Original Article

J. Phys. Ther. Sci.
26: 915–920, 2014

Intra-rater and Inter-rater Reliability of Mandibular Range of Motion Measures Considering a Neutral Craniocervical Position

Hector Beltran-Alacreu, PT, MSc[1–3], Ibai López-de-Uralde-Villanueva, PT, MSc[1–3], Alba Paris-Alemany, PT, MSc[1–4], Santiago Angulo-Díaz-Parreño, MSc[2, 3, 5], Roy La Touche, PT, MSc[1–4]*

[1)] *Department of Physiotherapy, Faculty of Health Science, The Center for Advanced Studies University La Salle, Universidad Autónoma de Madrid: Calle La Salle 10, Madrid, Madrid 28023, Spain*
[2)] *Research Group on Movement and Behavioral Science and Study of Pain, The Center for Advanced Studies University La Salle, Universidad Autónoma de Madrid, Spain*
[3)] *Institute of Neuroscience and Craniofacial Pain (INDCRAN), Spain*
[4)] *Hospital La Paz Institute for Health Research, Spain*
[5)] *Faculty of Medicine, Universidad San Pablo CEU, Spain*

Abstract. [Purpose] The aim of this study was to determine the inter-rater and intra-rater reliability of the mandibular range of motion (ROM) considering the neutral craniocervical position when performing the measurements. [Subjects and Methods] The sample consisted of 50 asymptomatic subjects. Two raters measured four mandibular ROMs (maximal mouth opening (MMO), laterals, and protrusion) using the craniomandibular scale. Subjects alternated between raters, receiving two complete trials per day, two days apart. Intra- and inter-rater reliability was determined using intra-class correlation coefficients (ICCs). Bland-Altman analysis was used to assess reliability, bias, and variability. Finally, the standard error of measurement (SEM) and minimal detectable change (MDC) were analyzed to measure responsiveness. [Results] Reliability was good for MMO (inter-rater, ICC= 0.95−0.96; intra-rater, ICC= 0.95−0.96) and for protrusion (inter-rater, ICC= 0.92−0.94; intra-rater, ICC= 0.93−0.96). Reliability was moderate for lateral excursions. The MMO and protrusion SEM ranged from 0.74 to 0.82 mm and from 0.29 to 0.49 mm, while the MDCs ranged from 1.73 to 1.91 mm and from 0.69 to 0.14 mm respectively. The analysis showed no random or systematic error, suggesting that effect learning did not affect reliability. [Conclusion] A standardized protocol for assessment of mandibular ROM in a neutral craniocervical position obtained good inter- and intra-rater reliability for MMO and protrusion and moderate inter- and intra-rater reliability for lateral excursions.

Key words: Reliability, Range of motion, Temporomandibular joint

(This article was submitted Nov. 25, 2013, and was accepted Jan. 8, 2014)

INTRODUCTION

The temporomandibular disorders (TMD) are frequent clinical issues. In Western countries, they affect approximately 25–85% of the population[1, 2)]. Furthermore, the probability of suffering this kind of dysfunction is higher between the ages of 13 and 35, especially in females[3)].

The most prevalent signs and symptoms associated with these disorders are tenderness of the mastication muscles, pain (head, face, inner ear, etc.), joint noise, and a decrease in the range of movement (ROM) of the temporomandibular joint (TMJ)[4, 5)]. These signs and symptoms may favor the presence of functional limitations such as difficulty of mastication and in vocalization[6)].

Therefore, it is convenient to measure the ROM in a dynamic way to quantify the TMD[7, 8)]. In addition, physicians can better comprehend the type of pathology of TMD thanks to evaluation of the ROM, final position, and the jaw movement pattern[9)]. Due to this, a variety of mandibular movements have been measured, although the most noted are the maximal mouth opening (MMO) and the lateral excursions[4)].

To determine if there is a limitation in mandibular movement, it is necessary to know the physiological ROM[10)]. However, there is controversy about the values of the MMO considered normal[11)]. According to a study carried out by Zawawi et al.[12)], the values are between 42 and 68 mm in males, whilst in females, they are between 40 and 57 mm. One of the possible causes of this lack of agreement is the absence of a protocol that refers to the craniocervical posture while measurement of the mandibular movement is done, since the head posture affects the MMO in asymptomatic subjects and subjects with TMD[10)]. Recent scientific evidence has shown an interaction between motor activity of the head and neck and jaw function[13, 14)], and these findings support the use of protocols to control the craniocer-

*Corresponding author. Roy La Touche (E-mail: roylatouche@yahoo.es)

vical position while mandibular dynamics and specifically mandibular ROM are measured.

The changes in the MMO can reflect the impact of the TMD as well as the therapeutic success of the intervention[15]. To detect changes in the results, it is necessary to have a consistent measurement procedure. This consistency in the measurement procedure is often expressed as an inter-evaluator, intra-evaluator, or test/re-test reliability coefficient. However, these coefficients are not measured based on quantitative changes[16, 17]. Recently, the minimum detectable change (MDC) with an appropriate measurement for a significant quantitative and statistical change has been proposed[18–20]. Furthermore, it is worth mentioning that there are numerous instruments that can be used to measure mandibular movements, all of them showing a high intra-rater and inter-rater reliability[8, 10, 18, 21–27].

The aim of this study was to determine the inter-rater and intra-rater reliability of the mandibular ROM considering the neutral craniocervical position when performing measurement.

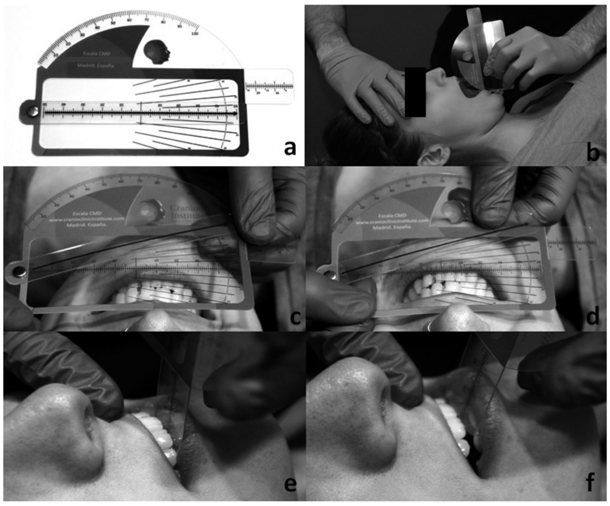

Fig. 1. a) Craniomandibular Scale; b) Maximal mouth opening measure; c) Lateral Excursion measure, first phase; d) Lateral Excursion measure, second phase; e) Protrusion measure, first phase; f) Protrusion measure, second phase.

SUBJECTS AND METHODS

We employed a prospective study with a single-group repeated measures for inter- and inter-rater reliability design. This study was planned and conducted in accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS)[28].

Sample size was calculated using the method described by Walter et al.[29], which estimates the sample size based on the intraclass correlation coefficient (ICC). For sample size estimates, we used the Power Analysis & Sample Size Software (PASS 12). A sample size of 47 subjects with 16 observations per subject (two sets of four measurements were performed each day for two days) achieves 80% power ($\beta = 0.2$) to detect an ICC of 0.80 under the alternative hypothesis when the ICC under the null hypothesis is 0.70 using an F-test with a significance level of 0.05. To allow for possible dropout from the study, we aimed to recruit at least 50 subjects.

A convenience sample of asymptomatic volunteers was obtained from our university campus and the local community through flyers, posters, and social media. The sample consisted of 50 asymptomatic subjects (30 females and 20 males, mean ± SD age 38.35 ± 10.31 years old) who were between 18 and 65 years old (the inclusion criterion); the exclusion criteria were (1) history of neck or face pain in data collection, (2) history of neck or face pain in the last 6 months, (3) being toothless, and (4) signs and symptoms of TMD. Prior to their participation, all individuals agreed to participate in the study and signed a statement of informed consent. All of the procedures used in this study were planned under the ethical norms of the Helsinki Declaration and were approved by the ethics committee of the Center for Advanced Studies, University La Salle.

Measurements were made by two physiotherapists trained to measure with a craniomandibular (CMD) scale. Both had more than 3 years of clinical experience.

The CMD scale is a thin plastic device that allows assessment of the TMJ movements with two parts (Fig. 1a). The first part has three-millimeter rulers (mm scale), which are used to measure the opening, lateral excursion, and the protrusion of the mouth. The second part is a plastic arm attached to the scale with a black line painted on it. This allows measurement of the mandibular deviations. The unit of measure of the scale is millimeters (mm).

Each subject came to the center on 2 days with an interval between sessions of 48 hours. On the first day, rater A performed the first measurement followed by rater B. On the second day, rater B performed the first measurement followed by rater A.

On the first day, the study protocol was explained to the subject, and he/she read and signed an informed consent form. Before starting the measurements, the subjects took off glasses, caps, and any jewelry. They all wore short-sleeved shirts, and those with long hair put it up. The movements were explained to the subjects, including how to find the neutral craniocervical position of the head in a supine position. The movements that were measured in this study and the order in which they were performed were as follows:

The subjects were positioned supine on a stretcher with their head in a neutral position and a small rolled towel under the occipital, which was reviewed by the rater. All measurements were performed with the scale in the same head and body position and were repeated three times. Using a pencil, the therapist drew a line crossing from the superior left central incisor to the inferior left central incisor. Raters were blinded to the measurements.

Maximal Mouth Opening (MMO). We took the CMD scale with the right hand, and the left hand was placed on the forehead to prevent loss of the subject's neutral position. The following verbal command was given: "open your mouth as wide as you can without moving your head". The assessor placed the scale on the incisal edge of the maxillary central incisor that was most vertically oriented, measured vertically to the labioincisal edge of the opposing mandibular incisor, and recorded this as the MMO measurement (Fig. 1b).

Lateral Excursions. The first step was the first verbal command: "show me your teeth". After the assessor placed the CMD scale with the 0 on the horizontal ruler on the mark of the incisors, second verbal command was given: "open your mouth slightly and move the mandible as far as possible to the right" (Fig. 1c). The upper incisor's mark remained at 0 mm, and the lower incisor's mark was used to determine the measurement in millimeters (Fig. 1d). The right lateral excursion (RLE) was always performed first, and then the left lateral excursion (LLE) was performed.

Protrusion. To measure the protrusion, the subject was given the following direction: "place your mouth in a comfortable position". The assessor then separated downwards the lower lip, and placed the small ruler on the inferior incisors, and took a measurement (x) of the upper incisors. The assessor then gave the subjects the following direction: "slide your jaw straight out in front of you as far as you can without moving your neck" (Fig. 1e). Then the CMD scale was held by the upper incisors, and the measurement was taken with respect to the lower incisors (y) (Fig. 1f). If the subject had a deep overbite, they were asked to open their mouth wider so that there was no interference from the maxillary incisors. The evaluator added together both measurements ($x + y = z$ mm). If the subject had lower incisors in front of the upper incisors, then the evaluator subtracted instead of adding measurements ($x - y = z$ mm).

For each of the measurements, the total mean is reported together with the standard deviation (SD). The Kolmogorov-Smirnov test was used to analyze the normal distribution of the variables ($p > 0.05$).

The intra-rater and inter-rater reliability was evaluated using the ICC[30] and Bland-Altman analysis with limits of agreement (LOA)[17]. The $ICC_{2,1}$ was designated as the two-way analysis of variance (ANOVA) random model for absolute agreement of measures. Interpretations of the $ICC_{2,1}$ was performed according to previously established categories for expressing levels of reliability: > 0.75 = good level of agreement, 0.50 to 0.75 = moderate level of agreement, and < 0.50 = poor level of agreement[31].

Bland-Altman Plots were constructed by calculating the mean difference between two measurements. The mean difference between trials is reported as bias. The bias ± the coefficient of reliability (the standard deviation of the bias multiplied by 1.96) is reported as the LOA[30, 32], and these plots were formed to give a graphical interpretation of the data as well as to determine reproducibility bias. Calculation of the occurrence of systematic or random changes in the data mean was performed through calculation of 95% confidence intervals of the mean differences between the data values of two measurements.

The closer the mean difference was to 0 and the smaller the SD of this difference, the better was the agreement[17]. Bland-Altman analysis was performed using MedCalc for Windows, version 12.7.1.0 (MedCalc Software, Mariakerke, Belgium).

Measurement error was expressed as the standard error of measurement (SEM), which was calculated as $SD \times \sqrt{1 - ICC}$, where SD is the SD of values from all participants and ICC is the reliability coefficient[33]. Measurement error was the systematic and random error of a patient's score that was not attributable to true changes in the construct to be measured[34].

Responsiveness was assessed with the MDC. The MDC_{90} expresses the minimal magnitude of change required to be 90% confident that the observed change between two measurements reflects real change and not just measurement error[35]. It was calculated as $SEM \times \sqrt{2} \times 1.96$[35, 36].

Data were analyzed with the SPSS statistical package (SPSS v.20.0; IBM Corp., Armonk, NY, USA).

RESULTS

No subjects were excluded from the study based on the inclusion and exclusion criteria.

Intra-rater reliability descriptive statistics, ICCs and associated 95% CIs, SEMs, and MDC_{90} values for trials 1 and 2 for each rater are presented in Table 1. The $ICC_{2,1}$ values were good for MMO and protrusion ($ICC_{2,1}$ range, 0.99−0.93); for both lateral excursion measurements, the $ICC_{2,1}$ values were moderate ($ICC_{2,1}$ range, 0.77−0.62).

Inter-rater reliability descriptive statistics, ICCs and associated 95% CIs, SEMs, and MDC_{90} values for trials 1 and 2 for each rater are presented in Table 2. The ICC values for MMO and protrusion measurements were good ($ICC_{2,1}$ range, 0.96−0.92); for both lateral excursions, the $ICC_{2,1}$ values were moderate ($ICC_{2,1}$ range, 0.71−0.51).

The average mean differences in measurements by raters A and B are reported (Tables 3 and 4), together with the SD and the 95% CI. Bland-Altman analysis for the inter-rater and intra-rater performances are shown with the LOA for each MMO, protrusion LLE, and RLE assessment (Tables 3 and 4).

DISCUSSION

This study provides new information about the reliability of four measurements of mandibular ROM in asymptomatic subjects with emphasis on the process of keeping the neutral craniocervical position during the different measurements. The results of this study demonstrate that measurement of the mandibular ROM with the CMD scale using a standardized measurement protocol performed by two experienced evaluators shows adequate intra- and inter-rater reliability.

The MMO and protrusion showed better intra- and inter-rater reliability results than the RLE or LLE. These results are similar to those observed in other studies with adult patients, adolescents with TMD, and asymptomatic subjects[21, 22, 37, 38]. Most lateral excursion measures have lower reliability than other mandibular ROM measures; however, the study by Walter et al.[29] found excellent intra- and inter-rater reliability for lateral excursions, protrusion, and MMO in healthy subjects and patients with TMD.

According to the classification of reliability levels proposed by Portney and Watkins, our results for both intra- and inter-rater reliability would be good for the MMO and protrusion and moderate for the lateral excursions (intra-rater, $ICC_{2,1}$ range, 0.77−0.62; inter-rater, $ICC_{2,1}$ range, 0.71−0.51); however, despite this classification, some au-

Table 1. Intra-rater reliability and descriptive statistics for measurements of mandibular ROM

Measurements	Rater A						Rater B					
	Trial 1 (Mean ± SD)	Trial 2 (Mean ± SD)	$ICC_{3,1}$	95% CI* for ICC	SEM	MDC	Trial 1 (Mean ± SD)	Trial 2 (Mean ± SD)	$ICC_{3,1}$	95% CI* for ICC	SEM	MDC
MMO (mm)	43.48 ± 4.11	43.5 ± 4.03	0.96	0.93–0.98	0.77	1.79	43.88 ± 3.86	43.84 ± 3.78	0.95	0.92–0.97	0.82	1.91
RLE (mm)	9.88 ± 0.82	9.98 ± 0.67	0.62	0.41–0.76	0.46	1.07	9.88 ± 0.72	9.92 ± 0.72	0.72	0.56–0.83	0.38	0.88
LLE (mm)	9.9 ± 0.8	10 ± 0.65	0.77	0.63–0.86	0.35	0.83	9.96 ± 0.81	10 ± 0.72	0.70	0.52–0.81	0.42	0.99
Protrusion (mm)	7.59 ± 1.66	7.66 ± 1.70	0.96	0.94–0.98	0.29	0.69	7.62 ± 1.85	7.72 ± 1.66	0.93	0.88–0.96	0.46	1.07

SD=standard deviation; ICC=interclass correlation coefficient; CI=confidence interval; SEM=standard error of measure; MDC=minimal detectable change; RLE=right lateral excursion; LLE=left lateral excursion; mm=millimeters

Table 2. Inter-rater reliability and descriptive statistics for measurements of mandibular ROM

	Trial 1						Trial 2					
	Rater A (Mean ± SD)	Rater B (Mean ± SD)	$ICC_{3,1}$	95% CI* for ICC	SEM	MDC	Rater A (Mean ± SD)	Rater B (Mean ± SD)	$ICC_{3,1}$	95% CI* for ICC	SEM	MDC
MMO (mm)	43.4 ± 4.11	43.88 ± 3.86	0.95	0.91–0.97	0.80	1.88	43.50 ± 4.03	43.84 ± 3.78	0.96	0.93–0.97	0.74	1.73
RLE (mm)	9.88 ± 0.82	9.88 ± 0.72	0.51	0.27–0.69	0.54	1.27	9.98 ± 0.67	9.92 ± 0.72	0.62	0.42–0.77	0.42	0.99
LLE (mm)	9.96 ± 0.83	9.96 ± 0.81	0.71	0.55–0.83	0.44	1.03	10.03 ± 0.65	10 ± 0.72	0.67	0.49–0.80	0.39	0.92
Protrusion (mm)	7.59 ± 1.66	7.62 ± 1.85	0.92	0.87–0.95	0.49	1.14	7.66 ± 1.70	7.72 ± 1.66	0.94	0.90–0.97	0.39	0.91

SD=standard deviation; ICC=interclass correlation coefficient; CI=confidence interval; SEM=standard error of measure; MDC=minimal detectable change; RLE=right lateral excursion; LLE=left lateral excursion; mm=millimeters

Table 3. Statistical metrics from Bland-Altman analysis of the intra-rater measurements

Measurements	Intra-rater					
	Rater A			Rater B		
	Mean differences ± SD	95% CI for mean differences	LOA (Lower limit-upper limit)	Mean differences ± SD	95% CI for mean differences	LOA (Lower limit-upper limit)
MMO (mm)	−0.02±1.08	−0.32 to 0.28	−2.2 to 2.1	0.04±1.16	−0.28 to 0.37	−2.2 to 2.3
RLE (mm)	−0.10±0.65	−0.29 to 0.07	−1.39 to 1.17	−0.04±0.53	−0.19 to 0.11	−1.09 to 1.01
LLE (mm)	−0.07±0.50	−0.21 to 0.06	−1.06 to 0.91	−0.04±0.60	−0.21 to 0.13	−1.22 to 1.14
Protrusion (mm)	−0.06±0.42	−0.18 to 0.05	−0.89 to 0.76	−0.10±0.65	−0.29 to 0.07	−1.38 to 1.17

SD=standard deviation; CI=confidence interval; RLE=right lateral excursion; LLE=left lateral excursion; LOA=limits of agreement; mm=millimeters

Table 4. Statistical metrics from Bland-Altman analysis of the inter-rater measurements

Measurements	Inter-rater					
	Trial 1			Trial 2		
	Mean differences ± SD	95% CI for mean differences	LOA (Lower limit-upper limit)	Mean differences ± SD	95% CI for mean differences	LOA (Lower limit-upper limit)
MMO (mm)	−0.39±1.14	−0.72 to 0.07	−2.6 to 1.8	−0.33±1.04	−0.63 to 0.03	−2.4 to 1.7
RLE (mm)	−0.006±0.77	−0.22 to 0.21	−1.5 to 1.5	0.05±0.60	−0.11 to 0.23	−1.12 to 1.24
LLE (mm)	0.00±0.62	−0.17 to 0.17	−1.22 to 1.22	0.03±0.55	−0.12 to 0.19	−1.06 to 1.13
Protrusion (mm)	−0.02±0.69	−0.22 to 0.17	−1.39 to 1.34	−0.06±0.55	−0.22 to 0.09	−1.15 to 1.02

SD=standard deviation; CI=confidence interval; RLE=right lateral excursion; LLE=left lateral excursion; LOA=limits of agreement; mm=millimeters

thors believe that the satisfactory level of reliability should be above 0.70[39, 40].

Previous evidence regarding MMO variations depending on the head position used to perform the measure[10, 15] led us to propose that this factor should be taken into account in the measurement protocols to reduce the variability in repeated measures outcomes and thus improve reliability values. It was observed in this study that, according to Bland-Altman analysis, the mean differences in mandibular ROM were very small (small bias), all of which were close to a value 0, which rules out the existence of reproducibility bias[41]. We should add that confirmation of the presence of systematic errors in any analysis was not performed, since zero was included in the 95% CI of the mean differences and the points were symmetrically distributed around zero in the Bland-Altman plot[17, 42]. The absence of systematic error indicates that the results of the measurements obtained on the second day were not influenced by behavioral factors or the learning effect. Moreover, these results show the absence of random error, as there was no tendency towards an increase or decrease in scattering points with increased values.

The results for SEM and MDC in the inter- and intra-rater analysis of mandibular ROM were very small considering the mean absolute measurement. This situation was reflected especially in the MMO MDC (intra-rater, range = 1.79−1.91 mm; inter-rater, range = 1.73−1.88 mm). Previous studies have found higher values for the MDC, between 3 mm and 5 mm in asymptomatic subjects[43] and between 6 and 9 mm in patients with TMD[24], suggesting that such wide differences in these results in connection with ours could be due to the protocols used in these studies apparently not taking into account the position of the head when performing measurements; it is possible that they did but did not explain it.

This study has several limitations. The first is that the population used in the study consisted of asymptomatic subjects; it would be interesting to replicate this study with a population of TMD patients, as they often present mandibular mobility alterations.

In this study, we evaluated the intra-rater reliability within two trials with only 48 hours between them, so it would be advisable to reproduce this protocol with more trials and a greater interval between them.

Another limitation is that the measurements were performed by only two raters who had experience with the operation of the scale. Not including inexperienced evaluators could be considered a limitation; however, previous evidence concerning reliability showed very similar results between experienced and inexperienced evaluators when measuring mandibular ROM[22].

In our protocol, when measuring the mandibular ROM, the neutral craniocervical position was controlled by explaining its importance in advance to the subjects, a verbal reminder, and manual clamping of the head, which could be performed using some instrumental fixation; thus, we consider that this could be a limitation in the study.

Due to the nature and design of this study, it was only possible to perform calculations of the MDC, and it will be necessary to determine whether the changes and evolution of the measurements are clinically relevant. This could be solved with a longitudinal clinical research design in which the minimum clinically important difference in mandibular ROM could be investigated.

Finally, we consider it necessary that future studies investigate whether the reliability of measurement of mandibular ROM is the same when using the neutral craniocervical position during outcome measurements compared with without using this position.

The results of this study demonstrate that applying a standardized protocol for the assessment of mandibular ROM in the neutral craniocervical position results in good inter- and intra-rater reliability for the MMO and protrusion, and moderate inter- and intra-rater reliability for lateral excursions. These results should be interpreted with caution, taking into account that the study was conducted with a sample of asymptomatic subjects and not patients with TMD.

REFERENCES

1) Hesse JR, Naeije M, Hansson TL: Craniomandibular stiffness toward maximum mouth opening in healthy subjects: a clinical and experimental investigation. J Craniomandib Disord, 1990, 4: 257–266. [Medline]
2) Tom BM, Rao VM, Farole A: Nondiscogenic causes of temporomandibular joint pain. Cranio, 1991, 9: 220–227. [Medline]
3) Helkimo M: Epidemiological surveys of dysfunction of the masticatory system. Oral Sci Rev, 1976, 7: 54–69. [Medline]
4) Phillips DJ Jr, Gelb M, Brown CR, et al.: Guide to evaluation of permanent impairment of the temporomandibular joint. American Academy of Head, Neck and Facial Pain; American Academy of Orofacial Pain; American Academy of Pain Management; American College of Prosthodontists; American Equilibration Society and Society of Occlusal Studies; American Society of Maxillofacial Surgeons; American Society of Temporomandibular Joint Surgeons; International College of Cranio-mandibular Orthopedics; Society for Occlusal Studies. Cranio, 1997, 15: 170–178. [Medline]
5) Chaput E, Gross A, Stewart R, et al.: The diagnostic validity of clinical tests in temporomandibular internal derangement: a systematic review and Meta-analysis. Physiother Can, 2012, 64: 116–134. [Medline] [CrossRef]
6) Helland MM: Anatomy and function of the temporomandibular joint. J Orthop Sports Phys Ther, 1980, 1: 145–152. [Medline] [CrossRef]
7) Griffith RH: Report of the president's conference on the examination, diagnosis, and management of temporomandibular disorders. J Am Dent Assoc, 1983, 106: 75–77. [Medline]
8) Danis CG, Mielenz TJ: Reliability of measuring active mandibular excursion using a new tool: the Mandibular Excursiometer. J Orthop Sports Phys Ther, 1997, 25: 213–219. [Medline] [CrossRef]
9) Kraus SL: TMJ Disorders: Management of the Craniomandibular Complex (Clinics in Physical Therapy). New York: Churchill Livingstone, 1978.
10) Higbie EJ, Seidel-Cobb D, Taylor LF, et al.: Effect of head position on vertical mandibular opening. J Orthop Sports Phys Ther, 1999, 29: 127–130. [Medline] [CrossRef]
11) Abou-Atme YS, Chedid N, Melis M, et al.: Clinical measurement of normal maximum mouth opening in children. Cranio, 2008, 26: 191–196. [Medline]
12) Zawawi KH, Al-Badawi EA, Lobo SL, et al.: An index for the measurement of normal maximum mouth opening. J Can Dent Assoc, 2003, 69: 737–741. [Medline]
13) Eriksson PO, Zafar H, Nordh E: Concomitant mandibular and head-neck movements during jaw opening-closing in man. J Oral Rehabil, 1998, 25: 859–870. [Medline] [CrossRef]
14) Eriksson PO, Häggman-Henrikson B, Nordh E, et al.: Co-ordinated mandibular and head-neck movements during rhythmic jaw activities in man. J Dent Res, 2000, 79: 1378–1384. [Medline] [CrossRef]
15) Stegenga B, de Bont LG, Dijkstra PU, et al.: Short-term outcome of arthroscopic surgery of temporomandibular joint osteoarthrosis and internal derangement: a randomized controlled clinical trial. Br J Oral Maxillofac Surg, 1993, 31: 3–14. [Medline] [CrossRef]
16) Cronbach JL, Gleser GC, Nanda H, et al.: The dependability of behavioural measurements: Theory of generalizability for scores and profiles. New York: John Wiley and Sons, 1972.
17) Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1986, 1: 307–310. [Medline] [CrossRef]
18) Kropmans TJ, Dijkstra PU, Stegenga B, et al.: Smallest detectable difference in outcome variables related to painful restriction of the temporomandibular joint. J Dent Res, 1999, 78: 784–789. [Medline] [CrossRef]
19) Kropmans TJ, Dijkstra PU, Stegenga B, et al.: Therapeutic outcome assessment in permanent temporomandibular joint disc displacement. J Oral Rehabil, 1999, 26: 357–363. [Medline] [CrossRef]
20) Wood GD, Branco JA: A comparison of three methods of measuring maximal opening of the mouth. J Oral Surg, 1979, 37: 175–177. [Medline]
21) Goulet JP, Clark GT, Flack VF, et al.: The reproducibility of muscle and joint tenderness detection methods and maximum mandibular movement measurement for the temporomandibular system. J Orofac Pain, 1998, 12: 17–26. [Medline]
22) Leher A, Graf K, PhoDuc JM, et al.: Is there a difference in the reliable measurement of temporomandibular disorder signs between experienced and inexperienced examiners? J Orofac Pain, 2005, 19: 58–64. [Medline]
23) Saund DS, Pearson D, Dietrich T: Reliability and validity of self-assessment of mouth opening: a validation study. BMC Oral Health, 2012, 12: 48. [Medline] [CrossRef]
24) Kropmans T, Dijkstra P, Stegenga B, et al.: Smallest detectable difference of maximal mouth opening in patients with painfully restricted temporomandibular joint function. Eur J Oral Sci, 2000, 108: 9–13. [Medline] [CrossRef]
25) Hoole J, Jenkins GW, Kanatas AN, et al.: Reliability and validity of a new method of nurse-led assessment of trismus. Br J Oral Maxillofac Surg, 2011, 49: 430–437. [Medline] [CrossRef]
26) Rivera-Morales WC, Goldman BM, Jackson RS: Simplified technique to measure mandibular range of motion. J Prosthet Dent, 1996, 75: 56–59. [Medline] [CrossRef]
27) Walker N, Bohannon RW, Cameron D: Discriminant validity of temporomandibular joint range of motion measurements obtained with a ruler. J Orthop Sports Phys Ther, 2000, 30: 484–492. [Medline] [CrossRef]
28) Kottner J, Audigé L, Brorson S, et al.: Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol, 2011, 64: 96–106. [Medline] [CrossRef]
29) Walter SD, Eliasziw M, Donner A: Sample size and optimal designs for reliability studies. Stat Med, 1998, 17: 101–110. [Medline] [CrossRef]
30) Krebs DE: Declare your ICC type. Phys Ther, 1986, 66: 1431. [Medline]
31) Portney LG, Watkins MP: Foundations of Clinical Research; Applications to Practice. 3. Pearson/Prentice Hall, Upper Saddle River, NJ, 2009.
32) Bland JM, Altman DG: Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet, 1995, 346: 1085–1087. [Medline] [CrossRef]
33) Weir JP: Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res, 2005, 19: 231–240. [Medline]
34) Mokkink LB, Terwee CB, Patrick DL, et al.: The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol, 2010, 63: 737–745. [Medline] [CrossRef]
35) Haley SM, Fragala-Pinkham MA: Interpreting change scores of tests and measures used in physical therapy. Phys Ther, 2006, 86: 735–743. [Medline]
36) Wyrwich KW: Minimal important difference thresholds and the standard error of measurement: is there a connection? J Biopharm Stat, 2004, 14: 97–110. [Medline] [CrossRef]
37) Wahlund K, List T, Dworkin SF: Temporomandibular disorders in children and adolescents: reliability of a questionnaire, clinical examination, and diagnosis. J Orofac Pain, 1998, 12: 42–51. [Medline]
38) Hatch JP, Rugh JD, Sakai S, et al.: Reliability of the craniomandibular index. J Orofac Pain, 2002, 16: 284–295. [Medline]
39) Kramer MS, Feinstein AR: Clinical biostatistics. LIV. The biostatistics of concordance. Clin Pharmacol Ther, 1981, 29: 111–123. [Medline] [CrossRef]
40) Andresen EM: Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil, 2000, 81: S15–S20. [Medline] [CrossRef]
41) Rankin G, Stokes M: Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil, 1998, 12: 187–199. [Medline] [CrossRef]
42) Lexell JE, Downham DY: How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil, 2005, 84: 719–723. [Medline] [CrossRef]
43) Kropmans TJ, Dijkstra PU, van Veen A, et al.: The smallest detectable difference of mandibular function impairment in patients with a painfully restricted temporomandibular joint. J Dent Res, 1999, 78: 1445–1449. [Medline] [CrossRef]