TITLE:

The Brighton musculoskeletal Patient Reported Outcome Measure (BmPROM): an assessment of validity, reliability and responsiveness.

ABSTRACT:

Background: In response for the need of a freely available, stand-alone, validated outcome measure for use within musculoskeletal physiotherapy practice, sensitive enough to measure clinical effectiveness, we developed a musculoskeletal patient reported outcome measure.

Objectives: This study examined the validity and reliability of the newly developed Brighton musculoskeletal Patient Reported Outcome Measure (BmPROM) within physiotherapy outpatient settings.

Methods: Two hundred and twenty four patients attending physiotherapy outpatient departments in South East of England with a musculoskeletal condition participated in this study. The BmPROM was assessed for user friendliness (rated feedback, *n*=224), reliability (internal consistency and test-retest reliability, *n*=42), validity (internal and external construct validity, *n*=224) and responsiveness (internal, *n*=25).

Results: Exploratory factor analysis indicated that a two-factor model provides a good fit to the data. Factors were representative of 'Functionality' and 'Wellbeing'. Correlations observed between the BmPROM and SF-36 domains provided evidence of convergent validity. Reliability results indicated that both subscales were internally consistent with alphas above the acceptable limits for both 'Functionality' (α = .85, 95% CI = .81- .88) and 'Wellbeing' (α = .80, 95% CI = .75- .84). Test-retest analyses (n= 42) demonstrated a high degree of reliability between 'Functionality' (ICC= .84; 95% CI = .72-.91) and 'Wellbeing' scores (ICC= .84; 95% CI = .72- .91). Further examination of test-retest reliability through the Bland-Altman analysis demonstrated that the difference between 'Functionality' and

'Wellbeing' test scores did not vary as a function of absolute test score). Large treatment effect sizes were found for both subscales (Functionality d = 1.10; Wellbeing 1.03).
Conclusion: The BmPROM is a reliable and valid outcome measure for use in evaluating physiotherapy treatment of musculoskeletal conditions.

Key words:
Musculoskeletal disorders, physiotherapy, validity, reliability, patient centred outcome measure

MAIN TEXT

Introduction

Musculoskeletal conditions in the UK are common, and are a major cause of ill-health, pain and physical disability. Almost one third of all primary care consultations are for musculoskeletal (MSK) conditions (30%), which are the most common reason for repeat consultations with a GP (DoH, 2006). A wide range of outcome measures are currently used to assess the clinical effectiveness of the management of these conditions. These outcome measures tend to be focused on a specific condition (Evans and Kagan, 1986; Fairbank et al., 1980; Roland and Morris, 1983; Ruta et al., 1994; Bolton and Breen, 1999; Hawker et al., 2011; Evans et al., 2004; Tan et al., 2004), or measure the generic health status of the patient (Brodin et al., 2014; Samsson and Larsson, 2015; Ware and Sherbourne, 1992; Ware et al., 2000).

At present there are limits to the validated outcome measures available for use within MSK physiotherapy practice. More recently tools have been designed to be used in conjunction with other measures (Hill et al., 2015). The need for a stand-alone outcome measure which is sensitive enough to measure clinical effectiveness in physiotherapy MSK settings, has

clinical utility and can be used universally across the full range of MSK conditions has been widely recognised (Kyte et al., 2015; The King's Fund, 2010). In response, we developed an MSK patient reported outcome measure, which is independent of other measures.

The aim of this study was to investigate the psychometric properties (reliability and validity) of the Brighton musculoskeletal Patient-Reported Outcome Measure (BmPROM) in outpatient physiotherapy departments. This paper outlines both the development and validation process of the BmPROM.

Methods

The BmPROM was devised following a detailed literature search and using a consensus nominal group involving senior MSK clinicians, and through patient focus groups. An iterative process was adopted with patients to ensure that the BmPROM captured all relevant domains and that all items were worded correctly. The BmPROM was modified accordingly and then piloted for acceptability, face validity and feasibility by a local private physiotherapy practice over a 3-month period with patients with MSK conditions.

The BmPROM is an 8-item questionnaire designed to include the major areas of importance to people with MSK problems including quality of life, activities of daily living, leisure/social activities, pain, medication, sleep, anxiety, and depression levels. Patients rate their response to each question using an 11-point Likert scale (ranging from 0 to 10) in relation to the problem that has brought them to seek physiotherapy. The wording response provided alongside the Likert scale reflected the focus of each question. In some questions the meaning of the response scale has been reversed in order to minimise the selection of the maximum score for each response. The scoring for these items was then reversed prior to the data analyses. The BmPROM has been designed to be as user-friendly as possible, for

patients' self-completion, taking approximately 5 minutes to complete. The BmPROM is available online in appendix 1.

At the end of the questionnaire, in order to provide a prompt for discussion between the physiotherapist and the patient, there is an optional section where the patient can list their expectations of both their physiotherapy treatment and of the physiotherapist. The optional data is not analysed in this study.

Ethical approval for all procedures was granted by the NHS NRES Committee North East (11/NE/0307).

Validity was established in a population of newly-referred patients attending physiotherapy departments in five NHS Trusts. Patients meeting the inclusion criteria were invited to participate. The inclusion criteria were: patients aged 16+ seeking/referred to physiotherapy treatment for a MSK condition; and fluent in spoken and written English. Participants were asked to complete the BmPROM and the Short Form-36 Health Evaluation Survey (SF-36) (Ware and Sherbourne, 1992) in the waiting room on their first visit to the physiotherapy department prior to treatment. Implied consent was obtained through the completion of the BmPROM and SF-36 which was returned to the site administrator.

Test-retest reliability was ascertained in a separate sample of new patients referred to one NHS Trust physiotherapy department. Patients who met the inclusion criteria described previously were invited to participate, completing the BmPROM on two occasions prior to commencing treatment. The site administrator sent newly referred patients a study information pack with participant information on the study and the BmPROM 10 days prior to attending the physiotherapy department. As stated in the study information provided to

patients, implied consent was obtained through the return of the BmPROM which was returned to the site administrator using a stamped addressed envelope. Consenting patients were asked to repeat completion of another BmPROM on their first visit to the physiotherapy department prior to commencing treatment.

The capacity of the instrument to measure change statistically (sensitivity to change) and the capacity of the instrument to detect clinically relevant changes (responsiveness) were established from participating patients in two physiotherapy departments. Patients were invited to complete the BmPROM on two occasions, once before commencement of treatment and again following their final treatment session.

To assess the user-friendliness of the BmPROM, participants were asked to provide feedback, and to rate how easy they found the questionnaire to complete. Responses were recorded on a 10-point Likert scale ranging from 1, 'very difficult', to 10, 'very easy'. Patients were also asked the following, 'Do you feel that the questionnaire covered all areas that are of interest to you in relation to the problem that brought you to physiotherapy today (please circle your response)?'. Patients could respond by circling 'Yes' or 'No'. Those patients who indicated 'No' were asked, 'If no, please write in the space below any additional areas that are of interest to you'.

Data Analysis

All analyses were performed using IBM SPSS Statistics for Windows (version 22.0). A preliminary item-level descriptive analysis was performed to examine the distribution of all items. Standard psychometric methods were used to evaluate the reliability and validity of the BmPROM as described below (Streiner and Norman, 2008).

Exploratory factor analysis (EFA) was used to examine internal construct validity (Tabachnick and Fidell, 2007). A number of well-recognised criteria for the factorability were used. Firstly, Pearson correlations between all items were examined with correlations of at least .30 suggesting reasonable factorability. Secondly, the Kaiser-Meyer-Olkin measure of sampling accuracy was examined with levels above the recommended value of 0.5 considered acceptable (Byrne, 2001). Third, Bartlett's test of sphericity was performed with a significant result indicating good factorability.

Model parameters were estimated using the Maximum Likelihood extraction method and oblique rotations. The choice of extraction method allows for the computation of a wide range of indices of the goodness of fit of the model and permits statistical significance testing of factor loadings and correlations among factors and the computation of confidence intervals (Fabrigar et al., 1999). The number of factors retained was determined by the Kaiser criterion (Kaiser, 1960) and a visual inspection of the scree plot.

Internal consistency of the measure was assessed using Cronbach's alpha. Alpha levels of .70 and above were considered acceptable (Nunnally, 1978). Test-retest reliability was analysed using the paired *t*-test, and intraclass correlation coefficient (ICC) (one-way mixed ANOVA model, adjusted for a single measure) (Shrout and Fleiss, 1979). Test-retest reliability was also assessed using the Bland-Altman method of assessing agreement between the test and retest measures (Bland and Altman, 1986). A plot of differences against the mean values of the test and retest was used to provide an indication of fixed bias, proportional bias and heteroscedastic bias.

External construct validity was examined by means of convergent and discriminant validity. This was performed through correlation of the BmPROM scores with the SF-36 scores. We

hypothesized that scales on the BmPROM that were considered to be conceptually related would be associated moderately ($r \geq 0.30$) to strongly ($r \geq 0.60$) (Campbell and Fiske, 1959) with those of the SF-36. Similarly, weaker associations were expected for scales between the two measures that were not considered to be conceptually related.

The responsiveness (Beaton et al., 2001) of the BmPROM was examined using a paired *t*-test, Cohen's *D* and standardised response mean (SRM), which compares change to the standard deviation of change (Cohen, 1988). A SRM greater than .80 is interpreted as a large effect, a SRM greater than 0.50 is interpreted as a moderate effect and a SRM greater than 0.2 is interpreted as a small effect (Cohen, 1988). The standard error of measurement (SEM) was calculated as an indicator of the amount of variation in the BmPROM due to measurement error based on the following: baseline score standard deviation multiplied by the square root of 1-ICC.

Results

Participant characteristics for each study population are displayed in Table 1.

In our study sample of 224 patients, the mean scores of each of the items were between 4.26 and 5.77. These results therefore found no suggestion of floor or ceiling effects which can present when a considerable proportion of patients score the maximum or minimum scores on a scale rendering them unable to discriminate between subjects at either extreme of the scale (Table 2). The skewness and kurtosis were within the normal ranges for assuming distribution and this was supported by examination of the histograms.

Initially, factorability of the eight items was examined. A number of well-recognised criteria for the factorability of the correlation were used. Firstly, item correlations ranged from .27 to .74 suggesting reasonable factorability. Secondly, the Kaiser-Meyer-Olkin measure of sampling accuracy was .85, which is above the recommended value of .50 (Byrne, 2001). Third, Bartlett's test of sphericity was significant (χ^2 (28) = 805.37, p < .001).

The initial eigenvalue (4.17) showed that the first factor accounted for 52.11% of the variance and the second (1.06) explained a further 13.25% of the variance. The two-factor solution, which explained a total of 65.36% of the variance, had at least three items per factor, had common conceptual meaning and measured different constructs. These conclusions were also supported through the examination of the scree plot which indicated a levelling off of eigenvalues after two factors, and the Kaiser Criterion. All items had primary loadings over .40 and three factors had a cross loading above .30. However two of these items had strong primary loadings (i.e. >.50) and one had a moderate primary loading (i.e. >.44). The factor loading matrix is presented in Table 2.

Two subscales were formed based on composite scores for the two factors, labelled as 'Functionality' and 'Wellbeing'. Subscales were created based on the means of the items that loaded primarily on each factor (see Table 2). Higher scores indicated an increased amount of Functionality and Wellbeing. Descriptive statistics for the newly-created subscales are presented in Table 3. The skewness and kurtosis were within the normal range for assuming distribution and this was supported by examination of the histograms.

The Cronbach's alphas for each of the scales were above the acceptable limits for both Functionality (α = .85, 95% CI = .81- .88) and Wellbeing (α = .80, 95% CI = .75- .84). No

substantial increases in the alphas for either of the scales could have been achieved by eliminating items.

Significant and high correlations occurred between the Functionality subscale and domain scores of the SF-36 that were considered to be conceptually related (i.e. the Physical function score, $r = .66$, $p < .01$; the Role function score, $r = .60$, $p < .01$ and the Social function score, $r = .70$, $p < .01$).

Significant and high correlations also occurred between the Wellbeing subscale and the Bodily pain score of the SF-36 ($r = .69$, $p < .01$), a domain which is conceptually related. As expected, a significant but lower correlation was observed between the Wellbeing subscale and the Vitality scores of the SF-36 ($r = .44$, $p< .01$), a domain that was considered to be a partially conceptually related. These results are shown in Table 4.

Test-retest reliability data was collected from 42 patients. The mean and standard deviation for each occasion are shown in Table 5. The results of the *t*–test indicated a small but statistically significant difference between Functionality measurements ($t_{41} = 2.06$, $p < .05$, $d = 0.18$) and a non-significant difference between Wellbeing measurements ($t_{41} = 0.06$, $p > .05$, $d = 0.01$). High test-retest reliability was found between Functionality measurements (*ICC*= .84; 95% *CI* = .72-.91) and between Wellbeing measurements (*ICC*= .84; 95% *CI* = .72- .91). Correlation coefficients between the mean differences and the means of the two measures were small and non-significant for Functionality ($r = .09$, $p > .05$) and Wellbeing ($r = .12$, $p > .05$) indicating that proportional bias was not present in the data. The Bland-Altman analyses estimated that the 95% limits of agreement between the T1-T2 measures ranged from -2.37 to 1.70 for mean differences in Functionality (Figure 1) and from –2.13 to 2.15 for mean differences in Wellbeing (Figure 2).

Responsiveness data (pre- and post-treatment data) were collected for 25 participants. A paired-samples *t*-test indicated that Functionality scores measured after physiotherapy (M = 7.5, SD = 1.73) were significantly higher with large effect size (t = -5.53, p < .001, d = -1.09) than scores measured before physiotherapy treatment (M = 5.45, SD = 2.04). A paired-samples *t*-test indicated that Wellbeing scores measured after physiotherapy treatment (M = 6.67, SD = 2.30) were significantly higher with large effect size (t = -5.15, p < .001, d = -1.20) than scores measured before physiotherapy treatment (M = 4.33, SD = 1.60). Positive SRM scores with a large effect size were calculated for both subscales: Functionality (SRM = 1.10) and Wellbeing (SRM = 1.03). The SEM was 0.75 for Functionality and 0.87 for Wellbeing.

With regards to user friendliness, from the 224 participating participants, the majority reported finding the BmPROM easy to complete (median score = 9, interquartile range = 2). Ninety percent of patients felt the questionnaire covered all areas of interest to them in relation to their musculoskeletal problem. The responses from the remaining 10% who felt that all their interests were not covered were varied and no common theme was identified.

Discussion

This study presents the measurement properties of the BmPROM, a patient reported outcome measure for use in MSK physiotherapy practice. Overall, the measure displayed good psychometric properties for the two subscales (Functionality and Wellbeing) supporting the reliability and validity of the BmPROM as a patient-reported outcome measure. A distinct advantage of the tool over other widely used patient-reported outcome measures, is that the measure is not condition-specific and instead can be used to assess the efficacy of treatment in all MSK patients. The measure therefore provides clinicians with the opportunity to

monitor patient outcomes as well as benchmark and evidence the quality of treatment in their practice.

Examination of the underlying factor structure of the measure revealed two dimensions. The first, referred to as Functionality, related to the patient's quality of life, capacity for leisure/social activities and capacity to carry out activities of daily living. The second, referred to as Wellbeing, related to pain, medication, sleep, anxiety and depression. The distinction of these two dimensions is aligned to the definition of physiotherapy accepted by the World Confederation of Physical Therapy (WCPT) in that, according to the WCPT, physiotherapy 'provides individuals with services to develop, maintain, and restore maximum movement and functional ability throughout the lifespan' and that maximizing quality of life and movement potential 'encompasses physical, psychological, emotional, and social wellbeing' (WCPT, 2016). The capacity of practitioners to identify improvements made in regards to a patient's Functionality and Wellbeing in isolation is considered advantageous compared to standard generic PROMs and Quality of Life measures that are typically used and that they provide one overall score (e.g. EQ-5D-5L). Although such a quality could be compared to the capacity of individualised instruments (e.g. Patient Generated Index [Ruta et al, 1994], Patient Specific Function Scale [Stratford et al., 1995] and the MYMOP [Paterson, 1996]) the application of individualised measures does not tend to be favoured by policy makers in isolation given they are not standardised, and can only provide limited evidence to demonstrate the quality of care, an outcome which is a wider quality improvement agenda driven by commissioners of services (DoH, 2011).

Reliable and valid methods of assessing musculoskeletal patients' reported outcomes in physiotherapy settings are important and necessary if we are to assess the effectiveness of, and change (i.e. responsiveness) following treatment. The BmPROM demonstrated the

ability to discriminate between patients of varying severity (i.e. those at either extreme of the scale). The measure demonstrated change statistically as well as the capacity to detect clinically relevant changes.

The BmPROM demonstrated that it is both relevant to what patients consider important and user-friendly. Despite containing all aspects considered relevant, the design of the measure remained brief. This feature can be viewed as a strength in that its application in clinical practice will not over burden the patient or the clinician.

The findings from this study are very promising. The strengths of this study include the expert involvement and iterative process undertaken with patients used in the development stage of the measure and the statistical examination used to identify underlying dimensions of the measure (i.e. Functionality and Wellbeing).

Our study is not without limitations which also serve to provide direction for future research in this area. First, although the test-retest reliability evidence was acceptable, logistical difficulties concerning data collection meant that the sample size for this analysis was less than anticipated. Second, patients participated in this study on a voluntary basis; it is possible that patients with lower levels of literacy may have chosen not to participate in the study. Based upon this assumption, it is possible that our sample is not representative of all patients attending physiotherapy in this setting. This may be particularly important when the measure is used with individuals whose first language is not English or who have lower level of literacy. Third, our provision of evidence for the external validation of the BmPROM consisted of comparisons with the SF-36 (i.e. a measure of general health). A wealth of convergent and divergence evidence was provided through these analyses to support the external validation of the BmPROM. However, although these analyses provided some important findings, these

findings could be strengthened with the application of think-aloud studies and furthermore, through a Multitrait-Multimethod Matrix, as defined by Campbell and Fiske (1959) whereby construct validity is demonstrated simultaneously through the examination of the trait (i.e. Functionality/Wellbeing) through multiple methods (e.g. observation, test). Finally, examination of the internal (structural) validity of the BmPROM in this study was through the application of EFA. Although the strengths of EFA have been widely demonstrated in the psychometric literature, the approach is an exploratory technique which is used to represent observed data, and does not therefore include a formal *a priori* hypothesis testing. The use of Confirmatory Factor Analysis (CFA) in future research, a theory-driven rather than data-driven approach, to assess the internal validity of the BmPROM would allow for a more sophisticated and advantageous technique to examine and identify the underlying dimensions of the outcome measure (Tabachnick and Fidell, 2007).

In conclusion, this study has shown that the BmPROM is a valid, reliable and responsive generic clinical outcome measure for use in evaluating physiotherapy treatment of musculoskeletal conditions.

Implications for physiotherapy practice

There is widely documented need for a validated outcome measure which is sensitive enough to measure clinical effectiveness in physiotherapy MSK settings across a number of different conditions. The psychometric properties of the BmPROM have been examined and these results support its use in the assessment of the efficacy of physiotherapy treatment for people with MSK conditions.

References

Beaton, D.E., Bombardier, C., Katz, J.N., Wright, J.G. (2001). A taxonomy for responsiveness. *Journal of Clinical Epidemiology,* 54, 1204–1217.

Bland, J.M. and Altman, D.G. (1986). Statistical method for assessing agreement between two methods of clinical measurement. *Lancet*, 8,1(8476), 307-310.

Bolton, J.E. & Breen, A.C. (1999). The Bournemouth Questionnaire: a short-form comprehensive outcome measure. I. Psychometric properties in back pain patients. *Journal* of *Manipulative & Physiological Therapeutics*, 22, 503-510.

Brodin, N., Lohela-Karlsson, M., Swardh, E., Opava, C.H. (2015). Cost effectiveness of a one year coaching program for healthy physical activity in early rheumatoid arthritis. *Disability* and *Rehabilitation,* 37,757-762. doi:10.3109/09638288.2014.940429.

Byrne, B. (2001). *Structural equation modelling with AMOS: Basic concepts, applications and programming.* Mahwah, New Jersey: Erlbaum.

Campbell, D.T. and Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56: 81-105.

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, New Jersey: Lawrence Erlbaum Associates.

DoH. *The Musculoskeletal Services Framework - A joint responsibility: doing it differently.* London: Department of Health, 2006.

DoH. *Equity and excellence. Liberating the NHS*. 2011. London: Department of Health, 2011.

Evans, C.J., Trudeau, E., Mertzanis, P., Marquis. P,, Peña. B.M., Wong, J., Mayne, T. (2004). Development and validation of the Pain Treatment Satisfaction Scale (PTSS): a patient satisfaction questionnaire for use in patients with chronic or acute pain. *Pain*, 112: 254-266. DOI: 10.1016/j.pain.2004.09.005

Evans, J.H. and Kagan, A. (1986). The development of a functional rating scale to measure the treatment outcome of chronic spinal patients. *Spine,* 11: 277-281.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4: 272-299.

Fairbank, J.C., Couper, J., Davies, J.B., O'Brien, J.P. (1980). The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy,* 66: 271-273.

Hawker, G.A., Mian, S., Kendzerska, T., French, M. (2011) Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res,* 63 (S11): S240-S252. DOI:10.1002/acr.20543.

Hill, J.C., Thomas, E., Hill, S., Foster, N.E., van der Windt, D.A. (2015). Development and validation of the Keele Musculoskeletal Patient Reported Outcome Measure (MSK-PROM). PLoS ONE 10(4): e0124557. https://doi.org/10.1371/journal.pone.0124557

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement,* 20, 141-151.

Kyte, D.G., Calvert, M., van der Wees, P.J., ten Hove, R., Tolan, S., Hill, J.C. (2015). An introduction to Patient-Reported Outcome Measures (PROMs) in Physiotherapy. *Physiotherapy* 2015; 101, 119-125. DOI:10.1016/j.physio.2014.11.003

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Paterson, C. (1996). Measuring outcomes in primary care: a patient-generated measure, MYMOP, compared to the SF-36 health survey. *BMJ*, 312: 1016–1020.

Roland, M. and Morris, R. (1983). A study of the natural history of back pain, part I: development of a reliable and sensitive measure of disability in low back pain. *Spine*, 8, 141-144.

Ruta, D., Garratt, A.M., Leng, M., Russell, I.T., MacDonald, L.R. (1994). A new approach to

the measurement of quality of life: the patient generated index. *Medical Care*, 32,1109–1126.

Ruta, D.A., Garratt, A.M., Wardlaw, D., Russell, I.T. (1994). Developing a valid and reliable measure of health outcome for patients with low back pain. *Spine,* 19, 1887-1896.

Samsson, K.S. and Larsson, M.E. (2015). Physiotherapy triage assessments of patients referred for orthopaedic consultation – long term follow up of health related quality of life, pain related disability and sick leave. *Manual Therapy*, 20, 38-45. DOI:10.1016/j.math.2014.06.009

Shrout, P.E. and Fleiss, J.L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.

Stratford, P., Gill, C., Westaway, M., Binkley, J. (1995). Assessing disability and change on individual patients: a report of a patient specific measure. *Physiotherapy Canada*, 47, 258–263. https://doi.org/10.3138/ptc.47.4.258

Streiner, D.L. and Norman, G.R. (2008). *Health measurement scales: a practical guide to their development and use* (4th ed.). Oxford: Oxford University Press.

Tabachnick, B.G. and Fidell, L.S. (2007). *Using Multivariate Statistics*. Boston: Pearson Education Inc.

Tan, G., Jensen, M.P., Thornby, J.I., Shanti, B.F. (2004). Validation of the Brief Pain Inventory for chronic non-malignant pain. The Journal of Pain, 5,133-137. DOI:10.1016/j.jpain.2003.12.005

The King's Fund. (2010). *Getting the most out of PROMs.* London, The King's Fund, 2010. Available at https://www.kingsfund.org.uk/sites/files/kf/Getting-the-most-out-of-PROMs-Nancy-Devlin-John-Appleby-Kings-Fund-March-2010.pdf (accessed 02-04-2015).

Ware, J.E. and Sherbourne, C.D. (1992). The MOS 36-item short-form health survey (SF-

36). I. Conceptual framework and item selection. *Medical Care*, 30,473-483.

Ware, J.E., Kosinski, M. and Gandek, B. (2000). *SF-36® health survey: manual and interpretation guide.* Quality Metric Incorporated, Lincoln, Rhodes Island.

World Confederation of Physical Therapy (WCPT). *Policy statement: Description of physical therapy.* Available at http://www.wcpt.org/policy/ps-descriptionPT (accessed on 15-12-2016).

TABLE 1 Descriptive statistics of patient characteristics (gender, age and body site of presenting problem) for each study population

Participant characteristics	Validity study (*n*=224)	Reliability study (*n*=42)	Responsiveness (*n*=25)
Gender [n (%)]			
Female	136 (61%)	25 (59%)	16 (64%)
Male	69 (31%)	15 (36%)	9 (36%)
Not stated	19 (8%)	2 (5%)	-
Age (years)			
Mean (SD) Range	51.7 (16.8) 17-88	50.0 (18.8) 23-83	59.5 (12.4) 40-82
General body site of presenting problem [n (%)]			
Lower limb	80 (36%)	6 (14%)	10 (40%)
Upper limb	62 (28%)	9 (21%)	6 (24%)
Spine	72 (32%)	15 (36%)	4 (16%)
Other	9 (4%)	12 (29%)	5 (20%)

TABLE 2 Descriptive statistics and factor loadings for items of the BmPROM (*n*=224).

	Min	Max	M	SD	Skewness	Kurtosis	Factor loading*	
							Wellbeing	Functionality
Item 1- Quality of life	0	10	5.77	2.13	-.36	-.57	(.49)	60
Item 2-Pain	0	10	4.26	2.29	.21	-.58	.64	
Item 3 - Leisure/Social activities	0	10	5.12	2.91	-.02	-1.03		.76
Item 4- Activities of daily living	0	10	5.40	2.75	-.17	-.99		.88
Item 5- Medication	0	10	5.60	3.47	-.14	-1.36	.63	
Item 6 - Sleep	0	10	5.17	3.13	.11	-1.19	.44	(.31)
Item 7- Anxiety	0	10	4.67	2.92	.21	-.94	.69	
Item 8- Depression	0	10	5.14	3.13	.07	-1.21	.70	(.39)

**Note.* Factor loadings < .03 are suppressed. Secondary factor loadings are in brackets.

TABLE 3 Descriptive statistics for the two subscales of the BmPROM (*n*= 224).

	No. of items	M (SD)	Skewness	Kurtosis
Functionality	3	5.46 (2.30)	-.20	-.79
Wellbeing	5	4.97 (2.24)	.003	-.72

TABLE 4 Pearson's correlation coefficient as determined by comparing the two BmPROM subscales to the corresponding SF-36 subscales (*n*=224).

	SF-36 Subscales							
	Physical function	Role function	Bodily pain	General health	Vitality	Social function	Role emotion	Mental Health
Functionality	.**66****	**.60****	.61**	.31**	.50**	.**70****	.44**	.38**
Wellbeing	.54**	.43**	.**69****	.34**	.44**	.44**	**.45****	**.49****

Note: **p< .05; **p < .01*

Text in bold highlights expected strong (i.e. >.60) correlations between conceptually related constructs.

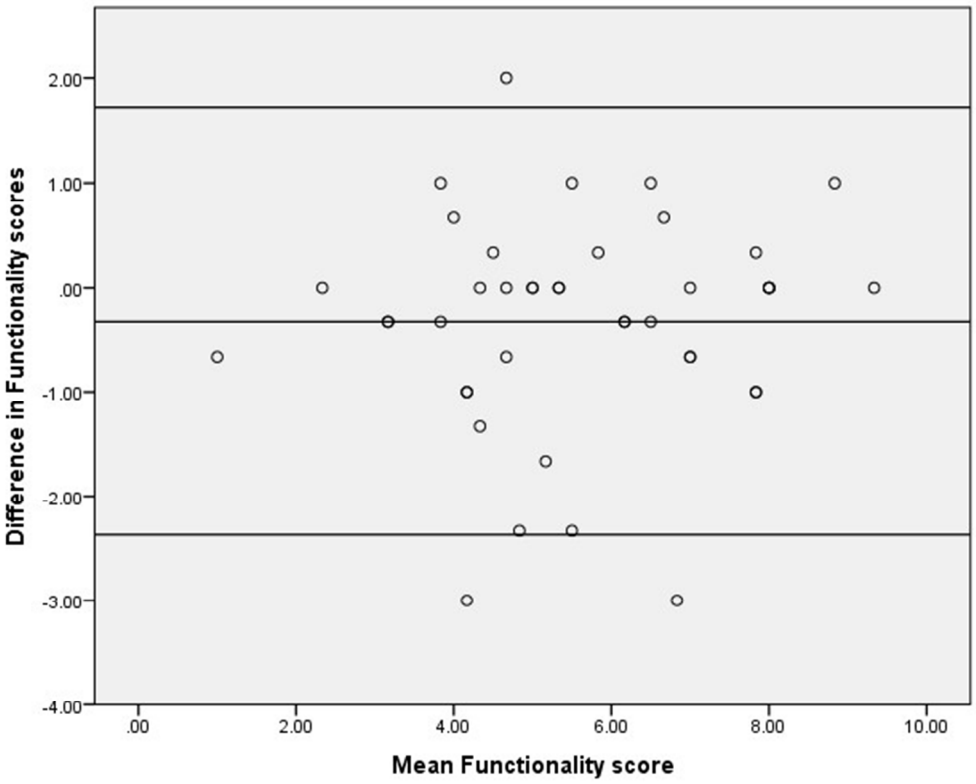
TABLE 5 Test-retest reliability of the BmPROM (*n*=42)

Dimension	Time 1 M (SD)	Time 2 M (SD)	T-test	ICC	n (% cases lying within 95% CI)
Functionality	5.40 (1.93)	5.73 (1.84)	-2.06*	.84 (.72-.91)	39 (82.86%)
Wellbeing	4.69 (1.95)	4.68 (1.82)	.06	.84 (.72- .91)	40 (95.24%)

Note. *$p < .05$

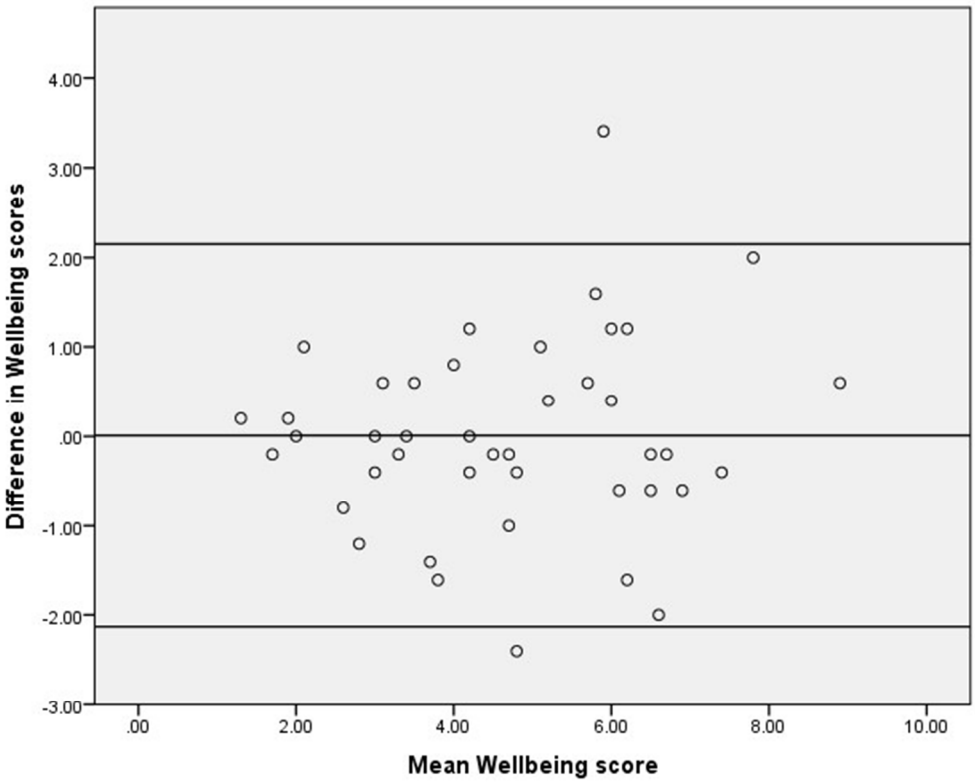
FIGURE 1 Bland and Altman scatterplot for differences in T1-T2 measures of Functionality with 95% limits of agreement (LOA).

FIGURE 2 Bland and Altman scatterplot for differences in T1-T2 measures of Wellbeing with 95% limits of agreement (LOA).



Bland and Altman scatterplot for differences in T1-T2 measures of Functionality with 95% limits of agreement (LOA). ‼ ┼

165x132mm (96 x 96 DPI)



Bland and Altman scatterplot for differences in T1-T2 measures of Wellbeing with 95% limits of agreement (LOA). ‼ ┼

165x132mm (96 x 96 DPI)