COMPARISON OF CERVICAL VERTEBRAE MATURATION STAGE USING LATERAL CEPHALOMETRIC RADIOGRAPHY VERSUS CONE-BEAM COMPUTED TOMOGRAPHY: A RETROSPECTIVE STUDY

Michael Cliff Wilson

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the School of Dentistry (Orthodontics).

Chapel Hill
2018

Approved by:

Tung T. Nguyen

Tate H. Jackson

Lorne D. Koroluk

ABSTRACT

Michael Cliff Wilson: Comparison of Cervical Vertebrae Maturation Stage using Lateral Cephalometric Radiography versus Cone-Beam Computed Tomography: A Retrospective Study
(Under the direction of Tung T. Nguyen)

Introduction: The Cervical Vertebrae Maturation(CVM) method has been proposed by Baccetti et al to be a useful technique in assessing skeletal maturity that does not require any additional radiation exposure to the patient other than the standard two-dimensional(2D) lateral cephalograph. Several studies have called into question the accuracy of this method. We explored the use of 3D imaging to assess if reliability is improved across examiners compared to the use of traditional 2D images with the implication that increased reliability might improve the timing of treatment with growth modification, and lead to an improved treatment response.

Methods: A sample of 25 previously treated orthodontic patients in the UNC Graduate Orthodontic Clinic were included. De-identified 2D cephalographs and 3D models of the patients' cervical vertebrae were evaluated by research personnel who are trained in using the CVM technique. Using an electronic survey, dental professionals used the CVM method to evaluate de-identified 2D images and 3D models and results were compared to a silver standard CVM stage. Radiographs were also manipulated to simulate sagittal head roll and evaluated by examiners to quantify its effect on CVM technique.

Results: There was a statistically significant difference in accuracy of the CVM technique when using 2D image vs 3D models ($p= <0.0001$). There was no significant difference in CVM technique accuracy with head roll of 10 degrees (p=0.89) or 20 degrees(p=0.80). Intra-rater reproducibility weighted kappa coefficients ranged from 0.50-0.78 in the 2D view, and 0.50-.76 in the 3D view.

Conclusion: The use of the CVM technique with 3D models of the cervical vertebrae to assess skeletal maturity appears to be less accurate when compared to the traditional 2D lateral cephalograph, and thus may not be advantageous to utilize in patient evaluation. Patient head tilt of up to 20 degrees in the coronal plane does not affect CVM staging accuracy.

ACKNOLEDGEMENTS

Thank you to my committee members, Dr. Nguyen, Dr. Koroluk, and Dr. Jackson, for your expertise, guidance, and advice throughout my project. Thank you to my fellow residents, for contributing their time to help in the completion of this research project. Thank you to my wife, Erica, and my family for your continued love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SMI	Skeletal Maturity Indicator
CBCT	Cone-Beam Computed Tomography
CVM	Cervical Vertebrae Maturation
CS	Cervical Stage
IRR	Intra-rater Reliability
3D	3-Dimensional
2D	2-Dimensional
NHP	Natural Head Position
VS	Versus
FHP	Frankfurt Horizontal Plane
T1	Time-point 1
T2	Time-point 2

LIST OF SYMBOLS

©	Copyright Symbol
®	Registered Trademark

REVIEW OF THE LITERATURE

It is well understood that skeletal growth occurs at different rates at different times during a human's life. Furthermore, it is obvious that different parts of the human body grow at different times relative to one another. When a child is born, its head makes up approximately thirty percent of its total size. By the time a human reaches full skeletal maturity, its head is nearer to twelve percent of its total size. Skeletal features closer to the cranial vault tend to grow more earlier in life than those further away from the head. These structures further away tend to grow relatively more later in life. This cephalocaudal gradient of growth is reflected in the growth of the maxilla and mandible, with the mandible reaching its mature size after the maxilla.[1] The timing of the peak growth of these two skeletal units is of particular interest in orthodontics, since the growth of these can have great effect on the treatment plan and outcome.

Since the time of Edward Angle, orthodontists have attempted to alter not only the position of the dental units, but also the skeletal bases in which they reside to achieve ideal occlusion. Dr. Angle created intermaxillary forces utilizing "elastic bands" with and without headgear to achieve proper "interlocking" of the upper teeth with the lower teeth.[2] Since the advent of cephalometric superimposition, it has become evident that these intermaxillary elastic bands, along with most other tooth-borne antero-posterior corrector will produce mostly only dental effects.[3] These dental effects are, in some cases, undesirable because they can result in an unacceptable outcome. In some instances, the treatment of choice, and only option, is to submit the patient for orthognathic surgery in conjunction with orthodontics. However, over the course of the last century, there have been innovative techniques presented that can produce orthopedic changes while limiting the amount of dental movement. Among these are the Herbst appliance, reverse pull headgear, and bone anchored maxillary protraction.[4]

In order for growth modification of the maxilla or mandible to be successful, it has to be timed to coincide with a period in which growth of the maxilla or mandible is actively occurring.[1,5] This leads us to the question: How do we predict when this growth is going to occur? In order to predict the timing of growth, one would need to be able to correlate a measurable and reliable physical change that occurs prior to the growth we are looking to predict.

Predictors of Growth

In the quest for a reliable predictor of the timing of growth, there have been several proposed methods. Hagg *et al* evaluated menarche and voice change as indicators of an impending pubertal growth spurt. They found that menarche occurred 1.1 years after peak growth velocity, which in the end excludes it from being a useful piece of information if it is not attainable until after the event one is trying to predict. Additionally, in their study of the voice change in males and females, it was discovered that the pubescent voice change can occur rather quickly in a few months, but it can also take place over the span of several years.[6] With the knowledge gained about menarche, one can start to make other correlations, such as the various stages leading up to menarche in the female physical growth pattern. This has led to practitioners basing their clinical assessment of when to expect the pubertal growth spurt on observable physical changes that occur prior to menarche.

Other studies have attempted to discern some of these observable changes, and have done so with some success. Nicolson *et al* found that there are visible changes as males and females obtain secondary sex characteristics that can be correlated with onset of the pubertal growth spurt.[7] Scales were set forth by Greulich, as well as Reynolds and Wines to delineate different stages an individual goes through with regards to sex characteristics.[8,9] Marshall and Tanner introduced the scale that has gained widespread use by physicians to evaluate secondary sex characteristics in females and correlate them with peak height velocity and menarche.[10] Even with these scales, the findings from these groups were still found that there wasn't a perfect correlation between their indices and peak growth velocity of the maxilla and mandible.[11–13]

Hand-Wrist Radiographs

As technology continued to progress with regards to healthcare, the evolution in the use of radiography in orthodontics had far-reaching effects on the profession as a whole. Evaluation of pre- and post- treatment dental and skeletal position were obviously the main focus, but it wasn't long before radiographs of other areas of the body became important in the treatment planning process. In an effort to correlate the pubertal growth to some reliable skeletal structure, Bjork and Helm made great strides. They investigated a sample of growing patients, and recorded their height changes over a period of several years. They also exposed radiographs of the subjects' hand and wrist at their annual visits. Bjork and his colleagues focused on the sesamoids of the metacarpophalangeal joint of the thumb, as they consistently ossify near puberty. They were able to correlate reliably the ossification of the sesamoid of the metacarpophalangeal joint in the thumb to one year before menarche.[11]

Although the critical time point in the growth of a female had been identified, it was still difficult translate this to clinical relevance. More information was needed to educate the patient on when they could expect this optimum time for orthodontic treatment to start. This led to further studies of the hand-wrist radiograph as a means for identifying different stages of growth leading up to and extending through the pubertal growth spurt. A more complete method of evaluating skeletal maturity was presented in 1982 by Fishman *et al*, utilizing six different sites in the hand.[14] In the midst of investigation of the proposed method, it was clearly proven that skeletal maturation was a more accurate and reliable indicator of facial and statural growth than chronologic age. There is much variation in amount of growth seen between individuals at any given chronologic age.

The method proposed by Fishman looked at eleven different "Skeletal Maturity Indicators" which are used to categorize the patient (Figure 1). These eleven different SMIs are as follows:

Width of epiphysis as wide as diaphysis

1) Third Finger-proximal phalanx
2) Third Finger- middle phalanx
3) Fifth finger- middle phalanx

Ossification

4) Adductor sesamoid of thumb

Capping of epiphysis

5) Third finger- distal phalanx
6) Third finger- middle phalanx

7) Fifth finger- middle phalanx

Fusion of epiphysis and diaphysis

8) Third finger- distal phalanx

9) Third finger- proximal phalanx

10) Third finger- middle phalanx

11) Radius

It was suggested in the method that one should first look at the ossification of the adductor sesamoid of the thumb. From here, there is a dichotomy that helps to more quickly identify which stage they are in rather than the practitioner having to look at all sites at once to categorize the patient. Through collection of longitudinal hand-wrist and cephalometric radiographs, it was determined that the peak growth in the mandible occurs around SMIs six to seven. With this dichotomy, the method was improved for the clinician to apply it to everyday diagnosis and treatment planning of patients. However, with an increased exposure to radiation than would be otherwise necessary for treatment, the search continued for a solution that did not require any extra radiographs.[14]

Cervical Vertebrae Maturation Method

Currently, standard records taken for a new orthodontic patient consist of a set of intraoral and extraoral photos, a panoramic radiograph, and a cephalometric radiograph. The radiographs both give us images of skeletal structures that could possibly be used to assess skeletal maturity. The dental age can also be assessed on the panoramic radiograph, but Bjork *et al* investigated the strength of correlation between dental age and growth, and found there to be a poor correlation relative to other commonly used indices.[11] In addition, the skeletal structures cannot be predictably viewed and measured due to inherent distortion in a panoramic film. Thus, the panoramic radiograph is not of great help with trying to accurately predict growth status. However, the lateral cephalometric radiograph gives a view of most of the bony structures of the skull and face, and can be measured to obtain linear and angular dimensions accurately and reproducibly.[15,16]

The cervical vertebrae became of particular interest because they were readily viewable in a lateral cephalograph and not part of the craniofacial complex. Lamparski *et al* set forth the original method, which was then improved to allow for evaluation of both males and females regardless of chronological age by Baccetti *et al*.[17–19] It was realized by Lamparski and colleagues that there were consistent, measurable changes in the morphology of the cervical vertebrae that were readily viewed on a lateral cephalogram. These changes were shown to correlate reliably with certain stages of growth leading up to and through the pubertal growth spurt.[2017,19]

A clear view of the veterbrae in the sagittal view is needed to utilize the CVM method originally defined by Lamparski. The CVM method evaluates the general shape of the cervical vertebrae, as well as the lower border of each vertebrae in order to assign the appropriate developmental stage in each cephalogram. The CVM method presented by Baccetti et al evaluates cervical vertebrae 2-4 for these characteristics to assess staging. The stages are from CS 1- CS 6, and are defined as (Figure 3):

CS 1- "The lower borders of all the three vertebrae (C2-C4) are flat. The bodies of both C3 and C4 are trapezoid in shape (the superior border of the vertebral body is tapered from posterior to anterior). The peak in mandibular growth will occur on average 2 years after this stage."

CS 2- "A concavity is present at the lower border of C2 (in four of five cases, with the remaining subjects still showing a cervical stage 1). The bodies of both C3 and C4 are still trapezoid in shape. The peak in mandibular growth will occur on average 1 year after this stage."

CS3- "Concavities at the lower borders of both C2 and C3 are present. The bodies of C3 and C4 may be either trapezoid or rectangular horizontal in shape. The peak in mandibular growth will occur during the year after this stage."

CS 4- "Concavities at the lower borders of C2, C3, and C4 now are present. The bodies of both C3 and C4 are rectangular horizontal in shape. The peak in mandibular growth has occurred within 1 or 2 years before this stage."

CS 5- "The concavities at the lower borders of C2, C3, and C4 still are present. At least one of the bodies of C3 and C4 is squared in shape. If not squared, the body of the other cervical vertebra still is rectangular horizontal. The peak in mandibular growth has ended at least 1 year before this stage."

CS 6- "The concavities at the lower borders of C2, C3, and C4 still are evident. At least one of the bodies of C3 and C4 is rectangular vertical in shape. If not rectangular vertical, the body of the other cervical vertebra is squared. The peak in mandibular growth has ended at least 2 years before this stage."[17]

Since its inception, the CVM method has been used by many, as well as investigated by many. In the last twenty years, mixed results have been found with regards to the accuracy and reliability of this method and its application in clinical practice. Nestman *et al* found that the CVM method has poor reproducibility due to a difficulty in assessing properly the shape of the vertebrae and distinguishing between trapezoidal, rectangular horizontal, square, and rectangular vertical.[21] However, other groups like Perinetti *et al* have found that it is acceptably accurate and reproducible for clinical use, with about 1 in 3 cases being misclassified by one stage.[22] Still, others have recommended that while relying entirely on this method may not be advisable[21,23], it can be an important adjunct to other methods of evaluating a patient's growth status.[1,19,22]

Conclusion

Significant advances in the field of craniofacial radiology have occurred in the last decade, which have allowed for advances in precision in diagnosis as well as techniques used in various areas of dentistry. Our current method of evaluating the cervical vertebrae in order to assess skeletal maturity is based on 2-dimensional lateral cephalograms. It is only fitting to search for a way for orthodontics as a profession to utilize more comprehensive 3-dimensional information about the cervical vertebrae, and in turn increase our knowledge of our patients' growth status.

REFERENCES

1. Franchi L, Baccetti T, McNamara J. Mandibular growth as related to cervical vertebral maturation and body height. *Am J Orthod Dentofac Orthop*. 2000;118:335-340. doi:10.1067/mod.2000.107009.

2. Angle E. The Angle system of regulation and retention of the teeth, and treatment of fractures of the maxillae. 1987.

3. Jones G., Bushang P., Kim K., Oliver D. Class II Non-Extraction Patients Treated with the Forsus Fatigue Resistant Device Versus Intermaxillary Elastics. *Angle Orthod*. 2008;(2):332-338.

4. Proffit WR, Fields HW, Sarver DM. *Contemporary Orthodontics*. 6th ed. Mosby Elsevier; 2007.

5. McNamara J. Neuromuscular and skeletal adaptations to altered function in the orofacial region. *Am J Orthod*. 64.

6. Hagg U, Taranger J. Menarche and voice change as indicators of the pubertal growth spurt. *Acta Odontol Scand*. 1980;(3):179-186. doi:10.3109/00016358009004718.

7. Nicolson A, Hanley C. Indices of Physiological Maturity: Derivation and Interrelationships. *Child Dev*. 1953;24:3-38.

8. Greulich WW et al. A handbook of methods for the study of adolescent children. *Monogr Soc Res Child Develpm*. 1938;3.

9. Reynolds EL, Wines J. Individual differences in physical changes associated with adolescence in girls. 1948;Amer. J. Dis. Child.(75):1-22.

10. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child*. 1969;44(235):291-303. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2020314/. Accessed June 1, 1969.

11. Bjork A, Helm S. Prediction of the Age of Maximum Puberal Growth in Body Height. *Angle Orthod*. 1967;37:134-143.

12. Mellion Z, Behrents R, Johnston L. The pattern of facial skeletal growth and its relationship to various common indexes of maturation. *Am J Orthod Dentofac Orthop*. 2013;143(6):845-854.

13. Nanda RS. The rates of growth of several facial components measured from serial cephalometric roentgenograms. *Am J Orthod.* 1955:658-673.

14. Fishman L. Radiographic Evaluation of Skeletal Maturation, A Clinically Oriented Method Based on Hand-Wrist Films. *Edw H Angle Soc Annu Meet*. September 1981.

15. Baumrind S, Frantz R. The reliability of head film measurements. *Am J Orthod*. 1971;60.

16. Broadbent BH. A New X-ray Technique and its Application to Orthodontia. *Angle Orthod*. 1931;1.

17. Lamparski DG. Skeletal age assessment utilizing cervical vertebrae. 1972.

18. Franchi L., Baccetti T., McNamara J. Mandibular growth as related to cervical vertebral maturation and body height. *Am J Orthod Dentofac Orthop*. 2000;(118):335-340.

19. Baccetti T, Franchi L, McNamara J. The Cervical Vertebral Maturation (CVM) Method for the Assessment of Optimal Treatment Timing in Dentofacial Orthopedics. *Semin Orthod*. 2005. doi:10,1053/j.sodo.2005.04.005.

20. Ball G, Woodside D, Tompson B, Hunter WS, Posluns J. Relationship between cervical vertebral maturation and mandibular growth. *Am J Orthod Dentofac Orthop*. 2011;139(5):455-461. doi:10.1016/j.ajodo.2010.01.035.

21. Nestman T, Marshall S, Qian F, Holton N, Fracisus R, Southard T. Cervical vertebrae maturation method morphological criteria: Poor reproducibility. *Am J Orthod Dentofac Orthop*. 2011;140(2):182-188. doi:10.1016/j.ajodo.2011.04.013.

22. Perinetti G, Caprioglio A, Contardo L. Visual assessment of the cervical vertebral maturation stages A study of diagnostic accuracy and repeatability. *Angle Orthod*. 2014;84:951-956. doi:10.2319/120913-906.1.

23. Engel T, Renkema A, Katsaros C, Pazera P, Pandis N, Fudalej P. The cervical vertebrae maturation (CVM)method cannot predict craniofacial growth in girls with Class II malocclusion. *Eur J Orthod*. 2016;38:1-7.

COMPARISON OF CERVICAL VERTEBRAE MATURATION STAGE USING LATERAL CEPHALOMETRIC RADIOGRAPHY VERSUS CONE-BEAM COMPUTED TOMOGRAPHY: A RETROSPECTIVE STUDY

Introduction

Since its inception, orthodontics has been the art and science of bringing the teeth into a harmonious relationship so that both function and esthetics are optimized. In most cases, this goal is achievable through altering the position of the dentition alone. However, in the instance where there exists a discrepancy between the skeletal bases that house the teeth, it can become necessary to accept a compromised dental relationship, or to alter the position of the skeletal bases to bring the upper and lower teeth into harmony. There are two options to change the position of the maxilla and mandible: 1) Growth modification and 2) Orthognathic surgery. Growth modification involves treatment of a patient during an active phase of growth in one or both of the skeletal units with the intention of creating more or less growth than would have occurred without intervention. Growth of the mandible is not constant throughout life, but increases during the pubertal growth spurt.[1,2] The importance of active growth of the mandible during treatment has been documented clearly by Pancherz and Hagg who found three times the amount of anterior-posterior change in position of the mandible can be gained if treatment occurs during the time of peak mandibular growth when compared with no treatment.[3,4]

The growth potential remaining at the circumaxillary sutures and the mandibular condyle, respectively are crucial to the success of growth modification. Accurate assessment of skeletal growth potential is crucial in the selection of the correct treatment modality for individual patients. Studies have shown a significant correlation between the craniofacial growth potential of a given patient and various characteristics, including attainment of secondary sex traits, change in height, menarche in females, voice change and calcification of wrist bones visible on radiographs, and changes of the cervical vertebrae on serial cephalograms.[2,4,5] While each has its merit, a combined approach has been reported as the most reliable way to predict the skeletal growth status of a given patient.[6,7]

Hand-wrist radiograph assessment has been shown to be an accurate assessment of skeletal growth potential.[5] However, this method requires additional radiation exposure to the patient beyond that required for routine diagnostic records. The cervical vertebrae maturation method(CVM), as proposed by Lamparski et al and then revised by Baccetti et al, has also been shown to be an accurate assessment of growth potential.[8,9] One advantage of the CVM method is that it uses anatomic markers included in a standard lateral cephalogram. While the CVM method has been validated in numerous studies and is widely used to stage growth potential, recent studies have questioned the accuracy and reliability of the method.[7,8,10–12] A key element in correctly classifying the stages of the CVM method requires accurate detection of subtle shape change of the vertebrae such as cupping of the lower border or relative changes in the ratio of width to height. It is possible that these three-dimensional (3D) changes do not translate well in two-dimensional (2D) radiographs or vice versa. Past studies found it was possible to perceive 3D from a combination of 2D views of an object.[13] Therefore, it is logical that simpler 2D perception could be developed from viewing 3D images. 3D models could overcome some of the limitations of 2D imaging such as superimposition of adjacent structures or head positioning error by allowing the observer the ability to view the vertebrae from multiple views and detect subtle anatomic changes. With CBCT becoming more and more popular among orthodontists, the use of this information to help determine the growth status of a patient would be of great benefit in clinical practice. Another possible error introduced in 2D that could potentially be eliminated in 3D is the presence of head roll in the coronal plane during image capture. No data currently exists on the effect of head roll on the accuracy of CVM stage classification.

The primary aim of this study was to compare the accuracy and reliability of the 2D-based CVM method using a conventional 2D image versus a novel 3-dimensional model of the cervical vertebrae. The secondary aim is to assess whether head roll in the coronal plane has any effect on the accuracy and reliability of the CVM method.

Materials and Methods

Sample

This retrospective study was approved by the the Institutional Review Board at the University of North Carolina at Chapel Hill. All CBCT volumes were obtained between 1/1/2007 and 1/1/2017, as a standard of care such as localization of pathology in the UNC School of Dentisty Department of Radiology. They were reviewed using selection and inclusion until a sufficient sample generated.

Inclusion Criteria for the Selection of Radiographs Used

- Clear view of cervical vertebrae 1 through 4
- Non-contributory health history

Exclusion Criteria

- Presence of craniofacial/cervical abnormality

- Skeletal metabolic disorders
- Pathology involving the vertebrae
- Poor Image Quality

After the inclusion and exclusion criteria were applied to 50 potential CBCT volumes, five CBCTs representing each stage of the CVM method (2-6) were selected for a total of 25 CBCTs. The convenience sample of images were selected by a panel of three experts assigning a silver standard CVM stage to the each of the images, and then concurrently collecting enough images for each stage of the CVM method until we had five for each stage of the CVM method (2-6).

Examiners in the study were 17 orthodontics residents within the UNC School of Dentistry that had prior training in the CVM method using conventional 2D lateral cephalograms and consented to taking the survey.

Image Creation

Four different views were created for each of the 25 CBCTs. These views were 1)3D model, 2)Frankfurt Horizontal Plane(FHP), 3)10 Degrees of Roll 4) 20 Degrees of Roll. A brief illustration of presentation of these images to observers is given in Figure 1. A silver standard CVM stage was assigned to each of the 25 CBCT files using the conventional 2D view by a panel of three experts trained in the use of the CVM Method. There was perfect agreement of the panel in CVM classification of each of the subjects.

To create the 3D models of the patients' cervical vertebrae, the de-identified CBCT volumes were segmented and cropped using ITK-Snap (www.itksnap.org) to only include the cervical vertebrae 1 through 4 including lower borders and exclude any facial structures. These models were then oriented using 3D Slicer (www.slicer.org) to achieve uniformity in orientation and resolution across all models.

Conventional 2D lateral cephalograms were extracted from these CBCT volumes (Dolphin). Brightness and contrast were adjusted for uniformity across images. Lateral cephalograms were captured in three different orientations: 1) FHP(right Orbitale- left Orbitale, right and left Porion.) 2) 10 Degrees roll relative to FHP 3) 20 degrees roll relative to FHP.

The Survey: Data Collection

All images were randomized and embedded in an electronic Qualtrics™ survey. The 3D models were able to be re-oriented by examiners for viewing from any perspective(viewstl.com).

Observers were calibrated with a five question calibration survey where they had to correctly classify each of five validated 2D CVM images prior to starting the study survey. Participants were not calibrated or trained in 3D.

Observers were asked to evaluate each image and select the correct CVM stage. Each observer took the survey on the same computer for consistency in viewing each radiographic image. Survey duration for each session was recorded by the Qualtrics software.

Intra-rater reliability data was collected by having eight of the observers take the survey a second time after a three month wash out period. The two time points were defined as time point 1 (T1) and time point 2 (T2).

Statistical Analysis

Accuracy in judging the CVM stage was defined as the difference between the Silver standard and participant's responses across all views and was calculated using a Cochran-

Mantel-Haenzel Mean Score test. Modified ridits were used to account for the ordinal nature of the outcome/ratings. The stratification by subject accounted for association of examiners within subjects. Additionally, the same test was carried out pairwise to compare the 2D view versus 3D model, 2D versus 10 Degree head roll, and 2D versus 20 Degree head roll. Level of significance was set at p=0.05.

To compare intra-rater reliability at T1 and T2, two statistical tests were completed. A Bowker's Test of Symmetry to measure discordance between the two timepoints for each subject. To account for the ordinal outcomes, a weighted kappa coefficient was calculated for 2D and also for 3D intra-rater reliability between T1 and T2 using Cicchetti-Allison weights, per standard SAS procedure. Weighted Kappa coefficient intervals for intra-examiner reliability were defined based on standards for strength of agreement proposed: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61– 0.80, substantial; and .0.80 almost perfect. [14]

Results

No participants reported issues with functionality or usability of the survey software.

The summary of responses at T1 in the 2D view with regard to the silver standard CVM stage are shown in the Table 1. The summary of responses across all examiners in the 3D view with regard to the silver standard are shown in Table 2. This table helps us to appreciate that when using the 3D view, as the CVM stage increased, so did the exact agreement with the silver standard. The mean of differences by silver standard is further illustrated in Table 3. In this figure, the closer the numerical value is to zero, the more accurate the examiners were in using the CVM method. This table illustrates that examiners were more accurate as a whole when evaluating images with silver standard CVM stages 4 and 5. Comparison of accuracy of responses in conventional 2D versus 3D view are illustrated in Table 4. Comparison of accuracy in responses in conventional 2D view, 10 Degree Head roll, 20 Degree head roll, stratified by CVM stage, are illustrated in Table 5.

In statistical comparison of examiner responses versus the silver standard across all views, there was a statistically significant difference between the 4 views of the cervical vertebrae ($p= <0.0001$). Furthermore, the Mean Score test also showed a statistically significant difference between the 2D view and the 3D view($p= <0.0001$). There was no statistically significant difference between the 2D view versus 10 Degree head roll($p= 0.8905$) or 20 Degree Head Roll ($p= 0.7959$).

Percent inter-rater agreement for each of the four views was calculated and are as follows: 69.18 agreement in conventional 2D, 55.29 agreement in 3D, 70.59 in 10 Degrees Head Roll, and 68.00 in 20 Degrees Head Roll.

Intra-Examiner Results

The range for Bowken's Test of Symmetry p values for each subject in the 2D and 3D views were 0.974-1.000 and 0.878-0.999, respectively, which shows there is no significant evidence of discordance in any of the subjects from T1 to T2. A summary of percent accuracy for T1 and T2 is illustrated in Table 7. A summary of weighted Kappa coefficients is illustrated in Table 8. Weighted Kappa coefficients ranges for the 2D and 3D views were 0.501-0.786 and 0.493-0.766, respectively. Intra-examiner reliability between T1 and T2 combined for all participants compared to the silver standard had a weighted Kappa coefficient of 0.708 in the 2D view(substantial agreement), and 0.632(moderate agreement) in the 3D view.

Discussion

The goal of this study was to determine if having more information in the form of a 3D model would improve the accuracy of CVM stage identification. This study found that there was a statistically significant difference in accuracy of CVM identification using 2D versus 3D imaging, with higher accuracy in 2D. This is not unexpected since the CVM technique was devised to be used with a 2D lateral cephalogram, and lacks specific criteria that could be defined and assessed with 3D models of the cervical vertebrae. The current CVM method guidelines described by Baccetti et al appear to be less accurate when used with a novel 3D model of the vertebrae.

It was interesting that the accuracy in this study was significantly lower than reported by previous studies.[7,8,10] On average, the participants' CVM stage assessment agreed with the silver standard only 36% of the time. This relatively low accuracy most likely was an indication that the training and calibration between the experts that determined the silver standard and the participants in this study was inconsistent. Observers in this study were in more exact agreement with the silver standard in the middle CVM stages, and less exact with their responses as we move away towards CVM stages 1 and 6. The coefficient of intra-rater reliability between T1-T2 was found to be .708(substantial agreement), which is slightly lower than the "almost perfect" agreement found by Perinetti et al.[10] These findings show that the participants were still able to reproduce with acceptable consistency the staging of each image of the vertebrae.

One possible explanation for the low CVM technique accuracy resulting from this investigation has to do with the how the images were presented versus how they are normally viewed in every day practice. In past visual cognition studies, it was found that perception of the target image is more accurate when presented along with some contextual information, and less accurate when presented in isolation.[15] The images in this study were presented with only the vertebrae visible, rather than having the rest of the skeletal head and neck structures included. This could have caused a decrease in the ability of the participants to accurately perceive the images and subsequently derive the shapes that are needed to utilize the CVM technique. It is very unlikely the fact that the 2D cephalograms used in this study were synthesized from a midsagittal view of the head and neck structures had any effect on the results of this study. Kumar *et al* found that there was no statistically significant difference in the precision and accuracy of geometry between traditional 2D lateral cephalograms and those synthesized from 3D CBCT images.[16]

Accuracy in this study was measured by evaluating agreement with the silver standard. When compared with previous studies, the accuracy in this study was relatively low.[7,8] However, it is of equal importance to assess whether the observers agreed with each other in their CVM classification. The panel of experts' training in CVM classification was accomplished at a different time and by a different instructor than that of the observers in this study, which introduces potential bias. Since all observers received the same training in use of the CVM method, their staging should be consistent with one another. This study found that the inter-observer percent agreement was 69 in the conventional 2D view, indicating that the CVM staging in this study was of similar reproducibility in 2D as showed in previous studies.[10] In addition, there is a marked difference in inter-observer agreement between 2D and 3D, with CVM staging in 3D being reproducible only half of the time at 55 percent inter-rater agreement.

One potential source of bias in this study was the possibility that the participants would spend their time and energy more on manipulation of orientation of the 3D models rather than assessing them. This concern arose because many of the observers did not have previous experience using the software that allowed for manipulation of the virtual 3D models. To address this potential bias, the electronic survey was constructed so the initial orientation of the 3D models with a lateral view of the vertebrae in an upright orientation. This reduced the amount of image re-orientation so that the participants could focus on classification the models with the CVM technique and less on the manipulation of the model itself.

A secondary aim of this study was to determine if roll of the head laterally during radiographic image capture affected the accuracy of the CVM staging. Roll of the vertebrae may create the illusion of a longer vertebrae in 2D images since the right and left borders of the vertebrae are projected onto a different vertical position in the cephalograms. Our results show orienting the vertebrae with 10 or 20 degreesroll did not affect the accuracy of CVM classification even though the lower border of the vertebrae and relative height to width dimension of the vertebrae were slightly altered with the re-orientation. Interestingly, some observers showed a 20% increase in accuracy of the CVM classification when the head roll was simulated at 20 degrees. While this study evaluated CVM accuracy when head orientation was changed in the coronal plane, Torres et al looked at the effects of CVM classification with changes in head and neck orientation in the sagittal plane. They found the observers were less accurate when the patient head position was tipped forward or backwards was a significant difference when the head was tipped forward or backward. [18] Future studies could evaluate whether rotation of the vertebrae around the sagittal plane(yaw) could affect the accuracy of the CVM method.

The 3D morphological guidelines to classify CVM needs additional refinements. These guidelines could describe the shape not only from the lateral aspect of the vertebrae, but from multiple angles. Furthermore, it could describe angular and subtle curvature changes to the inferior border of the vertebrae. Automated calculations of vertebrae height and width ratios could be calculated from 3D imaging software. This additional information could improve the accuracy and reliability of the method, especially if software could be designed to automatically analyze these anatomic differences and classify the CVM stage. Until then, the current guidelines can be used more accurately with a conventional 2D lateral cephalograms as opposed to a 3D model.

Conclusions:

This study rejected the null hypothesis, finding there is a statistically significant difference in the accuracy of CVM classification with 2D images versus 3D models of the cervical vertebrae. The main conclusions that can be made from this study are as follows:

- CVM staging is significantly more accurate with 2D lateral cephalograms than 3D models.
- A change in head roll to the patient's right or left of up to 20 degrees during lateral cephalogram capture did not effect accuracy of the CVM technique.

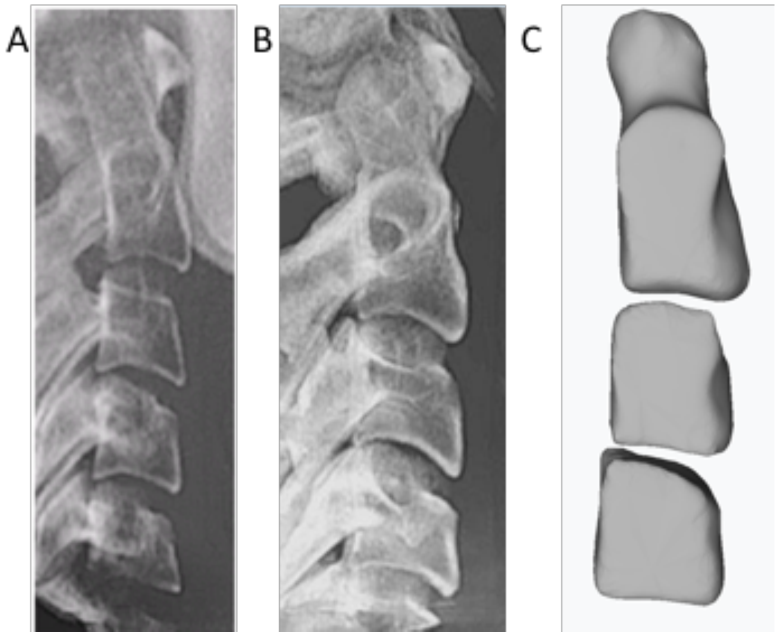

Figure 1. Survey Image Presentation. Three lateral views of the cervical vertebrae as seen by participants. A) Traditional view with patient in NHP; B) Lateral Cephalogram with 20 Degrees of simulated head tilt; C) Lateral view of manipulatable 3D model.

2D Score Frequencies by CVM Stage								
	Assigned Stage	1	2	3	4	5	6	All
All	N	29	21	81	129	118	47	425
	RowPctN	6.82	4.94	19.06	30.35	27.76	11.06	100
Correct Stage								
2	N	26	14	29	16	0	0	85
	RowPctN	30.59%	16.47%	34.12%	18.82%	0.00%	0.00%	100.00%
3	N	2	3	44	34	2	0	85
	RowPctN	2.35%	3.53%	51.76%	40.00%	2.35%	0.00%	100.00%
4	N	1	4	1	46	28	5	85
	RowPctN	1.18%	4.71%	1.18%	54.12%	32.94%	5.88%	100.00%
5	N	0	0	2	24	37	22	85
	RowPctN	0.00%	0.00%	2.35%	28.24%	43.53%	25.88%	100.00%
6	N	0	0	5	9	51	20	85
	RowPctN	0.00%	0.00%	5.88%	10.59%	60.00%	23.53%	100.00%

Table 1. All responses given during evaluation of 2D conventional cephalographs displayed as *n,* and row percent of *n* for each of the CVM stages. Light gray boxes show number of accurate responses and percent accuracy.

3D Score Frequencies by CVM Stage								
	Assigned Stage	1	2	3	4	5	6	All
All	N	17	25	58	100	104	121	425
	RowPctN	4.00	5.88	13.65	23.53	24.47	28.47	100
Correct Stage								
2	N	11	16	21	15	9	13	85
	RowPctN	12.94%	18.82%	24.71%	17.65%	10.59%	15.29%	100.00%
3	N	11	16	21	15	9	13	85
	RowPctN	12.94%	18.82%	24.71%	17.65%	10.59%	15.29%	100.00%
4	N	4	2	5	31	25	18	85
	RowPctN	4.71%	2.35%	5.88%	36.47%	29.41%	21.18%	100.00%
5	N	1	0	0	11	30	43	85
	RowPctN	1.18%	0.00%	0.00%	12.94%	35.29%	50.59%	100.00%
6	N	1	3	0	5	29	47	85
	RowPctN	1.18%	3.53%	0.00%	5.88%	34.12%	55.29%	100.00%

Table 2. All responses given during evaluation of 3D conventional cephalographs displayed as *n,* and row percent of *n* for each of the CVM stages. Light gray boxes show number of accurate responses and percent accuracy.

Mean of Differences by Silver Standard				
Silver Standard	2D	3D	10 Deg Roll	20 Deg Roll
2	0.41	1.40	0.27	-0.06
3	0.36	0.66	0.72	0.73
4	0.31	0.47	0.38	0.41
5	-0.07	0.33	-0.28	-0.21
6	-0.99	-0.66	-0.99	-0.91

Table 3. Mean of differences by silver standard for each of the 4 views of the cervical vertebrae. The closer to zero, the better the agreement between participants' responses and the silver standard.

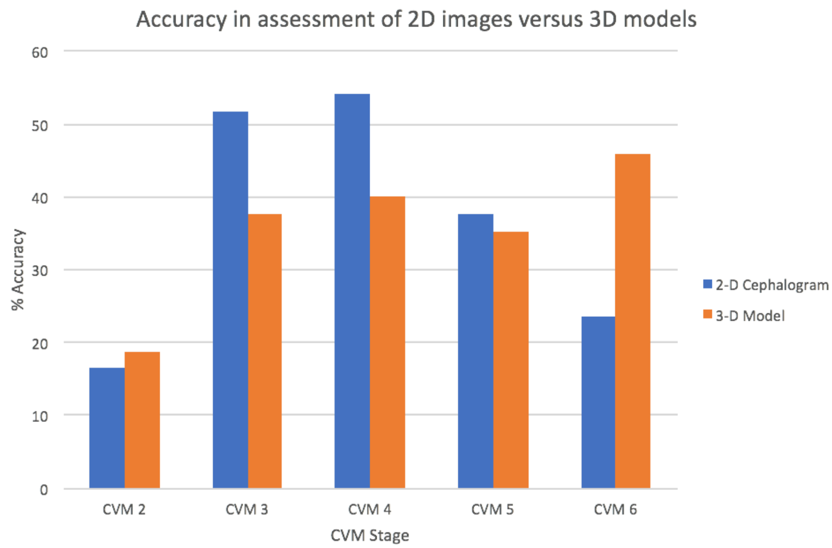

Table 4. Comparison of accuracy of responses in 2D versus 3D for the 17 observers.

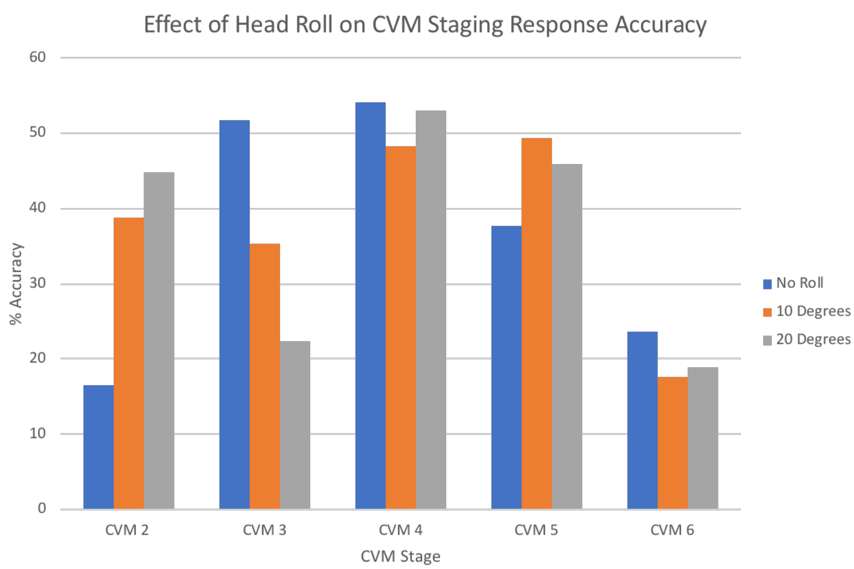

Table 5. Comparison of accuracy of responses in conventional 2D(blue), 10 Degrees head roll(orange) and 20 degrees head roll(gray) for the 17 observers.

Intra-Examiner Reliability in 2D versus 3D		
	2D	3D
Participant	Weighted Kappa Coefficient	Weighted Kappa Coefficient
1	0.79 (.61-.96)	0.67(.49-.85)
2	0.71 (.48-.94)	0.49(.22-.77)
3	0.50(.27-.74)	0.61(.47-.75)
4	0.75(.60-.90)	0.70(.52-.88)
5	0.76(.59-.93)	0.77(.63-.91)
6	0.78(.65-.92)	0.50(.23-.76)
7	0.62(.41-.83)	0.55(.34-.77)
8	0.54(.36-.73)	0.52(.35-.70)
Overall	0.7(.65-.77)	0.63(.57-.70)

Table 6. Weighted Kappa coefficients with ranges in parentheses for intra-examiner reliabilty in conventional 2D versus 3D for each observer as well as overall kappa.

Intra-Examiner Reliability Statistical Analysis						
	2D			3D		
Observer	Test of Symmetry p-value	Weighted Kappa Coefficient	P-value for Weighted Kappa (Two-sided)	Test of Symmetry p-value	Weighted Kappa Coefficient	P-value for Weighted Kappa (Two-sided)
1	1.00	0.79 (.61-.96)	.00	1.00	0.67(.49-.85)	.00
2	.98	0.71 (.48-.94)	.00	.97	0.49(.22-.77)	.00
3	.99	0.50(.27-.74)	.00	1.00	0.61(.47-.75)	.00
4	.97	0.75(.60-.90)	.00	1.00	0.70(.52-.88)	.00
5	.99	0.76(.59-.93)	.00	.98	0.77(.63-.91)	.00
6	.98	0.78(.65-.92)	.00	.97	0.50(.23-.76)	.00
7	.99	0.62(.41-.83)	.00	1.00	0.55(.34-.77)	.00
8	.99	0.54(.36-.73)	.00	.88	0.52(.35-.70)	.00
Overall	-	0.7(.65-.77)	-	-	0.63(.57-.70)	-

Table 7. Statistical analyses of intra-examiner data. Bowken's Test of Symmetry p-values (confidence interval=0.95) used to calculate discordance within observers. Weighted kappa with ranges for each observer and overall kappa were calculated for 2D and 3D. P-value for weighted kappa(two-sided) was calculated for 2D and 3D(95% confidence interval).

REFERENCES

1. Nanda RS. The rates of growth of several facial components measured from serial cephalometric roentgenograms. *Am J Orthod*. 1955:658-673.

2. Bjork A, Helm S. Prediction of the Age of Maximum Puberal Growth in Body Height. *Angle Orthod*. 1967;37:134-143.

3. Pancherz H. The mechanism of Class II correction in Herbst appliance treatment. *Am J Orthod Dentofac Orthop*. 1982;82(2):104-113.

4. Hagg U, Taranger J. Menarche and voice change as indicators of the pubertal growth spurt. *Acta Odontol Scand*. 1980;(3):179-186. doi:10.3109/00016358009004718.

5. Fishman L. Radiographic Evaluation of Skeletal Maturation, A Clinically Oriented Method Based on Hand-Wrist Films. *Edw H Angle Soc Annu Meet*. September 1981.

6. Mellion Z, Behrents R, Johnston L. The pattern of facial skeletal growth and its relationship to various common indexes of maturation. *Am J Orthod Dentofac Orthop*. 2013;143(6):845-854.

7. Baccetti T, Franchi L, McNamara J. The Cervical Vertebral Maturation (CVM) Method for the Assessment of Optimal Treatment Timing in Dentofacial Orthopedics. *Semin Orthod*. 2005. doi:10,1053/j.sodo.2005.04.005.

8. Franchi L, Baccetti T, McNamara J. Mandibular growth as related to cervical vertebral maturation and body height. *Am J Orthod Dentofac Orthop*. 2000;118:335-340. doi:10.1067/mod.2000.107009.

9. Lamparski DG. Skeletal age assessment utilizing cervical vertebrae. 1972.

10. Perinetti G, Caprioglio A, Contardo L. Visual assessment of the cervical vertebral maturation stages A study of diagnostic accuracy and repeatability. *Angle Orthod*. 2014;84:951-956. doi:10.2319/120913-906.1.

11. Nestman T, Marshall S, Qian F, Holton N, Fracisus R, Southard T. Cervical vertebrae maturation method morphological criteria: Poor reproducibility. *Am J Orthod Dentofac Orthop*. 2011;140(2):182-188. doi:10.1016/j.ajodo.2011.04.013.

12. Engel T, Renkema A, Katsaros C, Pazera P, Pandis N, Fudalej P. The cervical vertebrae maturation (CVM)method cannot predict craniofacial growth in girls with Class II malocclusion. *Eur J Orthod*. 2016;38:1-7.

13. Ullman S. Three-dimensional object recognition on the combination of views. *Cognition*. 1998;67(1-2):21-44.

14. Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. *BMJ*. 1997;315:540-543.

15. Williams A, Weisstein N. Line segments are perceived better in a coherent context than alone: An object-line effect in visual perception. *Mem & Cogn*. 1978;6.

16. Kumar V, Ludlow JB, Mol A, Cevidanes L. Comparison of conventional and cone beam CT synthesized cephalograms. *Dentomaxillofacial Radiol*. 2007:263-269.

17. McNamara J. Neuromuscular and skeletal adaptations to altered function in the orofacial region. *Am J Orthod*. 64.

18. Torres FC et al. Evaluation of the Cervical Vertebrae Maturation Index in Lateral Cephalograms Taken in Different Head Positions. *Braz Dent J [online]*. 2013;24:462-466.