The Reproducibility and Responsiveness of a Patient-Specific Approach: A New Instrument in Evaluation of Treatment of Temporomandibular Disorders

Annemiek Rollman, PT
PhD Student

Machiel Naeije, PhD
Biophysicist, Professor, and Chair

Corine M. Visscher, PT, PhD
Assistant Professor and Epidemiologist

Department of Oral Kinesiology
Academic Centre for Dentistry Amsterdam (ACTA)
Research Institute Move
University of Amsterdam
and VU Univeristy
Amsterdam, The Netherlands

Correspondence to:
Dr C. M. Visscher
Department of Oral Kinesiology
Academic Centre for Dentistry Amsterdam (ACTA)
Louwesweg 1
1066EA Amsterdam
The Netherlands
Fax: +31-20-5188414
Email: c.visscher@acta.nl

Aims: *To evaluate the choice of activities on the Patient Specific Approach (PSA) in a sample of temporomandibular disorder (TMD) patients and to determine the clinimetric properties of the visual analog scale (VAS) scores of the PSA, in terms of reproducibility and responsiveness.* ***Methods:*** *At treatment start, TMD patients reported the PSA activity which represents the most important activity that is impaired due to their TMD complaints. The amount of hindrance during this activity was rated on a VAS. During two follow-up measurements, patients used the VAS to rate the amount of hindrance and appraised their overall complaints in terms of "much worsened," "slightly worsened," "remained stable," "slightly improved," or "much improved." To determine the reproducibility and responsiveness, an intraclass correlation coefficient and receiver operating characteristics-curve were then calculated.* ***Results:*** *Of the 132 patients who fulfilled baseline measurements, 13% reported an activity that is not included in existing TMD-disability questionnaires. The reproducibility of the VAS scores of the 78 patients who reported that their complaints had "remained stable" at second measurement was good (intraclass correlation coefficient = 0.73). The responsiveness of the PSA was high, and the cutoff score for important improvement, where sensitivity (0.85) and specificity (0.84) were as much as possible the same, was 58%.* ***Conclusion:*** *The PSA for TMD patients is a new and easy-to-use tool in treatment evaluation. Moreover, the VAS score of the PSA has good reproducibility and responsiveness.* J OROFAC PAIN 2010;24:101–105

Key words: patient specific approach, reproducibility, responsiveness, TMD, treatment effect

Musculoskeletal disorders are characterized by pain and dysfunction.[1–3] Traditionally, success of treatment in these disorders is measured with physical findings such as range of motion, muscle pain, and muscle strength. Nowadays, patient's self-report on the relief of complaints is considered to be increasingly important in medical decision-making and determining the success of treatment.[4]

One way to measure complaints is by the use of questionnaires. For example, functional (dis-) abilities can be evaluated by disorder-specific questionnaires, such as the Oswestry Low Back Pain Disability Questionnaire[5] or the Neck Disability Index.[6] For temporomandibular disorders (TMD), amongst others the Mandibular Function Impairment Questionnaire (MFIQ) and the Jaw Disability Index (JDI) can be used to measure disability in functions such as chewing and yawning.[7,8] A consequence of using

this type of questionnaire is that the activity that is of greatest importance to the patient may not be included in the questionnaire. To overcome this problem, in the treatment of low-back pain patients, a measure is introduced that is tailored to the patient's specific complaints.[9,10] In this "Patient Specific Approach" (PSA), the patient is asked to report the most important activity that is difficult to perform because of the back pain. The amount of hindrance the patient experiences when performing this activity is measured by a visual analog scale (VAS). The change in hindrance of the activity during treatment is then used as a measure for treatment effect. Such an approach, in which improvement is measured only in the activity that is relevant to the individual patient, may not only be promising in low-back pain patients, but also in treatment evaluation of TMD patients. Therefore, the aim of this study was to evaluate the choice of activities on the PSA in a sample of TMD patients and to determine the clinimetric properties of VAS scores of the PSA, in terms of reproducibility and responsiveness.

Materials and Methods

Study Population

Between summer 2006 and winter 2008, TMD patients from the department of Oral Kinesiology of the ACTA, were invited to participate in the study. Patients were excluded if they were aged < 18 years or unable to read Dutch. Written informed consent of all participants was obtained. In total, 132 TMD-patients (114 female) with a mean age of 39 years (SD = 14) participated. Based on the Research Diagnostic Criteria for TMD,[7] Group I (muscle) disorder was found in 78% of these patients; Group II (disc displacement) disorder was found in 35%; Group III (arthralgia, arthritis, arthrosis) disorder was found in 24% of the patients. The medical ethical review board of the VU University Amsterdam approved this study.

Procedure

Prior to the patient's first visit to the department, as part of a routine set of questionnaires, patients received written instructions by mail regarding the use of the PSA. They were asked to consider the activity that is most important and difficult to perform because of their complaints, without examples given. The dentist discussed this consideration and gave further verbal instructions on the use of the PSA at the first visit. Then, the patients made a final choice on the activity that was most important to them, and the amount of hindrance performing this activity was rated on a 100 mm VAS. The left anchor of this VAS represented "no hindrance," and the right anchor represented "worst possible hindrance." This activity and its VAS score were considered the baseline measurement.

The participant was asked to rate on the VAS the amount of hindrance again a second time (at the second visit, before treatment started) and a third time (6 to 8 weeks after treatment started), without insight in the earlier VAS score. Moreover, at these follow-up measurements, patients were also asked to judge whether their general TMD complaints had changed by answering the following question: *"Since my initial visit at ACTA my TMD complaints: worsened much, worsened slightly, remained stable, improved slightly, improved much."*

Statistical Analyses

Reproducibility. The baseline- and second-measurement VAS scores of those patients who reported that their TMD complaints 'remained stable' were used to calculate the intraclass correlation coefficient (ICC). A two-way effects model, based on absolute agreement measures was used. The (conservative) single measure ICCs were calculated. An ICC lower than 0.40 suggests poor agreement; 0.40 to 0.75 fair to good agreement; > 0.75 to 1 excellent agreement.[11]

Responsiveness. First, Spearman's rank correlation coefficients were used to evaluate whether change in hindrance (baseline to third measurement) was associated with the amount of hindrance at baseline. The correlation coefficient was calculated both for the change in hindrance expressed in millimeters and for the relative change in hindrance expressed as a percentage of the baseline score. Then, the change in hindrance (expressed either in millimeters or as a percentage) with the lowest association with the baseline score was used in a Receiver Operating Characteristics analysis (ROC) curve. Patients who judged their complaints as "improved much" on the third measurement were classified as "improved," while patients who judged their complaints as "worsened much," "worsened slightly," "remained stable," or "improved slightly" were classified as "nonimproved." In this ROC analysis, sensitivity represents the percentage of correctly classified "improved" patients, whereas specificity represents the percentage of correctly classified "nonimproved" patients. The ROC curve

displays sensitivity versus 1-specificity for each change in hindrance. The area under the curve (AUC) can be interpreted as the probability of correctly identifying improved patients.[12] The AUC is a measure for responsiveness; an AUC of 0.50 to 0.70 represents low discriminative power; 0.70 to 0.90 moderate and > 0.90 represents high discriminative power.[13] From this curve, the change in hindrance where sensitivity and specificity were as much as possible the same, was considered the cutoff that best discriminates between improved and non-improved patients.[9,14,15] This indicates that a patient that has a decrease in hindrance ≥ cutoff has improved. The 95% confidence intervals (CIs) were determined using the Wilson-score method.[16] The Statistical Package for Social Sciences (SPSS 16.0) was used to analyze the data ($\alpha = 0.05$).

Table 1 Frequency of Patient Specific Activities Selected at Baseline (n = 132)

Activity	Frequency (%)
Eating	83 (63)
Opening mouth	14 (11)
Yawning	12 (9)
Sleeping	5 (4)
Talking	5 (4)
Relaxing	4 (3)
Others (eg, singing, playing an instrument)	9 (6)
Total	132 (100)

Results

One hundred and thirty-two patients completed the baseline measurement. They reported a wide variety of activities that were most important and difficult to perform to them (Table 1). Thirteen percent of the patients reported activities (relaxing, sleeping, and others) that were not included in other disability questionnaires for TMD complaints.[7,8] The amount of hindrance patients experienced while performing the activities, rated on the VAS, varied widely (mean ± standard deviation [SD] = 57 ± 20 mm).

Reproducibility

One hundred and twenty-three patients completed the second measurement (response rate = 93%), and 78 patients reported that their TMD complaints had "remained stable." They were included in the reproducibility analysis, which showed that the reproducibility of the PSA measurement was good (ICC = 0.72; 95% CIs = 0.57–0.82).

Responsiveness

The third measurement data were collected from 109 participants (response rate = 83%). Forty patients judged their complaints as "improved much" on third measurement and were classified as "improved," while 69 patients judged their complaints as "worsened much" (n = 1), "worsened slightly" (n = 2), "remained stable" (n = 26), or "improved slightly" (n = 39), and were classified as "nonimproved" (n = 40). A 64-year-old woman judged her complaints as "improved slightly," while her VAS score worsened from 9 mm on the baseline measurement to 41 mm on the third measurement, which suggests that she misinterpreted the instructions. The data of this outlier were excluded from the responsiveness analysis.

Figure 1 plots the change in hindrance and Figure 2 plots the relative change in hindrance against the amount of hindrance at baseline (n = 108). While no significant association was found between the amount of hindrance at baseline and the relative change in hindrance ($r_s = 0.15$; $P = .12$), a strong association was found with the change in hindrance expressed in millimeters ($r_s = 0.44$; $P = .00$). Therefore, the relative change in hindrance was used for the ROC analysis.

The responsiveness of the PSA, as illustrated in the ROC curve (Fig 3), was high (AUC = 0.91; 95% CIs = 0.86–0.97). A relative change of 58% was considered the optimal cutoff to discriminate between "improved" and "nonimproved patients" (sensitivity = 0.85; 95% CIs = 0.79–0.91; specificity = 0.84; 95% CIs = 0.77–0.91).

Discussion

In this study, the PSA is introduced as a new instrument for treatment evaluation of TMD patients. It is designed to evaluate the amount of hindrance a patient perceives from the activity that is most important and difficult to perform because of his/her complaints. The concept of this patient-tailored instrument was found in low-back pain literature.[9,10] In this study, an earlier suggestion for improvement of the methods was adopted[10]: of the initial two intake visits, one visit was replaced by written instructions patients received by mail.

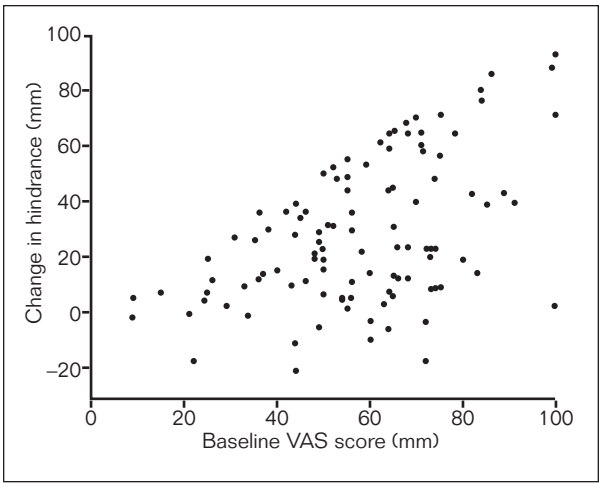

Fig 1 Scatterplots of the change in hindrance expressed in millimeters plotted against the baseline VAS scores (n = 108).

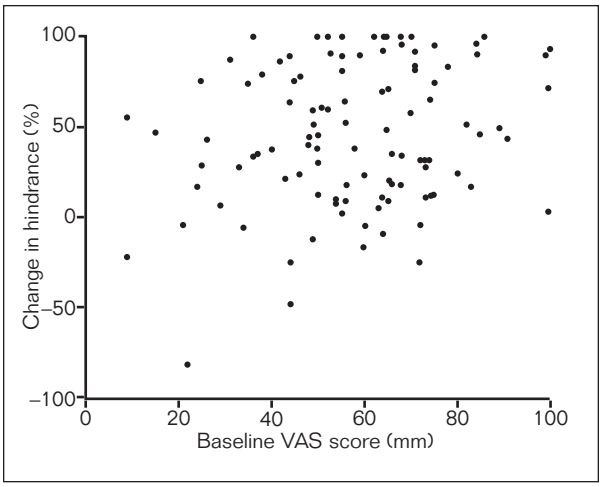

Fig 2 Scatterplots of the relative change in hindrance expressed in millimeters plotted against the baseline VAS scores (n = 108).

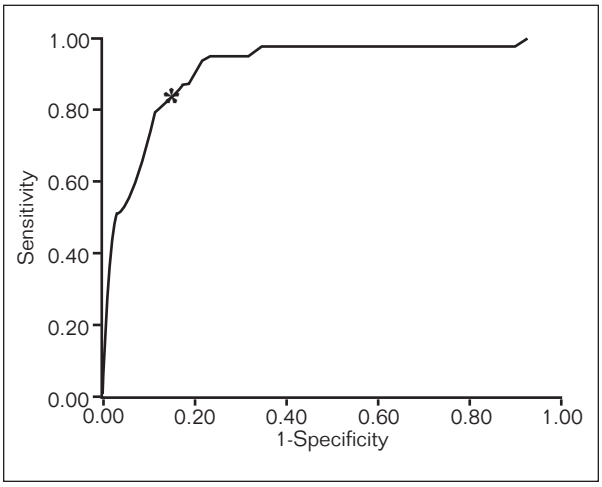

Fig 3 The ROC curve for relative change in amount of hindrance on the PSA. * = cutoff (58%; sensitivity = 0.85; specificity = 0.84).

So far, disability from mandibular activities can be measured by questionnaires such as the MFIQ or the JDI.[7,8] In these questionnaires, the amount of disability from a predetermined list of mandibular activities is used as the outcome measure. The PSA focuses on the most important activity that is hindered in the patient. In this study, 13% of the patients reported an activity that is not included in the before-mentioned lists. A possible weakness of the PSA may be that the patient's choice on the most important activity cannot be revised after baseline measurement. This issue has also been addressed in the low-back pain literature.[10] However, whether this happens and how this would influence the usefulness of the PSA is unknown. Further research is needed to explore this issue.

To determine the reproducibility, only the outcomes of those patients who indicated that their TMD complaints had not changed between the baseline and the second visit were used. The agreement between these measurements was good (it almost reached the level of excellent agreement), and is comparable to the reproducibility of the PSA in knee-dysfunction patients.[17] Moreover, it is also comparable to the reproducibility found for other uses of the VAS, such as pain intensity or health-related quality of life.[18,19] To determine the responsiveness, the TMD patients were asked to rate their hindrance from the PSA again 6 to 8 weeks after treatment started. At the third measurement, 40 patients showed that their complaints "improved much." The responsiveness reached the level of high discriminative power (AUC = 0.91), and was comparable to the responsiveness found in other studies[9,14,15,20,21] (AUC = 0.80–0.89).

The correlation analyses showed that the change in hindrance during treatment expressed in millimeters was positively associated with the amount of hindrance at baseline. Consequently, the cutoff to recognize patients who have improved is lower for patients with a low baseline score than for patients with a high baseline score. For practical use it is preferable to use one cutoff for all patients regardless of their baseline score. Since the relative change in amount of hindrance was not associated with the amount of hindrance at baseline, this change was used as measure for treatment effect.

To be able to use the PSA in the decision to end treatment, a cutoff is needed to recognize important improvement. Therefore, the patients who judged their complaints as "improved much" were discriminated from those who had not yet reached this level of improvement. Since the authors considered that "to unnecessarily continue treatment" is as harmful as "to prematurely end treatment," that cutoff was chosen for which sensitivity and specificity were the same. The estimate for the cutoff found in this study was 58% (sensitivity = 0.85; specificity = 0.84). Other studies on treatment evaluation that used pain intensity as outcome measure found slightly lower cutoffs (47 to 55%).[14,15,20,21]

Thus, the outcome of this study indicates that when a TMD patient shows a decrease in hindrance of at least 58%, a clinically important improvement is achieved. For practical reasons a cutoff of 60% can be used.

Conclusions

The PSA for TMD patients is a new and easy-to-use tool in treatment evaluation. Moreover, the VAS score of the PSA for TMD patients has good reproducibility and responsiveness.

Acknowledgments

The authors thank all colleagues at the Department of Oral Kinesiology who facilitated this study. This study was supported by the Institute for Dental Sciences in the Netherlands.

References

1. Feine JS, Widmer CG, Lund JP. Physical therapy: A critique. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1997;83:123–127.
2. Ferrari R, Russell AS. Regional musculoskeletal conditions: Neck pain. Best Pract Res Clin Rheumatol 2003; 17:57–70.
3. Ostelo RW, van Tulder MW, Vlaeyen JW, Linton SJ, Morley SJ, Assendelft WJ. Behavioural treatment for chronic low-back pain. Cochrane Database Syst Rev 2005:CD002014.
4. Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best Pract Res Clin Rheumatol 2005; 19:593–607.
5. Beurskens AJ, de Vet HC, Koke AJ, van der Heijden GJ, Knipschild PG. Measuring the functional status of patients with low back pain. Assessment of the quality of four disease-specific questionnaires. Spine 1995;20:1017–1028.
6. Vernon H, Mior S. The neck disability index: A study of reliability and validity. J Manipulative Physiol Ther 1991;14:409–415.
7. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: Review, criteria, examinations and specifications, critique. J Craniomandib Disord 1992;6:301–355.
8. Stegenga B, de Bont LG, de Leeuw R, Boering G. Assessment of mandibular function impairment associated with temporomandibular joint osteoarthrosis and internal derangement. J Orofac Pain 1993;7:183–195.
9. Beurskens AJ, de Vet HC, Koke AJ. Responsiveness of functional status in low back pain: A comparison of different instruments. Pain 1996;65:71–76.
10. Beurskens AJ, de Vet HC, Koke AJ, et al. A patient-specific approach for measuring functional status in low back pain. J Manipulative Physiol Ther 1999;22:144–148.
11. Shrout PE, Fleiss J. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86:420–428.
12. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. J Chronic Dis 1986;39:897–906.
13. Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240:1285–1293.
14. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149–158.
15. Giraudeau B, Rozenberg S, Valat JP. Assessment of the clinically relevant change in pain for patients with sciatica. Ann Rheum Dis 2004;63:1180–1181.
16. Newcombe RG. Two-sided confidence intervals for the single proportion: Comparison of seven methods. Stat Med 1998;17:857–872.
17. Chatman AB, Hyams SP, Neel JM, et al. The patient-specific functional scale: Measurement properties in patients with knee dysfunction. Phys Ther 1997;77:820–829.
18. Crossley KM, Bennell KL, Cowan SM, Green S. Analysis of outcome measures for persons with patellofemoral pain: Which are reliable and valid? Arch Phys Med Rehabil 2004;85:815–822.
19. Green C, Brazier J, Deverill M. Valuing health-related quality of life. A review of health state valuation techniques. Pharmacoeconomics 2000;17:151–165.
20. Lee JS, Hobden E, Stiell IG, Wells GA. Clinically important change in the visual analog scale after adequate pain control. Acad Emerg Med 2003;10:1128–1130.
21. ten Klooster PM, Drossaers-Bakker KW, Taal E, van de Laar MA. Patient-perceived satisfactory improvement (PPSI): Interpreting meaningful change in pain from the patient's perspective. Pain 2006;121:151–157.