Original article

The reliability of clinical decisions based on the cervical vertebrae maturation staging method

Aydin Sohrabi*, Sahar Babay Ahari*, Hossein Moslemzadeh*, Ali Rafighi* and Zahra Aghazadeh**

Departments of *Orthodontics and **Oral Medicine, Faculty of Dentistry, Tabriz University of Medical Sciences, Tabriz, Iran

Correspondence to: Sahar Babay Ahari, Avenida Sao Goncalo, No. 1034, 3rd andar ESQ, Guimaraes, Portugal. E-mail: saharb65@yahoo.com

Summary

Objectives: Of the various techniques used to determine the optimum timing for growth modification treatments, the cervical vertebrae maturation method has great advantages, including validity and no need for extra X-ray exposure. Recently, the reproducibility of this method has been questioned. The aim of this study was to investigate the cause of poor reproducibility of this method and to assess the reproducibility of the clinical decisions made based on it.

Materials/Methods: Seventy lateral cephalograms of Iranian patients aged 9‒15 years were observed twice by five experienced orthodontists. In addition to determining the developmental stage, each single parameter involved in this method was assessed in terms of inter- and intra-observer reproducibility. In order to evaluate the reproducibility of clinical decisions based on this method, cervical vertebrae maturation staging (CVMS) I and II were considered as phase 1 and CVMS IV and V were considered as phase 3.

Results: By considering the clinical approach of the CVMS method, inter-observer reproducibility of this method increased from 0.48 to 0.61 (moderate to substantial) and intra-observer reproducibility enhanced from 0.72 to 0.74.

Limitations: 1. Complete visualization of the first four cervical vertebrae was an inclusion criterion, which also limits the clinical application of CVMS method. 2. These results can be generalized when determining growth modification treatments solely for Class II patients.

Conclusions: Difficulty in determining the morphology of C3 and C4 leads to poor reproducibility of the CVMS method. Despite this, it has acceptable reproducibility in determining the timing of functional treatment for Class II patients.

Introduction

Timing is one of the crucial aspects of orthodontic treatments that must be considered in every growth modification treatment plan. In this regard, different techniques have been used to assess skeletal maturity, of which chronological age, body height, incidence of secondary sexual characteristics, dental development, skeletal maturation of the hand and wrist, and maturation of cervical vertebrae are the most common (1).

Recently, there has been a growing emphasis on the cervical vertebrae maturation staging (CVMS) method as a determinant of skeletal age in orthodontic patients. Most studies have concluded that the accuracy of this method is as much as that of the hand and wrist skeletal maturity method (2). On the other hand, the cervical vertebrae are visible in lateral cephalograms that are typically used in routine diagnostic processes by orthodontists and there is no need for extra X-ray exposure (3). Accordingly, the British Orthodontic Society has emphasized use of the CVMS method to assess skeletal maturation and has stated that there is no orthodontic indication for the hand and wrist radiographs (4, 5).

Principally, any diagnostic test must have high sensitivity, high specificity, and predictive values. In addition, the reproducibility
must be high so that the results of the test can be cited (6). Despite the reported advantages of the CVMS method, in some studies the reproducibility of this method has been questioned and it has not been recommended as a strict clinical guideline for the timing of orthodontic treatment (7).

The major clinical application of the CVMS method is to determine the optimum timing for growth modification treatments. In addition, on top of being a rating system to classify patients into different maturational groups, it serves as a clinical guideline for the recognition of the optimum time at which to begin the orthodontic treatment. Considering this point of view, the aim of this study was to evaluate the reproducibility of this method as an indicator of optimum treatment timing in clinical applications.

Furthermore, as developmental stages of cervical vertebrae are determined by evaluating their shape and lower borders, and since the reproducibility of the CVMS method depends on the clinician’s ability to distinguish the morphologic features of the first four vertebrae, the reproducibility of assessing the morphologic characteristics of each vertebra was studied.

### Subjects and methods

A pilot study was conducted initially with 10 radiographs and three observers, in which the inter-observer reproducibility was 0.71. On the basis of this pilot study, the required sample size and the number of observers were calculated as 62 and 5, respectively, by utilizing the sample size calculation formula for agreement studies (8). Alpha significance level of 0.05, power = 80 per cent, \( \rho_s = 0.7 \), and \( \rho_i = 0.8 \) were used for this study. To increase the power of the study, the sample size was increased to 70.

The samples consisted of pre-treatment lateral cephalometric radiographs of 70 patients. Inclusion criteria consisted of an age range of 9–15 years and good quality radiographs with complete visualization of the first four cervical vertebrae. The cephalograms were selected randomly. The observers were asked to complete a table for each cephalogram, which included the following features:

1. Lower border of C2 as straight or concave;
2. Lower border of C3 as straight or concave;
3. Lower border of C4 as straight or concave;
4. Morphology of C3 as trapezoid, rectangular horizontal, square, or rectangular vertical; and
5. Morphology of C4 as trapezoid, rectangular horizontal, square, or rectangular vertical.

The observers were then instructed to rate the cephalograms according to the CVMS method of Baccetti et al. (9) (Figure 1). After 2 weeks, the process was repeated by the same observers. The observers did not have prior knowledge of the patients’ developmental stage, chronological age, and dentition phase (the dentition in each cephalogram was covered with black cardboard). In addition, the lateral cephalograms were placed in a sequence different from the first rating to avoid expectation bias. The observers viewed only 15 cephalograms each time to decrease errors due to fatigue. After analysing the data obtained from the first and second observations, the combined data set was also analysed.

As stated before, taking into account the clinical applications of the CVMS method, the maturation stages of cervical vertebrae might be divided into three phases: before the optimum treatment timing, during the optimum treatment timing, and after the optimum treatment timing. With respect to this classification, CVMS I and CVMS II were considered as phase 1, CVMS III as phase 2, and CVMS IV and CVMS V as phase 3.

### Statistical analysis

Data were analysed with MedCalc, Version 8. Fleiss’ kappa statistical test was used to assess the intra-observer agreement and weighted Cohen’s kappa statistical test was used to assess the inter-observer agreement. Fleiss’ kappa test was performed using the online kappa calculator website at http://justusrandolph.net/kappa/.

### Results

Seventy cephalograms were evaluated in this study and the chi-square test showed no statistically significant inclination towards any age group, with a random frequency distribution (\( P = 0.62, \chi^2 = 12.0 \)). Inter-observer agreement was 0.48 (0.45 and 0.51 in the first and second ratings, respectively), and intra-observer agreement values were calculated at 0.59–0.85 for the five observers using the CVMS method.

By considering the clinical approach of the CVMS method as an indicator of proper timing for growth modification treatments, and therefore dividing the CVMS method into three phases as discussed above, inter-observer reproducibility of this method was 0.61 (0.47 and 0.65 in first and second observations, respectively). According to the Landis and Koch classification (10) this result is in a substantial range, with a minor increase in reproducibility from first to second observation (Table 1). Table 2 provides intra-observer agreement for the five observers measured by the Fleiss’ kappa.

Inter-observer agreement values in the evaluation of the shape of the third and fourth vertebrae (C3 and C4) are demonstrated in Table 3. According to Landis and Koch (10) these amounts are considered to be in moderate agreement.

Table 4 provides the results of inter-observer agreement for lower borders of the second, third, and fourth vertebrae, which are all in the substantial agreement range (10).

![Figure 1](image.png)

**Figure 1.** The improved cervical vertebrae maturation staging (CVMS) method (five developmental stages, CVMS I through CVMS V). Reprinted from Baccetti et al. (1).
On the other hand, intra-observer agreement was obtained in the range of 0.64–0.77 (mean = 0.70, standard deviation (SD) = 0.06) for assessing the shape of the body of C3, and in the range of 0.61–0.77 (mean = 0.70, SD = 0.07) for assessing the shape of the body of C4. Intra-observer agreements were at ranges of 0.65–0.85 (mean = 0.71, SD = 0.14) for the lower border of C2, 0.64–0.97 (mean = 0.84, SD = 0.14) for the lower border of C3, and 0.72–0.94 (mean = 0.84, SD = 0.10) for the lower border of C4.

**Discussion**

The major application of the CVMS method is to differentiate the pre-mandibular from post-mandibular growth peak phases. That there is no need for extra X-ray exposure, the simplicity of the method for non-radiologists (like dentists and orthodontists) to learn and that it can be applied with excellent accuracy makes this method a popular choice that is likely to evolve (11).

The CVMS method classifies subjects into five categories, CVMS I to V. These do not necessarily correspond with a clinical decision but, when Class II problems are focussed on, these stages gain specific clinical meanings. In this regard, CVMS I and II are the periods early to intervention; CVMS III is the optimum timing to start a growth modification treatment for a Class II patient, and in CVMS IV and V the most opportune time has been passed. So it is clear that discriminating between CVMS I and II and also distinguishing between CVMS IV and V have no significant impact on clinicians’ decisions in determining the peak of mandibular growth and instituting treatment (i.e. both CVMS I and II implicate the phase before the peak and CVMS IV and V show the post-peak course). In this study, in addition to evaluating the reliability of the CVMS method generally, the reproducibility of such clinical judgments made based on the CVMS method was evaluated.

Reproducibility of the CVMS method was estimated at 48 per cent (moderate agreement), which was increased by taking the clinical approach of the CVMS method into account (61 per cent). According to Landis and Koch classification (10), this is in the substantial range of reliability. Thus, the CVMS method shows almost acceptable reproducibility for determining optimum treatment timing in growth modification treatments.

Franchi et al. (12) reported inter-observer agreement levels of 98.6 per cent. They performed their investigation on cephalogram tracings, which might have affected the results. Özer et al. (13) found 98 per cent agreement between observers, in which the authors also served as observers. The reproducibility of the CVMS method was reported to be 95–98 per cent in a study by Uysal et al. (14). The Spearman Brown formula was used to assess the inter- and intra-observer agreement which is one of the shortcomings of this study. It is known that the correlation coefficient is not a measure of agreement; it merely shows association and any observed correlation is highly influenced (and increased) by the high between-subject variability, hence is misleading. Therefore, agreement tests are recommended for reproducibility investigations (15–17). In the present study, a free-marginal Fleiss’ kappa statistical test was used to assess inter-observer agreement and a weighted Cohen’s kappa statistical test was used to assess intra-observer agreement. Gabriel et al. (7) used Kendall’s coefficient of concordance to assess the inter-observer agreement of the CVMS method. Since the purpose of agreement studies is to investigate the inter-observer agreement of more than two observers, tests that are based on the agreement of only two observers, such as Kendall’s coefficient, can lead to inaccurate results (17, 18). On the other hand, if the observers are unaware of the number and characteristics of different stages in the sample, free-marginal statistical analysis should be used instead of fixed-marginal tests such as Cohen’s kappa or Scott’s pi to obtain accurate results (19).

Intra-observer reproducibility of 100 per cent was reported by Franchi et al. (20) for the CVMS method. In their study, the researchers also served as observers. In addition, the researchers were the inventors of the CVMS method and this might have affected the results. Flores-Mir et al. (21) examined 10 cephalograms with the CVMS method three times, and reported intra-observer agreement levels of 89 per cent by using an intra-class correlation coefficient formula. The small sample size and use of a quantitative statistical method for the ordinal data might have led to high intra-observer agreement in their study.

In a recent study (7), the overall percentage of inter-observer agreement for the CVM method was reported to be 50 per cent, which is similar to the results obtained in the current study (58 per cent). Avoiding bias by using different sequences of cephalograms in two ratings, each observer assessing only 15 cephalograms each time and having no prior knowledge about the biological indicators of the subject’s skeletal age are the technical similarities between these studies, which might have led to the similarity of the results.

As we know, distinguishing the shape of cervical vertebrae plays a crucial role in the determination of skeletal age. Since the reproducibility of the CVMS method depends on a clinician’s ability to determine the morphological features of cervical vertebrae, the reproducibility of assessing the shape and convexity of the lower border of each vertebra might affect the overall reproducibility of this method. It is obvious that primary clinical application of the

<table>
<thead>
<tr>
<th>Table 1. Weighted Cohen’s kappa values for inter-observer agreement. CVMS, cervical vertebrae maturation staging.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original CVMS</strong></td>
</tr>
<tr>
<td>Statistical level of agreement</td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td>First rating (original CVMS)</td>
</tr>
<tr>
<td>Second rating (original CVMS)</td>
</tr>
<tr>
<td>Pooled of first and second ratings</td>
</tr>
<tr>
<td><strong>Table 2. Fleiss’ kappa values for intra-observer agreement of the clinical approach of the cervical vertebrae maturation staging (CVMS) method (with three phases). SD, standard deviation.</strong></td>
</tr>
<tr>
<td>Observer number</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>Average</td>
</tr>
</tbody>
</table>
CVMS method is to clarify the optimum timing for growth modification treatments rather than classifying patients into developmentally age groups. Application of this clinical approach will decrease the negative impact of the low reproducibility of determining the shape of the vertebrae and enhance the overall reproducibility of the method. Indeed, high reproducibility in assessing the convexity of the lower borders of the vertebrae, unlike the relatively lower reproducibility of determining the shape of the vertebrae, could be the main reason for the high clinical reproducibility of the CVMS method. It is concluded from the investigation of Baccetti et al. (1) that the only factor to distinguish CVMS I, CVMS II, and CVMS III is the convexity of the lower borders of cervical vertebrae and possible changes in the shape of the second and third vertebrae, while the distinguishing feature between CVMS IV and CVMS V entirely depends on determining the shape of C3 and C4.

According to the results of the present study, inter-observer reproducibility for evaluating the shapes of C3 and C4 was moderate but it was substantial for assessing the convexity of the lower borders of C2, C3, and C4. It seems that difficulty in determining the correct shape of the third and fourth cervical vertebrae might lead to the low reproducibility of the cervical vertebrae maturation method, which was also reported in a study of Nestman et al. (22).

In a recent study by Nestman et al. (22), the reproducibility of the CVMS method on 30 cephalograms was evaluated by 10 orthodontists. In addition to assessing the cervical vertebrae maturation stage of each cephalogram, morphological characteristics of cervical vertebrae were also determined and analysed. The results indicated that inter-observer agreement for the convexity of lower borders of the vertebrae was high, while the agreement was low for the shapes of the vertebrae. On the other hand, intra-observer agreement was estimated at 44 per cent (average value) by considering the inconclusive cases and 62 per cent (average value) without considering them. They concluded that poor reproducibility in determining the shape of the cervical vertebrae resulted in the overall poor reproducibility. We obtained similar results. Nestman et al. (22) reported that one of the reasons for the low reproducibility of determining the shapes of the vertebrae was the multiplicity of the options (four options: trapezoid, rectangular horizontal, square, and rectangular vertical).

Limitations

1. Since radiographs with complete visualization of the first four cervical vertebrae were considered as an inclusion criteria, some samples were excluded, which also limit the clinical application of CVMS method.
2. These results can be generalized when determining growth modification treatments solely for Class II patients. An additional study undertaken to examine the reliability of CVMS method for determining the optimum treatment timing for other malocclusions (growth modification for Class III problems and maxillary expansion treatments) is in progress.

Conclusion

Inter-observer reproducibility of the CVMS method in determining the skeletal age was low to moderate, and intra-observer reproducibility of this method was substantial.

Reliability of the clinical decisions made based on the CVMS method was higher than when it was used solely as a grading system (to classify patients into five different maturational stages not related to clinical decisions). Although this enhancement is not great, it seems almost acceptable in determining the optimum treatment timing for Class II malocclusion.

Difficulty in determining the morphology of C3 and C4 leads to poor reproducibility of the CVMS method. Furthermore, inter-observer reproducibility of the method enhanced the second observation due to the experience gained in the first rating by the observers.

Funding

This study was supported by a Grant for Scientific Research (No. 334) from Tabriz University of Medical Sciences, Research Deputy.

References


