RELIABILITY OF THE CVMS METHOD TO ASSESS THE CRANIOFACIAL GROWTH SPURT

Holly Ingersoll
Marquette University

RELIABILITY OF THE CVMS METHOD TO ASSESS THE CRANIOFACIAL GROWTH SPURT

by

Holly L. Ingersoll D.D.S

A Thesis submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin

August 2024

ABSTRACT

RELIABILITY OF THE CVMS METHOD TO ASSESS THE CRANIOFACIAL GROWTH SPURT

Holly L Ingersoll, D.D.S

Marquette University, 2024

Objective:

Cervical vertebral maturation staging (CVMS) is a popular method to determine the growth potential of a child utilizing the changing shapes of C2, C3, and C4. Staging is separated into six stages where the first two stages are before the growth spurt, the second two stages are during the growth spurt, and the last two stages are after the growth spurt. This method is highly practical because it is read on a lateral cephalogram, a radiograph taken by the orthodontist for diagnostic records. However, many clinicians question its reliability and validity as a test. The objective of this study was to test the reliability of CVMS between observers and time periods.

Methods:

This retrospective study had two observers look at 925 lateral cephalograms from the Burlington Growth sample. Each lateral cephalogram chosen had a correlating hand wrist radiograph for a later study. The sample observed was randomly sorted, renamed, and cropped so only the cervical vertebrae could be visualized. Each observer read "The cervical vertebral maturation method: A user's guide" before the staging process (McNamara & Franchi, 2018). The article was used as the only reference for analyzing radiographs. The two observers ranked each cephalogram in one of the six CVM stages. Working periods were limited to 20 minutes with at least a 30-minute break to ensure observer focus. All observations were done in a one-week period. Observations were repeated, at least one month apart. CVM staging trials conducted by each observer were reported on separate excel documents.

Results:

Cohen's kappa was used to test inter-reliability between observers. Intra-reliability was assessed using ICC. Statistical agreement was high for inter-reliability and intra-reliability with a p-value of <0.0001. However, percent agreement for inter-reliability and intra-reliability was only moderate.

Conclusion:

The cervical vertebral maturation method was found to have a high agreement level between trained observers at various timepoints. Even so, percent agreement was only moderate. Orthodontists are advised to use CVMS in conjunction with other growth indicators, such as secondary sexual characteristics, when evaluating patients for orthodontic treatment.

ACKNOWLEDGEMENTS

Holly L. Ingersoll, D.D.S

I would like to thank Dr. Marinho Del Santo, Dr. Shengtong Han, and Mr. Tom Wirtz for their dedicated help in completing this thesis. Their advice, guidance, and expertise played an integral role in the completion of this project. Thank you to the AAOF, who provided the Burlington Growth Study sample used in this project. Their dedication to the further knowledge and development of the orthodontic community is greatly appreciated. I would also like to thank Dr. Ardita Ajvazi who spent hours being the second observer and obtaining the data for this project. Thank you to our program director, Dr. Dawei Liu, the Marquette faculty, orthodontic staff, and the Marquette Graduate School for their dedication to the enrichment of my education. Thank you to my co-residents, who are such a collaborative and supportive group, for a wonderful two years of residency. In the wise words of Elle Woods, “We did it!” Finally, I want to thank my husband, Dr. Jordan Ingersoll, for his unending support throughout this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Analyzing growth and development has always been an important subject in orthodontics. Timing in relation to growth is essential for orthodontics and can lower potential morbidity during treatment. Some procedures, like maxillary expansion, require treatment before the adolescent growth spurt when the suture is smooth and broad (Baccetti et al., 2005). Other procedures, like skeletal class II, are best treated when a patient is at their peak growth spurt. Such malocclusions are common, and treatment is often coordinated closely with the growth of the mandible. Finally, some procedures, like orthognathic surgery, are highly dependent on the discontinuation of growth.

While onset of the pubertal growth spurt often coordinates with the pre-teen and teenage years, timing varies widely with every child. It can be modified by genetic, environmental, and gender factors (Franchi et al., 2021; Soliman et al., 2014). Sexual dimorphism is evident in human development, causing differences in pubertal timing. Sex hormones, such as estrogen, are thought to play a role in adolescent growth. With the differences in growth timing, orthodontists tend to treat females too late and males too soon (Proffit, 2013). Similarly, variance in BMI indices can also play a role in the growth spurt of an individual, as higher BMIs during development lead to maturing soonerx (Soliman et al., 2014). Due to all these factors, orthodontists must adjust treatment timing to coordinate with a patient's growth and development.

Since the pubertal growth spurt is closely tied to sexual maturity, many orthodontists utilize secondary sexual characteristic development to identify the growth spurt and craniofacial growth potential (Mitani & Sato, 1992). While most of the changes are seen in genital development, there are some outward signs an orthodontist can view. In girls, the peak growth spurt will display breast development, while boys will have a redistribution of fat as well as mustache growth. The pubertal growth spurt tends to last 3.5 years in girls and 5 years in boys (Proffit, 2013). These secondary sexual characteristics may be subtle, and it may be difficult to determine what stage of puberty a patient is in.

Orthognathic growth has been shown to be highly correlated with long bone development, following the bodily growth spurt by 6 – 12 months (Hassel & Farman, 1995). Observations of epiphyseal plate ossification in the hand and wrist region were first noted by Hellman in 1928 (Hassel & Farman, 1995). In 1959, Greulich and Pyle composed an atlas of hand-wrist radiographs of children from birth to 19 years old (WW & SI, 1959). Growth plates in the hand undergo mineralization during the maturation process, changing the finger and wrist bones' shapes in a systematic way (Szemraj et al., 2018). Once the adductor sesamoid bone appears, it is predicted that the pubertal growth spurt will occur in a year (Hassel & Farman, 1995). In 1972, Fishman found that using hand-wrist radiographs was helpful in assessing growth potential for dentofacial orthopedics and developed the Skeletal Maturation Assessment (SMA). This assessment utilizes six easily identified indicators to determine the degree ossification present in the wrist radiographs which correlates with growth maturity (Fishman, 1982). While this method has developed into the gold standard in orthodontics to assess growth potential, it does have drawbacks (Rana et al., 2023). It requires the clinician to take an extra lateral cephalogram of the hand-wrist. This induces 0.0001 – 0.1 mSv of radiation to the patient and increases chair time (Manzoor Mughal et al., 2014; Subramanian & Viswanathan, 2023).

In his 1972 thesis, Lamparski first introduced the idea of using the cervical vertebrae to evaluate skeletal maturity (Lamparski, 1972). Hassel and Farman further developed Lamparski's idea and identified six stages of cervical vertebral maturation – initiation, acceleration, transition, deceleration, maturation, and completion (Hassel & Farman, 1995). They laid out the changing shapes of the cervical vertebrae as well as the indenting of the inferior vertebral border during adolescent puberty. Hassel and Farman's method was found to be highly comparable to Fishman's SMA (Baccetti et al., 2007; Gandini et al., 2006). In this methodology, C2, C3, and C4 are used to judge skeletal maturity. This is advantageous because it can be viewed on a lateral cephalogram, a standard diagnostic orthodontic radiograph, even when the patient is wearing a thyroid collar (Szemraj et al., 2018). CVM staging demonstrates a correlation with the maturing skeleton, where the middle stages CS 3 and CS 4 demonstrate the pubertal growth spurt (Franchi et al., 2000). According to Franchi et al., almost 95% of North American patients' CVM growth peaks coincided with their pubertal peak and mandibular growth (Franchi et al., 2021). If the reliability and validity of the CVM method can be confirmed, it would prove the test to be advantageous. Less radiographs may be taken, and the patient is exposed to less radiation.

The purpose of this study was to assess the reliability of the cervical vertebral maturation staging method between intra- and inter-observations. This project will be followed by a study assessing the validity of the cervical vertebral maturation staging method compared to the hand wrist maturation method.

CHAPTER 2
LITERATURE REVIEW

Pubertal Growth Trends

Human development includes two large growth spurts – the juvenile growth spurt and pubertal growth spurt. Research has shown that chronological age and skeletal age do not necessarily coincide, meaning the timing varies per individual (Montasser et al., 2017). Sex, race, genetics, and environmental factors all play a role in the timing and duration of puberty. In particular, the gender differences seen in pubertal growth is fairly evident. When younger, both sexes follow a similar growth pattern until around 8 years old (Kimura, 1977). At that point, the sexes diverge in development. Girls experience a pubertal growth spurt about two years earlier than boys, reaching full maturity around 16 years old. Meanwhile, males experience growth later and will typically reach maturity around 18 years old (Kimura, 1977).

Before adolescent growth, females tend to display a pre-pubertal growth spurt 1 – 2 years before puberty. This is evident through maxillary and mandibular growth seen in females, a trend not typically seen in males (Mellion et al., 2013). The female adolescent growth spurt occurs in 3 stages and lasts for about 3.5 years (Proffit, 2013). The changing stages are denoted by the development of secondary sex characteristics. Stage I is characterized by breast buds and the growth of early pubic hair. Peak growth is predicted to start one year from this point. The first stage typically begins around 9.3 years and will switch to the second stage around 10.9 years (Proffit, 2013). In stage II, breast development, armpit hair, and darker pubic hair develops. These changes coincide with the maximum accelerated growth (Proffit, 2013). Maxillary and mandibular growth occur slightly after the skeletal growth spurt and is generally shorter in duration (Mellion et al., 2013). The final stage is marked by the onset of menstruation which typically occurs 1 – 1.5 years after maximum accelerated growth (Proffit, 2013).

The male pubertal growth spurt lasts longer, about 5 years, and can be more difficult to define (Proffit, 2013). Stage I is marked by the "fat spurt" and an increase in scrotum size which lasts about one year. In stage II, the fat becomes redistributed, so boys will become thinner. Pubic hair appears and penis growth begins as well. This stage lasts about 8 – 12 months. In stage III, armpit hair and upper lip facial hair appears, pubic hair increases, the genitals almost fully developed, and muscular growth increases. This stage lasts anywhere from 15 – 24 months and represents peak height velocity (Proffit, 2013). Peak facial and mandibular growth occur 2.5 years after height growth onset (Mellion et al., 2013). The final stage of male puberty marks the end of height growth. Facial hair develops more, and muscle tone becomes even more defined (Proffit, 2013).

Racial, genetic, and environmental background also play a pivotal role in growth potential and pubertal onset. For example, the beginning of menstruation occurs at 17% body fat. If a girl falls below this level, she will not enter into the third stage of puberty (Baker, 1985). Growth is also closely tied to seasonal changes. It tends to occur faster in the spring and summer months compared to the fall and winter months. The location where children grow up can also play a role in development as well. Kids who live in cities tend to mature more quickly than those who grow up in the country (Proffit, 2013).

The pubertal growth spurt is characterized by growth at the epiphyseal plates of long bones. The ossification process at these sites is endochondral ossification, in which cartilage is converted into bone. This includes the mandible, which undergoes endochondral ossification at the condyle. Sex hormones stimulate cartilage proliferation in the epiphyseal plates, which correlates peak height velocity closely to sexual development. Skeletal maturation is also triggered by sex hormones and occurs more quickly than cartilage transformation. If pubertal onset occurs early, bone maturation overcomes cartilage development. This closes the epiphyseal plates, ending the growth in height. Females are typically shorter than males due to their earlier onset of puberty (Proffit, 2013). Facial structures, such as the maxilla, undergo intramembranous ossification for their calcification process. Since they develop in a different way, the facial bones show some variance from the pubertal growth spurt (Moore, 1997).

The maxilla and mandible are tied closely to the pubertal growth spurt, although not nearly as much as long bones are. The cephalocaudal gradient is evident during adolescence as the mandible grows more than the maxilla. Children begin with a convex profile, but their faces slowly straighten as the mandible's accelerated growth surpasses that of the maxilla (Proffit, 2013).

Maxillary Growth and Its Timing

Maxillary growth is closely tied to neural development. Like many of the facial and cranial bones, it develops through intramembranous ossification in fetal development and grows via the apposition of bone. In early childhood, the maxilla is translated anteriorly as the cranial base grows. Once the cranial base synchondroses fuse around age seven, maxillary growth continues to grow anteriorly and inferiorly through bone apposition (Proffit, 2013). This growth moves all facial features downward and forward in a clockwise rotation (Björk & Skieller, 1972). The effects from this growth can be analyzed in its three dimensions – transverse, sagittal, and vertical planes.

The width of the maxilla is gained through sutural growth at the intermaxillary suture. Greater growth is seen in the posterior portion of the suture compared to the anterior (Björk & Skieller, 1977). It is often thought as the first dimension to complete its growth within the maxillary complex. Around 10 – 13 years old, the suture begins to interdigitate and becomes tightly woven around 11 – 17 years old (Angelieri et al., 2013). In addition to the halting of sutural growth, surface growth begins to conclude as well. Inter-jugal growth has been found to complete around 13 years old (Edwards et al., 2007). Because of this, maxillary expansion is recommended at younger ages, when the mid-palatal suture is open. Surgical intervention to increase maxillary width is necessary after suture closure.

Most of the maxilla's length occurs from the growth at its posterior border. The apposition at the site pushes the maxilla forward, translating it into the anterior space. In addition, the maxillary alveolar process grows through apposition at the maxillary tuberosities. In contrast, remodeling of the anterior surface leads to minor bone resorption (Björk & Skieller, 1977). The amount of deposition and resorption at these surfaces is dependent of the amount of rotation of the craniofacial complex. In Nanda and Ghosh's study, A point overall was seen to move forward about 6.07 mm in females and 9.49 mm in males between the ages 6 – 24 years old (Nanda & Ghosh, 1995). Similarly, Edwards found that A point stopped translating forward around 17 years old (Edwards et al., 2007). When looking at growth rates, Manabe found that the length of the maxilla (ANS – PNS) significantly increased in size between CVM 2 and 3 in males (Manabe et al., 2022). Because of this, it is recommended that Class III malocclusions and transverse deficiencies be treated early to stimulate maxillary growth. The circummaxillary and palatal sutures are most responsive to orthodontic appliances at younger ages because the sutures have not begun interdigitation (Baccetti et al., 2007).

The last dimension to generally complete its growth is the vertical dimension. The posterior nasal spine stops translating around 16 years old, while the anterior nasal spine stops its movement around 17 years old (Edwards et al., 2007). The height of the maxilla is gained through apposition of bone to the alveolar processes and palatal roof. In fact, twice as much apposition occurs on the palatal floor than the orbital floor, making many of the changes occurring in the inferior part of the bone (Björk & Skieller, 1977). Through these processes, the maxilla and its teeth are carried downward and forward (Proffit, 2013).

Mandibular Growth and Its Timing

Unlike the maxilla, mandibular development in fetal growth occurs from endochondral and intramembranous ossification. Most of the mandible's growth, especially in height, occurs at the condyle where endochondral ossification takes place (Björk & Skieller, 1972; Mitani & Sato, 1992). The condylar head is part of the temporomandibular joint and has a cartilaginous layer. Growth of the condylar head occurs superiorly and slightly posteriorly in most individuals (Buschang & Gandini Júnior, 2002). The condylar growth contributes to even greater clockwise rotation of the mandible compared to the maxilla (Björk & Skieller, 1972). Buschang found that the mandible is translated 9.6 – 12.7 mm inferiorly and 1.9 – 2.7 mm anteriorly through growth between 10 – 15 years of age (Buschang & Gandini Júnior, 2002).

About half of this rotation is hidden from the remodeling of the mandibular borders. The largest amount of growth during pubertal growth occurs in the ramus area and is highly variable (Buschang & Gandini Júnior, 2002). Appositional growth occurs at the posterior border of the ramus and is responsive to the amount condylar growth. On the inferior side of the mandibular angle, resorption occurs. The appositional growth of the posterior ramus is greater than the resorption of the inferior border, which leads to a decrease in the gonial angle during growth. Although the posterior ramus adds new bone, the ramus inclination is mostly maintained. This is in part from the anterior border of the ramus experiencing some resorption, which contributes to the development of the mandibular dental arch length (Björk & Skieller, 1972).

In the anterior mandible, appositional growth occurs around the symphysis in the posterior, superior, and inferior areas (Buschang & Gandini Júnior, 2002). Meanwhile, the chin is fairly inactive in growth (Proffit, 2013). Buschang's study displayed the tendency of B point to shift posteriorly and superiorly indicating resorption of the anterior mandible (Buschang & Gandini Júnior, 2002). Even though slight resorption is occurring on the anterior surface, the general effect is all mandibular features are translated anteriorly. In Nanda and Ghosh's study, B point and pogonion grew on average 11.65 mm and 16.21 mm in males and 7.53 mm and 11.17 mm in females between the ages of 6 – 24 years old (Nanda & Ghosh, 1995). These values fluctuated greatly, a testament to the variability of mandibular growth in individuals (Nanda & Ghosh, 1995).

Mandibular growth is more closely tied to pubertal growth than maxillary growth. In Buschang's 1999 article, condylar growth began to accelerate at 12.2 years old in females and 14.3 years old in males (Buschang et al., 1999). It often occurs after maximum growth of the maxilla and exhibits more growth. In Silveria's study, the mandibular growth spurt occurred at 12.1 years in females with a spurt probability of 59%, while the male spurt was around 13.2 years with a spurt probability of 61% (Silveira et al., 1992). This was also confirmed by Nanda and Ghosh's study where the greatest change in sagittal growth occurred between 6 – 12 years in females and 12 – 18 years in males (Nanda & Ghosh, 1995). At its peak, the mandible grew an average of 5.4 mm in Baccetti's study (Baccetti et al., 2005; Franchi et al., 2021).

In concordance with the CVM method, the greatest length and height of mandibular changes occurs between CVM 3 and 4 (Baccetti et al., 2005; Manabe et al., 2022). In Franchi's study, 78% of cases showed the greatest mandibular length increase after CVM 3. In fact, growth spurt probability of over 50% is expected in females ages 11.3 – 13.1 years and males ages 12.2 – 14.2 years when their lateral cephalograms demonstrate CVM 3 (Franchi et al., 2021).

It is recommended that retrognathic individuals begin treatment early during CVM 3 (Manabe et al., 2022). Class II correction is highly dependent on condylar cartilage responsiveness. During stages CVM 1 and 2, maximum growth potential has yet to be obtained. Re-evaluation should be conducted one year after for functional appliance treatment. Once a concavity is seen on C2 and C3, representing CVM 3, functional appliance therapy can begin. In patients that receive Class II functional therapy, mandibular growth has shown to range from 2.4 - 4.7 mm compared to the untreated growth of 0.4 - 1.8 mm (Baccetti et al., 2005). In opposition to the class II recommendation, class III malocclusion treatment should be started early with maxillary protraction. However, if a patient has entered the CVM 3 and 4 stage before skeletal interception, treatment must turn its focus on to mandibular growth restriction (Baccetti et al., 2005; Manabe et al., 2022).

Cervical Vertebral Anatomy and Development

The human spine is composed of five regions containing 33 vertebrae. The regions consist of seven cervical, twelve thoracic, five lumbar, five sacrum, and five coccyx vertebrae. Within the seven cervical vertebrae, C1 – C2 are unique in shape and function, while C3 – C7 display similar anatomical landmarks. C1 is called the atlas and articulates with the occipital condyles. It allows the head to rock anteriorly and posteriorly. C2 is called the axis, and it allows for horizontal rotation of the head. C3 – C7 are similar in shape and display unique characteristics during maturation (Hassel & Farman, 1995). Fibrocartilaginous disks sit between the cervical vertebrae, which act as shock absorbers for the spine. They are attached to ligaments that affix to the head and produce various muscular functions. These bones develop and mature closely with the nervous tissue (Mitani & Sato, 1992).

In fetal development, ossification of the vertebral spine begins in the cervical region. The first ossification site appears in the lower cervical/upper thoracic region and ascends quickly to the upper cervical area. Once the cervical vertebrae have ossified, the lower thoracic area begins to ossify and ascends to the cervical vertebrae. Ossification will later descend into the lumbar and sacral regions. The ossification pattern has been found to be closely correlated with the muscular activity in fetal movement (Bagnall et al., 1977).

In the pubertal growth spurt, the vertebral growth in height occurs through endochondral ossification. The superior and inferior borders of the vertebrae have a cartilaginous layer where growth takes place. The vertebrae elongate vertically in puberty due to the location of these growth plates, creating a repeatable and predictable pattern of growth. After the completion of endochondral ossification, the anterior and lateral surfaces of the vertebrae grow through periosteal apposition (Hassel & Farman, 1995; Rana et al., 2023). Like other effects seen during puberty, cervical vertebrae are generally larger in boys compared to girls at every maturation level (Chatzigianni & Halazonetis, 2009).

There are three normal variations of cervical vertebrae – displacement of vertebrae that appear as subluxation, curvatures that resemble ligamentous injury, and skeletal growth centers that appear as fractures. Orthodontists can help monitor and identify asymptomatic pathology in patients through their view with the lateral cephalogram. Posterior arch defects, congenital defects, and odontoid anomalies can be common in patients (Cattell & Filtzer, 1965).

Staging of Cervical Vertebral Maturation

The different stages within the cervical vertebral maturation method are based on the changing shapes of vertebrae C2, C3, and C4. This is based on observer visualization. In McNamara's guide, the first step to staging is to assess the inferior borders of the vertebral bodies. The borders start flat, but notching appears as the skeleton matures. The second step is to analyze the shape of C3 and C4. The vertebrae start trapezoidal where the posterior border is slightly taller than the anterior border. Once a patient reaches puberty, the vertebrae will morph into horizontal rectangles and eventually vertical rectangles (McNamara & Franchi, 2018).

CVM 1 and 2 represent the pre-pubertal stages of development. In CVM 1, the inferior borders of C2 – C4 are flat and C3 – C4 are trapezoidal in shape. The posterior border of the vertebrae is slightly taller than the anterior border. This stage is seen in children until about two years before their mandibular growth spurt. In CVM 2, a slight notch appears on the inferior border of the C2 while the other vertebrae maintain their smooth border. The shape continues to be trapezoidal. Once CVM 2 is reached, 90% of patients will reach their mandibular growth peak in a year (Baccetti et al., 2005; Gandini et al., 2006; McNamara & Franchi, 2018; Morris et al., 2019).

CVM stages 3 and 4 represent maximum growth velocity during pubertal development. In CVM 3, C2 – C3 have concavities on the inferior border but C4 remains flat. In most instances, C3 – C4 are trapezoidal in shape, but can change to be more rectangular horizontally. This stage represents the maximum growth velocity. In CVM 4, C2 – C4 all have concavities on the inferior border and C3 – C4 are horizontally rectangular in shape. This stage represents the later part of peak growth velocity. The greatest change in bodily height has already occurred between CVM 3 – 4. CVM 4 represents the nearing of the end of the growth spurt (Baccetti et al., 2005; Gandini et al., 2006; McNamara & Franchi, 2018; Morris et al., 2019).

CVM stages 5 and 6 represent the post pubertal phases of human development. In CVM 5, C3 – C4 will become squarer in shape. This indicates that maximum growth potential has already been achieved and the majority of expected craniofacial growth has occurred. In CVM 6, C3 or C4 become vertically rectangular or square in shape, and the cortical bone becomes more distinct. It is estimated that up to 17% of females will never reach this phase (Baccetti et al., 2005; Gandini et al., 2006; McNamara & Franchi, 2018; Morris et al., 2019).

Gender and racial background influence the chronological timing of CVM changes. In Montasser's study, Hispanic patients reached CVM 3 before Caucasian and African American patients, indicating they begin maturity sooner. Even so, Caucasian children tended to transition between the stages and reach full maturity the fastest (Montasser et al., 2017).

Sex differences in CVM timing are seen during skeletal maturation as well. Kimura found that both sexes follow a similar growth pattern until 8 years old. At that point, the sexes diverge, and girls experience an earlier pubertal growth spurt, usually starting around one year earlier than boys (Kimura, 1977). For CVM 3, the average age of females was 8.6 – 11.5 years, while males were on average 10 – 14 years (Franchi et al., 2000). Females will ultimately finalize pubertal growth around 16 years old (Kimura (Kimura, 1977). Males are generally chronologically older when they begin to show vertebral changes. However, CVM 4, 5, and 6 occur at similar ages between males and females. Because of this, males display a shorter time between CVM 1 – 6 (Montasser et al., 2017). Growth is finalized around 18 years old (Kimura, 1977).

Other factors affect skeletal maturation such as environmental conditions, climate, socioeconomic status, and nutrition. Significant differences in CVM staging have been seen between continents in CVM 2 – 6. This may be from nutrition, which can play a role in maturation (Magalhães et al., 2022).

Although these stages have been distinctly defined, CVM staging is not always as clear. Growth is a perpetual process, and the changes denoted occur on a continuum. The vertebrae viewed on lateral cephalograms can often appear as a blend between two different stages. Head positioning, anatomy, and exposure can affect image quality and play a role in the analysis of the cervical vertebrae. McNamara advised that CVM staging method should be used as an adjunct to other assessments of pubertal growth and maturation (McNamara & Franchi, 2018).

Reliability of Cervical Vertebral Staging Method

The reliability of the cervical vertebral method has come into question due to the subjectivity of the method. Observers are asked to assess the changing shapes of the vertebrae and whether notching on the inferior border is apparent. The accuracy and reproducibility of the CVM staging is affected by these subjective measurements and variation in skeletal shape. Two criteria must be found true for the test to be considered reliable. There must be a high degree of agreement by an observer each time they conduct the test, and multiple observers must agree with one another during a trial period. Statistical agreement tests should be used to test for predictability.

Several studies have assessed the reliability of the test in the past. Sohrabi et al. tested intra-observer agreement and used the kappa statistical test to assess statistical reliability. Agreement of the individual observers was found to range between 0.59 – 0.85. Inter-observer agreement was tested with Cohen's kappa statistical test. Moderate agreement was seen in the C3 and C4 shape. Substantial agreement was seen in the lower borders of C2, C3 an C4 (Sohrabi et al., 2016).

Zhao's team tested the validity and reliability of the staging process as well. Cervical staging agreement was assessed with Kendall's W for interobserver agreement and weighted kappa for intra-observer agreement. Kendall's W value was greater than 0.8, signifying there was a strong statistical agreement between observers. However, complete agreement was low – 39.3% during the first trial and 44.9% during the second trial. Weighted kappa denoted that there was favorable intra-observer agreement (0.53 – 0.86). Intra-observer agreement varied per observer ranging from 40.7% - 79.1%. Reliability of the CVM method was found to be statistically acceptable (Zhao et al., 2012).

Several studies have displayed how observer expertise played a large role in high reproducibility. Franchi's team found interobserver reproducibility to be 98.6% and intra-observer reproducibility to be 100% (Franchi et al., 2000). Similarly, Hassel and Farman displayed inter-observer agreement to be 90% (18 out of 20 cases) and intra-observer agreement to be 95% (19 out of 20) with significant agreement ($p < 0.001$) (Hassel & Farman, 1995).

Gabriel's study found that that CVM was too variable for the clinical setting. Interobserver agreement was tested with ten observers using Kendall's W and was found to be moderate at both time points – 0.74 and 0.72 respectively. When observers disagreed, it was mostly by one stage apart. However, a significant amount had a discrepancy of two or more stages apart. Only 10% of subjects had all ten observers agree on staging (3 of 30). Intra-observer agreement was tested using kappa. This was found to below 0.8, making intra-observer observations to be moderate in agreement as well. Observers only agreed with themselves 62% of the time (Gabriel et al., 2009).

Certain stages may be more reliable than others. In Zhao's study, CVM 3 had the lowest amount of agreement between the six stages. This is imperative because CVM 3 is one of the most clinically important stages due to its coordination with the pubertal growth spurt. Likewise, a higher percent of agreement was found at CVM 6 compared to the other stages (Zhao et al., 2012). This alludes to the fact that puberty is a transitive process, and "stages" may be difficult to distinguish in a period of rapid change. From this finding, Zhao et al. concluded this method was a clinically unacceptable way to determine craniofacial growth potential.

Validity of Cervical Vertebral Staging Method

Whereas reliability assesses whether a test is repeatable, validity assesses if a test is accurate. This can be done in a variety of ways including comparing its results to mandibular growth and the hand wrist maturation method (HWM). Like reliability, research has shown mixed results of whether the CVM method is a valid test.

Franchi's study assessed CVM staging regarding acceleration of body height. After verifying the reliability of the staging, the team found that there was an acceleration of body height growth from CVM 1 – 4. CVM 4 – 5 and CVM 5 – 6 showed deceleration in growth. The greatest acceleration of growth occurred between CVM 3 –

4. Interestingly, two subjects, both females, did not display the greatest acceleration of growth during this phase. Instead, they showed this growth spurt between CVM 4 – 5. Ramus height and S-Gn also showed significant deceleration in growth during the CVM 4 – 5 and CVM 5 – 6 phases, similar to growth (Franchi et al., 2000).

In contrast, Mellion's team found the hand-wrist maturation method to be the best modality to verify pubertal state, while the cervical vertebral method consistently was found to be the worst from ages 6 – 11 years old. Chronological age was almost as good as the hand wrist method in judging pubertal stage. HWM and CVM had moderate and weak correlation with timing of jaw growth, displaying that these modalities may not be significantly helpful in the treatment planning process (Mellion et al., 2013).

However, several studies found a high correlation between the cervical vertebral method and hand wrist. Mitani and Silviani found the size of the mandible, its height, hand-wrist, and cervical vertebrae development to be correlated very closely with one another early in development but weakened after 11 years old (Mitani & Sato, 1992; Silveira et al., 1992). In Wong's study, the CVM method had a high correlation with HWM when looking at children near their pubertal growth spurt (Wong et al., 2009). Ferrillo et al. also found that there was a correlation between CVM and HWM methods in their systematic review and may be a valid way to assess skeletal maturity when assessed with moderate to high quality studies (Ferrillo et al., 2021). Szemraj found high correlation between CVM and HWM within their systematic review (Szemraj et al., 2018). Gandini et al. assessed the CVM method against HWM Björk method. The concordance value was 0.783 ± 0.098, showing there was fairly good correlation (Gandini et al., 2006). Tekın and Aydin compared hand-wrist radiographs, lateral cephalometric radiographs, and CBCT images of cervical vertebrae. Using Spearman's coefficient, they found HWM, CVM, and CVM from CBCT to be 0.785, 0.875, and 0.791 respectively (Tekın & Cesur Aydın, 2020).

In Baccetti's longitudinal study, CVM staging was compared against mandibular growth changes during annual assessments. During T1, the inferior border of C2 was flat in 93% of the individuals, while C3 and C4 were flat 100% of the time and had a trapezoidal shape. During T2, C2 had a concave inferior border in 80% of the test subjects, while C3 and C4's inferior borders were flat 93% of the time. C3 and C4 displayed a rectangular horizontal shape 3% and 13% of the time, respectively. According to the cephalometric, no difference was seen between T1 and T2 measurements besides the depth of concavity on C2. During T3, the inferior border of C2 was concave 100% of the time, C3 was concave 93% of the time, and C4 was flat 90% of the time. Their shapes varied between trapezoidal or rectangular horizontal. This depth in C2 and C3 was seen to increase during T2 to T3, and the anterior wall of C3 and C4 increased at T3 as well. During T4, concavities were seen on the inferior border of C2, C3 (93% of the time), and C4 (87% of the time). C3 and C4 were rectangular horizontal in shape. At T4, C4's inferior concavity increased significantly and C3 and C4's anterior border increased again. During T5, C2, C3, and C4 had concave inferior borders nearly 100% of the time. C3 was square in 60% of subjects and rectangular horizontal in 40% of subjects. C4 appeared square in 53% of subjects and rectangular horizontal in the rest of subjects. During T6, all inferior surfaces were concave. C3 and C4 were square 50% of the time and vertical rectangle the other 50% of the time. In T5 and T6, the vertebral bodies continued to grow in height, thus making their horizontal rectangular shape into square or vertical rectangular. (Baccetti et al., 2005).

CHAPTER 3
MATERIALS AND METHODS

Study Design

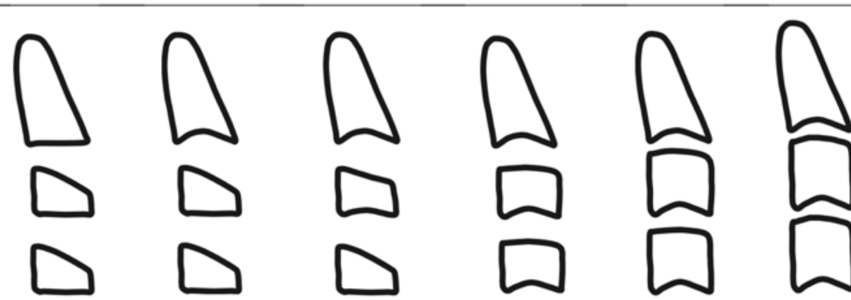
Lateral cephalogram samples were obtained from the AAOF's collection of the Burlington Growth sample, which consists of 1258 subjects. Longitudinal data from 98 of these subjects were chosen who had corresponding lateral cephalograms and hand wrist radiographs from ages 3 to 20 years old. This sample was the largest known by the research team to have a significant amount of correlating hand-wrist radiographs and lateral cephalograms. Comparison of lateral cephalograms to hand wrist radiographs will be used in a future project. 905 lateral cephalograms were extracted from the sample to be analyzed.

All lateral cephalograms were cropped so only the cervical vertebrae could be viewed. This was to eliminate any bias that may arise from viewing dentition or other skeletal features (Figure 1). The sample was then randomized and renamed by Mr. Tom Wirtz, an informatics specialist at Marquette University. Two observers, both current Marquette residents, were chosen to review the selected radiographs. Each observer read "The cervical vertebral maturation method: A user's guide" before staging the lateral cephalograms (McNamara & Franchi, 2018). The staging chart from this article was the only reference used during the observer staging process (Figure 2). Observers could reference the article during data collection as needed.

Figure 1. A sample of a cropped lateral cephalogram from the Burlington Growth Sample. This type of image was assessed by two observers to determine CVM staging.

Table 1. The Six Stages of Cervical Vertebral Maturation

Schematic representation	CS 1	CS 2	CS 3	CS 4	CS 5	CS 6



Inferior borders of C2, C3, and C4[a]	F, F, F	C, F, F	C, C, F	C, C, C	C, C, C	C, C, C
C3 morphology[a]	T	T	T	RH	S/RH	RV/RH
C4 morphology[a]	T	T	T/RH	RH	S/RH	RV/RH
Clinical implication	Prepubertal stage	Prepubertal ("get-ready") stage	Circumpubertal stage	Circumpubertal stage	Postpubertal stage	Postpubertal stage

[a] F= Flat; C= Concavity; T= Trapezoid; RH=Rectangular Horizontal; S=Square; RV=Rectangular Vertical

Figure 2. McNamara's staging chart from "The cervical vertebral maturation method: A user's guide" was used as reference before and during observation periods.

Experimental Procedure

Each observer was given a USB with a folder of the randomized, renamed, and cropped samples. Two separate excel documents were provided to record observations for two time points. All lateral cephalograms were ranked by each observer separately into one of the six CVMS stages using "The cervical vertebral maturation method: A user's guide" as reference. Observations were completed within one week for each viewing period. Observing periods were limited to 20 minutes with at least a 30-minute break between working periods to ensure observer focus. Observation periods were conducted twice, at least one month apart. Data obtained was labeled as T1 for the initial viewing and T2 for the final viewing. The observers were denoted at Observer A and Observer B.

Inter-observer Percent Agreement

Lateral cephalograms where the cervical vertebrae could not be assessed were eliminated from the data set. Inter-observer ratings were compared in excel. Four comparisons were made – Observer A vs B at T1, Observer A vs B at T2, Observer A at T1 vs Observer B at T2, and Observer A at T2 vs Observer B at T1. Comparisons were given a "1" if observers recorded the same value and a "0" if observers recorded different values. The four comparisons were then added up, giving a value of 0 – 4. This was labeled "Level of Agreement" (Table 1). The levels of agreement were added together for the entire data set. This value was divided by the value of all observances, representing overall percent agreement.

	RECORDINGS				COMPARISONS				
Ceph	Ob A T1	Ob A T2	Ob B T1	Ob B T2	Inter T1	Inter T2	OA T1 vs OB T2	OA T2 vs OB T2	Level of Agreement
C179.tif	5	5	5	5	1	1	1	1	4
C515.tif	5	6	5	5	1	0	1	0	2
C17284.tif	2	3	2	1	1	0	0	0	1
C11751.tif	2	2	1	3	0	0	0	0	0

Table 1: Example of how inter-rater observations were recorded and initially compared in excel. "1" was given if the observers agreed, and "0" was given if they disagreed. These comparisons were added together to equal the level of agreement, which was a number that ranged 0 – 4. 4 represents complete agreement, while 0 represents no agreement.

Intra-Observer Percent Agreement

Intra-observer ratings were compared in excel. Once again, comparisons were given a "1" if observations were the same value and a "0" if observations were different. This "level of agreement" was then added and compared against the total number of radiographs analyzed. An example of how observations were recorded in excel can be seen in Table 2.

	RECORDINGS				Comparisons	
Ceph	Ob A T1	Ob A T2	Ob B T1	Ob B T2	Ob A	Ob B
C179.tif	5	5	5	5	1	1
C515.tif	5	6	5	5	0	1
C17284.tif	2	3	2	1	0	0
C11751.tif	2	2	1	3	1	0

Table 2. Example of how intra-observer ratings were recorded and compared in excel. "1" was given if the observer agreed with themselves, while "0" was given if the observer disagreed with themselves.

Statistical Analysis

Data was collected and four statistical comparisons were able to be made. Two comparisons were the intra-observer recordings for Observer A and Observer B. The other two comparisons were inter-observer recordings at T1 and T2. Statistical analysis was conducted by Marquette University School of Dentistry's biostatistician Dr. Shengtong Han. Statistical analysis was conducted with statistical software R. Cohen's kappa was used to analyze inter-reliability. This statistical test accounts for the level of chance that two observers accidentally get the same result. Table 3 displays the interpretation guide for Cohen's kappa. Intra-reliability was assessed with intraclass correlation coefficient (ICC). ICC was analyzed using a single-measurement, absolute-agreement two-way mixed-effects model (Koo & Li, 2016). Table 4 displays the interpretation guide for ICC. Cohen's kappa and ICC display a higher level of agreement as the value nears 1.0 and have a poor level of agreement as the value nears 0.0.

Cohen's Kappa	Level of Agreement
<0.21	slight
0.21-0.40	fair
0.41-0.60	moderate
0.61-0.80	substantial
0.81-0.99	almost perfect
1	perfect

Table 3: Interpretation guide for Cohen's kappa value.

ICC	Level of Reliability
<0.50	poor reliability
0.5-0.75	moderate reliability
0.79-0.9	good reliability
>0.9	excellent reliability

Table 4: Interpretation guide to intraclass correlation coefficient (ICC).

CHAPTER 4
RESULTS

Inter-Observer Results

A total of 869 lateral cephalograms were analyzed by the two observers. The first set of observations occurred in September 2023. The second set of observations occurred in October 2023, at least one month after the first set of observations. Inter-observer percent agreement was 63.2% (Table 5).

Cohen's kappa was used to assess inter-observer reliability. Its value was 0.8966 and 0.8567 for the first and second trial respectively. Both had a p-value of <0.0001 (Table 6). This displays an almost perfect level of agreement. As Cohen's kappa accounts for chance, these results display that the level of agreement between observers was significant.

Sample	# of similar observances	# of total x-ray observances
Sums	2208	3496
Percent Agreement		0.631579

Table 5. Percent agreement for inter-reliability testing. The test had two observers, who both ran two trials. Therefore, overall total observances were 869 lateral cephalograms multiplied by 4 observances. The number of similar observances was assessed by finding level agreement for each radiograph.

	1st trial	**2nd trial**
Cohen's kappa	0.8966	0.8567
p value	<0.0001	<0.0001

Table 6. Cohen’s kappa statistically calculated for the first and second trial to assess inter-observer agreement. This displays almost perfect agreement.

Intra-Observer Results

The percent agreement for Observer A was 70.8%, while Observer B’s percent agreement was 64.9% (Table 7). ICC was used to test the statistical significance of intra-observer reliability. Observer A had a value of 0.8649 and Observer B had a value of 0.8422. Both displayed observances of a p-value of <0.0001 (Figure 3). Statistical interpretation of these results is that intra-observer ratings are highly reproducible. Results were replicated with multiple statistical tests – Cohen’s kappa, Kendall’s W, and CCC. For conciseness, only ICC was recorded in this study.

	Observer A	Observer B
Identical Answers	615	564
Total X-Rays	869	869
Percent Agreement	**0.7077**	**0.6490**

Table 7. Percent agreement for intra-observer readings. Each observer conducted two trials. These trials were compared, and identical observances recorded.

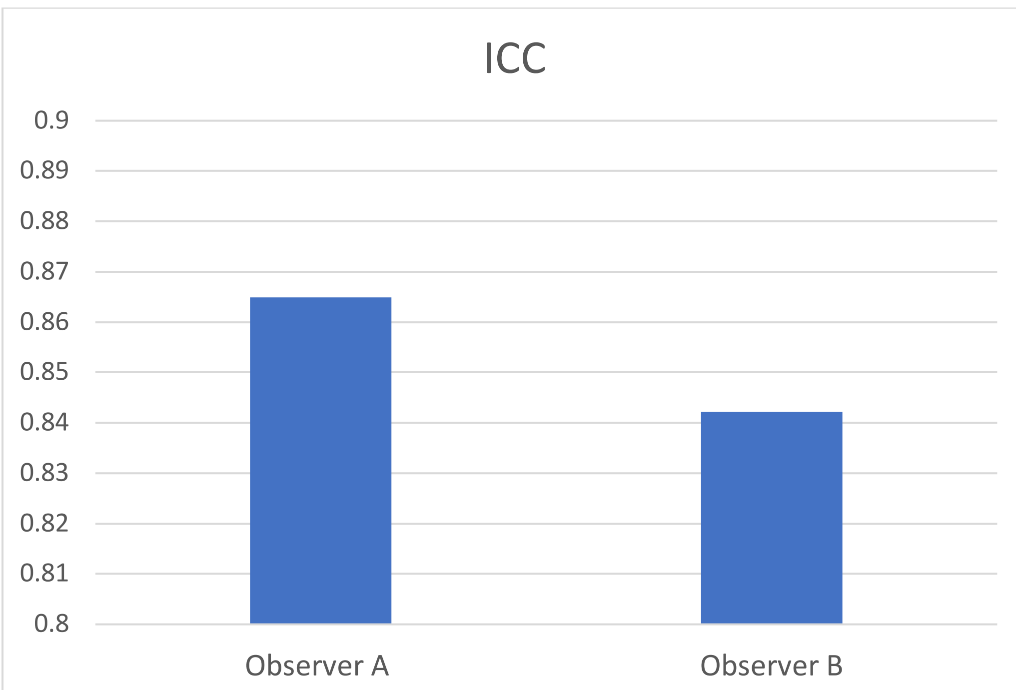


Figure 3. ICC was used to test the intra-reliability of both observers. Both fall into good reliability and display a p-value of <0.0001.

CHAPTER 5
DISCUSSION

The cervical vertebral maturation method was proposed to serve as a diagnostic measure to assess the growth potential of patients, but its precision and accuracy have since been questioned. The methodology of the staging process is a progressive classification, but its reading can be subjective as patients do not always fall into a specific category. In this study, observers staged lateral cephalograms into one of the six CVM stages to test the reliability of the CVM method.

Percent agreement was utilized to test the significance of observances in this study. Observer A expressed 70.8% agreement, while Observer B expressed 64.9% agreement. Both fell within good agreement. Previous literature confirms our results. Zhao's observers had varied intra-observer agreement between 40.7% - 79.1% (Zhao et al., 2012). Gabriel's intra-observer percent agreement was 62% (Gabriel et al., 2009).

The results from our study can conclude that the CVM method has good reliability with intra-observer staging and was verified using ICC. Sohrabi et al. found intra-observer agreement to be moderate to substantial through kappa (0.59 – 0.85) (Sohrabi et al., 2016). Similar results using weighted kappa found that there was moderate to substantial agreement in intra-observer agreement (0.53 – 0.86) in Zhao's study (Zhao et al., 2012). In Gabriel's study, kappa fell into moderate agreement (<0.8) (Gabriel et al., 2009). A higher accuracy may have been found in our results due to the significant sample assessed.

Inter-observer reliability was tested using percent agreement and Cohen's kappa in our study. Overall, percent agreement was found to be 63.2%. Franchi's study found a high-level inter-observer agreement of 98.6% (Franchi et al., 2000). However, Zhao and Gabriel found much lower percent agreement in their studies – 39.3% and 44.9% in Zhao's and 45% and 49% Gabriel's respectively (Gabriel et al., 2009; Zhao et al., 2012). Gabriel et al. included ten observers and assessed 60 radiographs. All ten observers only agreed completely 10% of the time. Most disagreements varied by one stage, but a significant amount had a discrepancy of two or more stages apart. Gabriel et al. concluded CVM had poor reproducibility due to the disagreements seen between viewers (Gabriel et al., 2009).

Our study used Cohen's kappa to statistically test the reliability of inter-observer agreement. Cohen's kappa was 0.8966 and 0.8567 for the first and second trial respectively, indicating almost perfect agreement. Literature results were highly variable when looking at inter-observer agreement. In Sohrabi's article, inter-agreement was 0.48 using Cohen's kappa (Sohrabi et al., 2016). In Zhao's study, inter-observer reliability was judged with Kendall's W and found to be 0.8 indicating substantial agreement (Zhao et al., 2012). Gabriel's Kendall's W was 0.74 and 0.72 respectively, indicating substantial agreement, although the study claimed to show poor reliability (Gabriel et al., 2009). Morris et al. found that inter-observer reproducibility to be high with kappa being 0.88. Sensitivity ranged from 0.25 – 1.0 and specificity ranged from 0.65 – 1.0 (Morris et al., 2019).

Nestman et al. attributed several factors in helping CVM achieve high reliability in previous literature. They note that several studies utilized previously traced lateral cephalograms, had expert observers, and a small sample size (Nestman et al., 2011). This was also articulated in Santiago's systematic review, where of the 23 articles selected, 17 were deemed to have poor quality due to their methodology (Santiago et al., 2012). In this study, there was a focus to limit any bias that may have been seen in these previous studies. Khajah reported that CVM staging accuracy is highly dependent on clinician experience (Khajah et al., 2020). Since this reading is somewhat subjective, observers may rely more on their previous experiences rather than methods provided in the staging process. To combat this bias, our observers were 0 – 1 years into their orthodontic residency. Both had learned about the staging process but had limited exposure before this study. Only one source was used as reference during the staging process. Another source of bias in CVM reading is tooth display. Santiago et al. found most studies reviewed did not report randomizing or blinding their data (Santiago et al., 2012). Seeing the dental age of a patient may sway an observer into guessing a certain developmental stage, especially if unsure on how to characterize the vertebrae. In this study, all areas of the lateral cephalogram were cropped except for viewing of the vertebral column. Finally, many previous studies did not address timing of when radiographs were viewed. Many had two trials one month apart, but that was the only stipulation. Observer fatigue can be notable, especially when viewing a significant number of radiographs. Observers were limited to 20-minute viewing increments to assess radiographs.

Several statistical tests can be used to test the reliability of the CVM method. From our literature review, a variety of tests were used, such as Kendall's W, Cohen's kappa, and ICC to assess inter-observer and intra-observer agreement. Koo reports that low ICC value can be reflective of low sample size, low variability within test subjects, or low raters (Koo & Li, 2016). When there are only two raters, also known as a test-retest for ICC, the linear mixed-effects model is used. It incorporates the mean of the answers, so differences in means will lower the ICC calculation. Other analyses must be conducted to understand if the poor agreement comes from bias or variability between the judges. (Liu et al., 2016). Our study used ICC as our primary test to assess intra-reliability. However, its results were confirmed by also calculating Kendall's W, Cohen's kappa, and CCC in the statistical report.

A related area of interest to CVM reliability is the reliability of staging. In a continuation study of Gabriel's reproducibility study, Nestman et al. tested for which parts of the CVM assessment were unreliable. They gave ten observers five questions, asking if they saw notching on C2, C3, and C4 and if C3 and C4 were trapezoidal, rectangular, square, or vertical rectangular in shape. Overall, inter-observer agreement was assessed using Kendall's W was 0.45. They found that observers generally had good agreement in terms of the notching question but had poor agreement when asked about the shapes of C3 and C4 (Nestman et al., 2011). This was confirmed by Sohrabi's study, where moderate agreement was seen in the C3 and C4 shape. Substantial agreement was seen in the lower borders of C2, C3, and C4. In their results, 88 of the 300 observations derived "inconclusive" staging based on how the questions were answered (Sohrabi et al., 2016). This occurred when the answers to the questions were incompatible with the staging definitions. For example, the observer could reach an inconclusive result if they answered C2 was not notched but C4 was notched, which does not align with any of the staging definitions. This was seen with some of the samples observed in our study as well (Figure 4). The best sensitivity and specificity were found near the ends of the age spectrums tested. The most variability came between ages 10 – 12 years, right when the growth spurt occurs (Morris et al., 2019).

Figure 4: Example of C4 showing a concavity before C3 within the Burlington sample set. By staging definition, this is an inconclusive result in the CVM method.

Although not assessed in previous studies found, head positioning may play a role in how well the cervical vertebrae can be assessed. During the early CVM stages, many children displayed a tendency to lift their head during the lateral cephalogram. Doing so caused the cervical spine to bend backward as seen in Figure 5. This caused some of the vertebrae to be not easily viewed. Such radiographs were assessed as best as possible during this study, but it may have increased the error rate due to unclear viewing of the vertebrae.

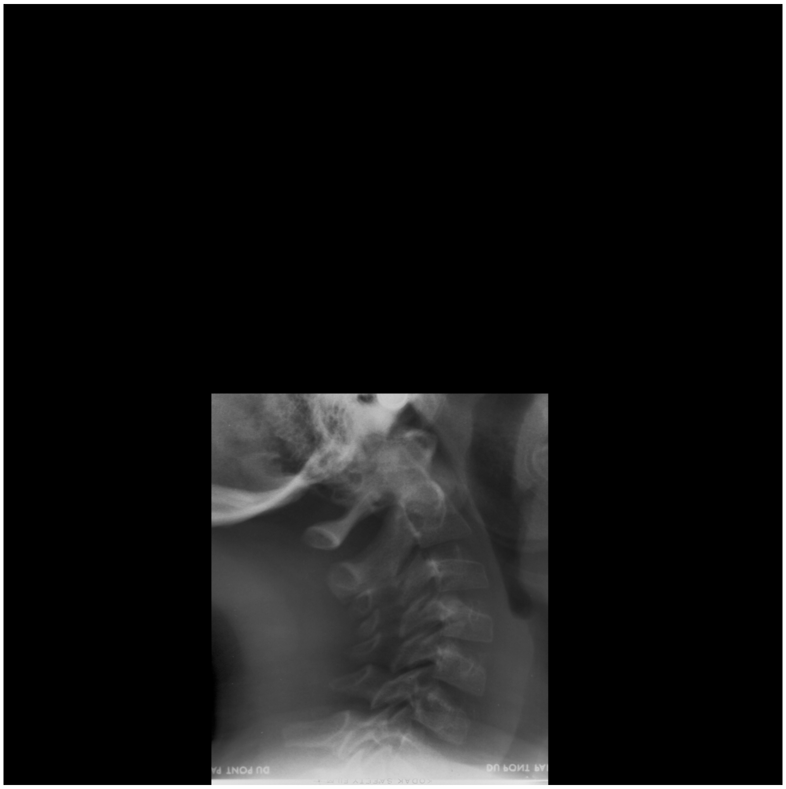


Figure 5. An example from the Burlington Growth Sample of the bent back positioning of cervical vertebrae from lifting the head. Many times, the positioning led to overlap of the cervical vertebrae which inhibited clear viewing.

Cervical anatomy may also play a role in the predictability of CVM staging. Variance in the size and shape of vertebrae was seen between patients. For example, at least one patient in the Burlington sample had severe concavities on all cervical vertebrae as seen in Figure 6. This is significant because the inferior border cannot accurately be assessed as the patient automatically begins at a CVM 4. This deviation in anatomy may be from a variance in genetics. Further investigation into genetic, racial, and environmental effects on cervical vertebrae anatomy may be indicated. The CVM theory was derived from predominantly Caucasian patient pools, which may be inaccurate to all patient backgrounds. Seeing how these various factors may affect vertebral anatomy, studying CVM staging in various backgrounds is becoming more and more critical as patient pools become ever more diverse.

Figure 6. An example of increased concavities on the inferior borders of the cervical vertebrae as a patient variant.

Other means, besides HWM and CVM, have been investigated for determining skeletal age. One promising method is by examining the clavicle on radiographs. The clavicle develops with endochondral ossification. In the pubertal stage, a second epiphyseal plate appears on the medial side of the bone. Complete fusion of this growth plate occurs around the age of 22 years old. Where hand wrist and cervical vertebrae maturation only show development until around age 18, observing the clavicle may prove useful in class III orthognathic cases or even dental implants. The clavicle can currently be observed through CT settings (Manzoor Mughal et al., 2014; Subramanian & Viswanathan, 2023).

The next step in assessing the CVM method will be using AI to identify vertebral stages. In their systematic review, Rana et al. assessed 13 papers looking into cervical vertebral assessment using machine learning. From their study, a positive agreement between CVM assessment and AI readings of cervical vertebrae was found (Rana et al., 2023). In Li's 2022 study, four popular convolutional neural networks (CNNs) were trained to assess cervical vertebrae. The CNNs had a total accuracy ranging between 61.15% - 67.06% and a weighted kappa of 0.796 – 0.826. While these results display good agreement between the four programs, it should also be noted that our human observers had very comparable results. CNNs are progressively improving, and greater accuracy may be able to be achieved in the future. (Li et al., 2022).

CHAPTER 6
CONCLUSION

The cervical vertebral maturation method has been identified as a test to determine skeletal age during adolescent development. It identifies the changing size and shape of C2, C3, and C4 on lateral cephalogram to determine how much growth can be predicted in a patient. Its reliability and validity have since been questioned due to the subjectivity of observer grading. Previous research has published mixed conclusions on the value of this test.

In this study, CVM was found to have a statistically high level of agreement between newly trained observers at two time points. Intra-observer agreement was assessed by the statistical test ICC. For Observer A, ICC was 0.8649, while Observer B performed slightly lower at 0.8422. All these values fell within the high level of agreement. Interobserver agreement was assessed using Cohen's kappa. Kappa value was found to be 0.8966 and 0.8567 for the first and second trials respectively. Both values displayed a high level of agreement between observers.

Although statistical agreement was high, percent agreement was only moderate. Interobserver agreement was found to be 63.2%. Intra-observer agreement was 70.8% for Observer A and 64.9% for Observer B. This admits that observers did not reach agreement about one third of the time with either themselves or each other, which clinically can be relevant. Clinicians are encouraged to use the CVM alongside other developmental predictors, such as secondary sexual characteristics, to estimate growth spurt potential. Utilizing all patient information available can help the orthodontist make a more accurate diagnosis about craniofacial growth potential.

While the CVM may always be flawed with human spectators, there is great potential in machine learning to give a more accurate diagnoses moving forward. Innovative and novel research has shown great promise in the predictability of the CVMS method.

A follow-up study will be done to assess the validity of the CVMS method as predictor test. It will be judged against hand wrist films from the Burlington sample. Similar study protocols will be utilized to limit bias introduced into the study.

REFERENCES

Angelieri, F., Cevidanes, L. H., Franchi, L., Gonçalves, J. R., Benavides, E., & McNamara, J. A., Jr. (2013). Midpalatal suture maturation: classification method for individual assessment before rapid maxillary expansion. *Am J Orthod Dentofacial Orthop*, *144*(5), 759-769. https://doi.org/10.1016/j.ajodo.2013.04.022

Baccetti, T., Franchi, L., & McNamara, J. A. (2005). The Cervical Vertebral Maturation (CVM) Method for the Assessment of Optimal Treatment Timing in Dentofacial Orthopedics. *Seminars in Orthodontics*, *11*(3), 119-129. https://doi.org/https://doi.org/10.1053/j.sodo.2005.04.005

Baccetti, T., Reyes, B. C., & McNamara, J. A., Jr. (2007). Craniofacial changes in Class III malocclusion as related to skeletal and dental maturation. *Am J Orthod Dentofacial Orthop*, *132*(2), 171.e171-171.e112. https://doi.org/10.1016/j.ajodo.2005.07.031

Bagnall, K. M., Harris, P. F., & Jones, P. R. (1977). A radiographic study of the human fetal spine. 2. The sequence of development of ossification centres in the vertebral column. *J Anat*, *124*(Pt 3), 791-802.

Baker, ER (1985). Body weight and the initiation of puberty. In (Vol. 28, pp. 573-579): *Clin Obstet Gynecol*.

Björk, A., & Skieller, V. (1972). Facial development and tooth eruption. An implant study at the age of puberty. *Am J Orthod*, *62*(4), 339-383. https://doi.org/10.1016/s0002-9416(72)90277-1

Björk, A., & Skieller, V. (1977). Growth of the maxilla in three dimensions as revealed radiographically by the implant method. *Br J Orthod*, *4*(2), 53-64. https://doi.org/10.1179/bjo.4.2.53

Buschang, P. H., & Gandini Júnior, L. G. (2002). Mandibular skeletal growth and modelling between 10 and 15 years of age. *Eur J Orthod*, *24*(1), 69-79. https://doi.org/10.1093/ejo/24.1.69

Buschang, P. H., Santos-Pinto, A., & Demirjian, A. (1999). Incremental growth charts for condylar growth between 6 and 16 years of age. *Eur J Orthod*, *21*(2), 167-173. https://doi.org/10.1093/ejo/21.2.167

Cattell, H. S., & Filtzer, D. L. (1965). Pseudosubluxation and other normal variations in the cervical spine in children. A study of one hundred and sixty children. *J Bone Joint Surg Am*, *47*(7), 1295-1309.

Chatzigianni, A., & Halazonetis, D. J. (2009). Geometric morphometric evaluation of cervical vertebrae shape and its relationship to skeletal maturation. *Am J Orthod Dentofacial Orthop*, *136*(4), 481.e481-489; discussion 481-483. https://doi.org/10.1016/j.ajodo.2009.04.017

Edwards, C. B., Marshall, S. D., Qian, F., Southard, K. A., Franciscus, R. G., & Southard, T. E. (2007). Longitudinal study of facial skeletal growth completion in 3 dimensions. *Am J Orthod Dentofacial Orthop*, *132*(6), 762-768. https://doi.org/10.1016/j.ajodo.2006.01.038

Ferrillo, M., Curci, C., Roccuzzo, A., Migliario, M., Invernizzi, M., & de Sire, A. (2021). Reliability of cervical vertebral maturation compared to hand-wrist for skeletal maturation assessment in growing subjects: A systematic review. *J Back Musculoskelet Rehabil*, *34*(6), 925-936. https://doi.org/10.3233/bmr-210003

Fishman, L. S. (1982). Radiographic evaluation of skeletal maturation. A clinically oriented method based on hand-wrist films. *Angle Orthod*, *52*(2), 88-112. https://doi.org/10.1043/0003-3219(1982)052<0088:reosm>2.0.co;2

Franchi, L., Baccetti, T., & McNamara, J. A. (2000). Mandibular growth as related to cervical vertebral maturation and body height. *American Journal of Orthodontics and Dentofacial Orthopedics*, *118*(3), 335-340. https://doi.org/https://doi.org/10.1067/mod.2000.107009

Franchi, L., Nieri, M., Lomonaco, I., McNamara, J. A., & Giuntini, V. (2021). Predicting the mandibular growth spurt. *Angle Orthod*, *91*(3), 307-312. https://doi.org/10.2319/080220-676.1

Gabriel, D. B., Southard, K. A., Qian, F., Marshall, S. D., Franciscus, R. G., & Southard, T. E. (2009). Cervical vertebrae maturation method: Poor reproducibility. *American Journal of Orthodontics and Dentofacial Orthopedics*, *136*(4), 478.e471-478.e477. https://doi.org/https://doi.org/10.1016/j.ajodo.2007.08.028

Gandini, P., Mancini, M., & Andreani, F. (2006). A comparison of hand-wrist bone and cervical vertebral analyses in measuring skeletal maturation. *Angle Orthod*, *76*(6), 984-989. https://doi.org/10.2319/070605-217

Hassel, B., & Farman, A. G. (1995). Skeletal maturation evaluation using cervical vertebrae. *American Journal of Orthodontics and Dentofacial Orthopedics*, *107*(1), 58-66. https://doi.org/https://doi.org/10.1016/S0889-5406(95)70157-5

Khajah, A., Tadinada, A., Allareddy, V., Kuo, C. L., Nanda, R., & Uribe, F. (2020). Influence of type of radiograph and levels of experience and training on reproducibility of the cervical vertebral maturation method. *Am J Orthod Dentofacial Orthop*, *157*(2), 228-239. https://doi.org/10.1016/j.ajodo.2019.03.025

Kimura, K. (1977). Skeletal maturity of the hand and wrist in Japanese children by the TW2 method. *Ann Hum Biol*, *4*(4), 353-356. https://doi.org/10.1080/03014467700002281

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Lamparski, D. (1972). Skeletal age assessment utilizing cervical vertebrae. Pittsburgh, PA.

Li, H., Chen, Y., Wang, Q., Gong, X., Lei, Y., Tian, J., & Gao, X. (2022). Convolutional neural network-based automatic cervical vertebral maturation classification method. *Dentomaxillofac Radiol*, *51*(6), 20220070. https://doi.org/10.1259/dmfr.20220070

Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C., & Tu, X. M. (2016). Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Arch Psychiatry*, *28*(2), 115-120. https://doi.org/10.11919/j.issn.1002-0829.216045

Magalhães, M. I., Machado, V., Mascarenhas, P., Botelho, J., Mendes, J. J., & Delgado, A. S. (2022). Chronological age range estimation of cervical vertebral maturation using Baccetti method: a systematic review and meta-analysis. *Eur J Orthod*, *44*(5), 548-555. https://doi.org/10.1093/ejo/cjac009

Manabe, A., Ishida, T., Kanda, E., & Ono, T. (2022). Evaluation of maxillary and mandibular growth patterns with cephalometric analysis based on cervical vertebral maturation: A Japanese cross-sectional study. *PLoS One*, *17*(4), e0265272. https://doi.org/10.1371/journal.pone.0265272

Manzoor Mughal, A., Hassan, N., & Ahmed, A. (2014). Bone age assessment methods: a critical review. *Pak J Med Sci*, *30*(1), 211-215. https://doi.org/10.12669/pjms.301.4295

McNamara, J. A., Jr., & Franchi, L. (2018). The cervical vertebral maturation method: A user's guide. *Angle Orthod*, *88*(2), 133-143. https://doi.org/10.2319/111517-787.1

Mellion, Z. J., Behrents, R. G., & Johnston, L. E., Jr. (2013). The pattern of facial skeletal growth and its relationship to various common indexes of maturation. *Am J Orthod Dentofacial Orthop*, *143*(6), 845-854. https://doi.org/10.1016/j.ajodo.2013.01.019

Mitani, H., & Sato, K. (1992). Comparison of mandibular growth with other variables during puberty. *Angle Orthod*, *62*(3), 217-222. https://doi.org/10.1043/0003-3219(1992)062<0217:comgwo>2.0.co;2

Montasser, M. A., Viana, G., & Evans, C. A. (2017). Racial and sex differences in timing of the cervical vertebrae maturation stages. *Am J Orthod Dentofacial Orthop*, *151*(4), 744-749. https://doi.org/10.1016/j.ajodo.2016.09.019

Moore, R. N. (1997). Principles of dentofacial orthopedics. *Semin Orthod*, *3*(4), 212-221. https://doi.org/10.1016/s1073-8746(97)80054-0

Morris, K. M., Fields, H. W., Jr., Beck, F. M., & Kim, D. G. (2019). Diagnostic testing of cervical vertebral maturation staging: An independent assessment. *Am J Orthod Dentofacial Orthop*, *156*(5), 626-632. https://doi.org/10.1016/j.ajodo.2018.11.016

Nanda, R. S., & Ghosh, J. (1995). Longitudinal growth changes in the sagittal relationship of maxilla and mandible. *Am J Orthod Dentofacial Orthop*, *107*(1), 79-90. https://doi.org/10.1016/s0889-5406(95)70159-1

Nestman, T. S., Marshall, S. D., Qian, F., Holton, N., Franciscus, R. G., & Southard, T. E. (2011). Cervical vertebrae maturation method morphologic criteria: poor reproducibility. *Am J Orthod Dentofacial Orthop*, *140*(2), 182-188. https://doi.org/10.1016/j.ajodo.2011.04.013

Proffit, W. R. (2013). *Contemporary orthodontics* (5th ed.). Elsevier/Mosby.

Rana, S. S., Nath, B., Chaudhari, P. K., & Vichare, S. (2023). Cervical Vertebral Maturation Assessment using various Machine Learning techniques on Lateral cephalogram: A systematic literature review. *J Oral Biol Craniofac Res*, *13*(5), 642-651. https://doi.org/10.1016/j.jobcr.2023.08.005

Santiago, R. C., de Miranda Costa, L. F., Vitral, R. W., Fraga, M. R., Bolognese, A. M., & Maia, L. C. (2012). Cervical vertebral maturation as a biologic indicator of skeletal maturity. *Angle Orthod*, *82*(6), 1123-1131. https://doi.org/10.2319/103111-673.1

Silveira, A. M., Fishman, L. S., Subtelny, J. D., & Kassebaum, D. K. (1992). Facial growth during adolescence in early, average and late maturers. *Angle Orthod*, *62*(3), 185-190. https://doi.org/10.1043/0003-3219(1992)062<0185:fgdaie>2.0.co;2

Sohrabi, A., Babay Ahari, S., Moslemzadeh, H., Rafighi, A., & Aghazadeh, Z. (2016). The reliability of clinical decisions based on the cervical vertebrae maturation staging method. *Eur J Orthod*, *38*(1), 8-12. https://doi.org/10.1093/ejo/cjv030

Soliman, A., De Sanctis, V., Elalaily, R., & Bedair, S. (2014). Advances in pubertal growth and factors influencing it: Can we increase pubertal growth? *Indian J Endocrinol Metab*, *18*(Suppl 1), S53-62. https://doi.org/10.4103/2230-8210.145075

Subramanian, S., & Viswanathan, V. K. (2023). Bone Age. In *StatPearls*. StatPearls Publishing. Copyright © 2023, StatPearls Publishing LLC.

Szemraj, A., Wojtaszek-Słomińska, A., & Racka-Pilszak, B. (2018). Is the cervical vertebral maturation (CVM) method effective enough to replace the hand-wrist maturation (HWM) method in determining skeletal maturation?—A systematic review. *European Journal of Radiology*, *102*, 125-128. https://doi.org/https://doi.org/10.1016/j.ejrad.2018.03.012

Tekın, A., & Cesur Aydın, K. (2020). Comparative determination of skeletal maturity by hand-wrist radiograph, cephalometric radiograph and cone beam computed tomography. *Oral Radiol*, *36*(4), 327-336. https://doi.org/10.1007/s11282-019-00408-y

Wong, R. W., Alkhal, H. A., & Rabie, A. B. (2009). Use of cervical vertebral maturation to determine skeletal age. *Am J Orthod Dentofacial Orthop*, *136*(4), 484.e481-486; discussion 484-485. https://doi.org/10.1016/j.ajodo.2007.08.033

WW, G., & SI, P. (1959). Radiographic atlas of skeletal development of the hand and wrist. In. Palo Alto, CA: Stanford University Press.

Zhao, X. G., Lin, J., Jiang, J. H., Wang, Q., & Ng, S. H. (2012). Validity and reliability of a method for assessment of cervical vertebral maturation. *Angle Orthod*, *82*(2), 229-234. https://doi.org/10.2319/051511-333.1