What is the minimal important difference of pain intensity, mandibular function, and headache impact in patients with temporomandibular disorders? Clinical significance analysis of a randomized controlled trial

ABSTRACT

Background: There are insufficient studies providing Minimal Clinical Important Difference (MCID) for outcomes related to temporomandibular disorders (TMD). **Objectives:** (1) To provide the MCID of outcomes related to TMD using the Global Rating of Change Scale (GRCS) as an anchor. (2) To verify which outcomes can predict a moderate or large response to the treatment. **Study Design:** Secondary analysis of a randomized controlled trial in subjects with TMD. **Methods:** Sixty-one women with TMD were divided into intervention and control groups. Visual Analogue Scale (VAS), Headache Impact Test (HIT-6), pressure pain thresholds (PPTs) of masticatory muscles, Mandibular Function Impairment Questionnaire (MFIQ), and Craniocervical Flexion Test (CCFT) were collected at baseline and 5-weeks follow-up. **Results:** Participants were divided based on their response to the treatment, according to the GRCS. MCID values were provided for subjects that moderately or largely improved to the treatment. MCID was between 0 and 1.90 for orofacial pain, around 2 points for the MFIQ, between 3 and 6.26 points for the HIT-6, around 0.2 kg/cm^2 for the PPTs on masticatory muscles, around 2.5 mm for MMO and between 60 and 68 points for CCFT. Orofacial pain and HIT-6 were the most discriminative variables at determining whether patients would largely/moderately improve or would not improve after treatment. **Conclusions:** The values of MCID could be used as guidance for both clinical practice and research. Pain intensity and headache impact were the most predictive outcomes for improvement of the general health status of women with TMD.

KEYWORDS: Minimal Clinical Important Difference, Temporomandibular disorders, Global Rating of Change Scale.

INTRODUCTION

The minimal clinically important difference (MCID) has been defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s health care management” by Jaeschke et al. 1989[1] (page 408, third paragraph). It is useful for guiding treatment decisions according to the magnitude of change found in clinical research, helping to determine the sample size for further studies, as well as emphasizing the primacy of the patient’s perspective and connects that perspective to the clinicians.[2]

Different methods to determine the MCID have been described.[3,4] Distribution-based approaches are based on the statistical characteristics of the sample. They express the observed change in a standardized metric, for example the effect size, the standardized response mean, and the standard error of the measurement, which links the reliability of the measurement instrument to the standard deviation of the population.[5] The disadvantage is that they do not provide a good indication of the importance of the observed change since they do not have the point of view of the patients.[2]

Anchor-based methods consider input from the patients in their assessments of MCID, using the Global Rating of Change Scale (GRCS) as an anchor, for example. In that way, the method takes into consideration how much change on the measurement instrument corresponds to a MCID defined on the anchor.[6,7] The advantage is that the concept of ‘minimal importance' as

perceived by the patients is explicitly defined and incorporated in these methods. One limitation of anchor-based approaches is that they are less precise, considering there is no information on whether an important change according to an anchor-based method lies within the error of the health status measurement. Furthermore, there is also a recall bias.[2,4,5]

Temporomandibular disorder (TMD) is a collective term for a number of clinical problems involving the masticatory musculature, temporomandibular joints, and associated structures.[8] At least one sign or symptom of TMD is present in 39% of the general population.[9] Furthermore, females presented higher risk of developing TMD than males, with a proportion of 2.3:1.[10] Temporomandibular disorders are considered to be a major public health problem as it is the main source of chronic orofacial pain and the most prevalent category of non-dental chronic pain conditions in the orofacial region.[8] There is abundant evidence that has shown the great impact that orofacial pain and specifically TMD pain has on women's quality of life.[11,12] These problems interfere with daily activities, diminishing patients' capacity for work and/or ability to interact with their social environment.[8] In addition, TMD is considered to have a great economic impact due to direct care and has been shown to have similar individual impact and burden as back pain and severe headache.[13]

Researchers are increasingly interested in demonstrating that trials results are not only statistically significant but also clinically meaningful.[14–17] There are some studies providing MCID values for some of the outcomes commonly used in clinical trials involving patients with TMD.[15,18–22] However, these MCID values have been generated from patients with general chronic pain, low back pain, or neck pain and not specifically from patients with TMD. Thus, the primary objective was to determine the MCIDs of clinical outcomes using the GRCS as an

anchor. As a secondary objective, this study aimed to verify if one (or a combination) of the outcomes could predict improvement after treatment.

METHODS

Study design

This is a study providing validity evidence and determining the MCID for outcomes in patients with TMD. Data was further analyzed from a previously published randomized controlled trial (registry …). Details on methods can be found elsewhere.[23]

Subjects

Sixty-one women with TMD were recruited through announcements in local and social media. Participants were included if they: were female; aged between 18 and 40 years old and had orofacial pain score ≥ 3 on a ten-point numerical pain rating scale for at least three months. The diagnosis of orofacial myalgia or mixed TMD was given according to the Research Diagnostic Criteria for TMD.[24] Participants self-reported the presence and intensity of neck pain and headache, not necessarily related to TMD. The exclusion criteria are described previously.[23]

This trial was approved by the ethics committee (CAE: …), and all subjects gave their written consent.

Outcomes

The outcomes (except the GRCS) were collected at baseline and at five weeks. Subjects were assigned to either the manual therapy plus exercise or the control group. The examiner was blinded to group allocation.[23]

Visual Analogue Scale (VAS)

The current pain, maximum orofacial pain in the last week, and minimum orofacial pain in the last week were measured using the VAS (0 to 10 cm). The reliability of VAS has been considered fair to good (Intraclass Correlation Coefficient - ICC of 0.55-0.83).[25] The MCID for pain has been reported to range from 1.5 to 3.2 points.[15,18–21] Furthermore, pain reduction of 30% has been considered clinically meaningful for individuals with chronic pain based on a distribution-based method of analysis.[15]

Pressure Pain Threshold (PPT) of masticatory muscles

Using a digital algometer, PPTs were measured on masseter and temporalis muscles, bilaterally.[26,27] This procedure has shown fair to good reliability in previous studies (ICC between 0.64 and 0.78).[28,29] Minimum detectable change for PPTs was considered to range from 0.45 to 1.13 kg/cm^2 for general spots of the body, based on a distribution-based method of analysis.[22]

Headache Impact Test (HIT-6)

The HIT-6 can be scored from 36 to 78 points.[30] The test-retest reliability is considered good (ICC from 0.76 to 0.80).[31] Considering the total range of 42 points, an anchor-based analysis was performed using a combination of a general measure of patient-perceived improvement tool and a headache-specific measure as the anchor, and the optimal cut-off point for the MCID was set at 8 points in patients with tension type headache.[32]

Mandibular Function Impairment Questionnaire (MFIQ)

The MFIQ has been used to assess the limitations of mandibular function in patients with TMD.[33,34] The Brazilian version of the MFIQ was used in this study considering the final score of 52 instead of 68, excluding items 1, 2, 6 and 7 from the final score.[35] The Smallest Detectable Difference for this questionnaire for the total score of 68 has been established for the original questionnaire to be 8 points, by a distribution-based method.[36] MCID for the Brazilian version has not been provided yet.

Maximum Mouth Opening (MMO)

Three active maximum mouth openings without pain were collected with a digital caliper.[24] The smallest detectable difference for this variable was established to be 6 mm, by a distribution-based approach.[37]

Craniocervical Flexion Test (CCFT)

The clinical CCFT[38,39] was applied using a visual feedback device (Stabilizer; Chattanooga Group Inc., Chattanooga, TN, USA) to evaluate the performance of the deep flexor muscles of the neck. Participants performed the nodding movement 10 times for 10 seconds in 5 progressive stages of increasing pressure (from 20mmHg to 22, 24, 26, 28, and 30 mmHg). They should maintain the contraction without using superficial muscles.[39]

The performance index[39,40] was calculated considering how many contractions were done by the patient at each level. The repetitions were multiplied by 2, 4, 6, 8, or 10 indicating the five levels of the test. The maximum score of the CCFT was 300.

The reliability of the procedure has been reported to be good/excellent (intra-rater ICC was 0.78 or 0.98, depending on the study)[38,41] but there was no previous study showing reference or MCID values for this test, using the performance index.

GRCS

The Global Rating of Change Scale was designed to quantify a patient's improvement or deterioration over time, either to determine the effect of an intervention or to chart the clinical course of a condition. This scale provides information about a person´s current health status.[42] It has been previously used on anchor-based approaches for establishing a clinically meaningful change in longitudinal studies,[17,43] providing the single best measure of the significance of the change from the individual perspective.[2] Subjects were asked to respond to the GRCS at 5-weeks follow-up evaluation.[2,42] The blind rater presented a scale ranging from -7 to +7 and asked the following: "considering that on the first evaluation you were at 'zero', use the scale to indicate if you are at the same point, better or worse than on the first evaluation, considering all the symptoms that you have in the orofacial region".

Intervention Protocol

The intervention group received 10 sessions of physical therapy for 5 weeks, applied by a physiotherapist and included non-manipulative techniques[44–49] for the upper cervical spine; and neck stabilization exercises with biofeedback.[23,34,50,51] The control group did not receive any study-specific intervention or advice for pain education for 5 weeks.[23]

Data Processing and statistical analysis

Subjects were classified as having no improvement when their responses to the GRCS were between -7 and 0, moderate improvement between 1 and 3 and large improvement between

4 and 7 after the treatment. Thus, the anchor-method approach was used to estimate the MCID of each of the outcome variables of interest (e.g. VAS, PPTs, HIT-6, CCFT) using the GRCS as the anchor. Changes from baseline to follow up were calculated for each one of the outcomes of interest and used for analysis. Correlations between the GRS and each of the variables of interest were tested in an exploratory fashion. Most of them were higher than 0.3 as recommended by the literature.[52]

All outcomes were analyzed using a Receiver Operating Characteristic analysis.[53] It compares the values of each measure from 2 groups according to the GRS: patients who showed an improvement after treatment and patients who did not. This method constructs a graph of the measures' performance in classifying the groups for each possible cut-off point and calculates the sensitivity and specificity for all the cut-offs.

For each outcome, two independent raters chose the selected cut-off in a way that both sensitivity and specificity could be maximized (i.e. high values); and thus, both values should be similar. In addition, the cut-off chosen should also maximize positive and negative likelihood ratios (LR) and the percentage of correctly classified patients. According to Cook et al.,[54] higher positive likelihood ratios and lower negative values are sought to maximize discrimination. Values higher than 2 for positive LR and lower than 0.5 for negative LR are recommended by the literature. A third rater was consulted when there were disagreements between the other raters, in order to provide a final decision on the best cut-off value.

A logistic regression model was followed for the secondary aim. First, a single logistic regression was done to examine the association between each independent variable (e.g. VAS, PPT, HIT-6, MFIQ, MMO, and CCFT) and groups involved (patients who improved and who did not as a dependent variable). The independent variables that were significant at $p\text{-}value \leq 0.20$ in

the univariate analysis were entered into a multiple logistic regression model, using a forward stepwise fashion. This p-value has been suggested by some as a conservative criterion to involve all potential variables that could be significant in a multivariable regression model. More traditional alpha levels can fail in identifying variables that could be important.[55]

After the addition of each one of the independent variables, a Receiver Operating Characteristic curves analysis was done to determine the discriminative ability of each model for distinguishing between subjects who improved and who did not.[53] They were compared and the model with the highest area under the curve was chosen as the best model.[56,57] The interpretation was done according to the recommended guideline.[58] The following guidelines are recommended to interpret the discriminatory performance of an AUC curve: excellent discrimination (AUC = 0.90 to 1.0); good discrimination (AUC = 0.80 to 0.90); fair discrimination (AUC = 0.70 to 0.80); weak discrimination (AUC = 0.60 to 0.70); and discrimination is no better than chance (AUC $\leq$0.50). The model that achieved statistical significance ($p<0.05$), presented a larger area under the curve and included fewer variables as predictive values as possible was considered to be the best predictive model. All data analyses were performed using STATA software and guided by a statistical expert.

RESULTS

General Results

At baseline, participants presented maximum pain of 6.1(±1.9) cm, minimum pain of 1.7(±1.6) cm, and current pain of 3.5(±2.7) cm. They scored 20.3(±9.4) points at MFIQ, 61.2(±6.8) points at HIT-6 and 51.2(±43.8) points at the TFCC. The maximum mouth opening was 34.7(±9.1) mm, masseters PPT was 1.1(±0.5) kg/cm^2 and temporalis PPT was 1.2(±0.5)

kg/cm^2. Difference tests were not significant between the groups for all variables. According to the GRCS, 30 subjects had no improvement (responses between -7 and 0), 18 subjects had moderate improvement (between 1 and 3) and 13 subjects had large improvement (between 4 and 7) after the treatment (Table 1).

Table 1

Minimal Clinically Important Differences

MCID for all outcomes of interest (e.g. cut-offs, sensitivity, specificity, percentage of correctly classified, positive and negative likelihood ratios) of subjects who moderately and largely improved after the treatment are shown in Tables 2 and 3. Cut-offs were higher on the comparison between subjects who did not improve and subjects who largely improved than on the comparison between subjects who did not improve and the ones who improved moderately. However, the CCFT showed the opposite behavior. Visual analog scales and HIT-6 outcomes correctly classified more than 75% of the subjects when subjects who did not improve were compared to the ones who improved largely.

Tables 2 and 3

Discriminative Analysis

The results from the single and multiple logistic regression analyses are described in Table 4.

In the **univariate analysis**, maximum, minimum and current pain, mandibular function (MFIQ) and headache (HIT-6) were able to discriminate patients who had moderate and large improvement from those who showed no improvement. However, the PPT from temporalis and CCFT were able to predict moderate improvement but not large improvement. On the other hand, PPT on masseters and MMO were able to predict large improvement but not moderate.

In the **multivariable analysis**, some of the combinations between variables were able to discriminate patients who had moderate and large improvement from those who showed no improvement after the therapy even better than the univariate analysis. For most of the combinations, the values regarding the area under the curve were higher for the large improvement, when compared to moderate improvement. Between the three pain scores, the minimum pain was chosen for all of the multivariable models due to the higher area under the curve for both comparisons. In both subgroups, the combination of minimum pain and headache impact were good predictors to determine if the patient will improve or not after the treatment (76% for moderate improvement and 91% on large improvement).

The combination of minimum pain and HIT-6 represented the best combination of 2 predictive factors for both comparisons. This combination was better than any other single variable. The best combination of 3 variables in order to predict if the patient would largely improve after the treatment or not, was the combination of HIT-6, MFIQ and minimum pain (93%). To predict moderate improvement, the best combination of 3 variables was the HIT-6, CCFT and minimum pain (77%). Figures 1 and 2 compare the area under the curve from those 2 best combinations.

Table 4

The Receiver Operating Characteristic curves compared which model better discriminates subjects who improved from subjects who did not (Figures 1 and 2). They are visually similar, and there is no significant statistical difference between them (p=0.83 for moderate improvement and p=0.64 for large improvement). Since the combination between HIT-6 and minimum pain was more parsimonious (i.e. simpler; including only 2 variables), it is the best combination to discriminate between subjects who will moderately (Figure 1) and largely (Figure 2) improve from those who did not.

Figure 1. Comparison between Receiver Operating Characteristic area of two regression models for moderate improvement: minimum pain + headache impact (blue line) against minimum pain + headache impact + CCFT (red line).

Figure 2. Comparison between Receiver Operating Characteristic area of two regression models for large improvement: minimum pain + headache impact (blue line) against minimum pain + headache impact + MFIQ (red line).

DISCUSSION

The MCID of all outcomes were determined for women with TMD using the GCRS as the anchor. Five of the nine outcomes were able to discriminate a participant who had large/moderate improvement from one who had no improvement, with the exception of CCFT, MMO and masseters´ and temporalis´ PPTs. The combination of pain scores and headache impact was the best combination of variables to discriminate subjects who moderately or largely improved from those who did not.

Pain Intensity

The MCID for general chronic pain has been reported to range from 1.5 to 3.2 cm on VAS,[15,18–21] or 30% of pain reduction.[15] No previous studies have been conducted in patients with TMD to determine the MCID of the VAS, therefore this study provides novel evidence. Women with TMD presented a large improvement of the general health status if they presented with a reduction of 1.2 cm on the maximum pain, 1.9 cm on the current pain and 0.9 cm on the minimum pain scales. Therefore, for patients with chronic TMD, it might not be necessary to reduce 3.2 cm on the VAS, after treatment to conclude that it was effective as reported in the literature.

Pressure Pain Thresholds

The minimum detectable change for PPTs has been considered to range from 0.45 to 1.13 kg/cm^2 in subjects with neck pain.[22] However, those values might not be applicable to the masticatory muscles. For this reason, it has been difficult to find clinically important differences in studies with PPT on the masticatory muscles, even when there is an improvement on other variables related to pain.[44,46,59,60]

The PPTs from masseters and temporalis have been previously validated to discriminate TMD from healthy patients, with a cut-off value of 1.78 kg/cm^2, sensitivity of 64% and specificity of 68%.[61] However, in general, the responsiveness of the PPTs is still unknown. The results of the present study showed that even patients who largely responded to treatment did not show more than 0.18 kg/cm^2 and 0.22 kg/cm^2 of difference on the PPT of masseters and temporalis, respectively.

HIT-6

The MCID for HIT-6 has been established for patients with chronic tension type headache at 8 points, on a total scale range of 42 points.[32] This is the first study to establish a cut-off for patients with TMD, which is very relevant considering that there is a high prevalence[62,63] and comorbidity[64,65] of headaches in this population. Although this study could not provide proper diagnosis of the type of headache, which is a limitation to be assumed, it can provide some conclusions regarding patients with TMD and concomitant headache. If there was a reduction of 3 points on the HIT-6 after treatment, this would also show moderate improvement of the general health status. Furthermore, a change of 7 points on the HIT-6 would represent a large improvement of the general health status after treatment.

MFIQ

Considering the original version of the MFIQ, (total score=68), the smallest detectable difference has been established to be 8 points in patients with painfully restricted temporomandibular joints.[36] However, the Brazilian version of the MFIQ utilizes a final score of 52 instead of 68.[35] Therefore, the previously established smallest detectable difference was too high to be achieved in the clinical trials using the Brazilian version of the questionnaire. According to our results, 2 points of difference after treatment on the Brazilian version of the MFIQ already represents clinically meaningful change.

MMO

The MCID for MMO has been established to be between 6 mm and 9 mm.[37] Considering that the patients who participated in this study had no severe limitation of maximum mouth opening (average of 35 mm at baseline), our results provide evidence that when a patient with no

limitation of the MMO presents more than 3 mm of improvement after an intervention, it might be expected a large improvement of his/her general health status.

CCFT

The CCFT has been used to determine the neck flexor muscles performance in subjects with neck and craniofacial involvement.[66–68] There is no previous study providing MCID for the performance index of the CCFT, thus, our findings cannot be compared to other studies. The present study provided evidence that increasing 60 points in the CCFT (total score=300) can be considered clinically relevant and represent a large improvement in the general health status of the patient with TMD. It is important to state that this test has been applied and scored in different ways across the literature and thus the MCID here provided is specifically for the performance index of the clinical CCFT.[38]

Implications for future research

Although the outcomes were able to discriminate a participant who had large/moderate improvement from one who had no improvement when assessed one by one, the combination of minimum pain (or any other pain score) and HIT-6 had better discrimination than single measures. Future studies that intend to verify the effect of interventions of the treatment of TMD should continue using VAS and HIT-6 since those outcomes were the best predictors of enhanced global health status.

Clinical implications

Clinicians may apply VAS and HIT-6 to verify the effects of their interventions. They may also use the MCID of all outcomes included in this study as guidance to consider their treatments' effectiveness to facilitate clinical decision-making when treating women from 18-40 years old with TMD and concomitant headache. However, once the MCIDs have been established for this determined group of patients, the distribution of the results should not be ignored by researchers and clinicians.[2] They may also consider the proportion of patients that achieved a small, moderate or large benefit, instead of considering only the mean difference.[2,6,7]

Strengths and Limitations

The anchor-based method is less precise, considering there is no information on whether an important change according to an anchor-based method, lies within the measurement error of the health status measurement. Although the literature presents evidence that MCID values provided by anchor-methods are better for people who are truly unchanged,[69] one of the limitations is that there is typically little information provided about the reliability and validity of the GRCS.[5] As a self-report measure of improvement, the GRCS reflects whether the patient is "feeling better", but not if the patient is actually "doing better" considering functional performance or socioeconomic improvement.[70] Also, this method did not account for the risks and costs of treatment but did define the effects of intervention in terms of the difference in outcome with and without intervention.[71] For future studies, it is recommend the use of the benefit-harm trade-off method[72] for determining the smallest worthwhile effect in order to assess the magnitude of effect that patients consider which justifies the costs, risks, and inconvenience of the intervention.

According to the literature,[73] it has been stated that estimated thresholds of MCID will

depend on a range of factors, including the instrument, patient population, selected anchors and the methods used to calculate it. Therefore, MCIDs values cannot be applicable without judgment in all situations. Thus, the information obtained from this study could provide some insights on how much a difference is perceived as clinically important, by patients with similar characteristics, in commonly used outcomes related to TMD after a treatment. These values can only be used as general guidelines and be applied judiciously to any particular clinical context.

The results of this study were obtained from a secondary analysis of a randomized controlled trial,[23] and thus it might not have been powered to answer all of our questions regarding MCID (e.g. calculated MCID by each group), and could explain in part the inconsistency on the cut-off for CCFT, for example. In addition, the low correlation of some of the variables with the GRS could have explained this behavior. Future studies should collect data from larger samples and apply the same calculations to verify the MCID in different populations, including other types of TMD, sexes, different ages, diagnoses, the severity of the dysfunction and interventions.

In conclusion, the study provided MCIDs for orofacial pain intensity, the sensitivity of the masticatory muscles, headaches and mandibular function in women with TMD. Those values can be used as general guidance in both clinical practice and future research with similar interventions and populations like the one used in this present study. Pain and headache impact were the most predictive outcomes for improvement on the general health status of women with TMD.

ACKNOWLEDGMENTS

…

REFERENCES

1. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415. http://www.ncbi.nlm.nih.gov/pubmed/2691207.
2. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77(4):371-383. doi:10.1016/S0025-6196(11)61793-X
3. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther*. 2012;20(3):160-166. doi:10.1179/2042618612Y.0000000001
4. Lydick E, Epstein RS. Interpretation of quality of life changes. *Qual Life Res*. 1993;2(3):221-226. http://www.ncbi.nlm.nih.gov/pubmed/8401458.
5. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol*. 2003;56(5):395-407. doi:10.1016/S0895-4356(03)00044-1
6. Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau LJ, Goodwin PJ. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. *J Clin Epidemiol*. 2007;60(5):448-455. doi:10.1016/j.jclinepi.2006.08.006
7. de Vet HCW, Terluin B, Knol DL, et al. Three ways to quantify uncertainty in individually applied “minimally important change” values. *J Clin Epidemiol*. 2010;63(1):37-45. doi:10.1016/j.jclinepi.2009.03.011

8. AAOP. *Orofacial Pain: Guidelines for Assessment, Diagnosis, and Management*. 5th ed. (de Leeuw R, Klasser GD, eds.). Chicago (IL): Quintessence Publishing Co, Inc; 2013.

9. Gonçalves DADG, Dal Fabbro AL, Campos JADB, Bigal ME, Speciali JG. Symptoms of temporomandibular disorders in the population: an epidemiological study. *J Orofac Pain*. 2010;24(3):270-278. http://www.ncbi.nlm.nih.gov/pubmed/20664828.

10. Bagis B, Ayaz EA, Turgut S, Durkan R, Özcan M. Gender difference in prevalence of signs and symptoms of temporomandibular joint disorders: a retrospective study on 243 consecutive patients. *Int J Med Sci*. 2012;9(7):539-544. doi:10.7150/ijms.4474

11. Riley JL, Gilbert GH. Orofacial pain symptoms: an interaction between age and sex. *Pain*. 2001;90(3):245-256.

12. LeResche L. Epidemiology of temporomandibular disorders: implications for the investigation of etiologic factors. *Crit Rev Oral Biol Med*. 1997;8(3):291-305.

13. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet (London, England)*. 2017;390(10100):1211-1259. doi:10.1016/S0140-6736(17)32154-2

14. van Grootel RJ, van der Glas HW. Statistically and clinically important change of pain scores in patients with myogenous temporomandibular disorders. *Eur J Pain*. 2009;13(5):506-510. doi:10.1016/j.ejpain.2008.06.002

15. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. *J Pain*. 2008;9(2):105-121. doi:10.1016/j.jpain.2007.09.005

16. Armijo-Olivo S, Warren S, Fuentes J, Magee DJ. Clinical relevance vs. statistical significance: Using neck outcomes in patients with temporomandibular disorders as an example. *Man Ther*. 2011;16(6):563-572. doi:10.1016/j.math.2011.05.006

17. Fuentes J, Armijo-Olivo S, Funabashi M, et al. Enhanced therapeutic alliance modulates pain intensity and muscle pain sensitivity in patients with chronic low back pain: an experimental controlled study. *Phys Ther*. 2014;94(4):477-489. doi:10.2522/ptj.20130118

18. Kovacs FM, Abraira V, Royuela A, et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine (Phila Pa 1976)*. 2007;32(25):2915-2920. doi:10.1097/BRS.0b013e31815b75ae

19. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical rating scale. *Pain*. 2001;94(2):149-158. doi:10.1016/S0304-3959(01)00349-9

20. van der Roer N, Ostelo RWJG, Bekkering GE, van Tulder MW, de Vet HCW. Minimal Clinically Important Change for Pain Intensity, Functional Status, and General Health Status in Patients With Nonspecific Low Back Pain. *Spine (Phila Pa 1976)*. 2006;31(5):578-582. doi:10.1097/01.brs.0000201293.57439.47

21. Maughan EF, Lewis JS. Outcome measures in chronic low back pain. *Eur Spine J*. 2010;19(9):1484-1494. doi:10.1007/s00586-010-1353-6

22. Walton DM, Macdermid JC, Nielson W, Teasell RW, Chiasson M, Brown L. Reliability, standard error, and minimum detectable change of clinical pressure pain threshold testing in people with and without acute neck pain. *J Orthop Sports Phys Ther*. 2011;41(9):644-650. doi:10.2519/jospt.2011.3666

23. Calixtre LB, Oliveira AB, Sena Rosa LR, Armijo-Olivo, S. Visscher, C. M. Alburquerque-

Sendín F. Upper cervical mobilizations and neck exercises effects on pain, mandibular function, and headache in women with TMD. A randomized, controlled trial. *J Oral Rehabil*. 2018;00:1-11. doi:doi.org/10.1111/joor.12733

24. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. *J Craniomandib Disord*. 1992;6(4):301-355. http://www.ncbi.nlm.nih.gov/pubmed/1298767. Accessed October 24, 2013.

25. Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. *Pain*. 1999;83(2):157-162. http://www.ncbi.nlm.nih.gov/pubmed/10534586. Accessed January 12, 2015.

26. Vanderweeën L, Oostendorp RAB, Vaes P, Duquet W. Pressure algometry in manual therapy. *Man Ther*. 1996;1(5):258-265. doi:10.1054/math.1996.0276

27. LaTouche R, Fernández-de-las-Peñas C, Fernández-Carnero J, et al. The effects of manual therapy and exercise directed at the cervical spine on pain and pressure pain sensitivity in patients with myofascial temporomandibular disorders. *J Oral Rehabil*. 2009;36(9):644-652. doi:10.1111/j.1365-2842.2009.01980.x

28. Costa YM, Morita-Neto O, de Araújo-Júnior ENS, Sampaio FA, Conti PCR, Bonjardim LR. Test-retest reliability of quantitative sensory testing for mechanical somatosensory and pain modulation assessment of masticatory structures. *J Oral Rehabil*. 2017;38(1):42-49. doi:10.1111/joor.12477

29. Gomes MB, Guimarães JP, Guimarães FC, Neves ACC. Palpation and pressure pain threshold: reliability and validity in patients with temporomandibular disorders. *Cranio J Craniomandib Pract*. 2008;26(3):202-210. doi:10.1179/crn.2008.027

30. Kosinski M, Bayliss MS, Bjorner JB, et al. A six-item short-form survey for measuring headache impact: the HIT-6. *Qual Life Res*. 2003;12(8):963-974. http://www.ncbi.nlm.nih.gov/pubmed/14651415. Accessed March 7, 2014.

31. Rendas-Baum R, Yang M, Varon SF, Bloudek LM, DeGryse RE, Kosinski M. Validation of the Headache Impact Test (HIT-6) in patients with chronic migraine. *Health Qual Life Outcomes*. 2014;12:117. doi:10.1186/s12955-014-0117-0

32. Castien RF, Blankenstein AH, Windt D a WM Van Der, Dekker J. Minimal clinically important change on the Headache Impact Test-6 questionnaire in patients with chronic tension-type headache. *Cephalalgia*. 2012;32(9):710-714. doi:10.1177/0333102412449933

33. Pereira LJ, Steenks MH, de Wijer a, Speksnijder CM, van der Bilt a. Masticatory function in subacute TMD patients before and after treatment. *J Oral Rehabil*. 2009;36(6):391-402. doi:10.1111/j.1365-2842.2008.01920.x

34. Calixtre LB, Grüninger BL da S, Haik MN, Alburquerque-Sendín F, Oliveira AB. Effects of cervical mobilization and exercise on pain, movement and function in subjects with temporomandibular disorders: a single group pre-post test. *J Appl Oral Sci*. 2016;24(3):188-197. doi:10.1590/1678-775720150240

35. Campos JDB, Carrascosa C, Maroco J. Validity and reliability of the Portuguese version of Mandibular Function Impairment Questionnaire. *J Oral Rehabil*. 2012;39(5):377-383. doi:10.1111/j.1365-2842.2011.02276.x

36. Kropmans TJ, Dijkstra PU, Stegenga B, Stewart R, de Bont LG. Smallest detectable difference in outcome variables related to painful restriction of the temporomandibular joint. *J Dent Res*. 1999;78(3):784-789. http://www.ncbi.nlm.nih.gov/pubmed/10096454.

37. Kropmans T, Dijkstra P, Stegenga B, Stewart R, de Bont L. Smallest detectable difference of maximal mouth opening in patients with painfully restricted temporomandibular joint function. *Eur J Oral Sci*. 2000;108(1):9-13. http://www.ncbi.nlm.nih.gov/pubmed/10706471.

38. Hudswell S, Von Mengersen M, Lucas N. The cranio-cervical flexion test using pressure biofeedback: A useful measure of cervical dysfunction in the clinical setting? *Int J Osteopath Med*. 2005;8(3):98-105. doi:10.1016/j.ijosm.2005.07.003

39. Jull GA, O'Leary SP, Falla DL. Clinical assessment of the deep cervical flexor muscles: the craniocervical flexion test. *J Manipulative Physiol Ther*. 2008;31(7):525-533. doi:10.1016/j.jmpt.2008.08.003

40. Jull G, Barrett C, Magee R, Ho P. Further clinical clarification of the muscle dysfunction in cervical headache. *Cephalalgia*. 1999;19(3):179-185. http://www.ncbi.nlm.nih.gov/pubmed/10234466.

41. Kotwani S, Bid DN, Ghatamaneni D, Alahmari KA, Ramalingam T, Paul Silvian S. Determining the reliability of craniocervical flexion test in asymptomatic individuals. *Hong Kong Physiother J*. 2018;38(1):33-40. doi:10.1142/S101370251850004X

42. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther*. 2009;17(3):163-170. doi:10.1179/jmt.2009.17.3.163

43. Christiansen DH, Frost P, Falla D, Haahr JP, Frich LH, Svendsen SW. Responsiveness and Minimal Clinically Important Change: A Comparison Between 2 Shoulder Outcome Measures. *J Orthop Sport Phys Ther*. 2015;45(8):620-625. doi:10.2519/jospt.2015.5760

44. Oliveira-Campelo NM, Rubens-Rebelatto J, Martín-Vallejo FJ, Alburquerque-Sendín F,

Fernández-de-Las-Peñas C. The immediate effects of atlanto-occipital joint manipulation and suboccipital muscle inhibition technique on active mouth opening and pressure pain sensitivity over latent myofascial trigger points in the masticatory muscles. *J Orthop Sports Phys Ther*. 2010;40(5):310-317. doi:10.2519/jospt.2010.3257

45. Aparicio EQ, Quirante LB, Rodríguez-Blanco C, Alburquerque-Sendín F. Immediate effects of the suboccipital muscle inhibition technique in subjects with short hamstring syndrome. *J Manipulative Physiol Ther*. 2009;32(4):262-269. doi:10.1016/j.jmpt.2009.03.006

46. La Touche R, Paris-Alemany A, Mannheimer JS, et al. Does mobilization of the upper cervical spine affect pain sensitivity and autonomic nervous system function in patients with cervico-craniofacial pain?: A randomized-controlled trial. *Clin J Pain*. 2013;29(3):205-215. doi:10.1097/AJP.0b013e318250f3cd

47. Hall T, Briffa K, Hopper D, Robinson K. Long-term stability and minimal detectable change of the cervical flexion-rotation test. *J Orthop Sports Phys Ther*. 2010;40(4):225-229. doi:10.2519/jospt.2010.3100

48. Mulligan B. *Manual Therapy, "NAGS", "SNAGS", and "MWMS".* (Services PV, ed.). Wellington; 2007.

49. González-Iglesias J, Cleland JA, Neto F, Hall T, Fernández-de-las-Peñas C. Mobilization with movement, thoracic spine manipulation, and dry needling for the management of temporomandibular disorder: a prospective case series. *Physiother Theory Pract*. 2013;29(8):586-595. doi:10.3109/09593985.2013.783895

50. O'Leary S, Falla D, Hodges PW, Jull G, Vicenzino B. Specific therapeutic exercise of the neck induces immediate local hypoalgesia. *J Pain*. 2007;8(11):832-839.

doi:10.1016/j.jpain.2007.05.014

51. Jull GA, Falla DL, Treleaven JM, Sterling M, O'Leary SP. *A Therapeutic Exercise Approach for Cervical Disorders*. 3rd ed. Edinburgh, UK: Churchill Livingstone - Elsevier Science.; 2004.

52. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102-109. doi:10.1016/j.jclinepi.2007.03.012

53. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes*. 2006;4(Mic):3-7. doi:10.1186/1477-7525-4-54

54. Cook C, Cleland J, Huijbregts P. Creation and critique of studies of diagnostic accuracy: Use of the STARD and QUADAS methodological quality assessment tools. *J Man Manip Ther*. 2007;15(2):93-102. doi:10.1179/106698107790819945

55. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. Second. Wiley; 2000. doi:10.1038/461726a

56. Cleves MA, Rock L. From the help desk: Comparing areas under receiver operating characteristic curves from two or more probit or logit models. *Stata J*. 2012;(3):301-13. doi:The Stata Journal

57. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther*. 2001;81(9):1546-1564. http://www.ncbi.nlm.nih.gov/pubmed/11688591.

58. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29-36.

doi:10.1148/radiology.143.1.7063747

59. Magri LV, Carvalho VA, Rodrigues FCC, Bataglion C, Leite-Panissi CRA. Effectiveness of low-level laser therapy on pain intensity, pressure pain threshold, and SF-MPQ indexes of women with myofascial pain. *Lasers Med Sci*. 2017:419-428. doi:10.1007/s10103-016-2138-x

60. Mansilla-Ferragut P, Fernández-de-Las Peñas C, Alburquerque-Sendín F, Cleland JA, Boscá-Gandía JJ. Immediate effects of atlanto-occipital joint manipulation on active mouth opening and pressure pain sensitivity in women with mechanical neck pain. *J Manipulative Physiol Ther*. 2009;32(2):101-106. doi:10.1016/j.jmpt.2008.12.003

61. Visscher CM, Lobbezoo F, Naeije M. Comparison of algometry and palpation in the recognition of temporomandibular disorder pain complaints. *J Orofac Pain*. 2004;18(3):214-219. http://www.ncbi.nlm.nih.gov/pubmed/15509000.

62. Franco AL, Gonçalves DAG, Castanharo SM, Speciali JG, Bigal ME, Camparis CM. Migraine is the most prevalent primary headache in individuals with temporomandibular disorders. *J Orofac Pain*. 2010;24(3):287-292. http://www.ncbi.nlm.nih.gov/pubmed/20664830.

63. Gonçalves D a G, Camparis CM, Speciali JG, Franco AL, Castanharo SM, Bigal ME. Temporomandibular disorders are differentially associated with headache diagnoses: a controlled study. *Clin J Pain*. 2011;27(7):611-615. doi:10.1097/AJP.0b013e31820e12f5

64. Speciali JG, Dach F. Temporomandibular dysfunction and headache disorder. *Headache*. 2015;55(S1):72-83. doi:10.1111/head.12515

65. van der Meer HA, Speksnijder CM, Engelbert RHH, Lobbezoo F, Nijhuis-van der Sanden MWG, Visscher CM. The Association Between Headaches and Temporomandibular

Disorders is Confounded by Bruxism and Somatic Symptoms. *Clin J Pain*. 2017;33(9):835-843. doi:10.1097/AJP.0000000000000470

66. Chiu TTWTHF, Law EYH, Chiu TTWTHF, et al. Performance of the Craniocervical Flexion Test in subjects with and without chronic neck pain. *J Orthop Sport Phys Ther*. 2005;35(9):567-571. doi:10.2519/jospt.2005.35.9.567

67. Galindez-Ibarbengoetxea X, Setuain I, Ramírez-Velez R, et al. Immediate Effects of Osteopathic Treatment Versus Therapeutic Exercise on Patients With Chronic Cervical Pain. *Altern Ther Health Med*. November 2017. http://www.ncbi.nlm.nih.gov/pubmed/29135458.

68. Luedtke K, Boissonnault W, Caspersen N, et al. International consensus on the most useful physical examination tests used by physiotherapists for patients with headache: A Delphi study. *Man Ther*. 2016;23:17-24. doi:10.1016/j.math.2016.02.010

69. Stratford PW, Riddle DL. When Minimal Detectable Change Exceeds a Diagnostic Test-Based Threshold Change Value for an Outcome Measure: Resolving the Conflict. *Phys Ther*. 2012;92(10):1338-1347. doi:10.2522/ptj.20120002

70. Gatchel RJ, Mayer TG, Chou R. What does/should the minimum clinically important difference measure?: A reconsideration of its clinical value in evaluating efficacy of lumbar fusion surgery. *Clin J Pain*. 2012;28(5):387-397. doi:10.1097/AJP.0b013e3182327f20

71. Ferreira ML, Herbert RD, Ferreira PH, et al. A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain. *J Clin Epidemiol*. 2012;65(3):253-261. doi:10.1016/j.jclinepi.2011.06.018

72. Barrett B, Brown D, Mundt M, Brown R. Sufficiently important difference: Expanding the

framework of clinical significance. *Med Decis Mak*. 2005;25(3):250-261. doi:10.1177/0272989X05276863

73. King MT. A point of minimal important difference (MID): A critique of terminology and methods. *Expert Rev Pharmacoeconomics Outcomes Res*. 2011;11(2):171-184. doi:10.1586/erp.11.9

Table 1. Number of subjects (n), mean and standard deviation (SD), minimum (Min) and maximum (Max) value of the differences between baseline and follow-up for subjects who did not improve, had moderate improvement and had large improvement respectively.

Outcome	No improvement				Moderately improved				Largely improved			
	n	Mean (SD)	Min	Max	n	Mean (SD)	Min	Max	n	Mean (SD)	Min	Max
Max. Pain (cm)	30	-0.2 (2.2)	-5.5	4.6	18	-1.3 (1.8)	-4.8	2.7	13	-3.1 (2.4)	-6.5	0.9
Min. Pain (cm)	30	0.3 (1.5)	-3.8	4	18	-0.5 (1.5)	-2.9	3.4	13	-1.7 (1.7)	-4.9	0
Curr. Pain (cm)	30	0.1 (2.2)	-5.4	5.4	18	-0.9 (1.8)	-3.9	2.6	13	-3.2 (2.4)	-6.1	0.1
MFIQ	30	-0.2 (5.5)	-12.0	15	18	-2.4 (6.1)	-12.0	13	13	-4.5 (7.6)	-15	12
HIT-6	30	-1.9 (7.5)	-24.0	11	18	-6.9 (9.2)	-33.0	8	13	-11.0 (6.9)	-21	3
PPT Temp (kg/cm^2)	30	0.2 (0.3)	-0.5	1.4	18	-0.1 (1.3)	-1.3	0.3	13	0.2 (0.2)	0.1	0.6
PPT Mass (kg/cm^2)	30	0.1 (1.3)	-0.5	0.6	18	0.0 (1.3)	-0.9	0.5	13	0.3 (0.2)	-1	0.8
MMO (mm)	30	0.9 (6.1)	-10.8	12.7	18	-0.3 (6.3)	-14.0	12.9	13	4.5 (7.3)	-9.7	20.5
CCFT	30	28.00 (100.5)	-230.66	288	18	71.74(98.1)	-134	260	13	52.00 (52.7)	-42	132

Max- Maximum, Min- Minimum, Curr- Current, MFIQ- Mandibular function impairment questionnaire, HIT-6 - Headache impact test, PPT- pressure pain threshold, Temp- Temporalis, Mass- Masseters, MMO- Maximum mouth opening, CCFT- Craniocervical flexion test.

Table 2. Number of subjects (n), cut-off value, best sensitivity, best specificity, correctly classified (CC), positive likelihood ratio (+LR) and negative likelihood ratio (-LR) when differentiating between subjects with no improvement from subjects with **moderate improvement** for all outcomes.

	Subjects				CC		
Outcome	**(n)**	**Cut-off**	**Sensitivity**	**Specificity**	**(%)**	**+LR**	**-LR**
Maximum Pain	48	-0.60	66.67	66.67	66.67	2.00	0.50
Minimum Pain	48	0.00	73.33	66.67	70.83	2.20	0.40
Current Pain	48	-0.60	70.00	55.56	64.58	1.58	0.54
MFIQ	48	-1.51	66.67	72.22	68.75	2.40	0.46
HIT-6	48	-3.00	73.33	61.11	68.75	1.89	0.44
PPT Temporalis	48	0.10	66.67	72.22	68.75	2.40	0.46
PPT Masseters	48	0.07	60.00	55.56	58.33	1.35	0.72
MMO	48	2.54	50.00	61.11	54.17	1.29	0.82
CCFT	48	68.00	55.56	83.33	72.92	3.33	0.53

MMO - Maximum mouth opening, CCFT - Craniocervical flexion test, HIT-6 - Headache impact test, PPT - pressure pain threshold, n - number of subjects, CC - correctly classified, +LR - Positive Likelihood Ratio, -LR - Negative Likelihood Ratio

Table 3. Number of subjects (n), cut-off value, best sensitivity, best specificity, correctly classified (CC), positive likelihood ratio (+LR) and negative likelihood ratio (-LR) when differentiating between subjects with no improvement from subjects with **large improvement** for all outcomes.

	Subjects				CC		
Outcome	**(n)**	**Cut-off**	**Sensitivity**	**Specificity**	**(%)**	**+LR**	**-LR**
Maximum Pain	43	-0.90	76.67	84.62	79.07	4.98	0.28
Minimum Pain	43	-0.20	76.67	92.31	81.40	9.97	0.25
Current Pain	43	-1.90	93.33	69.23	86.05	3.03	0.10
MFIQ	43	-2.00	70.00	76.92	72.09	3.03	0.39
HIT-6	43	-6.26	80.00	76.92	79.07	3.47	0.26
PPT Temporalis	43	0.22	36.67	53.85	41.86	0.79	1.18
PPT Masseters	43	0.18	40.00	38.46	39.53	0.65	1.56
MMO	43	2.69	69.23	53.33	58.14	1.48	0.58
CCFT	43	60.00	53.85	80.00	72.09	2.69	0.57

MMO - Maximum mouth opening, CCFT - Craniocervical flexion test, HIT-6 - Headache impact test, PPT - pressure pain threshold, n - number of subjects, CC - correctly classified, +LR - Positive Likelihood Ratio, -LR - Negative Likelihood Ratio

Table 4. P-value and area under the curve (AUC) of the regression analyses using each outcome as predictive value for GRCS and also some of the comparisons between the outcomes.

	Moderately improved		**Largely improved**	
	p-value	**AUC**	**p-value**	**AUC**
Single regression analysis				
PPT Temporalis	0.05#	0.670	0.50	0.605
PPT Masseters	0.33	0.569	0.06#	0.709
MMO	0.50	0.568	0.11#	0.646
CCFT	0.14#	0.670	0.41	0.662
Maximum Pain	0.09#	0.680	<0.01#	0.811
Minimum Pain	0.09#	0.699**	<0.01#	0.867**
Current Pain	0.14#	0.620	<0.01#	0.815
MFIQ	0.06#	0.627	0.05#	0.721
HIT-6	0.14#	0.664**	0.41	0.853**
Multiple regression analysis (2 outcomes)				
HIT-6 + MFIQ	0.02*	0.701	<0.01*	0.887
Min pain + MFIQ	0.04*	0.683	<0.01*	0.838
Min pain + HIT-6	0.02*	0.763**	<0.01*	0.915**
Min pain + PPT Temporalis	0.22	0.674	-	-
HIT-6 + PPT Temporalis	0.08	0.744	-	-
Min pain + PPT Temporalis	0.13	0.718	-	-
Min pain + PPT Masseters	-	-	<0.01*	0.871
MMO + PPT Masseters	-	-	<0.01*	0.764
MFIQ + PPT Masseters	-	-	<0.01*	0.779
MMO + MFIQ	-	-	<0.01*	0.725
Multiple regression analysis (>2 outcomes)				
HIT-6 + MFIQ + Min pain	0.01*	0.727	<0.01*	0.928**
HIT-6 + CCFT + Min pain	0.01*	0.772**	-	-
HIT-6 + Min pain + PPT Temporalis	0.05	0.777	-	-
MFIQ + Min pain + PPT Masseters	-	-	<0.01*	0.871
HIT-6 + MFIQ + Min pain + MMO	-	-	<0.01*	0.925

HIT-6 + CCFT + Min pain + MFIQ	<0.01*	0.764	-	-

AUC- Area under the curve, PPT- Pressure pain threshold, MMO - Maximum mouth opening, CCFT- Craniocervical flexion test, MFIQ- Mandibular function impairment questionnaire, HIT-6 - headache impact test. #statistically significant ($p<0.2$), *statistically significant ($p<0.05$), **highest area under the curve with statistical significance.