Cervical vertebral maturation as a biologic indicator of skeletal maturity

A systematic review

Rodrigo César Santiago; Luiz Felipe de Miranda Costa; Robert Willer Farinazzo Vitral; Marcelo Reis Fraga; Ana Maria Bolognese; Lucianne Cople Maia

ABSTRACT

Objective: To identify and review the literature regarding the reliability of cervical vertebrae maturation (CVM) staging to predict the pubertal spurt.

Materials and Methods: The selection criteria included cross-sectional and longitudinal descriptive studies in humans that evaluated qualitatively or quantitatively the accuracy and reproducibility of the CVM method on lateral cephalometric radiographs, as well as the correlation with a standard method established by hand-wrist radiographs.

Results: The searches retrieved 343 unique citations. Twenty-three studies met the inclusion criteria. Six articles had moderate to high scores, while 17 of 23 had low scores. Analysis also showed a moderate to high statistically significant correlation between CVM and hand-wrist maturation methods. There was a moderate to high reproducibility of the CVM method, and only one specific study investigated the accuracy of the CVM index in detecting peak pubertal growth.

Conclusions: This systematic review has shown that the studies on CVM method for radiographic assessment of skeletal maturation stages suffer from serious methodological failures. Better-designed studies with adequate accuracy, reproducibility, and correlation analysis, including studies with appropriate sensitivity-specificity analysis, should be performed. (Angle Orthod. 2012;82:1123–1131.)

KEY WORDS: Skeletal maturation; Cervical vertebrae

INTRODUCTION

The issue of optimal timing for orthodontic treatment is linked intimately to the identification of periods of craniofacial growth that can contribute significantly to the treatment of patients with skeletal discrepancies.1,2

The use of radiographic analysis to estimate skeletal maturation stage is a widely used method for predicting the timing of pubertal growth and for estimating growth velocity and the proportion of growth remaining.3

The hand-wrist radiograph is considered4–7 to be the most standardized method of skeletal maturation assessment based upon time and the sequence of appearance of the carpal bones and certain ossification events. Skeletal maturity is generally determined using stages in the ossification of bones of the hand-wrist, because of the relationship between the overall horizontal and vertical facial growth velocity and skeletal maturity determined by hand-wrist methods3 or because of the quantity of different types of bone available,8–11 or by evaluating the ossification onset of the sesamoid.9 The usual means by which to assess the hand-wrist radiograph are the comparison atlas of Greulich and Pyle8 and Tanner et al.9 and the processes that use specific indicators that relate skeletal maturation to the pubertal growth curve, such as the methods described by Bowden10 and Fishman.11

There are some limitations in the interpretation of skeletal maturity from hand-wrist radiographs.3 The ossification sequence and timing of skeletal maturity

Note: The above text is a review article on cervical vertebral maturation and its reliability as a biologic indicator of skeletal maturity. It discusses the methods, results, and conclusions of a systematic review on this topic. The introduction sets the context for the importance of skeletal maturation in orthodontic treatment timing. The abstract summarizes the objectives, materials and methods, results, and conclusions of the review. The introduction also highlights the limitations of hand-wrist radiographs in assessing skeletal maturity.
within the hand-wrist area show polymorphism and sexual dimorphism, which can limit the clinical predictive use of this method. Moreover, there are concerns about the extra radiation exposure resulting from use of this method, and its use must be questioned if other comparable methods of assessment are available. Finally, events in the hand and wrist are indicators of the peak and the end of the pubertal growth spurt, but these events do not signal the onset of the pubertal growth spurt.

The cervical vertebrae maturation (CVM) method was introduced by Lamparski for use in growth assessment, allowing skeletal age evaluation and eliminating the need for additional radiographic exposure since the vertebrae are already recorded in the lateral cephalogram taken as a pretreatment record. However, the reproducibility of the CVM method has been questioned.

In spite of extensive research about CVM methods applied to skeletal maturity assessment during the past years, statistical and biological details about the relationship between vertebral shape and skeletal maturation are still missing. Systematic reviews are useful tools with which to obtain evidence-based clinical information. The literature is controversial regarding the CVM method in terms of its utility for determining skeletal maturity. The aim of this systematic review was to identify and qualify the evidence and methodology of those reports and to evaluate the following question: How reliable is CVM staging in terms of predicting the pubertal spurt?

**MATERIALS AND METHODS**

To identify potentially relevant studies reporting data related to maturational indexes, specifically CVM methods, detailed search strategies were developed and executed. The search also included appropriate changes in the vocabulary and followed each database’s syntax rules. Citations to potentially relevant studies in journals, dissertations, and conference proceedings were located by searching the appropriate electronic database in an effort to minimize publication bias.

Table 1 shows the database search and outlines our search strategy. This electronic investigation was conducted through July 2010. We also searched by hand by checking the references of the retrieved articles to identify all possible articles to be included in this review. No language restriction was applied.

The selection criteria for considering studies included in this systematic review (SR) were the following: (1) cross-sectional or longitudinal descriptive studies in humans that evaluated qualitatively or quantitatively an established CVM method on lateral cephalometric radiographs to determine skeletal maturation; (2) studies that used hand-wrist radiographs as the standard method when evaluating correlation or comparison with the CVM method; and (3) studies that evaluated the reproducibility of the CVM method. Studies with inadequate sample sizes (ie, cleft lip/palate patients or repeated samples) that introduced a new or additionally modified version of the CVM method, editor’s summaries, reference-only works, patents, meeting abstracts, opinion articles, and reviews were excluded.

Initially, the titles and abstracts identified were reviewed. Duplicate articles appearing in more than one database search were considered only once. Each abstract was checked to determine whether it presented data related to patients’ skeletal maturation stage, as assessed by the CVM method. Any investigation not fulfilling this criterion was excluded from further evaluation. If the reviewer could not decide on a study’s eligibility by examining the title and abstract, its full text was retrieved.

The full texts of remaining articles were retrieved for further evaluation in duplicate by two reviewers. In addition, to document the methodological soundness of each article, a quality assessment modified from the
The cervical vertebrae are already shown on the lateral cephalogram film taken as a pretreatment record, and it is well known that the lateral view of the cervical vertebrae bodies changes with growth. In recent years, evaluation of the cervical vertebrae has been increasingly used to determine skeletal maturity.1,5,14,17,18

None of the studies included in this SR could be used for a meta-analysis because of the different methods of CVM and hand-wrist maturation (HWM) evaluation applied. The study of Imanimoghaddam et al.25 showed different correlation levels between four different CVM methods and the same HWM (TW3 method) in a unique sample. Thus, accuracy, correlation, and reproducibility may be influenced by the method.

There are a great variety of CVM methods, including simple qualitative analysis of the vertebral shape and

| Table 2. Criteria for Assessing Quality Components in the Studies Included |
|------------|----------------|
| Yes | No |
| A. Are the objectives clearly formulated? | 1 | 0 |
| B. Are there key elements of study design early in the paper? | 1 | 0 |
| C. Was the sample size calculated? | 1 | 0 |
| D. Does the study report demographic characteristics of the study population? | 1 | 0 |
| E. Were the sample selection criteria clearly described? | 1 | 0 |
| F. Does the study describe specifications of material and methods involved including how and when measurements were taken? | 1 | 0 |
| G. Was there a reliability assessment, with adequate level of agreement intraexaminer or/and interexaminer? | 1 | 0 |
| H. Were there blinding measurements? | 1 | 0 |
| I. Does the study give details of methods of assessment (measurements) for each variable of interest? | 1 | 0 |
| J. Was there a complete and adequate reporting of results, with self-explanatory tables and figures? | 1 | 0 |
| K. Was there a statistical analysis appropriate for data? | 1 | 0 |
| L. Was the P value stated or confidence intervals provided? | 1 | 0 |

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE),22 Standards for the Reporting of Diagnostic Accuracy Studies (SATRD),23 and Lagravere et al.24 was performed (Table 2). When two reviewers disagreed, a third investigator was called in, and consensus was reached.

One point was given to each criterion, if fulfilled. Quality assessment scores ranged from 0 to 12. The studies were classified as “low” (score 0 to 6), “moderate” (score 7 to 10), or “high” (score 11 to 12) according to this assessment. Studies with a low score (<7 points) were considered to be of poor methodological quality and were not considered at all in terms of the SR conclusions. Disagreements between the reviewers were resolved by reexamination of the article in question, with discussion until both researchers were satisfied with the decision.

RESULTS

The electronic searches identified 343 records after duplicated references were removed. From these, 109 were selected for abstract evaluation. A total of 39 abstracts were retrieved for complete text detailed evaluation, including two articles added from hand searching. Ultimately, our search yielded 23 articles that met the inclusion criteria (Figure 1). Once the articles were selected, we systematically assigned a methodological score to each study in order to characterize them as useful or not (Table 3). Based on quality assessment, we found that six studies3,6,19,25–27 were particularly useful, with a moderate to high score. As stated above, 17 studies4,7,14,28–41 with a low score (<6.5 points) were considered to be of low methodological quality and were not particularly useful. The characteristics of moderate- and high-quality studies are presented in Table 4.

DISCUSSION

One objective of orthodontic treatment in adolescents is to take advantage of potential skeletal growth to treat skeletal discrepancies. Sexual maturation characteristics, chronologic age, dental development, height, weight, and skeletal development are some of the more common methods that have been used to identify stages of growth. Peak growth velocity in standing height is the most valid representation of rate of overall skeletal growth. It forms a useful historic longitudinal measure of an individual’s growth pattern but has little predictive value in terms of future growth rate or percentage of total growth remaining. On the other hand, skeletal maturation staging from radiographic analysis is a widely used approach to predict timing of pubertal growth, to estimate growth velocity, and to estimate the proportion of growth remaining.3

According to a previous study,15 skeletal age offers no value over chronological age, either in assessing or predicting the time of pubertal growth. On the other hand, the hand-wrist radiograph is considered to be the most standardized method of skeletal assessment.4,7,18,27 Although some studies4,12,13 stated that the use of hand-wrist radiographs to predict growth spurt is not sufficiently accurate to be of value in clinical orthodontics, the validity of skeletal maturity assessment using the hand-wrist radiographs has been observed.23 However, to avoid taking additional radiographs, it is relevant to relate maturational stages to skeletal features other than the bones in the hand and wrist.

The cervical vertebrae are already shown on the lateral cephalogram film taken as a pretreatment record, and it is well known that the lateral view of the cervical vertebrae bodies changes with growth. In recent years, evaluation of the cervical vertebrae has been increasingly used to determine skeletal maturity.1,5,14,17,18

None of the studies included in this SR could be used for a meta-analysis because of the different methods of CVM and hand-wrist maturation (HWM) evaluation applied. The study of Imanimoghaddam et al.25 showed different correlation levels between four different CVM methods and the same HWM (TW3 method) in a unique sample. Thus, accuracy, correlation, and reproducibility may be influenced by the method.

There are a great variety of CVM methods, including simple qualitative analysis of the vertebral shape and
size\textsuperscript{6,16} to quantitative measurements of vertebral shape (some of which are limited to height and width distances and ratios) and the depth of the inferior concavity\textsuperscript{4,42} and to other, more specific measurements, rendered through geometric morphometric analysis\textsuperscript{20} or linear regression formulae.\textsuperscript{40,43}

Caution should be taken in the application of the results presented in this SR. Most of the studies investigating the use of cervical vertebral radiographic analysis of skeletal maturation did not report any kind of randomization, blinding, and sample size calculation. Articles that presented methodological deficiencies,
which would likely compromise the interpretation of reported results, were excluded from this study. We found 23 articles that met the inclusion criteria, but only six were considered to be of moderate to high methodological quality.

Initially, the research on the direct relationship between cervical vertebral anatomical alterations and facial growth was limited because of the problems of longitudinal radiographic recordings, although there are a few longitudinal databases that can be accessed that have hand-wrist and lateral cephalometric radiographs taken simultaneously. All of the moderate- to high-ranked studies selected in this review were based on cross-sectional data, and these studies’ designs have inherent limitations in terms of analyzing growth. Cross-sectional sampling is relatively insensitive to the individual variability seen easily in a longitudinal sample. In the study of growth and development, longitudinal research is an essential method for the detailed study of craniofacial growth, which can introduce bias in the staging results, and some studies used few rigorous measures of association for measuring agreement between judges. In a sample of 30 untreated subjects randomly selected, Gabriel et al. concluded that CVM was a poorly reproducible method. We do agree that the study of Gabriel et al. was the first specifically designed to eliminate the methodological errors observed in other studies of CVM reproducibility. However, the intraexaminer reproducibility outcome they found was interpreted as “low,” while a widely accepted scale of reproducibility would score it from “moderate” to “substantial.” Furthermore, during frequency analysis of interobserver agreement and disagreement at the second time point in the study, the percentage of agreement increased 10% (607 to 665), while disagreements decreased (743 to 685). At the second time point,
3 weeks after the first observation, the observers were retrained in the CVM method. Finally, the interexaminer reproducibility (Kendall’s W) was moderate at the initial and 3-week time periods.

Soegiharto et al., Chang et al., Lai et al., and Uysal et al. cited reproducibility results ranging from 85% to 98% using patients’ actual radiographs. To determine values of reproducibility, these studies used the Spearman rank correlation test or Cohen’s Kappa statistic. These are adequate measures of association used for ordinal data, recommended for measuring agreement between judges. However, two of these four studies used the authors themselves as observers in interobserver and intraobserver agreement. According to Gabriel et al., authors who serve as observers have a “research-level” understanding of the CVM method, and, because of this, reproducibility results might be overstated. We concluded that in studies in which the authors themselves serve as observers, both the discussion and conclusion sections should report

Table 4. Characteristics of Moderate- and High-Quality Included Studies

<table>
<thead>
<tr>
<th>Author/ Demography and Study Design</th>
<th>Sample Size, Male/Female</th>
<th>Age Range/ Mean Age (Male/Female), y</th>
<th>CVM* Evaluation Method</th>
<th>Standard Method (HWM* Evaluation Method)</th>
<th>CVM* Reproducibility Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chang et al., 6</td>
<td>244/259</td>
<td>8–18</td>
<td>Lamparski modified by Hassel and Farman (qualitative)</td>
<td>Fishman</td>
<td>Wilcoxon signed rank test</td>
</tr>
<tr>
<td>Gabriel et al., 19</td>
<td>15/15</td>
<td>Not cited</td>
<td>Baccetti, Franchi, and McNamara (2005) (quantitative)</td>
<td>Not evaluated</td>
<td>Kendall’s W Kappa</td>
</tr>
<tr>
<td>Imanimoghaddam et al., 25</td>
<td>20/27</td>
<td>10–15</td>
<td>Lamparski and Hassel and Farman (qualitative); San-Román and Mito et al. (quantitative)</td>
<td>Not evaluated</td>
<td>TW3</td>
</tr>
<tr>
<td>Lai et al., 26</td>
<td>330/379</td>
<td>8.00–18.00</td>
<td>Baccetti, Franchi, and McNamara (2005) (quantitative)</td>
<td>NTUH-SMI</td>
<td>Spearman rank order correlation coefficient</td>
</tr>
<tr>
<td>Soegiharto et al., 2</td>
<td>Indonesian: 648/774 UK: 303/442</td>
<td>8.00–18.00/10.00–17.00/8.00–15.00</td>
<td>Baccetti, Franchi, and McNamara (2002) (quantitative)</td>
<td>Fishman</td>
<td>Cohen’s Kappa</td>
</tr>
<tr>
<td>Uysal et al., 27</td>
<td>(Turkish), 213/290</td>
<td>5.3–24.1/12.00 ± 2.07/12.03 ± 3.03</td>
<td>Lamparski modified by Hassel and Farman (qualitative)</td>
<td>Bjork and Grave an Brown formula</td>
<td>Spearman Brown</td>
</tr>
</tbody>
</table>

a CVM* indicates cervical vertebral maturation; bTW3 indicates Tanner & Whitehouse method; cNTUH-SMI indicates National Taiwan University Hospital Skeletal Maturation Index; dROC indicates Receiver Operating Characteristics; HWM*, hand-wrist maturation.
clearly that the results were obtained by observers with a high level of expertise in the CVM method.

Three studies\textsuperscript{6,26,27} included in this review described significant correlation (variable correlation strength) between hand-wrist skeletal maturation and CVM. Uysal et al.\textsuperscript{27} reported moderate to high correlation between CVM and HWM for boys ($r = 0.78$) and girls ($r = 0.88$). Chang et al.\textsuperscript{6} and Lai et al.\textsuperscript{26} found a high degree of correlation between CVM and HWM for boys ($r = 0.97/0.91$) and girls ($r = 0.97/0.94$), respectively. This difference may be attributed to the variety of CVM methods and different sample sizes. The study of Imanimoghaddam et al.\textsuperscript{25} reported a different correlation between specific CVM methods and the same HWM (TW3 method) in a unique sample. This finding is consistent with those publications\textsuperscript{6,26,27} that identified...
correlations of varying strengths between CVM and HWM using different CVM methods.

Wong et al.\(^1\) stated that the CVM method is not sensitive for detecting maturity except in the growth-spurt period and that studies with wide age range, such as 5 to 18 years, might affect the correlation coefficient obtained because of the inclusion of subjects with skeletal maturity far from the pubertal growth. They used an age range from 10 to 17 years in a previous study. In four articles\(^2,6,25,26\) from our sample of five selected studies the age ranged from 8 to 18 years, but in two of the studies\(^6,26\) the correlation coefficient seemed not to have been affected by age variety.

According to Soegiharto et al.,\(^2\) correlation values are not able to demonstrate that one method is better than the other. They evaluated the effectiveness of the skeletal maturation index and the CVM index through receiver operating characteristics analysis. We judge that their results provided the unique and best-documented evidence that really introduced a reliable analysis of the effectiveness of the CVM method in detecting peak pubertal growth. They also determined significant differences in the ability of the CVM and HWM methods to predict pubertal growth. Although statistically significant differences were observed between the two methods, their findings indicate that both the skeletal maturation index and the CVM index are valid clinical diagnostic indices for the prediction of peak growth of the maxilla and the mandible.

Finally, the prediction of skeletal maturation methods improves as the time of the growth spurt is approached. As stated by Houston et al.,\(^4\) the use of individual ossification events is of limited use during pubertal growth-spurt prediction, and analysis that includes bone stages as well as ossification events is recommended.\(^8\)

CONCLUSION

- Although some studies indicate that the CVM method shows good correlation with the HWM method, with considerable levels of reproducibility, these parameters are not good enough for determining the validity of the CVM method. Furthermore, these conclusions were based on six articles, which points to a low level of evidence, and the biggest question remains unanswered: How reliable is CVM staging for predicting the pubertal spurt?

REFERENCES


