Acta Scientific MEDICAL SCIENCES (ISSN: 2582-0931)

Volume 6 Issue 7 July 2022

Research Article

The Interpretability of Change Scores of the Pain Disability Index and its Responsiveness in Chronic Musculoskeletal Pain Rehabilitation

Albère JA Köke[1,2,3]*, Laura WME Beckers[1], Alexandra NTD Pulles[1] and Rob JEM Smeets[1,5]

[1]Department of Rehabilitation Medicine, Care and Public Health Research Institute (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, location Eindhoven, The Netherlands

***Corresponding Author:** Albère JA Köke, Department of Rehabilitation Medicine, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands.

Received: March 04, 2022

Published: June 09, 2022

Abstract

Purpose: The Pain Disability Index (PDI) questionnaire assesses the reduction of disability, the main goal of chronic musculoskeletal pain (CMSP) rehabilitation. Moreover, the PDI is widely used and its clinimetric properties seem to be positive. This study aimed to provide reference values to interpret change scores of individuals on the PDI, for daily clinical practice, and to assess its responsiveness.

Methods: The smallest detectable change (SDC), and minimal clinically important change (MCIC) of the PDI were investigated in 374 patients with CMSP, participating in an interdisciplinary pain rehabilitation programme (IPRP). Two Global Perceived Effect (GPE) statements were used to investigate these properties and values.

Results: The hypotheses concerning correlations of the GPEs with the change scores on the PDI could not be confirmed. The SDC was evidently larger than the MCIC, independent of the GPE used.

Conclusion: The PDI appeared not to be able to distinguish clinically important change from measurement error in individual patients for a broad range of change scores. For now, the calculated SDC can be used as a cut-off score to be fairly certain of real improvement. Further research is necessary to enhance usability.

Keywords: Outcome; Disability; Chronic Pain; Clinically Relevant; Responsiveness

Abbreviations

WPN: Dutch Working group on Pain Rehabilitation; DDPR: Dutch Dataset Pain Rehabilitation; CMSP: Chronic Musculoskeletal Pain; SDC: Smallest Detectable Change; MCIC: Minimal Clinically Important Change; PROMs: Patient-reported Outcome Measures; PDI: Pain Disability Index; MUMC+: Maastricht University Medical Centre+; IRRP: Interdisciplinary Pain Rehabilitation Programme; GPE: Global Perceived Effect; ICC: Intraclass Correlation Coefficient; SEM: Standard Error of Measurement

Introduction

Public reporting of treatment outcomes in healthcare has become widespread [1]. Payers require the use of quality measures and are tying financial rewards to performance [2]. For this reason, the Dutch Working group on Pain Rehabilitation (WPN) has developed the Dutch Dataset Pain Rehabilitation (DDPR) to improve standardised measurement and reporting of outcomes in chronic musculoskeletal pain (CMSP) rehabilitation treatment and implemented it in thirty-two treatment facilities [3]. Ten of the facilities did not start or continue data collection [4]. One of the most mentioned reasons was the lack of reference values to interpret the change scores of individual patients [4].

Properties of an outcome measure like its interpretability and responsiveness to change are required to use it routinely in clinical practice and for the evaluation of the effectiveness of care [5]. The smallest detectable change (SDC) indicates how much change scores can vary in stable patients, reflecting measurement error [6]. The minimal clinically important change (MCIC) represents the smallest change in a score that patients perceive as important [6]. These values can help a health professional to interpret change scores. Only if the SDC is smaller than the MCIC, it is possible to distinguish clinically important change from measurement error in individual patients with a large amount of certainty [7]. Responsiveness reflects the ability of an instrument to detect change over time in the construct to be measured [8]. Only when an outcome measure is responsive, it can be used to objectify a difference over time.

The SDC and MCIC of the psychosocial patient-reported outcome measures (PROMs) of the DDPR have already been investigated [9]. These PROMs were shown to be unable to distinguish clinically important change from measurement error in individual patients for a broad range of change scores [9]. In the DDPR, the primary PROM to assess effectiveness of pain rehabilitation is the Pain Disability Index (PDI). The PDI assesses pain-related disability [10], which is in line with the main aims of pain rehabilitation to reduce pain-related disability and increase participation [3]. The PDI is widely used [11-13]. Moreover, the PDI is short and easy to comprehend [13]. Previous studies stated that the Dutch PDI is responsive to change [13,14] and that the PDI can detect real change [14]. However, due to different contexts (e.g., population) and methodology used (e.g., methods to calculate the SDC), the results of these studies cannot be generalised to the overall interdisciplinary pain rehabilitation setting. We aim to provide reference values to interpret change scores of individuals on the PDI and to assess its responsiveness, investigated in a heterogeneous group of patients with CMSP.

Materials and Methods

Methods

Setting and research design

This was a retrospective study with a single-group repeated-measures design. The participating tertiary care centres were Maastricht University Medical Centre+ (MUMC+) (Maastricht), rehabilitation centre Adelante (Hoensbroek), and Laurentius hospital (Roermond). Participants were included between December 2008 and April 2015. The participants had to report musculoskeletal pain for more than three months, and were referred theretofore to an interdisciplinary pain rehabilitation programme (IPRP). To start with the rehabilitation programme, participants had to be willing to improve their daily functioning despite the pain, while addressing psychosocial factors that seemed to contribute to the maintenance of pain-associated disability. Physiatrists assessed the eligibility of potential participants for this programme. If included, participants attended an IPRP during, in principle and mostly, twelve weeks. The programme involved a combination of physical, psychological (mostly graded activity or exposure in vivo), educational, and/or work-related components and was delivered by an interdisciplinary team of healthcare providers as a physiatrist, physiotherapist, occupational therapist, and psychologist.

Data collection

Before the first appointment with the physiatrist (T1), and after completing the IPRP (T2), data were routinely collected as described by the DDPR. Participants completed the questionnaires at home, either web-based or on paper. The data were stored in electronic patient records of the participating centres. In the Netherlands, no permission from a medical ethical committee is required for the evaluation of outcomes of care based solely on data derived from medical records (The Central Committee on Research Involving Human Subjects). All participants gave written informed consent, stating that the data could be used for scientific research.

Outcome measures

Pain disability index

The PDI is a seven-item questionnaire, which asks the respondent to rate the degree to which pain interferes with functioning in seven areas of activity: family/home responsibilities, recreation, social activity, occupation, sexual behaviour, self-care, and life-support activity [10]. For each of the seven areas, respondents rate their level of disability on a scale ranging from 0 (no disability) to 10 (total disability). The total score is computed by summing the scores on all items, and ranges from 0 to 70 [10,15]. Higher scores reflect higher interference of pain with daily activities. In this study, the PDI in the Dutch language is used. The test-retest reliability of the Dutch PDI is good, the intraclass correlation coefficient (ICC) is 0.78 [12], as well as its internal consistency. There is also evidence for the construct validity of the Dutch PDI [12].

Global perceived effect

Two Global Perceived Effect (GPE) questions were used for the participants to rate whether and how much their condition had improved or deteriorated after completing the IPRP. Participants 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'). These GPE questions reflect the 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. Participants rated the questions on a 5-point Likert scale with the options 1 'clearly improved'; 2 'improved'; 3 'unchanged (neither better, nor worse)'; 4 'worse'; and 5 'clearly worse'. Patient ratings of the extent to which one has improved or deteriorated are regarded as useful for both validation and exploring questionnaire interpretability [16] and have even been recommended as a core outcome measure for chronic pain research [17,18].

Data analysis

Participants who completed the entire PDI at T1 and T2 and answered at least one of the GPE questions, 'complete cases', were included in the data analyses. In case of a missing answer on one of the two GPEs, no analyses concerning the missing GPE were performed. Data analyses were based on the COSMIN guidelines [8].

To classify participants as 'improved', 'unchanged', or 'deteriorated', the answers to the GPE questions were trichotomised. Participants who indicated that they had '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 PDI for the whole group and for the 'improved', 'unchanged', and 'deteriorated' subgroups were calculated by subtracting the mean follow-up score (T2) from the mean baseline score (T1). By doing so, a positive change score means a reduced interference of pain with functioning in the seven areas.

To determine the SDC, first, the standard error of measurement (SEM) was calculated. The SEM agreement take the systematic difference between test and retest into account while the SEM consistency ignore systematic differences [19]. The SEM agreement was calculated by taking the square root of the within-subject variance (variance due to systematic differences between 'true' scores of patients σ^2_p + residual variance (i.e. random error variance) between test moments $\sigma^2_{\text{moments of measurement}}$ ($\sqrt{(\sigma^2_p+\sigma^2_{\text{moments of measurement}})}$) of participants categorised as 'unchanged' on the GPE [6]. Thus, this method only takes 'unchanged' participants into account. To be 95% confident that the observed improvement was a real improvement and not caused by measurement error, not taking type II errors into account, the SDC was calculated as $1.64\sqrt{2}$SEM agreement, in which 1.64 represents the z-score corresponding to 95% confidence interval (CI) (one-sided) [20,21]. If one takes type II errors into account, the SDC should be 4 x SEM [7]. We calculated the SDC using the first as well as the second method.

The MCIC was calculated from the mean change in 'improved' participants minus the mean change in 'unchanged' participants [7,22]. In this study, only the SDC and MCIC in terms of improvement were examined, because the group of participants who indicated to have experienced deterioration was too small to perform adequate analyses. All statistical analyses were performed using SPSS22 for Macintosh (IBM Corp, Armonk, NY). A construct approach, using hypotheses testing, was adopted to assess responsiveness

[6]. In the first place, floor and ceiling effects were examined and considered to be present if, at baseline (T1), more than 15% of the participants reached the maximum or minimum score respectively [23]. Secondly, the change scores on the PDI were compared with the self-indicated changes on the GPE's. Correlation coefficients were calculated by Spearman's rho, as the GPEs were measured on an ordinal scale and the outcomes of the PDI were continuous [6]. We formulated hypotheses regarding the expected correlations: the correlation coefficient between the change score on the PDI and the answer on the GPE would be moderate, between r = 0.6 and r = 0.8, for both GPEs. We took into account that both the outcome measure and the GPE answer would probably be accompanied by a degree of measurement error [6]. To confirm that the PDI was responsive to change, we decided that the hypothesis had to be confirmed for both GPEs.

Results and Discussion

Characteristics of the participants

The number of participants with complete data for the GPE 'physical activity' was 374 and for the GPE 'coping' 372. The mean age was 45.5 years (SD = 11.0). Participants indicated a wide variety of pain sites. The mean pain intensity was 6.6 on an 11-NRS (SD = 1.8). In 51% of the participants, the pain existed for more than five years. Extensive information concerning the population is displayed in table 1.

Age (years) (mean (SD))	**45.5 (11.0)**
Sex (%)	
Male	34.1
Female	65.9
Marital status (%)	
Single	23.5
Married/in a relationship	76.5
Education (%)	
Higher level	20.4
Average level	38.2
Lower level	41.4
Site of pain referred for[a] (%)	
Head	38.8
Face/throat	16.9
Neck	64.5
Shoulder(s)/upper back	67.0
Arm(s)	56.6
Hand(s)/finger(s)	52.2
Chest/stomach	24.7
Lower back	80.6
Hip(s)	56.1
Upper leg(s)/knee(s)	64.2
Ankle(s)/feet/foot	52.5
Elsewhere	36.8
Psychological counselling in the past (%)	
Yes	69.7
No	30.3
Work status (%)	
Employed/student	45.8
Unemployed/not a student	54.2
Self-rated health (%)	
Good or very good	32.4
Fair	42.9
Poor	24.6
Duration of complaints (%)	
Less than five years	49.6
More than five years	50.4
Pain intensity (NRS[b] (0-10)) (mean (SD))	6.6 (1.8)
Disability due to pain (PDI) (mean (SD))	40.1 (11.5)
[a]Participants could indicate more pain sites for which they were referred for treatment. [b]Numeric Rating Scale	

Table 1: Baseline characteristics of participants (n = 374).

Interpretability and responsiveness of the PDI

In table 2, the outcomes of the calculations concerning the SEM, SDC, and MCIC are displayed. The GPE with the highest correlation coefficient was 'physical activity'. Based on this GPE, the MCIC did not exceed the SDC: the SDC, calculated as $1.64*\sqrt{2}*SEM$, being 20.67 and the MCIC 5.36. Also using the GPE 'coping', the SDC was higher than the MCIC: the SDC being 24.12 and the MCIC 6.19. When the SDC was calculated as 4*SEM, the SDC was even higher for both GPEs.

GPE	SEM	4.0*SEM	1.64*√2*SEM	MCIC
Physical activity	8.9	35.6	20.7	5.4
Coping	10.5	42.0	24.1	6.2

Table 2: Outcomes for the PDI: SEMs, SDCs and MCICs according to both GPEs.

Floor and ceiling effects were considered not to be present, with 0.5% of the participants at the minimum and 2.3% at the maximum of the score range at baseline.

In table 3, the change scores and Spearman's rho, according to each GPE, are displayed. The mean change score on the PDI for the whole group was 11.0 (SD = 13.4). According to the GPE 'physical activity', 'improved' participants had a mean change score of 12.1 (SD = 13.4), and 'unchanged' participants had a mean change score of 6.7 (SD = 10.8). According to the GPE 'coping', mean change scores were, respectively, 11.8 (SD = 13.1) and 5.6 (SD = 13.8).

ΔT1-T2 (n = 374)	11.0 (13.4)	Spearman's rho
GPE Physical activity		r = 0.227
Improved (n = 317)	12.1 (13.4)	
Unchanged (n = 48)	6.7 (10.8)	
Deteriorated (n = 9)	-5.3 (11.5)	
GPE Coping		r = 0.174
Improved (n = 330)	11.8 (13.1)	
Unchanged (n = 37)	5.6 (13.8)	
Deteriorated (n = 5)	-1.0 (8.9)	
Values are mean (SD).		

Table 3: Outcomes for the PDI: change scores and Spearman's rho correlation coefficients according to both GPEs.

Spearman's rho correlation coefficient for the change scores on the PDI with the GPE 'physical activity' was 0.227, meaning that the first hypothesis was refuted. The correlation coefficient of the change scores on the PDI with the GPE 'coping' was 0.174, so the second hypothesis was refuted as well.

Discussion

The aim of this study was to provide reference values to interpret change scores on the PDI of individual patients with CMSP for daily clinical practice and to assess its responsiveness, following an IPRP. The PDI was shown unable to distinguish clinically important change from measurement error in individual patients for a broad range of change scores. Only as to patients with change scores higher than the calculated SDC of about 20 (when using the formula 1.64*√2*SEM) a healthcare professional can be almost certain that he or she has experienced real change as a result of the IPRP. The PDI was also not responsive according to the GPEs used.

Independent of the GPE used for the calculations, the SDC was evidently larger than the MCIC. Furthermore, the correlations between the change scores on the PDI and the change reported on the GPEs did not meet the prior stated hypotheses: the correlations were drastically lower than expected. To our knowledge, no study in a comparable patient population using comparable methods had been performed. A recently published study of Beemster., *et al*. indicated results conflicting with the results of our study: a higher MCIC and a lower SDC. They found a MCIC of 12.5 for the total study sample and a SDC of 3.4 for individuals, and concluded that the PDI can detect real change [14]. However, the population included patients with subacute musculoskeletal pain and this study concerned vocational rehabilitation. This may have resulted in a younger, homogenous population with unimodal problems, in which relevant change is achieved relatively quickly and in unequivocal terms (when returned to work), which can result in a smaller standard deviation of scores. Moreover, Beemster., *et al*. [14] used the optimal cut-off point to calculate the MCIC, which might result in unchanged patients being classified falsely as improved [24], potentially resulting in an overestimation of the MCIC. Furthermore, Beemster., *et al*. [14] calculated the standard error of measurement (SEM) using the ICC from another study, which could have led to an underestimation of the SDC and an overestimation of the MCIC [25]. In addition, Beemster., *et al*. [14] defined patients who had no statistically significant PDI-change, (change scores from -6 to +6) as 'unchanged', not patients who defined themselves as 'unchanged' on the GPE. This resulted in a group of 'unchanged' patients with a very low standard deviation of PDI change scores, which automatically results in a low SDC, regardless of the ICC used in the formula. Lastly, the GPE used was different: the patients were

asked to indicate how much their (pain) complaints had changed. In 2012 Soer., *et al*. [13] concluded that the total score of the PDI is responsive to change and that change is clinically important if the PDI score changes 9.5 points, based on a GPE 'complaints', and 8.5 points when using a GPE 'self-care'. This study concerned patients with neck and back pain only. The optimal cut-off point was used to determine the MCIC in this study as well. In 2013, Soer., *et al*. [12] reported a SDC of 17.9 in patients with acute back pain, chronic low back pain and widespread pain. In 2020 Borghuis., *et al*. [26] performed interviews with patients with CMSP, which showed that the PDI may not fully capture all relevant outcomes and may underestimate meaningful outcomes of pain rehabilitation. PDI items as sexual behaviour, self-care and life-support activities were often described as irrelevant. Furthermore, high disability scores on one or more items that are highly relevant for patients may be underestimated by using the total score of the PDI [26]. These issues, indicated by patients themselves, might contribute to the low responsiveness.

In our opinion, a strength of the current study is the conscientious way in which the (calculation) methods were established, in which we strictly followed the COSMIN guidelines. The GPEs, particularly the GPE 'physical activity', at face value, better match the construct that the PDI assesses, than the GPEs used in previous studies. We included two GPE questions to assess whether or not the difference in the GPE itself would influence the results and as such improved our insight in the possible range of change scores and its interpretability. Another strength is the large sample size (n = 317 and n = 330) of the 'improved' groups and the nearly recommended number (n = 50) of 'unchanged' participants for the GPE 'physical activity' (n = 48) [23]. Furthermore, generalisability is high as the participants showed a wide variety of pain sites, which resembles the variety seen in clinical practice. However, this study also has its limitations. To start, the standard deviation of the change scores of self-proclaimed 'unchanged' participants is large. We cannot immediately attribute this to low responsiveness. This may also be due to the use of a 5-point instead of 7-point Likert scale, possibly making the choice between 'unchanged' and 'improved' too substantial for some participants, and for whom the option 'a little improved' would have better suited their experience. Taking participants with 'little improvements' into the 'improved' group could have changed the results. Also, only 'complete cases' were included. This might have caused selection bias of very dissatisfied or very satisfied participants, because these participants might have been more willing to take the effort to fill in the questionnaires at both the start and end of the IPRP.

A manner to use the outcomes of this study in daily clinical practice, would be to use a change score of 20 as a benchmark for real change 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 [7]. As mentioned in earlier articles, PROMs can be helpful in the communication between healthcare providers and patients [5,27], and the PDI can be continued to be used for this purpose. However, we have to state that our results call into question its use in situations in which the effectiveness of rehabilitation programmes in patients with CMSP is to be quantified by external parties. For future research, first, we recommend a study in which more methods that are currently used to calculate the SDC and MCIC are compared, for instance the calculation method we used and the limits of agreement/optimal cut-off point method, to evaluate the difference in outcomes. However, caution is needed when interpreting the results when using a GPE to determine the SDC and MCIC, as it has been known that current status and themes related to the treatment process are likely to influence the rating of the GPE [28-31]. If there is a preference to keep using the PDI in clinical practice, and to get more insight in how PDI scores behave naturally, it could be of added value to perform a study that evaluates the natural fluctuation of PDI scores in patients with CMSP when they do not participate in a rehabilitation programme. Furthermore, the baseline score will have an influence on the possible and expected magnitude of change of that score and it's MCIC. For healthcare professionals, it may be inspiring to have an overview of how to interpret change scores according to a baseline score. Beemster., *et al*. [14] already took baseline scores into account, looking at quartile subgroups, for their population. However, in line with the conclusion of Borghuis., *et al*. [26] we also suggest further research regarding other outcome measures than the PDI that try to capture personal, relevant pain rehabilitation goals. It is advisable to examine whether more objective outcome measures (for example measurable behaviour of patients as return to work, doctor visits, hours of social contact, or engagement in hobbies) are more responsive and represent the experienced effectiveness of pain rehabilitation programmes more adequately. It is imaginable that factual differences in work and social contact

are taken into account by patients when they rate the effectiveness of an IPRP. Furthermore, this can give insight into the cost-effectiveness of the IPRP. We also continue to recommend involving the honest opinion of clinicians in the assessment procedure of effectiveness of therapy because of their medical knowledge, experience and personal contact with patients.

Until a more appropriate (set of) outcome measure has been developed, the change scores of the PDI as an outcome of an IPRP should be used with great caution. Further research has to be conducted to enhance the usability of the PDI and to develop (a set of) outcome measures that better represent the experience of patients.

Conclusion

Until a more appropriate (set of) outcome measure has been developed, the change scores of the PDI as an outcome of an IPRP should be used with great caution. Further research has to be conducted to enhance the usability of the PDI and to develop (a set of) outcome measures that better represent the experience of patients.

Acknowledgements

No acknowledgements.

Conflict of Interest

All authors state that there are no conflicts of interests.

Bibliography

1. McGlynn EA and EA Kerr. "Creating Safe Harbors for Quality Measurement Innovation and Improvement". *JAMA* 315.2 (2016): 129-130.

2. McGlynn EA., *et al*. "Reimagining quality measurement". *The New England Journal of Medicine* 371.23 (2014): 2150-2153.

3. Köke AJA., *et al*. "Dutch Dataset Pain Rehabilitation in daily practice: Content, patient characteristics and reference data". *European Journal of Pain* 21.3 (2017): 434-444.

4. Köke AJA., *et al*. "Implemetatie en uitbreiding Nederlandse Dataset Pijnrevalidatie". OPR, Editor (2012).

5. Higginson IJ and AJ Carr. "Measuring quality of life: Using quality of life measures in the clinical setting". *BMJ* 322.7297 (2001): 1297-300.

6. de Vet. "Henrica Measurement in Medicine: a practical guide". Cambridge University Press (2011).

7. Terwee CB., *et al*. "Linking measurement error to minimal important change of patient-reported outcomes". *Journal of Clinical Epidemiology* 62.10 (2009): 1062-1067.

8. Mokkink LB., *et al*. "The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study". *Quality Life Research* 19.4 (2010): 539-549.

9. Pulles ANTD., *et al*. "The responsiveness and interpretability of psychosocial patient-reported outcome measures in chronic musculoskeletal pain rehabilitation". *European Journal of Pain* 24.1 (2020): 134-144.

10. Tait RC., *et al*. "The Pain Disability Index: psychometric and validity data". *Archives of Physical Medicine and Rehabilitation* 68.7 (1987): 438-441.

11. McKillop AB., *et al*. "Measuring participation in patients with chronic back pain-the 5-Item Pain Disability Index". *Spine Journal* 18.2 (2018): 307-313.

12. Soer R., *et al*. "Extensive validation of the pain disability index in 3 groups of patients with musculoskeletal pain". *Spine (Phila Pa 1976)* 38.9 (2013): E562-568.

13. Soer R., *et al*. "Responsiveness and minimal clinically important change of the Pain Disability Index in patients with chronic back pain". *Spine (Phila Pa 1976)* 37.8 (2012): 711-715.

14. Beemster T., *et al*. "The interpretation of change score of the pain disability index after vocational rehabilitation is baseline dependent". *Health Quality and Life Outcomes* 16.1 (2018): 182.

15. Chibnall JT and RC Tait. "The Pain Disability Index: factor structure and normative data". *Archives of Physical Medicine and Rehabilitation* 75.10 (1994) 1082-1086.

16. Guyatt GH., *et al*. "A critical look at transition ratings". *Journal of Clinical Epidemiology* 55.9 (2002): 900-908.

17. Dworkin RH., *et al*. "Core outcome measures for chronic pain clinical trials: IMMPACT recommendations". *Pain* 113.1-2 (2005): 9-19.

18. Fischer D., *et al*. "Capturing the patient's view of change as a clinical outcome measure". *JAMA* 282.12 (1999): 1157-1162.

19. de Vet HC., *et al*. "When to use agreement versus reliability measures". *Journal of Clinical Epidemiology* 59.10 (2006): 1033-1039.

20. Andersson EI., *et al*. "Performance tests in people with chronic low back pain: responsiveness and minimal clinically important change". *Spine (Phila Pa 1976)* 35.26 (2010): E1559-1563.

21. de Vet H., *et al*. "Three ways to quantify uncertainty in individually applied "minimally important change" values". *Journal of Clinical Epidemiology* 63.1 (2010): 37-45.

22. Jaeschke R., *et al*. "Measurement of health status. Ascertaining the minimal clinically important difference". *Control Clinical Trials* 10.4 (1989): 407-415.

23. Terwee CB., *et al*. "Quality criteria were proposed for measurement properties of health status questionnaires". *Journal of Clinical Epidemiology* 60.1 (2007): 34-42.

24. Ostelo RW and HC de Vet. "Clinically important outcomes in low back pain". *Best Practice and Research: Clinical Rheumatology* 19.4 (2005): 593-607.

25. Beckers L and R Smeets. "Methodological and interpretive concerns about Beemster *et al.*'s article 'The interpretation of change score of the pain disability index after vocational rehabilitation is baseline dependent': a letter to the editor". *Health Quality Life Outcomes* 18.1 (2020): 301.

26. Borghuis EMC., *et al*. "Assessing discrepancies in outcomes of pain rehabilitation: "these questionnaires don't measure results that are relevant to me". *Disability and Rehabilitation* 42.16 (2020): 2374-2380.

27. Deyo RA and WB Carter. "Strategies for improving and expanding the application of health status measures in clinical settings. A researcher-developer viewpoint". *Medical Care* 30.5 (1992): MS176-186; discussion MS196-209.

28. Evans R., *et al*. "I know it's changed": a mixed-methods study of the meaning of Global Perceived Effect in chronic neck pain patients". *European Spine Journal* 23.4 (2014): 888-897.

29. Grovle L., *et al*. "Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status". *Journal of Clinical Epidemiology* 67.5 (2014): 508-515.

30. Kamper SJ., *et al*. "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* 63.7 (2010): 760-766 e1.

31. Schmitt J and RP Di Fabio. "The validity of prospective and retrospective global change criterion measures". *Archives of Physical Medicine and Rehabilitation* 86.12 (2005): 2270-2276.