The responsiveness and interpretability of psychosocial patient-reported outcome measures in chronic musculoskeletal pain rehabilitation

Received: 23 May 2018 | Revised: 31 July 2019 | Accepted: 8 August 2018

DOI: 10.1002/ejp.1470

ORIGINAL ARTICLE

EJP European Journal of Pain WILEY

The responsiveness and interpretability of psychosocial patient-reported outcome measures in chronic musculoskeletal pain rehabilitation

Alexandra N. T. D. Pulles[1] | **Albère J. A. Köke**[1,2,3] | **Robin P. Strackke**[4] | **Rob J. E. M. Smeets**[1,5]

[1]Department of Rehabilitation Medicine, CAPHRI, Maastricht University, Maastricht, The Netherlands

[2]Adelante Centre of Expertise in Rehabilitation and Audiology, Hoensbroek, The Netherlands

[3]Zuyd University for Applied Sciences, faculty Health and Technology, Heerlen, The Netherlands

[4]Adelante Centre of Expertise in Rehabilitation and Audiology, Maastricht University Medical Centre+, Maastricht, The Netherlands

[5]CIR Revalidatie, Eindhoven/Zwolle, The Netherlands

Correspondence
Alexandra N. T. D. Pulles, Department of Rehabilitation Medicine, CAPHRI, Maastricht University, Maastricht, The Netherlands.
Email: antdpulles@gmail.com

Abstract

Background: For several widely used patient-reported outcome measures (PROMs) in chronic musculoskeletal pain (CMSP) rehabilitation, it is still not known whether they are responsive to change, and what the smallest detectable change (SDC) and minimal clinically important change (MCIC) are. Knowledge of these values can be used to accurately interpret change scores in research and clinical practice.

Methods: In this retrospective cohort study, the responsiveness, the SDC and the MCIC of the mental components of the Research and Development 36-Item Health Survey (RAND-36), the Pain Catastrophizing Scale (PCS) and the Tampa Scale of Kinesiophobia (TSK) were investigated in CMSP patients. Responsiveness, the SDC and MCIC were determined by using both anchor and distribution-based methods.

Results: For all outcome measures, there was a progression from smallest to largest mean change scores between participants who did not perceive change and those who reported change after treatment. However, correlations of the Global Perceived Effect (GPE) with the change scores on the outcome measures were low. For all outcome measures, the SDC was larger than the MCIC.

Conclusions: For this population, the questionnaires were shown not to be responsive. Furthermore, the questionnaires appeared not to be able to distinguish clinically important change from measurement error in individual patients. The finding of large measurement errors of PROMs is in line with previous research in pain rehabilitation. Using generic PROMs only, to examine changes in psychosocial status due to a pain rehabilitation programme, is therefore questionable.

Significance: This study shows that widely used generic psychosocial PROMs might not be responsive and not able to distinguish clinically important change from measurement error in individual chronic musculoskeletal pain patients. It therefore seems reasonable to reconsider the (compulsory) use of these PROMs for assessing the quality of pain rehabilitation programmes, and necessary to consider other, more objective, outcome measures for this purpose in this population.

1 | INTRODUCTION

The assessment of quality of care is becoming increasingly important in health care (Eindhoven et al., 2015). Many medical and non-medical parties assign a high value to objective outcomes of health care interventions. Furthermore, evidence shows that the systematic use of patient-reported outcome measures (PROMs): (a) leads to better communication between health care providers and patients (Chen, Ou, & Hollis, 2013; Deyo & Carter, 1992; Santana & Feeny, 2014); (b) is helpful in the analysis of problems (Chen et al., 2013; Deyo & Carter, 1992; Higginson & Carr, 2001; Marshall, Haywood, & Fitzpatrick, 2006); (c) can be a key tool when monitoring response to treatment (Chen et al., 2013; Deyo & Carter, 1992; Higginson & Carr, 2001); (d) improves patient satisfaction (Chen et al., 2013; Santana & Feeny, 2014). For a PROM to be used in research and clinical practice, besides validity and reliability, its responsiveness to clinical change and its interpretability have to be ensured (Higginson & Carr, 2001).

Responsiveness can be considered as longitudinal validity (Angst, 2011; De Vet, Terwee, Mokkink, & Knol, 2011) and reflects "the ability of an instrument to detect change over time in the construct to be measured" (Mokkink et al., 2010b). Once responsiveness is examined, attention has to be paid to the interpretability of the scores (De Vet et al., 2011). The smallest detectable change (SDC) score gives an indication of by how much change scores can vary in stable patients, reflecting measurement error (De Vet et al., 2011). The minimal clinically important change (MCIC) represents "the smallest change that is important to patients" (Ostelo & de Vet, 2005; Stratford, Binkley, Riddle, & Guyatt, 1998) and is determined by comparing patients who have not experienced change due to treatment with those who have (Jaeschke, Singer, & Guyatt, 1989). Only if the SDC is smaller than the MCIC is it possible to distinguish clinically important change from measurement error in individual patients with a large amount of certainty (Terwee, Roorda, Knol, Boer, & De Vet, 2009).

The IMMPACT initiative has recommended six core outcome domains to evaluate chronic musculoskeletal pain (CMSP) treatment, including "emotional functioning" (Chiarotto, Ostelo, Turk, Buchbinder, & Boers, 2017). Enjoyment of life and emotional wellbeing are important areas affected in CMSP patients, and patients consider functioning and wellbeing to be appropriate targets of treatment (Turk et al., 2008). In the Dutch Dataset Pain Rehabilitation, the mental components of the Research and Development 36-Item Health Survey (RAND-36), the Pain Catastrophizing Scale (PCS) and the Tampa Scale of Kinesiophobia (TSK) are included to gain insight into this outcome domain (Köke et al., 2017). In this study, the responsiveness and interpretability of PROMs concerning the domain "emotional functioning" are investigated in CMSP patients. To our knowledge, it is not known whether these questionnaires are responsive to change, and what the SDC and MCIC of these questionnaires are in the overall group of CMSP patients treated in a pain rehabilitation programme. For some of these questionnaires, this was either never examined or only examined in studies of subgroups, such as low back pain patients (Campbell et al., 2006; Lundberg, Grimby-Ekman, Verbunt, & Simmonds, 2011; Ostelo, Swinkels-Meewisse, Vlaeyen, Knol, & de Vet, 2007; Roelofs, Goubert, Peters, Vlaeyen, & Crombez, 2004; Taylor, Taylor, Foy, & Fogg, 1999; Woby, Roach, Urmston, & Watson, 2005). The outcomes of our study can be used to accurately interpret change scores on the aforementioned questionnaires in all CMSP patients. However, generic measures as these may not capture the most relevant issues for this heterogeneous patient population. Therefore, the measurement error of each of these PROMs is expected to be large. When a PROM appears not to be responsive to change or is not able to detect clinically important changes, its use in clinical practice should be reconsidered.

2 | METHODS

2.1 | Setting and research design

This is a retrospective cohort study, with a single-group repeated-measures design, in chronic (lasting more than 3 months) musculoskeletal pain patients referred to a pain rehabilitation programme in one of three rehabilitation centres in the Netherlands. Rehabilitation physiatrists selected patients for this treatment programme between December 2008 and April 2015. To start the treatment, patients had to be willing to improve their daily functioning despite pain, as pain reduction is not a primary goal in this rehabilitation programme, while addressing psychosocial factors that seem to contribute to the maintenance of pain-associated disability. Included patients attended a multidisciplinary biopsychosocial rehabilitation programme for chronic pain which involved a combination of physical, psychological, educational and/or work-related components, delivered by a multidisciplinary team of health care providers (Kamper et al., 2015).

2.2 | Data collection

Before the intake appointment (T1), and after completing the rehabilitation programme, approximately 12 weeks later (T2), data were routinely collected as described by the Dutch Dataset Pain Rehabilitation and stored in electronic patient records. Patients completed the questionnaires at home, either web-based or on paper. Data from the database of Maastricht University Medical Centre+ (Maastricht), rehabilitation centre Adelante (Hoensbroek) and Laurentius

hospital (Roermond) were used. In the Netherlands, no permission from a medical ethics committee is required for the evaluation of outcomes of care based solely on anonymous data derived from medical records. All patients gave written informed consent, stating that the data could be used anonymously for analyses of the outcomes of care.

2.3 | Outcome measures

The RAND-36 is a derivative of the Medical Outcomes Study 36-Item Short Form Health Survey (SF-36) and measures health-related quality of life at physical and mental levels (Hays, Sherbourne, & Mazel, 1993). It comprises 36 items that assess eight health concepts with multi-item scales: physical functioning (10 items); role limitations caused by physical health problems (four items); role limitations caused by emotional problems (three items); social functioning (two items); emotional wellbeing (five items); energy/fatigue (four items); pain (two items); general health perceptions (five items); and it contains an additional item about perceived health change (Hays et al., 1993). In this study, only subscales regarding mental health ("role limitations due to emotional problems," "social functioning," "emotional wellbeing" and "energy/fatigue") were taken into account. As described in the introduction, these are the subscales used in the core outcome domain "emotional functioning". Some raw scores have to be recoded and then every item has to be transformed linearly to a 0–100 score (percent of total possible score) finally, all items are averaged in the same scale together (Hays & Morales, 2001; Hays et al., 1993). Higher scores indicate a better health status. Internal consistency between all eight scales is high (Cronbach's α varies from 0.71 to 0.92; VanderZee, Sanderman, & Heyink, 1996). The correlation between the scales varies between $r = 0.12$ and $r = 0.71$ (Van der Zee & Sanderman, 1993). Test–retest reliability over 2 months per subscale varies between moderate and strong ($r = 0.58$–$r = 0.82$; Van der Zee & Sanderman, 1993).

The PCS is a 13-item questionnaire that is developed to identify catastrophic thoughts or feelings in relation to painful experiences (Lamé, Peters, Kessels, Van Kleef, & Patijn, 2008; Sullivan, Bishop, & Pivik, 1995). Patients have to indicate the degree to which they experienced each of the 13 thoughts or feelings when experiencing pain. Items are scored on a 5-point Likert scale, with scoring possibilities ranging from "not at all" (0) and "always" (4) (Sullivan et al., 1995). The total score is computed by summing all items, and ranges from 0 to 52 (Lamé et al., 2008). High scores indicate that more catastrophic thoughts or feelings are experienced (Lamé et al., 2008). Internal consistency is adequate to excellent (Cronbach's α for the total PCS ranges from 0.87; Sullivan et al., 1995 to 0.95 Osman et al., 2000). Test–retest reliability was shown to be moderate ($r = 0.67$; Lamé et al., 2008). The psychometric properties of the Dutch version of the PCS are adequate (Severeijns, Vlaeyen, van den Hout, & Weber, 2001). Reduction of pain catastrophizing is known to mediate the improvement of functioning in patients with low back pain (Smeets, Vlaeyen, Kester, & Knottnerus, 2006). Moreover, pain catastrophizing is a potent predictor of fear of movement or (re)injury (Woby, Watson, Roach, & Urmston, 2004a).

The TSK is a 17-item questionnaire that measures the fear of (re)injury due to movement (Crombez, Vlaeyen, Heuts, & Lysens, 1999). Items are scored on a 4-point Likert scale, with scoring possibilities ranging from "strongly disagree" (1) to "strongly agree" (4) (Lamé et al., 2008). The total score is computed by summing all items after inversion of four of the items and ranges from 17 to 68 (Lamé et al., 2008). High scores indicate more fear of pain/(re)injury due to movement or activities (Lamé et al., 2008). Internal consistency of the Dutch version is fair (Cronbach's $\alpha = 0.76$; Crombez et al., 1999). Test–retest reliability in chronic back pain patients was shown to be moderate ($r = 0.63$; Lamé et al., 2008). Fear of movement leads to increased avoidance (Vlaeyen, Kole-Snijders, Rotteveel, Ruesink, & Heuts, 1995) and avoidance behaviour is postulated to be one of the mechanisms in sustaining chronic pain disability (Vlaeyen, Kole-Snijders, Boeren, & Van Eek, 1995). Fear of movement/(re)-injury is also an important predictor of self-reported disability levels (Vlaeyen, Kole-Snijders, Rotteveel, et al., 1995). A reduction in fear-avoidance beliefs about work and physical activity are shown to be related to reductions in disability (Woby, Watson, Roach, & Urmston, 2004b).

The Global Perceived Effects (GPEs) were measured as follows. After completing the rehabilitation programme, patients answered the questions "To what extent do you experience a difference in your daily activities compared to the situation before participation in the rehabilitation programme?" (GPE "physical activity") and "To what extent do you experience a difference in the way you cope with problems, compared to the situation before participation in the rehabilitation programme?" (GPE "coping"). The GPEs were selected by rehabilitation physiatrists of the participating treatment centres before this study was commenced. These two GPEs were specifically chosen because they reflect two main goals of pain rehabilitation: improving the daily activity level despite being in pain; and learning to cope with problems emerging from experiencing chronic pain in daily life. Patients rated this on a 5-point Likert scale with options 1 "clearly improved;" 2 "improved;" 3 "unchanged" (not better, not worse); 4 "worse;" 5 "clearly worse."

2.4 | Patient population

Included in the data analysis were patients who completed the (subscale of the) questionnaire under study at T1 and T2 and answered at least one of the GPEs. Thus, different subgroups of the cohort were formed. The number of available data per questionnaire is displayed in the corresponding tables.

2.5 | Methods of data analysis

Data analysis was based on the COSMIN criteria (Mokkink et al., 2010a, 2010b). First, of all outcome measures, floor and ceiling effects were examined and considered to be present when, at baseline (T1), more than 15% of the patients reached the maximum or minimum score (Terwee et al., 2007). Floor and ceiling effects lead to limited responsiveness because change cannot be measured accurately in these patients (Terwee et al., 2007). To classify patients as improved or not, the answer to the GPE was trichotomised. Patients who indicated that they were "clearly improved" or "improved" were labelled as "improved;" those who indicated that they did not experience any change were considered "unchanged;" and those who responded that their status was "worse" or "clearly worse" were labelled as "deteriorated". Mean change scores on the measurement instruments for the whole group, and for the "improved," "unchanged" and "deteriorated" subgroups were calculated by subtracting the mean follow-up score from the mean baseline score (baseline score (T1) - follow-up score (T2)).

Responsiveness was examined using distribution-based and anchor-based methods.

For the distribution-based method, the standardized response mean (SRM) was calculated from the mean score difference (baseline score (T1) – follow-up score (T2)) of the total cohort, divided by the standard deviation of the difference of this score (Angst, Verra, Lehmann, & Aeschlimann, 2008; Husted, Cook, Farewell, & Gladman, 2000; Norman, Wyrwich, & Patrick, 2007). The SRM is one of the most common distribution-based measures of responsiveness (Angst, 2011; Angst et al., 2008). When the expected magnitude of the treatment effect is given, the SRM is an appropriate measure to estimate responsiveness (Mokkink, Terwee, Knol, & de Vet, 2011). The effect of a multidisciplinary biopsychosocial pain rehabilitation programme on reducing disability was estimated to be moderate (an effect size of 0.50), based on the results of a meta-analysis (Kamper et al., 2015). The instrument with an effect size similar to this, is considered to be the most responsive, so the level of the SRM was set at 0.50 (Mokkink et al., 2011).

For the anchor-based method, responsiveness was examined by comparing changes on the instrument (PROMs) and changes on the GPEs. Correlations between change scores is the preferred method of assessing responsiveness when a "gold standard" is available (Mokkink et al., 2010a). In this study, the GPEs were considered to be the gold standard. Correlation coefficients were calculated using Spearman's rho, as the gold standard is measured on an ordinal scale (and trichotomised) and the outcomes of the measurement instruments are continuous (De Vet et al., 2011). An α value of 0.05 and lower was considered statistically significant. A correlation coefficient of 0.60 was set as an acceptable indication of adequate responsiveness, based on the fact that both the outcome measure and the gold standard are accompanied by a certain degree of measurement error (De Vet et al., 2011). To confirm that a questionnaire is responsive to change, we decided that it had to be found responsive by both methods. We therefore defined a questionnaire to be responsive to change when the SRM lay between 0.40 and 0.60 and the Spearman's rho correlation coefficient was higher than 0.60.

To determine the SDC, first the standard error of measurement (SEM) was calculated. The $SEM_{agreement}$ was calculated by taking the square root of the within-subject variance ($\sqrt{(\sigma^2_{error} + \sigma^2_{moments\ of\ measurement})}$) of patients categorized as "unchanged" on the GPE (de Vet, Terwee, Knol, & Bouter, 2006). To be 95% confident that the observed improvement was a real improvement and not caused by measurement error, the SDC was calculated as $1.64*\sqrt{2}*$ SEM (Andersson, Lin, & Smeets, 2010; Terwee et al., 2009; de Vet et al., 2010). Changes greater than the SDC were consequently considered to indicate real change because only the data of "unchanged" patients were used to calculate the SDC (Pool, Ostelo, Hoving, Bouter, & de Vet, 2007). This SDC, however, neglects type II errors (Terwee et al., 2009). When taking type II errors into account, the SDC should be $4 \times$ SEM (Terwee et al., 2009). In this study, the SDC is the score calculated by using the formula $1.64*\sqrt{2}*$ SEM. However, to give an insight into the magnitude of type II errors, the scores calculated with $4 \times$ SEM are displayed as well in the outcome tables.

The MCIC was calculated from the mean change in "improved" patients minus the mean change in "unchanged" patients (Jaeschke et al., 1989; Terwee et al., 2009). The GPE group "deteriorated" was not analysed because the sample size was below the required 50 (Terwee et al., 2007). In this study, therefore, only minimal change in terms of improvement was examined.

All calculations were performed using complete case analysis, a case being "complete" when the patient completed the relevant (subscale of the) questionnaire at T1 and T2 and answered at least one of the GPEs. All statistical analyses were performed using SPSS22 for Macintosh (IBM Corp).

3 | RESULTS

3.1 | Patient characteristics

The number of data suitable for analysis was the highest for the subscale "role limitations due to emotional problems" of

the RAND-36. This cohort consisted of 359 patients. Patients had a mean age of 45.28 years ($SD = 11.02$) and 66.3% were female. Patients indicated a wide variety of musculoskeletal pain sites. Pain intensity at baseline was on average 6.63 on the numeric rating scale (NRS: 0–10). In 48.2% of the patients, the pain had existed for more than 5 years. Extensive information concerning the patient population is displayed in Table 1. Statistical analysis showed no significant differences in characteristics between the subgroups of the cohort for the different (subscales of the) questionnaires under study.

3.2 | The RAND-36

Only for the RAND-36 subscale "role limitations due to emotional problems" were there significant floor and ceiling effects, with 39.6% of patients at the minimum and 36.5% at the maximum score. For this reason, the responsiveness of this subscale was not further investigated.

In Tables 2–5, the outcomes for the mental subscales of the RAND-36 in terms of change scores, the SRM and Spearman's rho, according to each anchor (GPE), are displayed. All subscales of the RAND-36 showed a progression from smallest to largest mean change scores between participants who did not perceive change and those who perceived change. The SRMs of the subscales "emotional wellbeing" (0.46) and "social functioning" (0.56) lay between 0.4 and 0.6. The change scores on the subscale "emotional wellbeing" had the highest correlation ($r = -0.290$), using the GPE "coping" as the anchor.

In Tables 6–9, the outcomes of the calculations of the SEM, SDC and MCIC are displayed for each subscale of the RAND-36, separately. The SDC of the subscale "energy/fatigue" was the smallest (−21.34), when using the "physical activity" GPE. The SDCs of the other subscales ranged from −22.94 (subscale "emotional wellbeing") to −79.55 (subscale "role limitations due to emotional problems"). The MCIC of all subscales did not exceed the SDC of any of the corresponding subscales, independent of the anchor used.

3.3 | The PCS

In Table 10, the outcomes for the PCS in terms of change scores, the SRM and Spearman's rho, according to each anchor (GPE), are displayed. The mean change score on the PCS for the whole cohort was 8.70 ($SD = 8.48$). This resulted in a SRM of 1.03. Spearman's rho correlation coefficient for the PCS with the "physical activity" GPE was 0.248. According to this anchor, improved patients had a mean change score of 9.62 ($SD = 8.22$) and unchanged patients had a mean change score of 3.84 ($SD = 8.50$). According to the "coping" GPE, change scores were, respectively, 9.26 ($SD = 8.27$) and 4.65 ($SD = 9.19$). The correlation coefficient of the PCS with this GPE was 0.172.

In Table 11, the outcomes of the calculations of SEM, SDC and MCIC from the PCS are displayed. The GPE with the highest correlation coefficient was that for "physical activity." Based on this anchor, the MCIC did not exceed the SDC, the SDC being 15.15 and the MCIC 5.78, on a scale of 0–52.

3.4 | The TSK

In Table 12, the outcomes for the TSK in terms of change scores, the SRM and Spearman's rho, according to each anchor (GPE), are displayed. The mean change score on the TSK for the whole cohort was 5.48 ($SD = 7.70$). This resulted in a SRM of 0.71. Spearman's rho correlation coefficient for the TSK with the "physical activity" GPE was 0.235. According to this anchor, improved patients had a mean change score of 6.40 ($SD = 7.36$) and unchanged patients had a mean change score of 1.90 ($SD = 7.28$). According to the "coping" GPE, change scores were, respectively, 6.21 ($SD = 7.30$) and 1.06 ($SD = 9.06$). The correlation coefficient of the TSK with this GPE was 0.222.

In Table 13, the outcomes of the calculations of the SEM, SDC and MCIC from the TSK are displayed. The GPE with the highest correlation coefficient was that for "physical activity." Based on this anchor, the MCIC did not exceed the SDC, the SDC being 12.22 and the MCIC 4.50, on a scale of 17–68.

4 | DISCUSSION AND CONCLUSIONS

The objective of this study was to determine the responsiveness, the SDC and the MCIC of PROMs that examine psychosocial elements (the mental subscales of the RAND-36, the PCS and the TSK) in CMSP patients in a multidisciplinary biopsychosocial rehabilitation programme. To our knowledge, the responsiveness and interpretability of these widely used questionnaires had not yet been examined in this heterogeneous group of CMSP patients.

The results showed that in general there was a progression from smallest to largest mean change scores on the PROMs between participants who did not perceive change and those who did. This suggests that the PROMs reflect changes as measured by the GPEs. The SRM of some of the PROMs lay near the expected value of 0.50. However, correlations of the GPEs with the change scores were very low for all PROMs, ranging from $r = -0.129$ to $r = -0.290$. Therefore, no questionnaire appeared to be responsive. This implies that these PROMs were not able to measure perceived change in patients. Furthermore, for all PROMs, the SDC was larger than the MCIC, independent of the GPE used. This means that these questionnaires were not able to distinguish clinically important change from measurement error in individual

TABLE 1 Baseline characteristics of patients. ($n = 359$)

Variable	
Age (years) (mean, *SD*)	45.28 (10.97)
Sex	
Male	118 (32,9)
Female	238 (66,3)
Marital status	
Single	85 (23,7)
Married/in a relationship	262 (73,0)
Education	
Higher level	75 (20,9)
Average level	128 (35,7)
Lower level	136 (37,9)
Site of pain referred for[a]	
Head	86 (24,0)
Face/throat	35 (9,7)
Neck	180 (50,1)
Shoulder(s)/upper back	181 (50,4)
Arm(s)	139 (38,7)
Hand(s)/fingers(s)	126 (35,1)
Chest/stomach	52 (14,5)
Lower back	249 (69,4)
Hip(s)	150 (41,8)
Upper leg(s)/knee(s)	174 (48,5)
Ankle(s)/feet/foot	131 (36,5)
Elsewhere	78 (21,7)
Psychological counselling in the past	
Yes	240 (66,9)
No	107 (29,8)
Work status	
Employed/student	163 (45,4)
Unemployed/not a student	183 (51,0)
Self-rated health	
Good or very good	103 (28,7)
Fair	137 (38,2)
Poor	75 (20,9)
Duration of complaints	
Less than five years	171 (47,6)
More than five years	173 (48,2)
Pain intensity (NRS[b] [0–10]; mean, *SD*)	6.63 (1.82)
Disability (PDI[c] [0–70]; mean, *SD*)	39.90 (11.30)
Score RAND-36-physical functioning (mean, *SD*)	41.80 (21.10)
Scores on PROMs (mean, *SD*)	
RAND-36-social functioning	48.43 (25.01)
RAND-36-role limitations emotional problems	48.56 (44.40)
RAND-36-emotional wellbeing	50.53 (14.46)
RAND-36-energy/fatigue	38.47 (15.93)
PCS	20.44 (10.23)
TSK	35.04 (8.15)

Note: Number of missing data varies between variables and varies from 3 (sex) to 44 (self-rated health). Values are total number and percentage unless indicated otherwise.

Abbreviations: NRS, numeric rating scale; PCS, pain catastrophizing scale; PROMs, patient-reported outcome measures; TSK, Tampa Scale of Kinesiophobia

[a]Patients could indicate more pain sites for which they were referred for help.

[b]Numeric rating scale.

[c]Pain disability index.

patients. The hypothesis that the measurement error of these generic questionnaires would be substantial has thus been confirmed.

The strengths of this study are the large sample sizes and an amply sufficient number of self-perceived unchanged patients to be able to adequately determine the results. Moreover, the patients in this study appeared to be comparable to patients in previous chronic pain studies as to age, gender, pain intensity and scores on the TSK and RAND-36 at baseline (Pool et al., 2007; van der Roer, Ostelo, Bekkering, van Tulder, & de Vet, 2006; Soer et al., 2013). However, this study also has its limitations. First, the use of GPEs as gold standards. A GPE has high face validity and is therefore frequently considered a reasonable gold standard for PROMs, provided that the GPE assesses the same construct as the instrument under study (De Vet et al., 2011). Although previous research has shown a relationship between pain catastrophizing, avoidance behaviour, health-related quality of life on a mental level and functioning in CMSP patients, it is assumable that the PROMs and GPEs used in this study have measured different constructs. The data of this study must therefore be interpreted with caution. Moreover, GPE scores are dependent on personal interpretation and may be biased by current status (Kamper et al., 2010) or may be influenced by such factors as mood, life events and the perceived need for socially desirable answers. The use of GPEs in outcome assessments therefore needs further study. Furthermore, in retrospect, we would have formulated both GPEs differently. The GPE "coping," should be rephrased into "To what extent do you experience a difference in the way you cope with problems evolving from your pain?," so as to more closely address a specific aim of treatment. Moreover, we should have indicated in the GPE "physical activity", that the personally relevant activities that the patient has indicated to improve at the start of the rehabilitation programme are meant. In addition, the decision concerning which group is

TABLE 2 Outcomes for the RAND-36 "social functioning:" change scores, SRM and Spearman's rho correlation coefficients according to each anchor (GPE)

ΔT1-T2 ($n = 358$)	−15.22 (27.00)	
SRM	0.56	
		Spearman's rho
Anchor (GPE) Physical activity		$r = -0.186^{b}$
Improved ($n = 298$)	−17.45 (26.68)	
Unchanged ($n = 50$)	−7.25 (24.37)	
Deteriorated ($n = 8$)	15.63 (30.44)	
Anchor (GPE) Coping		$r = -0.176^{b}$
Improved ($n = 310$)	−17.10 (26.97)	
Unchanged ($n = 40$)	−4.06 (23.75)	
Deteriorated ($n = 6$)	4.17 (29.23)	

Note: Values are mean (*SD*) unless indicated otherwise.
Abbreviations: GPE, global perceived effect; SRM, standardized response mean.
[a]$P < .05$;
[b]$P < .01$;
[c]$P < .005$.

TABLE 3 Outcomes for the RAND-36 "role limitations due to emotional problems:" change scores, SRM and Spearman's rho correlation coefficients according to each anchor (GPE)

ΔT1-T2 ($n = 359$)	−14.21 (49.51)	
SRM	/	
		Spearman's rho
Anchor (GPE) Physical activity		/
Improved ($n = 297$)	−17.17 (49.71)	
Unchanged ($n = 51$)	−2.61 (48.92)	
Deteriorated ($n = 9$)	14.81 (33.79)	
Anchor (GPE) Coping		/
Improved ($n = 311$)	−17.04 (49.86)	
Unchanged ($n = 40$)	5.83 (45.85)	
Deteriorated ($n = 6$)	−5.56 (32.77)	

Note: Values are mean (*SD*) unless indicated otherwise.
Abbreviations: GPE, global perceived effect; SRM, standardized response mean.

defined as "clinically changed" is arbitrary. In this study, the threshold was set at "improved" because it was thought that, from this point, patients would have experienced therapy as effective. The options "slightly improved" and "slightly deteriorated" were not included in order to prevent patients who had not really experienced a change indicating a slight change because of gratitude for the effort made by therapists or for other reasons. Furthermore, only including complete cases in the analyses may have caused a form of selection bias. Patients who filled in the PROMs completely and answered the GPEs might have been predominantly those who experienced an effect from the treatment, had more time or energy to fill in the PROMs or were more likely to give socially desirable answers. This would have the most effect on the SRM, as treatment effect contributes to effect sizes. However, a missing data analysis was performed and there was no statistically significant difference between patients with or without missing data on work status or level of pain on the NRS, at either start or end of the treatment programme. Finally, the absence of outcomes for deteriorated patients. It is recommended that separate MCICs be assessed for improved and deteriorated patients, as previous studies have shown different MCICs for improvement and deterioration (De Vet et al., 2011). To adequately determine the MCIC for a group, COSMIN recommends a sample size of at least 50 patients (Terwee et al., 2007), the group of deteriorated patients in this study was smaller.

To reiterate, little research has been done on the responsiveness and interpretability of these questionnaires. The responsiveness of the SF-36, of which the RAND-36 is a derivative, has previously been examined in subgroups of CMSP patients. A number of these studies showed floor effects for role limitation subscales (Campbell et al., 2006; Suarez-Almazor, Kendall, Johnson, Skeith, & Vincent, 2000; Taylor et al., 1999), and low SRMs for these subscales (Angst et al., 2008; Taylor et al., 1999), in line with our results. Taylor et al. found that six subscales of the SF-36 had a correlation coefficient lower than −0.50 when comparing change scores with the answer on the GPE. They concluded that only the subscale "social functioning," out of all the mental subscales, showed responsiveness for patients with low back pain (Taylor et al., 1999). The results of the current study do not clearly indicate which subscale is most responsive. No comparable study that examined the responsiveness of the PCS or TSK in any kind of chronic pain patients has, as far as we are aware, been published. Concerning the interpretability of the questionnaires under study, other studies only calculated the SDC of the TSK. The SDC, measured in acute non-specific low back pain patients with no treatment in the period between the measurements, was 9.2 (Ostelo et al., 2007). This is lower than the SDC calculated for the TSK in this study, but still higher than the calculated MCIC. An additional literature research showed that also for other generic PROMs used in pain rehabilitation, the SDC appears to be larger than the MCIC. Examples are the pain self-efficacy questionnaire (Maughan & Lewis, 2010) and the EuroQol (van der Roer et al., 2006) in low back pain patients. In 1993, Guyatt and colleagues already concluded that generic profiles may be unresponsive to changes in specific conditions (Guyatt et al., 1993). In 2000, Suarez-Almazor and colleagues advised conducting additional research to evaluate the role of generic

TABLE 4 Outcomes for the RAND-36 "emotional wellbeing:" change scores, SRM and Spearman's rho correlation coefficients according to each anchor (GPE)

ΔT1-T2 ($n = 139$)	−6.91 (15.10)	
SRM	0.46	
		Spearman's rho
Anchor (GPE) Physical activity		$r = -0.133$
Improved ($n = 113$)	−8.14 (14.48)	
Unchanged ($n = 20$)	−3.80 (17.14)	
Deteriorated ($n = 5$)	0.80 (7.16)	
Anchor (GPE) Coping		$r = -0.290^{b}$
Improved ($n = 118$)	−9.05 (14.14)	
Unchanged ($n = 17$)	4.71 (13.58)	
Deteriorated ($n = 3$)	−1.33 (18.04)	

Note: Values are mean (*SD*) unless indicated otherwise.
Abbreviations: GPE, global perceived effect; SRM, standardized response mean.
[a]$P < .05$;
[b]$P < .01$;
[c]$P < .005$.

TABLE 5 Outcomes for the RAND-36 "energy/fatigue:" change scores, SRM and Spearman's rho correlation coefficients according to each anchor (GPE)

ΔT1-T2 ($n = 262$)	−12.71 (17.93)	
SRM	0.71	
		Spearman's rho
Anchor (GPE) Physical activity		$r = -0.250^{b}$
Improved ($n = 220$)	−14.52 (18.07)	
Unchanged ($n = 35$)	−3.29 (12.77)	
Deteriorated ($n = 6$)	1.67 (17.22)	
Anchor (GPE) Coping		$r = -0.195^{b}$
Improved ($n = 230$)	−13.85 (18.17)	
Unchanged ($n = 29$)	−3.28 (13.32)	
Deteriorated ($n = 2$)	−10.00 (14.14)	

Note: Values are mean (*SD*) unless indicated otherwise.
Abbreviations: GPE, Global Perceived Effect; SRM, standardized response mean.
[a]$P < .05$;
[b]$P < .01$;
[c]$P < .005$.

measures of quality of life in patients with low back pain, before wide implementation in clinical settings or outcomes research, as these measures seemed insensitive to change in those dimensions of importance to the patient (Suarez-Almazor et al., 2000). Yet, at the moment, generic PROMs are widely implemented in pain rehabilitation programmes

TABLE 6 Outcomes for the RAND-36 "social functioning:" SEMs, SDCs and MCICs according to each anchor (GPE)

Anchor (GPE)	SEM	4.0 × SEM	SDC	MCIC
Physical activity	−17.81	−71.24	−41.31	−10.20
Coping	−16.83	−67.31	−39.03	−13.04

Abbreviations: GPE, global perceived effect; MCIC, minimal clinically important change; SDC, smallest detectable change; SEM, standard error of measurement.

TABLE 7 Outcomes for the RAND-36 "role limitations due to emotional problems:" SEMs, SDCs and MCICs according to each anchor (GPE)

Anchor (GPE)	SEM	4.0 × SEM	SDC	MCIC
Physical activity	−34.30	−137.20	−79.55	−14.56
Coping	−32.27	−129.08	−74.84	−22.87

Abbreviations: GPE, global perceived effect; MCIC, minimal clinically important change; SDC, smallest detectable change; SEM, standard error of measurement.

TABLE 8 Outcomes for the RAND-36 "emotional wellbeing:" SEMs, SDCs and MCICs according to each anchor (GPE)

Anchor (GPE)	SEM	4.0 × SEM	SDC	MCIC
Physical activity	−12.12	−48.48	−28.11	−4.34
Coping	−9.89	−39.56	−22.94	−13.76

Abbreviations: GPE, Global Perceived Effect; MCIC, minimal clinically important change; SDC, smallest detectable change; SEM, standard error of measurement.

TABLE 9 Outcomes for the RAND-36 "energy/fatigue:" SEMs, SDCs and MCICs according to each anchor (GPE)

Anchor (GPE)	SEM	4.0 × SEM	SDC	MCIC
Physical activity	−9.20	−36.80	−21.34	−11.23
Coping	−9.54	−38.16	−22.13	−10.57

Abbreviations: GPE, global perceived effect; MCIC, minimal clinically important change; SDC, smallest detectable change; SEM, standard error of measurement.

and in some countries, the collection of these data is even compulsory.

The comments in the introduction about the favourable effects of using PROMs continue to apply, independent of the results of this study. Professionals can continue using these questionnaires in their daily practice as a source of information. However, our results call into question their use in situations in which the effectiveness of rehabilitation programmes in CMSP patients is to be quantified. Based on our results, it seems reasonable to reconsider the implementation and use of solely generic PROMs to measure the effect of treatment in CMSP patients, since our study shows the inadequate

TABLE 10 Outcomes for the PCS: change scores, SRM and Spearman's rho correlation coefficients according to each anchor (GPE)

ΔT1-T2 ($n = 280$)	8.70 (8.48)	
SRM	1.03	
		Spearman's rho
Anchor (GPE) Physical activity		$r = 0.248$[b]
Improved ($n = 237$)	9.62 (8.22)	
Unchanged ($n = 38$)	3.84 (8.50)	
Deteriorated ($n = 5$)	2.00 (5.48)	
Anchor (GPE) Coping		$r = 0.172$[b]
Improved ($n = 245$)	9.26 (8.27)	
Unchanged ($n = 31$)	4.65 (9.19)	
Deteriorated ($n = 4$)	5.50 (8.74)	

Note: Values are mean (*SD*) unless indicated otherwise.
Abbreviations: GPE, global perceived effect; PCS, pain catastrophizing scale; SRM, standardized response mean.
[a]$P < .05$;
[b]$P < .01$;
[c]$P < .005$.

TABLE 11 Outcomes for the PCS: SEMs, SDCs and MCICs according to each anchor (GPE)

Anchor (GPE)	SEM	4.0 × SEM	SDC	MCIC
Physical activity	6.53	26.12	15.15	5.78
Coping	7.19	28.76	16.68	4.61

Abbreviations: GPE, global perceived effect; PCS, pain catastrophizing scale; MCIC, minimal clinically important change; SDC, smallest detectable change; SEM, standard error of measurement.

TABLE 12 Outcomes for the TSK: change scores, SRM and Spearman's rho correlation coefficients according to each anchor (GPE)

ΔT1-T2 ($n = 344$)	5.48 (7.70)	
SRM	0.71	
		Spearman's rho
Anchor (GPE) Physical activity		$r = 0.235$[b]
Improved ($n = 286$)	6.40 (7.36)	
Unchanged ($n = 48$)	1.90 (7.28)	
Deteriorated ($n = 8$)	−3.88 (10.12)	
Anchor (GPE) Coping		$r = 0.222$[b]
Improved ($n = 300$)	6.21 (7.30)	
Unchanged ($n = 36$)	1.06 (9.06)	
Deteriorated ($n = 6$)	−2.00 (6.51)	

Note: Values are mean (*SD*) unless indicated otherwise.
Abbreviations: GPE, global perceived effect; SRM, standardized response mean, TSK, Tampa Scale of Kinesiophobia
[a]$P < .05$;
[b]$P < .01$;
[c]$P < .005$.

TABLE 13 Outcomes for the TSK: SEMs, SDCs and MCICs for the TSK according to each anchor (GPE)

Anchor (GPE)	SEM	4.0 × SEM	SDC	MCIC
Physical activity	5.27	21.08	12.22	4.50
Coping	6.36	25.44	14.75	5.15

Abbreviations: GPE, global perceived effect; MCIC, minimal clinically important change; SDC, smallest detectable change; SEM, standard error of measurement; TSK, Tampa Scale of Kinesiophobia.

ability of these to reflect change in the individual patient. The question now arises of how to use these questionnaires in daily clinical practice in CMSP patients until more appropriate ways to assess the effects of rehabilitation programmes are developed. For rehabilitation professionals, an option could be to use the calculated SDCs as a benchmark in individual patients. When the change score on a questionnaire is higher than the calculated SDC, there is a 95% certainty that the patient has experienced a real improvement (Terwee et al., 2009). In the future, it would be interesting to examine the responsiveness and interpretability of PROMs that focus on physical, conceivable more concrete aspects of disease in the CMSP population. Perhaps, the responsiveness of these PROMs will be higher and the measurement error will be lower. The results of this study however, compel us to consider other types and more objective outcome measures to assess the effectiveness of CMSP rehabilitation programmes. Examples could include measurable behaviour of patients (the number of doctor visits, use of pain medication, return to work) or physical tests, preferably associated with the for the patient important goals of the rehabilitation programme. Furthermore, we want to emphasize the importance of the opinion of the clinician in the assessment, because of his/her (medical) knowledge, experience and personal contact with the patients. Concluding, measuring outcomes of CMSP rehabilitation needs a broader assessment, in which PROMs are integrated as well as clinical/clinician based measurements.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

R.J.E.M. Smeets and A.J.A. Köke constructed the concept and design of the work. Data collection was performed by A.J.A. Köke, R.J.E.M. Smeets and R.P. Strackke. Data analysis and interpretation was performed by A.N.T.D.

Pulles. A.J.A. Köke, R.P. Strackke and R.J.E.M. Smeets and A.N.T.D. Pulles discussed the results of this analysis. The draft of the article was made by A.N.T.D. Pulles. The article was reviewed and commented on by R.J.E.M. Smeets and A.J.A. Köke. The final version of the manuscript to be published was read, revised and approved by A.J.A. Köke, R.P. Strackke and R.J.E.M. Smeets and A.N.T.D. Pulles.

REFERENCES

Andersson, E. I., Lin, C. C., & Smeets, R. J. (2010). Performance tests in people with chronic low back pain: Responsiveness and minimal clinically important change. *Spine (Phila Pa, 1976)*, *35*, E1559–E1563. https://doi.org/10.1097/BRS.0b013e3181cea12e

Angst, F. (2011). The new COSMIN guidelines confront traditional concepts of responsiveness. *BMC Medical Research Methodology*, *11*, 152. https://doi.org/10.1186/1471-2288-11-152

Angst, F., Verra, M. L., Lehmann, S., & Aeschlimann, A. (2008). Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. *BMC Medical Research Methodology*, *8*, 26. https://doi.org/10.1186/1471-2288-8-26

Campbell, H., Rivero-Arias, O., Johnston, K., Gray, A., Fairbank, J., & Frost, H. (2006). Responsiveness of objective, disease-specific, and generic outcome measures in patients with chronic low back pain: An assessment for improving, stable, and deteriorating patients. *Spine (Phila Pa, 1976)*, *31*, 815–822. https://doi.org/10.1097/01.brs.0000207257.64215.03

Chen, J., Ou, L., & Hollis, S. J. (2013). A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res*, *13*, 1. https://doi.org/10.1186/1472-6963-13-211

Chiarotto, A., Ostelo, R. W., Turk, D. C., Buchbinder, R., & Boers, M. (2017). Core outcome sets for research and clinical practice. *Brazilian Journal of Physical Therapy*, *21*(2), 77–84. https://doi.org/10.1016/j.bjpt.2017.03.001

Crombez, G., Vlaeyen, J. W., Heuts, P. H., & Lysens, R. (1999). Pain-related fear is more disabling than pain itself: Evidence on the role of pain-related fear in chronic back pain disability. *Pain*, *80*, 329–339. https://doi.org/10.1016/S0304-3959(98)00229-2

de Vet, H. C., Terluin, B., Knol, D. L., Roorda, L. D., Mokkink, L. B., Ostelo, R. W., … Terwee, C. B. (2010). Three ways to quantify uncertainty in individually applied 'minimally important change' values. *Journal of Clinical Epidemiology*, *63*, 37–45. https://doi.org/10.1016/j.jclinepi.2009.03.011

de Vet, H. C., Terwee, C. B., Knol, D. L., & Bouter, L. M. (2006). When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, *59*, 1033–1039. https://doi.org/10.1016/j.jclinepi.2005.10.015

de Vet, H. C., Terwee, C. B., Mokkink, L. B., Knol, D. L. (2011). *Measurement in medicine: A practical guide*. Cambridge, UK: Cambridge University Press.

Deyo, R. A., & Carter, W. B. (1992). Strategies for improving and expanding the application of health status measures in clinical settings: A researcher-developer viewpoint. *Medical Care*, *30*(Supplement), MS176–MS186. https://doi.org/10.1097/00005650-199205001-00015

Eindhoven, D., Wierda, E., de Bruijne, M., Amoroso, G., de Mol, B., Umans, V., … Borleffs, C. (2015). The year of transparency: Measuring quality of cardiac care. *Netherlands Heart Journal*, *23*, 457–465. https://doi.org/10.1007/s12471-015-0739-9

Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine*, *118*, 622–629. https://doi.org/10.7326/0003-4819-118-8-199304150-00009

Hays, R. D., & Morales, L. S. (2001). The RAND-36 measure of health-related quality of life. *Annals of Medicine*, *33*, 350–357. https://doi.org/10.3109/07853890109002089

Hays, R. D., Sherbourne, C. D., & Mazel, R. M. (1993). The rand 36-item health survey 1.0. *Health Economics*, *2*, 217–227. https://doi.org/10.1002/hec.4730020305

Higginson, I. J., & Carr, A. J. (2001). Using quality of life measures in the clinical setting. *BMJ*, *322*, 1297–1300.

Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: A critical review and recommendations. *Journal of Clinical Epidemiology*, *53*, 459–468. https://doi.org/10.1016/S0895-4356(99)00206-1

Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*, *10*, 407–415. https://doi.org/10.1016/0197-2456(89)90005-6

Kamper, S. J., Apeldoorn, A., Chiarotto, A., Smeets, R., Ostelo, R., Guzman, J., & van Tulder, M. (2015). Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. *BMJ*, *350*, h444. https://doi.org/10.1136/bmj.h444

Kamper, S. J., Ostelo, R. W., Knol, D. L., Maher, C. G., de Vet, H. C., & Hancock, M. J. (2010). Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *Journal of Clinical Epidemiology*, *63*, 760–766.e1. https://doi.org/10.1016/j.jclinepi.2009.09.009

Köke, A., Smeets, R., Schreurs, K., Baalen, B., Haan, P., Remerie, S., … Reneman, M. (2017). Dutch dataset pain rehabilitation in daily practice: Content, patient characteristics and reference data. *European Journal of Pain*, *21*, 434–444. https://doi.org/10.1002/ejp.937

Lamé, I. E., Peters, M. L., Kessels, A. G., Van Kleef, M., & Patijn, J. (2008). Test—retest stability of the pain catastrophizing scale and the tampa scale for kinesiophobia in chronic pain over a longer period of time. *Journal of Health Psychology*, *13*, 820–826. https://doi.org/10.1177/1359105308093866

Lundberg, M., Grimby-Ekman, A., Verbunt, J., & Simmonds, M. (2011). Pain-related fear: A critical review of the related measures. *Pain Research and Treatment*, *2011*, 1–26. https://doi.org/10.1155/2011/494196

Marshall, S., Haywood, K., & Fitzpatrick, R. (2006). Impact of patient-reported outcome measures on routine practice: A structured review. *Journal of Evaluation in Clinical Practice*, *12*, 559–568. https://doi.org/10.1111/j.1365-2753.2006.00650.x

Maughan, E. F., & Lewis, J. S. (2010). Outcome measures in chronic low back pain. *European Spine Journal*, *19*, 1484–1494. https://doi.org/10.1007/s00586-010-1353-6

Mokkink, L., Terwee, C., Knol, D., de Vet, H. (2011). The new COSMIN guidelines regarding responsiveness. Author's response.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … de Vet, H. C. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, *19*, 539–549. https://doi.org/10.1007/s11136-010-9606-8

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … de Vet, H. C. (2010b). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*, 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Norman, G., Wyrwich, K. W., & Patrick, D. L. (2007). The mathematical relationship among different forms of responsiveness coefficients. *Quality of Life Research*, *16*, 815–822. https://doi.org/10.1007/s11136-007-9180-x

Osman, A., Barrios, F. X., Gutierrez, P. M., Kopper, B. A., Merrifield, T., & Grittmann, L. (2000). The pain catastrophizing scale: Further psychometric evaluation with adult samples. *Journal of Behavioral Medicine*, *23*, 351–365.

Ostelo, R. W., & de Vet, H. C. (2005). Clinically important outcomes in low back pain. *Best Practice & Research Clinical Rheumatology*, *19*, 593–607. https://doi.org/10.1016/j.berh.2005.03.003

Ostelo, R. W., Swinkels-Meewisse, I. J., Vlaeyen, J. W., Knol, D. L., & de Vet, H. C. (2007). Assessing pain and pain-related fear in acute low back pain: What is the smallest detectable change? *International Journal of Behavioral Medicine*, *14*, 242–248. https://doi.org/10.1007/BF03002999

Pool, J. J., Ostelo, R. W., Hoving, J. L., Bouter, L. M., & de Vet, H. C. (2007). Minimal clinically important change of the neck disability index and the numerical rating scale for patients with neck pain. *Spine (Phila Pa, 1976)*, *32*, 3047–3051. https://doi.org/10.1097/BRS.0b013e31815cf75b

Roelofs, J., Goubert, L., Peters, M. L., Vlaeyen, J. W., & Crombez, G. (2004). The tampa scale for kinesiophobia: Further examination of psychometric properties in patients with chronic low back pain and fibromyalgia. *European Journal of Pain*, *8*, 495–502. https://doi.org/10.1016/j.ejpain.2003.11.016

Santana, M.-J., & Feeny, D. (2014). Framework to assess the effects of using patient-reported outcome measures in chronic care management. *Quality of Life Research*, *23*, 1505–1513.

Severeijns, R., Vlaeyen, J. W., van den Hout, M. A., & Weber, W. E. (2001). Pain catastrophizing predicts pain intensity, disability, and psychological distress independent of the level of physical impairment. *The Clinical Journal of Pain*, *17*, 165–172. https://doi.org/10.1097/00002508-200106000-00009

Smeets, R. J., Vlaeyen, J. W., Kester, A. D., & Knottnerus, J. A. (2006). Reduction of pain catastrophizing mediates the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. *The Journal of Pain*, *7*, 261–271. https://doi.org/10.1016/j.jpain.2005.10.011

Soer, R., Köke, A. J., Vroomen, P. C., Stegeman, P., Smeets, R. J., Coppes, M. H., & Reneman, M. F. (2013). Extensive validation of the pain disability index in 3 groups of patients with musculoskeletal pain. *Spine (Phila Pa, 1976)*, *38*, E562–E568. https://doi.org/10.1097/BRS.0b013e31828af21f

Stratford, P. W., Binkley, J. M., Riddle, D. L., & Guyatt, G. H. (1998). Sensitivity to change of the Roland-Morris back pain questionnaire: Part 1. *Physical Therapy*, *78*, 1186–1196. https://doi.org/10.1093/ptj/78.11.1186

Suarez-Almazor, M., Kendall, C., Johnson, J., Skeith, K., & Vincent, D. (2000). Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. *Rheumatology*, *39*, 783–790.

Sullivan, M. J., Bishop, S. R., & Pivik, J. (1995). The pain catastrophizing scale: Development and validation. *Psychological Assessment*, *7*, 524. https://doi.org/10.1037/1040-3590.7.4.524

Taylor, S. J., Taylor, A. E., Foy, M. A., & Fogg, A. J. (1999). Responsiveness of common outcome measures for patients with low back pain. *Spine (Phila Pa, 1976)*, *24*, 1805. https://doi.org/10.1097/00007632-199909010-00010

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., … de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*, 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Terwee, C. B., Roorda, L. D., Knol, D. L., De Boer, M. R., & De Vet, H. C. (2009). Linking measurement error to minimal important change of patient-reported outcomes. *Journal of Clinical Epidemiology*, *62*, 1062–1067. https://doi.org/10.1016/j.jclinepi.2008.10.011

Turk, D. C., Dworkin, R. H., Revicki, D., Harding, G., Burke, L. B., Cella, D., … Rappaport, B. A. (2008). Identifying important outcome domains for chronic pain clinical trials: An IMMPACT survey of people with pain. *PAIN®*, *137*, 276–285. https://doi.org/10.1016/j.pain.2007.09.002

van der Roer, N., Ostelo, R. W., Bekkering, G. E., van Tulder, M. W., & de Vet, H. C. (2006) Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine (Phila Pa 1976)*, *31*, 578–582.

Van der Zee, K., & Sanderman, R. (1993). *Measuring general health status with the RAND-36: Users manual (Het meten van de algemene gezondheidstoestand met de RAND-36: Een handleiding)*. Groningen: Northern Center for Healthcare Research, University of Groningen.

VanderZee, K. I., Sanderman, R., & Heyink, J. (1996). A comparison of two multidimensional measures of health status: The Nottingham health profile and the RAND 36-item health survey 1.0. *Quality of Life Research*, *5*, 165–174. https://doi.org/10.1007/BF00435982

Vlaeyen, J. W., Kole-Snijders, A. M., Boeren, R. G., & Van Eek, H. (1995). Fear of movement/(re) injury in chronic low back pain and its relation to behavioral performance. *Pain*, *62*, 363–372. https://doi.org/10.1016/0304-3959(94)00279-N

Vlaeyen, J. W., Kole-Snijders, A. M., Rotteveel, A. M., Ruesink, R., & Heuts, P. H. (1995). The role of fear of movement/(re) injury in pain disability. *Journal of Occupational Rehabilitation*, *5*, 235–252. https://doi.org/10.1007/BF02109988

Woby, S. R., Roach, N. K., Urmston, M., & Watson, P. J. (2005). Psychometric properties of the TSK-11: A shortened version of the Tampa Scale for Kinesiophobia. *Pain*, *117*, 137–144. https://doi.org/10.1016/j.pain.2005.05.029

Woby, S. R., Watson, P. J., Roach, N. K., & Urmston, M. (2004a). Adjustment to chronic low back pain—the relative influence of fear-avoidance beliefs, catastrophizing, and appraisals of control. *Behavior Research and Therapy*, *42*, 761–774. https://doi.org/10.1016/S0005-7967(03)00195-5

Woby, S. R., Watson, P. J., Roach, N. K., & Urmston, M. (2004b). Are changes in fear-avoidance beliefs, catastrophizing, and appraisals of control, predictive of changes in chronic low back pain and disability? *European Journal of Pain*, *8*, 201–210. https://doi.org/10.1016/j.ejpain.2003.08.002

How to cite this article: Pulles ANTD, Köke AJA, Strackke RP, Smeets RJEM. The responsiveness and interpretability of psychosocial patient-reported outcome measures in chronic musculoskeletal pain rehabilitation. *Eur J Pain*. 2020;24:134–144. https://doi.org/10.1002/ejp.1470