

Level of patient–physician agreement in assessment of change following conservative rehabilitation for shoulder pain

Stephanie D. Moore-Reed¹, W. Ben Kibler²,
Heather Bush³ and Tim L. Uhl⁴

Shoulder & Elbow
2017, Vol. 9(2) 127–132
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1758573216658799
journals.sagepub.com/home/sel
SAGE

Abstract

Background: Assessment of health-related status has been shown to vary between patients and physicians, although the degree of patient–physician discordance in the assessment of the change in status is unknown.

Methods: Ninety-nine patients with shoulder dysfunction underwent a standardized physician examination and completed several self-reported questionnaires. All patients were prescribed the same physical therapy intervention. Six weeks later, the patients returned to the physician, when self-report questionnaires were re-assessed and the Global Rating of Change (GROC) was completed by the patient. The physician completed the GROC retrospectively. To determine agreement between patient and physician, intra-class correlation (ICC) coefficient and Pearson's *r* using the 15-point GROC and weighted kappa using a consolidated three-point GROC were calculated.

Results: Utilizing the 15-point GROC, complete agreement was observed in 37 of 99 patients (37%). ICC and Pearson's *r* between patient and physician were 0.62 and 0.63, respectively. Utilizing a consolidated three-point GROC, complete agreement was observed in 76 of 99 patients (77%). Weighted kappa was 0.62.

Conclusions: Assessment of change reported by the patient demonstrates moderate to good agreement with physician assessment. These findings indicate that the GROC does reflect and represent similar assessment of change in health status by patients and physicians. This can aid discussion of both past treatment results and future treatment plans.

Keywords

conservative treatment, orthopaedics, rehabilitation, self-report questionnaires, shoulder pain

Date received: 25th November 2015; accepted: 14th June 2016

Introduction

Health-related assessment ratings have been shown to vary between patients and clinicians, resulting in patient–clinician discordance. This discordance has been reported in assessments of disease severity,^{1–6} physical functioning,^{5,7–12} pain^{8,11,13} and quality of life¹³ in a variety of acute and chronic pathologies and select musculoskeletal disorders. In general, patients tend to rate themselves as being more severely impacted compared to physician ratings.^{1,2,6,8} However, there is some evidence that this may vary depending upon the pathology being examined.^{5,9,13} The magnitude of disagreement and whether clinicians overestimate or underestimate impairments and disease severity appears to vary based on the disease.^{5,9,13} This may reflect clinicians tending to pre-determine the effects that a health condition will have on a

patient based on the perceived generalized severity of the condition, rather than the individual patient's characteristics.

Determining the most 'true' assessment of a patient's health or healing status can be challenging because physicians and patients are likely to factor different information into their judgment. Physicians are often thought to

¹Department of Kinesiology, California State University, Fresno, CA, USA

²Shoulder Center of Kentucky, Lexington, KY, USA

³Department of Biostatistics, University of Kentucky, Lexington, KY, USA

⁴Division of Athletic Training, University of Kentucky, Lexington, KY, USA

Corresponding author:

Stephanie D. Moore-Reed, ATC, 5275 N Campus Dr M/S SG28 Fresno, CA 93740, USA.

Email: sdmreed@csufresno.edu

consider pain as a secondary result of a pathology or anatomic abnormality.¹⁴ Evidence suggests physicians use their clinical experience,^{1,2} the patient's disease duration¹⁵ and objective findings (e.g. clinical signs and symptoms and laboratory tests)^{1,3,4,14–16} to determine their assessment. Patients, on the other hand, may not understand the abnormalities explained by laboratory tests or diagnostic imaging,¹ and judge the severity of their injuries on their individual experience.¹⁷ Patients also sense pain in a multifactorial manner that may be experienced even in the absence of pathology¹⁴ and factor pain into their assessment.^{4,6} A study of patients with lupus identified that patient-reported pain accounted for 20% of the variance in patient-reported disease activity, although it was not a significant predictor of physician-reported disease activity.⁴ Additionally, patients and physicians may have different expectations with regard to the progression or outcome of the intervention, or what constitutes a satisfactory progression in treatment or a good outcome.¹⁸

The impact of patient–physician discordance in orthopaedics and shoulder surgery is relatively unknown. Although discordance has been examined in one-time assessments of disease state or impairment, few studies have examined the effect that these differences in the perception of the results of treatment may have on the assessment of change over time or outcome following an intervention.^{3,18,19} Two studies have reported fair to good patient–clinician agreement in patients with low back pain¹¹ and disorders of the neck–shoulder region.¹² Agreement on the assessment of outcome has been examined relative to pain and overall satisfaction, although only in a cohort of post-operative patients following total hip arthroplasty.¹⁸ The need to examine the agreement or discordance is important because this is not an issue of whose assessment is right or wrong; each perspective is equally valid. The patient's perspective should always be considered by the healthcare provider because the patient is actually experiencing the treatment and is affected by the results. However, clinicians are responsible for the content, timing and direction of treatment and are therefore most influential in guiding the patients' course of medical treatment. Determining whether a patient has improved is an important factor when making treatment decisions for both the physician and patient. If there is discordance in patient-reported and physician-reported assessment of change among patients seeking medical care for shoulder pain, strategies should be developed to improve agreement and communication, aiming to manage the discordance and perhaps develop other tools to assess change that will minimize discordance. Therefore, the purpose of the present study was to examine the level of agreement between patient and physician assessment of change using a Global Rating of Change (GROC) scale in a

cohort of patients being treated for shoulder problems. We hypothesized that there would be moderate (66% to 75%) agreement between patient-reported and physician-reported assessment of change.

Materials and methods

Subjects

Data from 99 subjects were used in this analysis [mean (SD) age = 41 (12) years; height = 175 (10) cm; weight = 84 (19) kg, 66 males]. These data come from a larger study in which patients were enrolled prospectively over 2 years. Of the 191 eligible subjects (220 patients approached, 211 enrolled, 20 withdrew), 99 had all of the data required for the present analysis. Patients reporting to the Lexington Clinic Orthopaedics and Sports Medicine Center with shoulder pain were identified as potential subjects. Patients were eligible for enrollment if they presented with a clinical history consistent with dysfunction as a result of musculoskeletal shoulder injury, reported pain with overhead activity and were aged between 15 years and 60 years. Patients were excluded if they demonstrated signs and symptoms consistent with cervical radiculopathy,²⁰ adhesive capsulitis,²¹ glenohumeral arthritis²² or reported tingling/numbness in the upper extremity, surgery on the involved shoulder within the past year, or steroid injection within the last month.

Patients who met the criteria and consented to participate underwent a full standardized examination by the physician and completed a battery of self-reported questionnaires including a numeric pain rating scale (NPRS; 0 = no pain, 10 = worst pain) and the Quick Disabilities of the Arm, Shoulder and Hand (QuickDASH; 0 = no disability, 100 = severe disability). All patients read and signed an informed consent form prior to their enrollment in the study that was approved by the institutional review boards of the University of Kentucky and Lexington Clinic. Subjects were prescribed physical therapy and provided with a standardized rehabilitation protocol to take to the therapist of their choosing.

Procedures

Patient-oriented assessment of change was collected prospectively at the time of physician follow-up [mean (SD) 6 (1) weeks]. Subjects completed a GROC to assess perceived improvement. The GROC is a 15-item scale ranging from 'a very great deal worse' to 'a very great deal better' (Fig. 1).²³ Subjects were instructed to select the statement that best represented their perceived change in functional status subsequent to the initial evaluation. Physician assessment of change was carried out retrospectively at the end of the enrollment period. The

	<u>0</u> About the same	
<u>-1</u> Almost the same, hardly any worse at all		<u>1</u> Almost the same, hardly any better at all
<u>-2</u> A little worse		<u>2</u> A little better
<u>-3</u> Somewhat worse		<u>3</u> Somewhat better
<u>-4</u> Moderately worse		<u>4</u> Moderately better
<u>-5</u> A good deal worse		<u>5</u> A good deal better
<u>-6</u> A great deal worse		<u>6</u> A great deal better
<u>-7</u> A very great deal worse		<u>7</u> A very great deal better

Figure 1. Global Rating of Change scale.

treating physician, an orthopaedic surgeon (WBK), was provided with clinical notes from both the initial evaluation and follow-up visit for each patient and completed the same 15-point GROC scale. Intra-rater reliability was excellent [intra-class correlation (ICC)=0.929] and was established by having the physician rate the same 10 subjects at two separate times, with a minimum of 1 week between ratings.

Data reduction

The 15-point GROC was further consolidated into a three-point scale by collapsing response options into ‘better’ (GROC score ≥ +3), no change (−2 to +2) and ‘worse’ (GROC score ≤ −3) based on previously reported cut-offs used to identify clinically meaningful improvement.²⁴ Providing patients (or clinicians) with too many options may be of concern because the individual may have difficulty attaching meaning to each separate response choice.²³ By treating the 15-point scale as continuous, ICC and Pearson’s *r* calculations could be conducted, whereas the consolidated three-point scale allowed for confirmation of the findings with weighted kappa using a more simplified scale of better/no change/worse.

Statistical analysis

To assess patient–physician agreement, ICC, Pearson’s *r* correlation coefficient and linear weighted kappa were calculated. ICC and Pearson’s *r* were calculated using the responses on the 15-point GROC. ICCs were interpreted according to: <0.40 Poor, 0.40–0.75 Fair to Good and >0.75 Excellent.²⁵ Linear weighted kappa was calculated using the consolidated three-point scale (better, no change, worse). The strength of agreement for kappa was interpreted according to: <0.00

Poor, 0.00–0.20 Slight, 0.21–0.40 Fair, 0.41–0.60 Moderate, 0.61–0.80 Substantial and 0.81–0.99 Almost Perfect.²⁶ Maximum kappa was calculated according to Sim and Wright.²⁷ The maximum kappa value provides a more meaningful reference value for interpretation because inadequate variation in the data can result in artificially low kappa values.²⁷

Results

Utilizing the 15-point GROC scale, complete agreement between patient-reported and physician-reported GROC score was observed in 37 of 99 patients (37%). ICC and Pearson’s *r* were 0.62 and 0.63, respectively. Utilizing the consolidated three-point scale (better, no change, worse), complete agreement was observed in 76 of 99 patients (77%). Weighted kappa was 0.62 with a maximum weighted kappa was determined to be 0.95. A bivariate relationship between patient-reported and physician-reported GROC scores is depicted in a scatterplot (Fig. 2).

Discussion

One of the keystones of the doctor–patient relationship is that they are both in agreement regarding the results of treatment and the direction of future care. This requires agreement on the status of these treatment efforts. One key element would be the change in functional status as a result of treatment. The present study examined the patient–physician agreement or discordance related to assessment of change following rehabilitation in patients with shoulder pain. Our hypothesis of moderate agreement was supported, indicating that the GROC scale appears to reflect and represent the same degree of change perceived by each group. Overall, we observed moderate to good agreement. Our findings

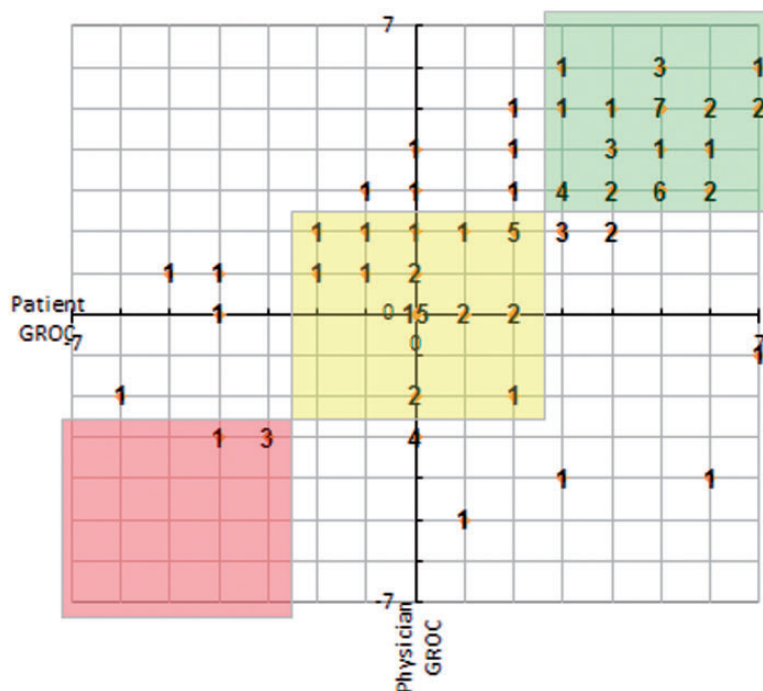


Figure 2. Patient–physician agreement plot. Points that fall within the green ('better', $n = 37$), yellow ('no change', $n = 35$) and red ('worse', $n = 4$) boxes indicate that the patient and physician both rated the patient in the same category. Points that fall outside of the boxes indicate disagreement between the patient and physician ($n = 23$). The values represent the number of patients represented by that data point. GROCC, Global Rating of Change.

indicate similar patient–physician agreement compared to previous research. Patient–physician agreement reported in the literature ranges from 58% to 77%.^{1,4,6,7,10,16,28} Our finding of 37% complete patient–physician agreement on the 15-point GROCC scale was expected to be lower because complete agreement was necessary. Using the three-point scale, we were able to examine more global agreement (i.e. did the patient and physician agree that the patient was better, the same or worse?) rather than matching exactly to a particular point on a 15-point scale. Complete patient–physician agreement using this three-point scale was 77%, which is at the high end of (although consistent with) previous studies. Our assessment of weighted kappa (0.62) is also higher than previous reports aiming to assess patient–clinician agreement (range 0.09 to 0.39).^{3,10,12}

Our findings of a higher agreement than previous literature may be because our patients did not report high pain severity or disability. Discordance between patients and physicians is known to be greater and more common in patients with more severe ratings of disease activity, impairment or pain.^{2,4,10,18} The mean (SD) rating of current pain on the NPRS was 4 (2) at initial evaluation and 3 (2) at follow-up. The mean (SD) QuickDASH at initial exam was 38 (18), indicating that our patients were approximately 40% disabled at initial evaluation. At follow-up, patients improved by a mean (SD) of 8 (15) points on the QuickDASH.

Our sample appears to represent the typical population of shoulder pain patients because our levels of pain and disability are consistent with patients with shoulder pain seeking care from an orthopaedic surgeon.^{29–31}

Limited research has explored the agreement or discordance in ratings of change over time with respect to functional health status. Patients with rheumatoid arthritis rated their change in global function over 3 months, as did their treating physician.¹⁹ The patient–physician relationship observed ($ICC = 0.64$, $r = 0.63$) was very similar to the findings of the present study.¹⁹ Patients with heart disease were asked to use a seven-point 'transition index scale' that appears to be quite similar to the GROCC to assess change in health-related quality of life in patients with heart disease.³ Poor agreement ($k = 0.09$ to 0.23) was identified between patients and physicians. The low agreement may be a result of the type of data collected and compared. A single global assessment made by the physician was compared with multiple domains assessed by the patients.³ In the present study, the same global assessment was performed by both the patient and physician, which appears to result in a higher agreement in the present study and in previous research.¹⁹

In the only previous study assessing patient–physician agreement in change over time involving an orthopaedic population, patient-reported assessment of pain and overall satisfaction following total hip arthroplasty was compared with physician assessment

using a visual analogue scale (VAS).¹⁸ Differences in patient and physician ratings of pain were statistically significantly different [mean (SD) 1.7 (2.6) cm and 1.1 (1.8) cm, respectively], although the difference was only 0.6 cm on the VAS. Reports of overall satisfaction between patient and physician were not significantly different [8.6 (2.1) cm and 8.8 (1.7) cm, respectively]. It was noted that the patient–physician agreement was notably worse among the patients with high pain or low satisfaction. Although this was an orthopaedic population, the cohort was postsurgical and no assessment of agreement (e.g., kappa, ICC) was provided, making it difficult to make direct comparisons with the present study. Our data provide the first examination of the assessment of change following conservative rehabilitation in an orthopaedic population.

We used a 15-point GROC to assess perceived change. The ‘global’, less specific nature of the GROC allows the patient to base their response on what is most important to them.²³ This was ideal for addressing the purpose of the present study in that we aimed to identify whether differences existed between perceptions of patients and clinicians. Test–retest reliability of the GROC within 24 hours was excellent in patients with musculoskeletal disorders (ICC range 0.90 to 0.99).³² One limitation of a global rating of change assessment is that it requires the patient to recall their previous condition with respect to their current status.²³ It has been suggested that GROC scores may be influenced by current status as the follow-up time increases.³² The three-point GROC showed a much higher percentage of complete agreement between patient and physician evaluations, probably as a result of limiting the available options. It may serve as a better basis for discussion between the patient and physician regarding the results of treatment, and therefore help to guide the discussion about future treatment plans.

Limitations

A few limitations of the present study should be noted with respect to interpreting these results accurately. First, patients completed the GROC at the time of their visit, whereas the physician completed the GROC retrospectively at the end of the enrollment period. The physician had his own notes to refer to when completing the GROC, although it may have been more timely to have the physician rate the patient using the GROC scale immediately after the visit. However, it was felt that a longer time interval could provide a more objective analysis of the amount of change, and performing the evaluations at one time would improve the consistency of the ratings. Additionally, the inclusion of a single physician may limit the generalizability of the results and a validation of the findings with additional physicians should be performed.

Our assessments examined change over time from baseline to follow-up. Although all patients were prescribed a standardized physical therapy intervention, several variables could have factored into the results including the expectation of treatment success, patient satisfaction with outcome or physician services¹⁸ and adherence to therapy. Future studies should account for those variables to further explain the patient–clinician relationship with regard to agreement on health-related assessment.

Conclusions

The results of the present study indicate that the physician-reported assessment of change demonstrated moderate to good agreement with the patient-reported assessment of change in patients with orthopaedic shoulder pain, which supports our hypothesis of moderate agreement between the two parties. This indicates that both the patient and physician are for the most part in agreement with respect to how the patient is responding to a non-operative intervention, suggesting that there is limited discordance in the treatment of orthopaedic conditions. The results suggest that the GROC can be used to represent both patient and physician assessments of the results of treatment. It can also serve as an effective means to facilitate the patient–physician dialogue, linking both stakeholders’ perceptions of the treatment so that both can understand the perceptions of the treatment, the results of the treatments, and the need for (as well as direction of) future treatments.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: The Lexington Clinic entered into a contract to pay the University of Kentucky for a doctoral student, S.D.M-R., as a part-time research assistant as she completed her doctoral degree.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical review and patient consent

All patients read and signed an informed consent form prior to their enrollment in the study that was approved by the institutional review boards of the University of Kentucky and Lexington Clinic.

References

1. Alarcon GS, McGwin G Jr, Brooks K, et al. Systemic lupus erythematosus in three ethnic groups. XI. Sources of discrepancy in perception of disease activity: a comparison of physician and patient visual analog scale scores. *Arthritis Rheum* 2002; 47: 408–13.
2. Dandorfer SW, Rech J, Manger B, Schett G and Englbrecht M. Differences in the patient’s and the

- physician's perspective of disease in psoriatic arthritis. *Semin Arthritis Rheum* 2012; 42: 32–41.
3. Wyrwich KW, Metz SM, Kroenke K, Tierney WM, Babu AN and Wolinsky FD. Triangulating patient and clinician perspectives on clinically important differences in health-related quality of life among patients with heart disease. *Health Serv Res* 2007; 42: 2257–74.
 4. Yen JC, Abrahamowicz M, Dobkin PL, Clarke AE, Battista RN and Fortin PR. Determinants of discordance between patients and physicians in their assessment of lupus disease activity. *J Rheumatol* 2003; 30: 1967–76.
 5. Sewitch MJ, Abrahamowicz M, Bitton A, et al. Psychosocial correlates of patient–physician discordance in inflammatory bowel disease. *Am J Gastroenterol* 2002; 97: 2174–83.
 6. Khan NA, Spencer HJ, Abda E, et al. Determinants of discordance in patients' and physicians' rating of rheumatoid arthritis disease activity. *Arthritis Care Res (Hoboken)* 2012; 64: 206–14.
 7. Berkanovic E, Hurwicz ML and Lachenbruch PA. Concordant and discrepant views of patients' physical functioning. *Arthritis Care Res* 1995; 8: 94–101.
 8. Dobkin PL, De Civita M, Abrahamowicz M, et al. Patient–physician discordance in fibromyalgia. *J Rheumatol* 2003; 30: 1326–34.
 9. Hidding A, van Santen M, De Klerk E, et al. Comparison between self-report measures and clinical observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. *J Rheumatol* 1994; 21: 818–23.
 10. Kwoh CK, O'Connor GT, Regan-Smith MG, et al. Concordance between clinician and patient assessment of physical and mental health status. *J Rheumatol* 1992; 19: 1031–7.
 11. Perreault K and Dionne CE. Patient–physiotherapist agreement in low back pain. *J Pain* 2005; 6: 817–28.
 12. Perreault N, Brisson C, Dionne CE, Montreuil S and Punnett L. Agreement between a self-administered questionnaire on musculoskeletal disorders of the neck–shoulder region and a physical examination. *BMC Musculoskel Disord* 2008; 9: 34.
 13. Chassany O, Le-Jeunne P, Duracinsky M, Schwalm MS and Mathieu M. Discrepancies between patient-reported outcomes and clinician-reported outcomes in chronic venous disease, irritable bowel syndrome, and peripheral arterial occlusive disease. *Value Health* 2006; 9: 39–46.
 14. Turk DC and Okifuji A. Assessment of patients' reporting of pain: an integrated perspective. *Lancet* 1999; 353: 1784–8.
 15. Hudson M, Impens A, Baron M, et al. Discordance between patient and physician assessments of disease severity in systemic sclerosis. *J Rheumatol* 2010; 37: 2307–12.
 16. Neville C, Clarke AE, Joseph L, Belisle P, Ferland D and Fortin PR. Learning from discordance in patient and physician global assessments of systemic lupus erythematosus disease activity. *J Rheumatol* 2000; 27: 675–9.
 17. Platt FW and Keating KN. Differences in physician and patient perceptions of uncomplicated UTI symptom severity: understanding the communication gap. *Int J Clin Pract* 2007; 61: 303–8.
 18. Lieberman JR, Dorey F, Shekelle P, et al. Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty. *J Bone Joint Surg Am* 1996; 78: 835–8.
 19. Hanly JG, Mosher D, Sutton E, Weerasinghe S and Theriault D. Self-assessment of disease activity by patients with rheumatoid arthritis. *J Rheumatol* 1996; 23: 1531–8.
 20. Wainner RS, Fritz JM, Irrgang JJ, Boninger ML, Delitto A and Allison S. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine* 2003; 28: 52–62.
 21. Griggs SM, Ahn A and Green A. Idiopathic adhesive capsulitis. A prospective functional outcome study of nonoperative treatment. *J Bone Joint Surg Am* 2000; 82A: 1398–407.
 22. Kelley MJ and Ramsey ML. Osteoarthritis and traumatic arthritis of the shoulder. *J Hand Ther* 2000; 13: 148–62.
 23. Kamper SJ, Maher CG and Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther* 2009; 17: 163–70.
 24. Young IA, Cleland JA, Michener LA and Brown C. Reliability, construct validity, and responsiveness of the neck disability index, patient-specific functional scale, and numeric pain rating scale in patients with cervical radiculopathy. *Am J Phys Med Rehabil* 2010; 89: 831–9.
 25. Fleiss JL. *The design and analysis of clinical experiments*. New York, NY: Wiley, 1986.
 26. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–74.
 27. Sim J and Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005; 85: 257–68.
 28. Scheuer E, Steurer J and Buddeberg C. Predictors of differences in symptom perception of older patients and their doctors. *Fam Pract* 2002; 19: 357–61.
 29. Polson K, Reid D, McNair PJ and Larmer P. Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. *Man Ther* 2010; 15: 404–7.
 30. Mintken PE, Glynn P and Cleland JA. Psychometric properties of the shortened disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and Numeric Pain Rating Scale in patients with shoulder pain. *J Shoulder Elbow Surg* 2009; 18: 920–6.
 31. Gummesson C, Ward MM and Atroshi I. The shortened disabilities of the arm, shoulder and hand questionnaire (QuickDASH): validity and reliability based on responses within the full-length DASH. *BMC musculoskeletal disorders* 2006; 7: 44.
 32. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC and Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010; 63: 760–6e1.