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

Hector Beltran-Alacreu, PT, MSc1–3), Ibai López-de-Uralde-Villanueva, PT, MSc1–3), Alba Paris-Alemany, PT, MSc1–4), Santiago Angulo-Díaz-Parreño, MSc2, 3, 5), Roy La Touche, PT, MSc1–4)*

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 craniofacial 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 craniofacial 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

INTRODUCTION

The temporomandibular disorders (TMD) are frequent clinical issues. In Western countries, they affect approximately 25–85% of the population1, 2). Furthermore, the probability of suffering this kind of dysfunction is higher between the ages of 13 and 35, especially in females3).

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 vocalization6).

Therefore, it is convenient to measure the ROM in a dynamic way to quantify the TMD7, 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 pattern9). Due to this, a variety of mandibular movements have been measured, although the most noted are the maximal mouth opening (MMO) and the lateral excursions4).

To determine if there is a limitation in mandibular movement, it is necessary to know the physiological ROM10). However, there is controversy about the values of the MMO considered normal11). 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 craniofacial posture while measurement of the mandibular movement is done, since the head posture affects the MMO in asymptomatic subjects and subjects with TMD10). Recent scientific evidence has shown an interaction between motor activity of the head and neck and jaw function13, 14), and these findings support the use of protocols to control the craniofacial posture while measurement of the mandibular movement is done, since the head posture affects the MMO in asymptomatic subjects and subjects with TMD10).
Studies (GRRAS) the Guidelines for Reporting Reliability and Agreement. This study was planned and conducted in accordance with repeated measures for inter- and intra-rater reliability design.

The changes in the MMO can reflect the impact of the TMD as well as the therapeutic success of the intervention\(^5\). 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 cranio- cervical position when performing measurement.

**SUBJECTS AND METHODS**

We employed a prospective study with a single-group repeated measures for inter- and intra-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 cranio- cervical 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)[31]. The ICC was designated as the twoway analysis of variance (ANOVA) random model for absolute agreement of measures. Interpretations of the ICC[2,1] values were 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[31]. 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

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Rater A</th>
<th></th>
<th></th>
<th></th>
<th>Rater B</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trial 1</td>
<td>Trial 2</td>
<td>ICC</td>
<td>95% CI* for ICC</td>
<td>SEM</td>
<td>MDC</td>
<td>Trial 1</td>
<td>Trial 2</td>
</tr>
<tr>
<td>MMO (mm)</td>
<td>43.48 ± 4.11</td>
<td>43.5 ± 4.03</td>
<td>0.96</td>
<td>0.93–0.98</td>
<td>0.77</td>
<td>1.79</td>
<td>43.88 ± 3.86</td>
<td>43.84 ± 3.78</td>
</tr>
<tr>
<td>RLE (mm)</td>
<td>9.88 ± 0.82</td>
<td>9.98 ± 0.67</td>
<td>0.62</td>
<td>0.41–0.76</td>
<td>0.46</td>
<td>1.07</td>
<td>9.88 ± 0.72</td>
<td>9.92 ± 0.72</td>
</tr>
<tr>
<td>LLE (mm)</td>
<td>9.9 ± 0.8</td>
<td>10 ± 0.65</td>
<td>0.77</td>
<td>0.63–0.86</td>
<td>0.35</td>
<td>0.83</td>
<td>9.96 ± 0.81</td>
<td>10 ± 0.72</td>
</tr>
<tr>
<td>Protrusion (mm)</td>
<td>7.59 ± 1.66</td>
<td>7.66 ± 1.70</td>
<td>0.96</td>
<td>0.94–0.98</td>
<td>0.29</td>
<td>0.69</td>
<td>7.62 ± 1.85</td>
<td>7.72 ± 1.66</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Rater A</th>
<th></th>
<th></th>
<th></th>
<th>Rater B</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trial 1</td>
<td>Trial 2</td>
<td>ICC</td>
<td>95% CI* for ICC</td>
<td>SEM</td>
<td>MDC</td>
<td>Trial 1</td>
<td>Trial 2</td>
</tr>
<tr>
<td>MMO (mm)</td>
<td>43.4 ± 4.11</td>
<td>43.88 ± 3.86</td>
<td>0.95</td>
<td>0.91–0.97</td>
<td>0.80</td>
<td>1.88</td>
<td>43.5 ± 4.03</td>
<td>43.84 ± 3.78</td>
</tr>
<tr>
<td>RLE (mm)</td>
<td>9.88 ± 0.82</td>
<td>9.88 ± 0.72</td>
<td>0.51</td>
<td>0.27–0.69</td>
<td>0.54</td>
<td>1.27</td>
<td>9.98 ± 0.67</td>
<td>9.92 ± 0.72</td>
</tr>
<tr>
<td>LLE (mm)</td>
<td>9.96 ± 0.83</td>
<td>9.96 ± 0.81</td>
<td>0.71</td>
<td>0.55–0.83</td>
<td>0.44</td>
<td>1.03</td>
<td>10.03 ± 0.65</td>
<td>10 ± 0.72</td>
</tr>
<tr>
<td>Protrusion (mm)</td>
<td>7.59 ± 1.66</td>
<td>7.62 ± 1.85</td>
<td>0.92</td>
<td>0.87–0.95</td>
<td>0.49</td>
<td>1.14</td>
<td>7.66 ± 1.70</td>
<td>7.72 ± 1.66</td>
</tr>
</tbody>
</table>

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
authors believe that the satisfactory level of reliability should be above 0.7039, 40.

Previous evidence regarding MMO variations depending on the head position used to perform the measure16, 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 bias41. 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.

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 performed using some instrumental fixation; thus, we recommend that this could be a limitation in the study.

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 subjects43) and between 6 and 9 mm in patients with TMD24), 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 ROM21).

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.

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

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Rater A</th>
<th>Rater B</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMO (mm)</td>
<td>Mean differences ± SD</td>
<td>95% CI for mean differences</td>
</tr>
<tr>
<td></td>
<td>−0.02±1.08</td>
<td>−0.32 to 0.28</td>
</tr>
<tr>
<td>RLE (mm)</td>
<td>−0.10±0.65</td>
<td>−0.29 to 0.07</td>
</tr>
<tr>
<td>LLE (mm)</td>
<td>−0.07±0.50</td>
<td>−0.21 to 0.06</td>
</tr>
<tr>
<td>Protrusion (mm)</td>
<td>−0.06±0.42</td>
<td>−0.18 to 0.05</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Trial 1</th>
<th>Trial 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMO (mm)</td>
<td>Mean differences ± SD</td>
<td>95% CI for mean differences</td>
</tr>
<tr>
<td></td>
<td>−0.39±1.14</td>
<td>−0.72 to 0.07</td>
</tr>
<tr>
<td>RLE (mm)</td>
<td>−0.006±0.77</td>
<td>−0.22 to 0.21</td>
</tr>
<tr>
<td>LLE (mm)</td>
<td>0.00±0.62</td>
<td>−0.17 to 0.17</td>
</tr>
<tr>
<td>Protrusion (mm)</td>
<td>−0.02±0.69</td>
<td>−0.22 to 0.17</td>
</tr>
</tbody>
</table>

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

36) Wyrwy KW: Minimal important difference thresholds and the standard error of measurement: is there a connection? J Biopharm Stat, 2004, 14: 97–110. [Medline] [CrossRef]