and 3 - severe).

- The *sensory* descriptors (the first 11) are summed to give a sensory score. Maximum possible score is 33.

- The *affective* descriptors (the last 4) are summed to give an affective score. Maximum possible score is 12.

- *Total pain score* by summing *sensory* and *affective scores.* Maximum possible score is 45.

- The *VAS average pain scale* is scored continuously 0 –10, to the nearest mm, along a 10cm line. Maximum possible score is 10.0.

- Allocation of a rank score to the *current pain intensity* verbal rating scale descriptors (0 - no pain; 1 - mild; 2 - discomforting; 3 - distressing; 4 - horrible; and 5 - excruciating). Maximum possible score is 5.

Figure 3: Scoring the SFMPQ [30]

1. ICC (1,1) = $\underline{BMS - WMS}$

           BMS + (k-1) WMS

                                                                                                      .

2. SEM = $\sqrt{WMS}$

3. CoR = $1.96\sqrt{2}WMS$

         = 2.77SEM

where:

BMS is between subjects mean square

WMS is within subjects mean square (taken from the one-way ANOVA,

residual source of variation)

k is the number of measurements per subject

SEM is standard error of measurement

CoR is coefficient of repeatability

**Figure 4: Equations used to calculate the intraclass correlation coefficient,
the standard error measurement and the coefficient of repeatability for the
data analyses.**

8

6

Level of the CoR

4

2

0

-2

0   10   20   30   40

Mean total of SFMPQ scores (test 1 and test 2)

8

6

95% limit of agreement

4

2

mean difference

0

-2

-4

95% limit of agreement

-6

-8

0   10   20   30   40

Mean total of SFMPQ scores (test 1 and test 2)
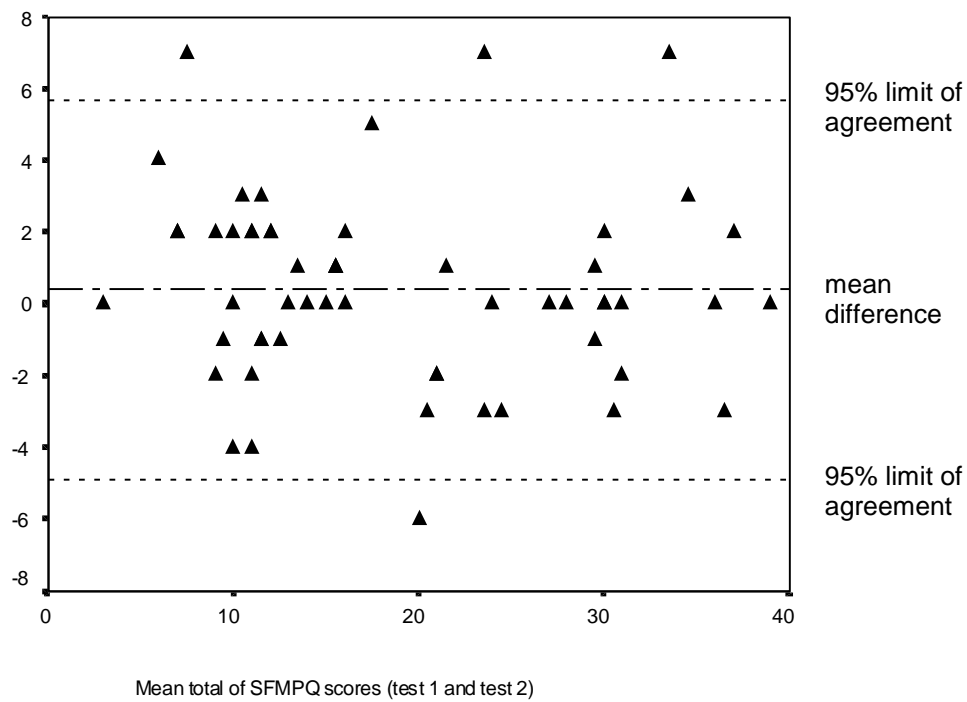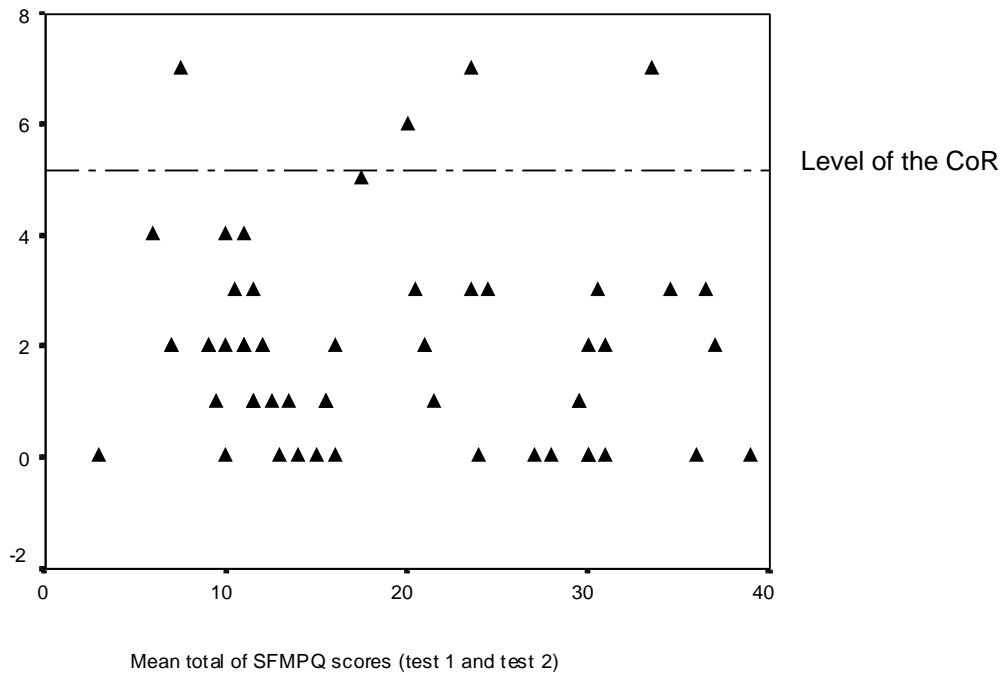
Figure 5: Plots of the absolute difference and the mean difference of the mean of the two scores (test 1 and 2) against the mean of the total score of the SFMPQ

The absolute difference is the difference between the scores ignoring the minus sign. The mean difference is the difference between the scores taking into account the minus sign, the direction of the difference.

23

Table 1: Summary of errors made by the sample on completion of the SFMPQ

Errors in completion	Subjects recruited in clinic (n=51) n (%)	Subjects recruited by post (n=20) n (%)	**Total sample (n=71)** n (%)
No errors made in completion of SFMPQ	**37 (72.5)**	**5 (25.0)**	**42 (59.1)**
Multiple errors / not easily clarified over the phone (subjects excluded)	**1 (2.0)**	**4 (20.0)**	**5 (7.0)**
Omitted VAS for average pain (subjects included, except for this component)	**3 (5.9)**	**2 (10.0)**	**5 (7.0)**
Errors in completion of the SFMPQ that could be clarified (subjects included)	**10 (19.6)**	**9 (45.0)**	**19 (26.8)**
Failed to indicate in the 'none' box if the descriptor did not apply	5 (9.8)	8 (40.0)	**13 (18.3)**
Minor error; missed out 1 or 2 descriptors	4 (7.8)	1 (5.0)	**5 (7.0)**
Omitted VAS for average pain and only specified their descriptor	1 (2.0)*	0 (0.0)	**1 (1.4)***

*The rank score of the descriptors that did not apply to this subject's pain, was clarified as 'none' but his average pain VAS score could not be clarified.

Table 2: The differences between the SFMPQ scores over the two time-points

	Mean scores times 1 and 2	Mean difference over 2 time points*	95% CI of mean difference	Absolute difference over 2 time points	95% CI of absolute difference	95% limits of agreement
Sensory component (n=57)	14.36	-0.04	-0.58 to 0.69	1.54	1.09 to 2.00	**-4.66 to 4.59**
Affective component (n=57)	4.54	0.38	-0.01 to 0.78	1.00	0.72 to 1.28	**-2.48 to 3.24**
Total sensory and affective score (n=57)	**18.90**	**0.40**	**-0.30 to 1.11**	**1.95**	**1.46 to 2.43**	-4.90 to 5.71
Average pain (cm) (n=52)	6.65	-0.12	-0.33 to 0.08	0.58	0.25 to 0.56	**-1.6 to 1.35**
Current pain (n=57)	2.66	-0.02	-0.22 to 0.18	0.40	0.45 to 0.71	**-1.47 to 1.44**

CI – confidence interval

*The negative values indicate that the scores were higher at time-point 2.

Table 3: Results for test-retest reliability statistical analyses

	ICC (95% CI)	SEM	CoR
Sensory component (n=57)	**.95** (.92 to .97)	1.64	**4.54**
Affective component (n=57)	**.88** (.81 to .93)	1.01	**2.80**
Total sensory and affective score (n=57)	**.96** (.94 to .98)	1.87	**5.18**
Average pain (n=52)	**.89** (.82 to .94)	0.52	**1.44**
Current pain (n=57)	**.75** (.61 to .84)	0.51	**1.41**

ICC - intraclass correlation coefficient   CI - confidence interval

SEM - standard error of the measurement   CoR - coefficient of repeatability

Reference List

1.   Simmonds MJ. Pain and disability: a challenge for patients and practitioners. *Physiotherapy Theory and Practice*. 1997; 13: 1-2.

2.   Wright A. Editorial. *Man Ther*. 1999; 4: 185-6.

3.   Main CJ, Watson PJ. Psychological aspects of pain. *Man Ther*. 1999; 4: 203-15.

4.   Jadad AR, McQuay HJ. The measurement of pain. In: Pynsent P, Fairbank J, Carr A, eds. *Outcome Measures in Orthopaedics*: Butterworth Heinemann, 1993: 16-29.

5.   Keefe FJ, Caldwell DS, Qeen K, et al. Osteoarthritic knee pain: a behavioural analysis. *Pain*. 1987; 28: 309-21.

6.   Melzack R, Katz J. Pain measurement in persons in pain. In: Melzack R, Wall P, eds. *Textbook of Pain*. 3rd ed. Edinburgh: Churchill Livingstone, 1994: 337-51.

7.   McDowell I, Newell C, *Measuring Health - A Guide to Rating Scales and Questionnaires*, Oxford: Oxford University Press. 1996.

8.   Rothstein JM. Measurement and clinical practice: theory and application. In: Rothstein JM, ed. *Measurement in Physical Therapy*. Edinburgh: Churchill Livingstone, 1985: 1-46.

9.   Williams ACdeC, Davies HTO, Chandury Y. Simple pain rating scales hide complex idiosyncratic meanings. *Pain*. 2000; 85: 457-63.

10.   Bowling A, *Measuring Disease: A Review of Disease-Specific Quality of Life Measurement Scales*, Buckingham: Open University Press. 1995.

11.   Piotrowski C. Assessment of pain: a survey of practicing clinicians. *Percept Mot Skills.* 1998; 86: 181-2.

12.   Melzack R, Katz J. The McGill Pain Questionnaire: appraisal and current status. In: Turk DC, Melzack R, eds. *Handbook of Pain Assessment.* London: The Guilford Press, 1992: 152-68.

13.   Grushka M, Sessle B. Applicability of the McGill pain questionnaire to the differentiation of 'toothache' pain. *Pain.* 1984; 19: 49-57.

14.   Burckhardt CS. The use of the McGill pain questionnaire in assessing arthritis pain. *Pain.* 1984; 19: 305-14.

15.   Huskisson EC. Measurement of pain. *J Rheumatol.* 1982; 9: 768-9.

16.   Reading AE. The internal structure of the McGill Pain Questionnaire in dysmenorrhea patients. *Pain.* 1979; 7: 353-8.

17.   Reading AE. A comparison of the McGill Pain Questionnaire in chronic and acute pain. *Pain.* 1982; 13: 185-92.

18.   Melzack R, Torgerson WS. On the language of pain. *Anesthesiology.* 1971; 34: 50-9.

19.   Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain.* 1975; 1: 277-99.

20.   Holroyd KA, Holm JE, Keefe FJ, et al. A multi-centre evaluation of the McGill Pain Questionnaire: results from more than 1700 chronic pain patients. *Pain.* 1992; 48: 301-11.

21. Lowe NK, Walker SN, McCallum RC. Confirming the theoretically structure of the McGill Pain Questionnaire in acute clinical pain. *Pain*. 1991; 46: 53-60.

22. Pearce J, Morley S. An experimental investigation of the construct validity of the McGill Pain Questionnaire. *Pain*. 1989; 39: 115-21.

23. Turk DC, Rudy TE, Salovey P. The McGill Pain Questionnaire reconsidered: confirming factor structures and examining appropriate uses. *Pain*. 1985; 21: 385-97.

24. Dworkin SF, Whitney CW. Relying on objective and subjective measures of chronic pain: guidelines for use and interpretation. In: Turk DC, Melzack R, eds. *Handbook of Pain Assessment*. New York: The Guilford Press, 1992: 429-46.

25. Love A, Leboeuf C, Crisp TC. Chiropractic chronic low back pain sufferers and self-report assessment methods. Part 1. A reliability study of the Visual Analogue Scale, the Pain Drawing and the McGill Pain Questionnaire. *J Manipulative Physiol Ther*. 1989; 12: 21-5.

26. Haas M. Statistical methodology for reliability studies. *J Manipulative Physiol Ther*. 1991; 14: 119-32.

27. Keating J, Matyas T. Unreliable inferences from reliable measurements. *Aust J Physiother*. 1998; 44: 5-10.

28. Graham C, Bond SS, Gerkovich MM, et al. Use of the McGill Pain Questionnaire in the assessment of cancer pain: replicability and consistency. *Pain*. 1980; 8: 377-87.

29.   Chapman CR, Casey KL, Dubner R, et al. Pain measurement: an overview. *Pain*. 1985; 22: 1-31.

30.   Melzack R. The Short-Form McGill Pain Questionnaire. *Pain*. 1987; 30: 191-7.

31.   Serrao JM, Marks RI, Morley SJ, et al. Intrathecal midazolam for the treatment of chronic low back pain: a controlled comparison with epidural steroid a pilot study. *Pain*. 1992; 48: 5-12.

32.   Stelian J, Gil I, Habot B, et al. Improvement of pain and disability in elderly patients with degenerative osteoarthritis of the knee treated with narrow-band light therapy. *J Am Geriatr Soc*. 1992; 40: 23-6.

33.   Zalon M. Comparison of pain measures in surgical patients. *J Nurs Meas*. 1999; 7: 135-52.

34.   Dechow E, Davies R, Carr A, et al. A randomised, double-blind, placebo-controlled trial of sclerosing injections in patients with chronic low back pain. *Rheumatology*. 1999; 38: 1255-9.

35.   Frost S, Grossfeld S, Kirkley A, et al. The efficacy of femoral nerve block in pain reduction for outpatient hamstring anterior cruciate ligament reconstruction: a double-blind, prospective, randomised trial. *Arthroscopy*. 2000; 16: 243-8.

36.   Bonaiuti D, Fontanella G. The affective dimension of low-back pain: its influence on the outcome of back school. *Arch Phys Med Rehab*. 1996; 77: 1239-42.

37.	Dudgeon D, Raubertas RF, Rosenthal SN. The Short Form McGill Pain Questionnaire in chronic cancer pain. *J Pain Symptom Manage*. 1993; 8: 191-5.

38.	Bland MJ, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1: 307-10.

39.	Gronblad M, Lukinmaa A, Konttinen YT. Chronic low back pain: intercorrelation of repeated measures for pain and disability. *Scand J Rehabil Med*. 1990; 22: 73-7.

40.	Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clinical Rehabil*. 1998; 12: 187-99.

41.	Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures - statistics and strategies for evaluation. *Controlled Clinical Trials*. 1991; 12: 142S-58S.

42.	Behi R, Nolan M. Reliability: consistency and accuracy in measurement. *British Journal of Nursing*. 1995; 4: 472-5.

43.	Streiner DL, Norman GR, *Health Measurement Scales*. 2nd ed, Oxford: Oxford University Press. 1995.

44.	Bruton A, Conway JH, Holgate ST. Reliability: what is it, and how is it measured? *Physiotherapy*. 2000; 86: 94-9.

45.	Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998; 17: 101-10.

46.	Heft MW, Parker SR. An experimental basis for revising the graphic rating scale for pain. *Pain*. 1984; 19: 153-61.

47.   Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979; 86: 420-8.

48.   Chinn S. Repeatability and method comparison. *Thorax*. 1991; 46: 454-6.

49.   Bland M, *An Introduction to Medical Statistics*. Vol. 3rd edition, Oxford: Oxford University Press. 2000.

50.   British Standards Institution. Precision of test methods. Part 1: Guide for determination and reproducibility far a standard test method. . BS5497, part1. British Standards Institution. 1979: 1-22.

51.   Jensen MP, McFarland A. Increasing the reliability and validity of pain intensity measurement in chronic pain patients. *Pain*. 1993; 55: 195-203.

52.   Papageorgiou AC, Badley EM. The quality of pain in arthritis: the words patients use to describe overall pain and pain in individual joints at rest and on movement. *J Rheumatol*. 1989; 16: 106-12.

53.   Hayes V, Morris J, Wolfe C, et al. The SF-36 Health Survey Questionnaire: is it suitable for use with older adults? *Age Ageing*. 1995; 24: 120-5.

54.   Roche PA. Memory for pain: a physiotherapy issue. *Physiotherapy Theory and Practice*. 1993; 9: 87-90.