Reliability and reproducibility of the method of assessment of midpalatal suture maturation: 
A tomographic study

Natália Maria Vieira Barbosa; Ana Cláudia de Castro Ferreira Conti; Leopoldino Capelozza-Filho; Renata Rodrigues de Almeida-Pedrin; Mauricio de Almeida Cardoso

ABSTRACT

Objectives: To assess reliability and reproducibility of the individual assessment of midpalatal suture maturation in computed tomography among orthodontists and radiologists for potential diagnosis application.

Materials and Methods: Sixty axial slices from cone-beam computed tomography and multi-slice CT scans of patients aged between 11 and 21 years old (33 females and 27 males) were selected. For the investigation of reliability and reproducibility of the method, two groups of examiners were established. The first group consisted of 11 orthodontists and the second consisted of 10 radiologists. Each group examined the images and performed individual assessment of the midpalatal suture maturation method twice within an interval of 21 days. During the first and second analyses, the sequence of images was randomized to reduce potential bias. Weighted Cohen’s kappa was performed to assess inter- and intra-examiners’ agreement. The percentage of perfect agreement and the number of stages apart for each disagreement were calculated. The significance level was $P < .05$.

Results: The overall inter-examiner agreement was satisfactory in the first (kappa: 0.37) and the second (kappa: 0.34) analyses. Intra-examiner agreement outcomes were similar between orthodontists (kappa: 0.44) and radiologists (kappa: 0.41). The percentage of perfect agreement was 43.2%.

Conclusions: The method for individual assessment of midpalatal suture maturation revealed potential reliability and reproducibility. However, the agreement rate observed in the present study was not high enough for a method designed for routine clinical applications. (Angle Orthod. 2019;89:71–77.)

KEY WORDS: Tomography; X-ray computed; Maxillary expansion; Sutures, Cranial; Diagnosis

INTRODUCTION

Accurate diagnosis, proper treatment planning and timing, and success of rapid maxillary expansion (RME) may depend on the assessment of craniofacial development and maturational parameters to predict a patient’s growth pattern. Craniofacial parameters are assessed radiographically. Currently, hand-wrist bone age and cervical vertebrae maturation method are the most common skeletal parameters studied and used as tools for timing orthopedic treatment such as RME. Predicting treatment approaches is easier and more consistent in the early stages of craniofacial developmental, while uncertainty remains when skeletal maturation is more advanced.

Histologic, radiographic, and tomographic studies in scientific literature in previous years provided...
information on the morphology, development, ossification, and closure of the midpalatal suture. Despite contributing to the prognosis of surgical approaches, correlations between chronological age and maturation stages of the midpalatal suture have not yet been established. Similarly, a threshold age for the RME was not determined based on analysis of the midpalatal suture.6,14–16

Computed tomography (CT) emerged as an efficient tool for image diagnosis in situ6 because it enabled the assessment of the skeletal maturation with a detailed registration of bone morphology. Consequently, the three-dimensional images obtained with CT may be essential for making clinical decisions that involve craniofacial development. Methods based on CT have been developed for the assessment of maturation stages in the midpalatal suture.5,17,18 The method of individual assessment proposed by Angelieri et al.5 was designed to allow a reliable and reproducible diagnostic pathway for clinical use.

Specifically, this method described the morphology of the palatal suture progressively through a graphic representation of the maturation stages. The maturation of the midpalatal suture was classified into five stages (from A to E) based on a sample of 140 subjects. In CT, these stages ranged from the lack of ossification (stage A) to the complete closure of the suture (Stage E). Despite supporting the clinical investigation of skeletal maturation, Angelieri et al.5 did not observe correlation between the maturation stages and chronological age.

Optimal diagnostic tests must be reliable, reproducible and practical.19–22 The method proposed by Angelieri et al.5 helps in determining the feasibility and prognosis of RME. However, this method may depend on the level of training and skills for image analysis and interpretation. The present study aimed at investigating the reliability and reproducibility of the individual assessment of midpalatal suture maturation through the quantification of intra- and inter-examiner agreement. The clinical scope of this study focused of the evaluation of this method as a diagnostic tool for planning RME.

MATERIALS AND METHODS

The present study was approved by the Committee of Ethics in Research of Sagrado Coração University, under the protocol number: 1.514.605. All patients consented for the use and analysis of CT images.

The present study was conducted with CT scans collected from databases of three different radiologic clinics. Cone-beam CT scans (CBCT) were obtained for dental diagnosis and multi-slice CT scans (MSCT) for the investigation of alterations in the paranasal sinuses. The patients were between the ages of 11 and 21 years old. None of the patients underwent radiation exposure exclusively for the present study aim. All of the images were stored as Digital Imaging and Communications in Medicine format files for further visualization and manipulation in software for medical image analysis. The sample consisted of 202 CT scans (n = 152 CBCT and 50 MSCT), obtained in a period of 12 months. The main author examined the scans and prepared the axial slices from the midpalatal suture as recommended by Angelieri et al.5 Using the software for image manipulation InVivoDental5 (Anatomage, San Jose, Calif), the CT scans were visualized and a single axial slice of the midpalatal suture was chosen and stored as a JPEG file. The sample exclusion criteria were patients with deep palate, the need for visualization of the midpalatal suture in two or more axial slices, and the interference of any teeth in an axial slice of the midpalatal suture. Out of the initial sample (n = 202), 60 axial slices (33 females and 27 males) were randomly selected by lot. The number of slices used in the present study was determined by sample size calculation for weighted Cohen’s kappa tests that indicated a minimum of 50 axial slices.23

The main researcher codified, randomized, and organized the 60 images in four presentations (Microsoft Office Power Point; Microsoft, Redmond, Wash), in sets of 15 slices each. The sets of images were sent to two groups of examiners, the first with 11 orthodontists and the second with 10 radiologists. Twenty-one examiners volunteered in this study, but none of them participated in the study design.

According to Angelieri et al.,5 the midpalatal sutures observed in each image were classified in five stages of maturation (Figure 1). Together with the sets of images, all the examiners received training material that consisted of (1) explanation on the method with detailed description of the maturation stages; (2) a diagram retrieved from the original study of Angelieri et al.5 used to support the classification; (3) a table designed from the diagram (Figure 2); and (4) CT images provided with their respective schematic drawings also retrieved from the original study of Angelieri et al.5 and an axial slice as an example of each stage.

The examiners had no previous knowledge about the method and were instructed to read and understand the training material prior to each round of evaluation. The examiners were asked to perform the analysis (T1) on each of the four sets of images within a 24-hour interval. Next, the 60 slices were codified and randomized again and underwent a new analysis 21 days later (T2). The examiners were between 27 and 63 years old (mean age: 35 years), and answered a questionnaire about time of experience in the field (as
orthodontist or radiologist) varying between 2 and 20 years (mean: 7.47 years), and concerning their experience with CT images analysis and diagnosis. The level of midpalatal suture maturation was classified by each examiner and tabulated in a research sheet designed by the main examiner. Because 41 images were obtained from CBCT scans and 19 from MSCT, the present research also performed comparisons on rating agreement between the different CT modalities.

**Figure 1.** The midpalatal suture maturation stages A, B, C, D, and E, according to Angelieri et al.5 (A) The suture is seen as a relatively straight radiopaque line, stage A. (B) The suture appears as a sinuous line of high-density, stage B. (C) Two radiopaque, winding, and parallel lines between themselves separated by areas of low radiographic density, stage C. (D) The palatine bones become more radiopaque and the suture is not visualized in this region, stage D. Note that it is still possible to observe the two parallel radiopaque lines in the palatal area. (E) It is no longer possible to see the suture along the maxillary and palatine bones, which indicates suture fusion, stage E.

**Statistical Analysis**

Intra- and inter-examiner agreement was assessed for each group at both T1 and T2 with weighted Cohen’s kappa24 and intraclass correlation coefficient (ICC) was applied between different CT modalities, using a confidence interval of 95%. Statistics were performed using SPSS (Statistics for Windows, Version 17.0; SPSS Inc, Chicago, Ill). The percentage of perfect agreement and each stage disagreement was also calculated for all examiners at T1 and T2.
RESULTS

In total, 2520 inter-examiner analyses were performed (60 images analyzed twice by 21 examiners). In general, the examiners agreed in 1088 (43.2%) analyses, and disagreed by one stage apart in 882 (35%). Among orthodontists, agreement was observed in 578 (43.8%) analyses out of 1320 while, for radiologists, agreement was observed in 511 (42.6%) out of 1200. Table 1 expresses the agreement and disagreement distributed according to the maturation stages. For the intra-examiner analyses, orthodontists reached $kappa_w: 0.44$, while radiologists reached $0.41$ (Table 2). According to ICC, rating agreement between the different CT modalities demonstrated no significant differences with values of 0.52 for CBCT and 0.47 for MSCT.

Table 3 shows inter-examiner agreement in the first and second analyses in the two groups, as well the total agreement. In the first analysis, orthodontists reached $kappa_w: 0.39$ and radiologists, $0.44$. Together, all the examiners reached 0.37. In the second analysis, the agreement reached 0.34 and 0.35 for orthodontists and radiologists, respectively. Total agreement reached 0.34. All these values ranged between 0.21–0.60 and, according to Hallgren, would be considered between fair and moderate agreement.

Table 4 reports the median of maturation stages scored in each group and the consequent agreement between groups. Good outcomes were observed ($kappa_w: 0.72$).

DISCUSSION

The accuracy and validation in categorical diagnostic tests depends on examiners’ agreement. Methods that are considered highly reproducible are also considered reliable. Reliability is the capacity of a method to result in identical or similar outcomes in different clinical or statistical experiments. More specifically, any test or procedure considered reliable will always result in similar outcomes regardless of the time, environment, or examiner. In clinical practice, reliability plays a relevant part in reducing the occurrence of diagnostic errors.

Due to the high variability of maturational stages found in several studies that evaluated the midpalatal suture, an individual and local analysis is justified. Recent developments, such as CT scans, have enabled more accurate assessment of craniofacial morphological structures, such as the midpalatal suture. Such examinations are an important tool in clinical practice to support decisions on the feasibility of RME. In the present study, the classification of axial slices from different CT devices (CBCT and MSCT), rated by different examiners, was designed to test the method of reliability and reproducibility for assessment.
in individuals. Although a more accurate image was expected in MSCT, ICC inter-examiner analysis showed higher agreement values for CBCT, but this difference was not statistically significant.

Apart from reliability and reproducibility, validation is also necessary. This was not an aim of the present study because the method was previously validated in the literature. Specifically, the individual assessment of midpalatal suture maturation was compared with hand-wrist and cervical vertebrae maturation and showed strong statistical association.

The two examiner groups in the current study consisted of orthodontists and radiologists with varied age and experience. The selection of examiners was done in this way to test the method for clinical practice. According to Obuchowski, a minimum of 10 examiners is necessary to evaluate methods for image diagnosis. Based on that, Rainey, Burnside, and Harrison performed a study to investigate the reliability of a method for the assessment of cervical vertebrae maturation. Twenty orthodontists with no experience with the method were used as examiners. Similarly, the present study used groups with >10 examiners who had no experience with the method. In this study, the agreement scores achieved were low, ranging from fair to moderate. As the examiners only received training material from the original study of Angeli et al., it is possible that training and experience are important factors for achieving reproducibility.

Examiners with different levels of experience and familiarity with CT in practice were selected in the present study to represent the clinical routine where the method must be reproducible independently of the clinician. Only 14 examiners out of the 21 had previous experience with CT exams. This may explain the fair-to-moderate agreement rates. As the different CT images demonstrated no statistically significant differences between them, CBCT images should be preferred, since they provide lower levels of radiation exposure.

In a systematic review, Santiago et al. reported that common methodological limitations in studies on reliability of diagnostic methods include lack of image randomization, blinding, and sample size calculation.

### Table 3. Weighted Kappa (CI95%) for Inter-Examiner Agreement in First (T1) and Second (T2) Analyses for Each Group of Examiners

<table>
<thead>
<tr>
<th>Analyses</th>
<th>Examiner</th>
<th>kappa,</th>
</tr>
</thead>
<tbody>
<tr>
<td>First (T1)</td>
<td>Radiologists</td>
<td>0.44 (0.40–0.48)</td>
</tr>
<tr>
<td></td>
<td>Orthodontists</td>
<td>0.39 (0.35–0.43)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.37 (0.33–0.41)</td>
</tr>
<tr>
<td>Second (T2)</td>
<td>Radiologists</td>
<td>0.35 (0.31–0.40)</td>
</tr>
<tr>
<td></td>
<td>Orthodontists</td>
<td>0.34 (0.30–0.38)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.34 (0.30–0.38)</td>
</tr>
</tbody>
</table>

### Table 4. Weighted Kappa (CI95%) Between the Different Examiners (Considering Their Median Score for Maturation Stages)

<table>
<thead>
<tr>
<th>Examiners</th>
<th>kappa,</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiologists × Orthodontists</td>
<td>0.72 (0.60–0.84)</td>
</tr>
</tbody>
</table>

Additionally, the paper points out that when the authors participate as examiners, the results must be carefully interpreted because the results represent outcomes from professionals experienced with the method being tested. This information strengthens the methodology of the present study, especially because agreement rates were obtained with basic training, sample randomization, and calculation, and by using 21 examiners with different experience in the field.

Between the two analyses performed using the 60 images, randomization was performed for assessing the intra-examiner agreement, which resulted in moderate agreement as well. In a similar study, Gabriel et al. provided training material to examiners with explanations regarding each stage of CVM. The weighted Kappa outcome reported by the authors was <0.8. In the present study, the outcomes of weighted Kappa were lower (0.41 for orthodontists and 0.44 for radiologists), statistically interpreted as moderate agreement. This can be explained by the fact that clinicians are more familiar with lateral cephalograms than CT scans. Additionally, the authors expressed percentages of agreement and disagreement for each stage of maturation. Percentages were also expressed in the present study. In general, the agreement rate was 43% (absolute agreement), whereas the disagreement within a difference of a single stage of maturation was 35% (relative agreement). The combination of absolute and relative agreements was 78%, showing a high level of similar scores between examiners. As a biological process, midpalatal suture maturation does not change suddenly from one stage to another. Thus, disagreements within a difference of a single stage may be expected and even accepted during the analysis. This phenomenon was considered with the application of weighted Kappa, which attributes less significance to disagreements of a single stage. In this context, it should be highlighted that the transition from stage C to stage D is extremely important because this transition may suggest a change in treatment plan and prognosis.

When the groups of examiners were compared, they had different weighted Kappa values. However, the differences may not be considered statistically significant because they were in the same interpretation range (fair agreement) for the weighted Kappa tests, except for radiologists in the first evaluation who reached moderate agreement (0.44). Radiologists were in agreement to a slightly greater degree in both
analyses, which may indicate more familiarity with the process of image interpretation and analysis. The reproducibility rate achieved in the present study with the two groups of examiners was considered fair-to-moderate. However, when the median scores were considered between orthodontists and radiologists, the concordance reached excellent values (kappa w: 0.72). This indicates that, in general, the examiners displayed better agreement. These outcomes are closer the agreement rates obtained in a study with previous calibration and training. The agreement values obtained in that study were considerably higher, possibly because the examiners were more familiar and calibrated with the method. This confirms the potential of this method for diagnostic purposes but does not allow its classification as reliable and reproducible. It can be used for research purposes since a calibration process is normally required but, for clinical application, it was not indicated because orthodontists are not sufficiently familiar with this method. When using CBCT images to identify a patient’s midpalatal sutural maturation status, it would be recommended that a report from the radiologist be requested. Further studies evaluating the diagnostic test for prediction of orthodontic treatment outcomes and clinical changes would be necessary. Future studies should also be done using a longitudinal sample for analysis and to compare the findings with those from a gold standard examiner.

CONCLUSIONS

- The method for individual assessment of midpalatal suture maturation revealed potential reliability and reproducibility. However, it was not reliable enough to be used as a method for routine clinical applications.
- An extensive training program is necessary for more reliable and reproducible application of this method.

REFERENCES


