UNIVERSITY OF LIVERPOOL

Doctorate of Dental Science

Reliability of Cervical Vertebrae Maturation Staging Method.

Billie-Jean Rainey

January 2014

Acknowledgements

This research project would not have been possible without the support of many people. I wish to express my gratitude to my supervisor, Dr. Jayne Harrison, who was abundantly helpful and offered invaluable assistance, support and guidance. Deepest gratitude is also due to Dr Jim McNamara and Dr Lorenzo Franchi, without whose knowledge and assistance this study would not have been successful.

I also wish to express my love and gratitude to my beloved families; for their understanding & endless love, through the duration of my studies.

Table of Contents

Abstract

Background

Knowledge of craniofacial growth and development is a prerequisite for the comprehensive and successful management of orthodontic patients. In orthodontic treatment during adolescence, craniofacial growth is often paramount to its success of treatment, especially in patients with skeletal discrepancies. The ultimate goal, in orthodontics, would be the ability to predict accurately the onset, duration and magnitude of the peak pubertal growth spurt, particularly in relation to the mandible.

The radiographic assessment of features of skeletal maturation has been extensively researched, as a means of determining an individual's growth potential. Historically, assessments of the ossification of the bones on the hand-wrist radiograph were evaluated. However for orthodontics, in the UK and some other parts of the world, this method has been superseded by assessment of morphological features of the cervical vertebrae, on the lateral cephalogram. This increase in popularity is because the cervical vertebrae assessment prevents additional radiation to the patient. It is, therefore, safer for the patient.

Aim

This study aimed to:

1. Determine the reliability and reproducibility of Cervical Vertebrae Maturation (CVM) stage assessment amongst orthodontists in training and specialist orthodontists, looking at a sample of consecutive lateral cephalograms taken at Liverpool University Dental Hospital.

2. Determine the reliability and reproducibility of CVM stage assessment amongst orthodontists in training and specialist orthodontists, looking at a sample of ideal images provided by co-author of the index, Dr J McNamara.
3. Compare the agreement of specialist orthodontists with orthodontists in training.
4. Determine whether increased experience with the index improved the agreement between observers.
5. Determine if the principal investigator (BJR) and research supervisor (JEH) agree with the experts and developers of the index (JMN/LF) and determine if they could be classified as experts.

Design

This was a two phase reliability study. A group of 20 orthodontic clinicians, none of whom had used a CVM staging method previously, were trained in the use of the improved version of the CVM method for the assessment of mandibular growth using McNamara's teaching programme.

They independently assessed a sample of 72 consecutive lateral cephalograms, taken at Liverpool University Dental Hospital, on two separate occasions. The cephalograms were presented in a random order and interspersed with 11 ideal images from McNamara for standardisation. The intra- and inter-observer agreements were evaluated, for both image samples, using the weighted kappa statistic.

The principal researchers also completed the two phase reliability study. Their results were analysed separately and compared to the findings for observers with no previous experience.

The principal investigators then mutually agreed on staging of each radiographs and compared these to the staging given by the developers of the index, to determine if the principal investigator and research supervisor could be classified as experts.

Results

The intra-observer and inter-observer agreements were substantial, (weighted kappa 0.6-0.8). The overall intra-observer agreement was 0.70 (SE 0.01) with average agreement 89%. The inter-observer agreement on the first occasion was 0.68 (SE 0.03) and 0.66 (SE 0.03) on the second occasion, with an average inter-observer agreement of 88%.

Conclusions

The intra-observer and inter-observer agreement of classifying CVM stages, using the improved version of the CVM method for the assessment of mandibular growth, were substantial.

Chapter 1: Introduction

Knowledge of craniofacial growth and development is a prerequisite for the comprehensive and successful management of orthodontic patients. It plays a crucial role in the diagnosis, treatment planning, result and overall stability of the outcome of a patents orthodontic treatment. Numerous methods have been investigated to identify the stage of growth and development and predict both the timing of onset, and potential of this growth. These include assessment by chronological age, skeletal age, skeletal maturation,[1,2] mandibular growth,[3] standing height,[2,4-6] menarche and voice changes[7] and cervical vertebrae maturation.[8] Of these, the use of hand-wrist radiographs to assess skeletal maturity and growth have been investigated by several authors.[1,4,9,10-19] Initially advocated by Bergersen,[2] Fishman[1] introduced the skeletal maturity index (SMI) in 1982, in response to, conflicting evidence from Houston[9] and Hagg.[10] The principal of using the skeletal maturity index varied in popularity, mostly because it always required additional radiation exposure and specific skill to interpret. As a result, alternatives to hand-wrist radiographs were developed. These included assessment of the relationship between skeletal maturity using hand-wrist radiographs and CVM staging using lateral cephalograms.[20-31] Lateral cephalograms are commonplace in orthodontics and more familiar to the orthodontist. Other authors have looked at the relationship between CVM and mandibular growth[32-38] and largely concluded that the CVM method is a valid indicator for the assessment of skeletal maturity, and is comparable with the use of hand-wrist radiographs. There have been conflicting reports in the literature regarding the reliability and reproducibility of the various methods of CVM staging.[11,20,22,24,26,28,29,31,34,37,87,88-91] This study aims to determine the reliability of the most common method of CVM staging in a group of orthodontists in training and orthodontic specialists.

Chapter 2: Literature review

2.1 Introduction

There has been an increasing trend for orthodontics in the adult population[39] however, the majority of treatment is still carried out on growing patients.[40] Orthodontists must, therefore, have an in depth understanding of craniofacial growth and development. They must use this knowledge to diagnose and plan treatment for their patients, so as to determine the most appropriate and most efficient options. It is important for orthodontists to be able to evaluate the stage of growth of a patient, and evaluate how much growth remains.[41,42] They may want to use future growth to facilitate treatment, attempting to modify or alter it, or to assess potential adverse effects of future growth. However, for some patients, it may be advantageous to delay treatment until growth is completed for optimal results.

The rate of growth varies throughout childhood and adolescence and not all systems in the body grow at the same rate.[43] Craniofacial growth is determined by genetics but is also influenced by environmental factors. Craniofacial growth and development, from birth to adulthood, occurs in a sequential and predictable pattern, however it demonstrates large individual variation in terms of the chronological age at which children reach similar developmental events such as sexual maturation, dental development and peak statural height.[44]

Many investigators have looked at different methods to assess the stage of growth and maturational development of an individual patient and predict the timing of future growth. Due to the wide variation in maturity with chronological age, assessment of physiological maturity is a better indicator in the assessment of developmental status.[45] Various methods of assessing developmental status have been reported in the literature, including those based on somatic, dental, skeletal and sexual maturity indicators as predictors.[1-48,50-84]

2.2 Prediction of Growth and Development

Growth and development in humans is not a uniform process. It is characterised by periods of acceleration and deceleration from birth to adulthood.[24] Maturation can be viewed as a "*series of successive transformations through time leading to the attainment of adult stature".*[46] An individual's stage of maturity can be evaluated by assessing whether various maturity criteria have been reached. Sexual maturation, chronological age, dental development, height, weight and skeletal development are some of the common ways that have been used to help determine the stage of development or the maturation of an individual.[1-48,50-73]

2.2.1 Chronological age

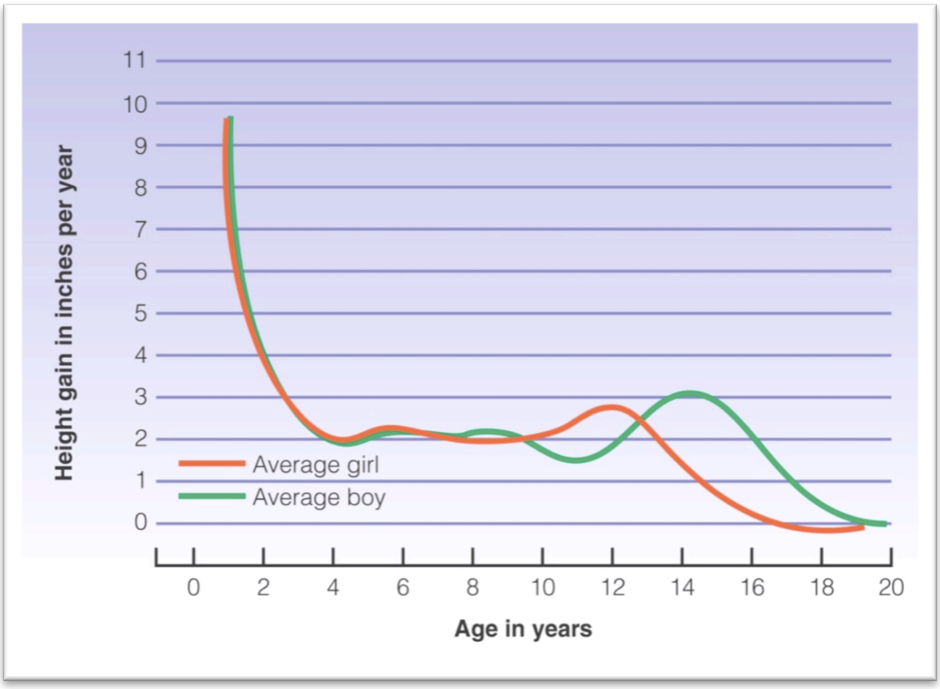
Among healthy children, the early signs of puberty can be seen as young as 8-9 years of age in girls and 9-10 years in boys, and range anywhere up to the late teens.[4] Developmentally, an individual may be advanced or delayed for their chronological age.[47] Therefore, the general stage of growth and development cannot be estimated accurately from chronological age alone,[48] and assessment of an individual's physiological age is a more valid means of determining their maturation.

2.2.2 Somatic development

Growth charts show expected height, weight, Body Mass Index (BMI) and height velocity for boys and girls. A number of growth charts, following the development of children from birth to adulthood, have been developed for assessment of children in the UK. They are designed to be used, either for assessment and monitoring of an individual, or for screening a population. They are based on longitudinal or cross-sectional population data. Both types of data have their flaws. Longitudinal data suffer from the influence of time-lag bias and are only applicable to the population from which the data are derived from. Cross-sectional charts are not strictly valid for monitoring growth over time. In 2002, the Royal College of Paediatrics and Child Health (RCPCH)[49] convened an expert group to provide guidance on the validity of available growth charts, including Tanner-Whitehouse,[50] Gairdner-Pearson,[51] Buckler-Tanner[52] and the UK 1990 growth references.[53] They concluded that the 'UK 1990 reference is the only suitable reference that can be recommended'. Wright et al., and also recommended further improvements and refinements in the area of chart design using the UK 1990.[54] More recently, the World Health Organisation commissioned a multicentre growth study to develop growth standards.[55,56] The WHO growth standards followed how children grew in ideal conditions, from birth to 24months. They were based on breast fed children, developing under ideal conditions, in different countries around the world. The UK 1990 forms a valuable reference for clinicians or researchers, as it combines growth data from several children in the UK from 1990, and describes the growth patterns at that time. The WHO data depicts the growth of children in a perfect environment, and not real life, and is therefore a standard describing how children should grow, rather than a reference chart describing how they do. This difference is highlighted if we compare the UK1990 and WHO data from 0-24 months. The WHO standard shows a slower rate of weight gain from 4-24 months. This slower weight gain has been shown to be indicative of a reduced risk of obesity later in life. However, the WHO growth standard may be more applicable to the 2012 multinational UK population compared to the UK 1990 population, accounting for geographical and ethic variation of the population. This has lead to the amalgamation of the UK 1990 from 24 months to adulthood, and the WHO standards, from 0-24 months,[57] to form the current recommended growth charts.

If we consider a growth chart, in terms of predicting the stage of maturation, one depicting height of an individual against their age, will show a steady increasing curve from birth to adulthood, until maximum height is reached. This tells us nothing of the rate/velocity of growth that can occur. However, If we plot height change against age of patient, we can see the rate of growth and therefore the individual's phases of growth. (Figure 1).[49] It is this height velocity chart which is of interest to the orthodontist when attempting to assess the stage of maturation of an individual.

Figure 1: Height Velocity Graph for Average Male and Female Children[49]



Generally, the rate of growth is highest in the first year of life, falling rapidly until 5-6 years of age. A slowly decelerating phase exists until adolescence, interrupted in some cases by a brief and variable juvenile growth spurt, around 6-8 years. There is then an accelerated phase of increased growth rate around 10-16 years. This is associated with puberty and often referred to as the '*circumpubertal growth spurt*'.[43,50,59] The average onset of the pubertal growth spurt is from 10-14 years in girls and 12-17 years in boys. In both sexes, the Peak Height Velocity (PHV) is seen approximately two years after the onset of the pubertal growth phase. Growth velocity then slows until adulthood. Onset, duration, velocity and direction of growth can vary significantly among individuals of the same chronological age, with individuals being categorised as early, late or average maturers, according the age they undergo their pubertal growth spurt.[46]

Sullivan et al.,[59] first reported the method of prediction of the pubertal growth spurt by measurement of standing height. This necessitates the standardised measurement of standing height over a period of time prior to commencing orthodontic treatment. After the second reading and for each reading thereafter, height measurements are translated into the height velocity values and plotted on a growth chart. Tanner developed an onlay chart as part of the Harpenden growth study, which was placed over the height velocity chart and allowed the operator to determine which growth chart the patient was following and make predictions.[60] This method is no longer current as the individuals it was based on, the Harpenden sample, is now historic. The validity of this method is further weakened by the fact the girls in the sample were in state care and many had emotional problems, which may have contributed to delays in pubertal onset and restricted growth. The current growth charts have a supplemental guide for assessing pubertal stage using Tanner's stage of puberty, however these require a patient interview regarding the development of secondary sex characteristics, which may be inappropriate for patients seeking orthodontic treatment.

A large body of research has suggested that growth of the facial dimensions is correlated to the growth pattern of standing height. Nanda[61] and Houston[62] reported that the maximum velocity in facial growth was reached after the maximum body height, whereas Hunter[4] and Moore[63] reported that stature and facial growth were coincident with each other. If it is accepted that these events occur at generally the same time, this information can assist the orthodontist to predict the effect of future growth. Overall, little strong evidence exists of a correlation between standing height and dentofacial growth with several studies arguing both sides.[2,4,16] Van der Beek[5] summarises the discussion highlighting the fact that a comparison is measurable, between standing height and mandibular growth, and is measurable based on specific aspects of a growth curve e.g. time of onset and duration of growth curves can be compared using correlative techniques.

The study by Mitani and Sato[3] aimed to explore a possible relationship between growth of the mandible compared to other clinical variables during puberty. A small sample was used to assess several variables including development of the hyoid bone, cervical vertebrae, hand bones and standing height. The study revealed that there was variability in the timing and magnitude of peak growth between the variables and mandibular growth seemed to express the most variability.

In conclusion, peak growth velocity in standing height is a valid representation of the rate of overall skeletal growth. It is a useful, historic, longitudinal measure of an individual's growth pattern but has little predictive value of future growth rate or percentage of total growth remaining due to the difficulty in identifying it practically.[64] Although this correlation has been scientifically proven, the practicality of using it creates a problem. Longitudinal, standardised height assessment would be required to identify accurately the onset of the accelerated phase of growth, and even with this, it may only be possible to detect the maximum growth event, when the velocity graph takes a downward turn, i.e. after the event has occurred.

2.2.3 Dental development

Many investigators have deliberated using the stage of dental development as an indication of the stage of maturation of an individual. Tanner, in 1962, reported a low level of correlation between the stage of dental eruption and skeletal maturity, and a wide individual variation.[65] Hagg et al.,[66] stated that dental development had only a weak correlation with the physiological maturity in females and was not an accurate predictive index. However, Sierra[67] aimed to correlate dental calcification with skeletal maturity. She concluded that calcification of the lower canine had the greatest correlation with the skeletal maturity. This theory was supported by Coutinho.[68] Subsequently, it has been shown that neither the late mixed dentition or early permanent dentition are valid indicators of the onset of the pubertal growth spurt.[69,70]

2.2.4 Pubertal sexual development

Onset of puberty varies with gender, generation, genetics and environment and varies considerably from one individual to another.[71-73] Sexual maturation and secondary sex characteristics have been shown to correlate well with stages of an individual's biological maturity. In females, the onset of the menarche usually follows their growth spurt by one year[7] therefore, although accurate at predicting this stage of development, it is of little benefit in identifying the period leading up to peak height velocity for girls.[6,11] The skeletal age of girls at the onset of breast development is reported to vary as much as their chronological age, therefore is not suitable for identification of stage of maturity a girl is at.[60] Pubic hair development has a close relationship with peak height velocity[60,74] although it does not correlate to the take-off point of pubertal growth spurt, and is therefore also of little benefit.

Voice changes in male adolescent take place throughout the period of circumpubertal growth spurt and are most commonly seen on the decelerating phase of growth in this stage. It, therefore, is also not of benefit in predicting the onset of the growth spurt, but may be a characteristic that an orthodontist would take into account as an indicator of having missed maximum growth phase.[10,66] The development of testes and penis in adolescent boys, is due to androgen hormones. The increase in stature follows development of testes therefore it can be used as an indicator of pending growth spurt.[60,75]

The questioning of patients regarding the development of secondary sexual characteristics to predict skeletal maturity, although accurate in some cases, can be seen as inappropriate and sensitive topic in the context of the orthodontic patient/parent interview as they may fail to see the significance of such a line of questioning. This means that it not included routinely in the orthodontic patient history.

2.2.5 Skeletal maturity

Skeletal maturity refers to the amount of ossification of a bone. The evaluation of skeletal maturation, using radiographs, allows the analysis of biological maturity of individuals. Maturation is marked by an orderly, reproducible sequence of recognisable changes in appearance of the skeleton.[12] Maturation staging from radiograph analysis is a widely used approach to predict the timing of pubertal growth, to estimate growth velocity and establish an estimate of how much growth is potentially remaining.[64] The two main methods of assessing skeletal maturity for the prediction of growth are the use of hand-wrist radiograph and the assessment of the cervical vertebrae from a lateral cephalogram.

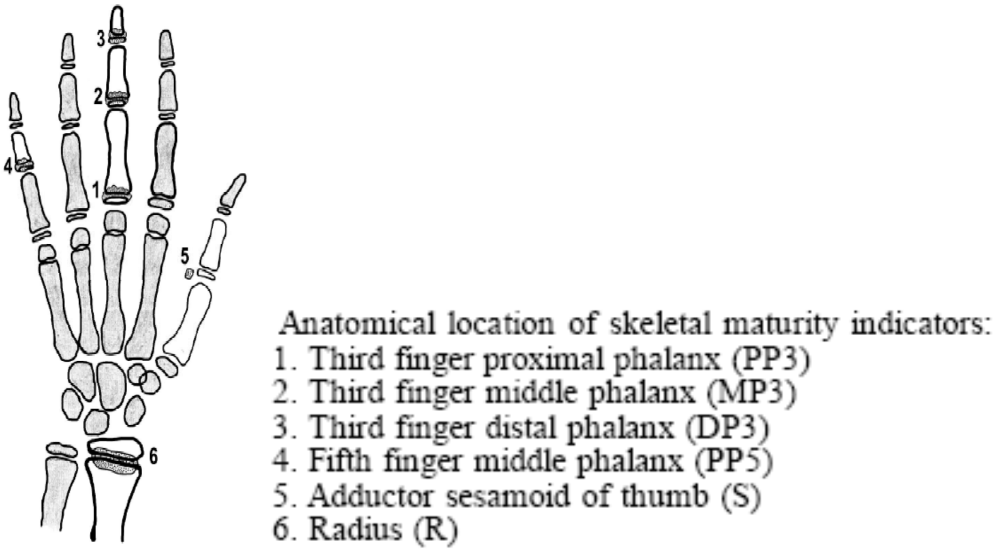
Hand-wrist radiograph to assess Skeletal Maturity

The mean peak velocity in skeletal growth has been shown to be approximately 14 years for boys, with a 2 year range, therefore can occur anywhere between 12-16 years. In girls, the range is 10-14 years.[46] This is approximately coincident with the pattern of statural growth velocity. This variability in development illustrates again how chronological age gives little insight into identifying the stages of developmental progression through adolescence to adulthood.[1,47] The hand-wrist radiograph is commonly used for skeletal developmental assessment. Greulich and Pyle initially described the method of assessing skeletal maturity using hand-wrist radiographs, by comparing it to a standardised collection of radiographic images in an atlas, to predict the developmental stage.[13] The use of the hand-wrist radiograph is based on the different types of bone available in the region, and their different stages of ossification. Greulich and Pyle derived their standards from a white American population of reasonably high socioeconomic status, and therefore are only accurately applicable to a similar population. Tanner and Whitehouse also developed a method of assessment of skeletal maturity, from hand-wrist radiographs. It is derived from a British population and therefore more applicable to our population, however these data are now out dated as a result of time-lag bias, and no longer current.[14]

Milner et al.,[15] compared the above two methods of assessment, on a population of British children, and found the Greulich method consistently underestimated the bone age compared to the Tanner and Whitehouse method. This may be due to the fact the standards were derived from two different populations, highlighting that maturation is influenced by the genetics and environment of a population. Therefore, this must always be considered when applying population standards to individuals.

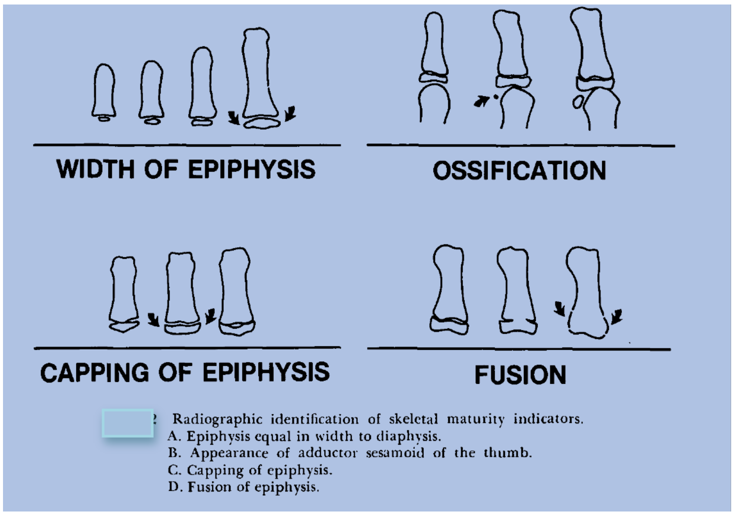
In the 1980s, Fishman developed an alternative index using the hand-wrist radiograph, called the Fishman Method of Prediction (FMP), to assess skeletal maturity based on 6 sites located on the thumb, third finger, fifth finger and radius[1] (Figure 2).

Figure 2: The six sites described in the FMP[1]



Fishman described 11 discrete Skeletal Maturation Indicators (SMI) using 4 stages of maturation at these 6 sites. The ossification stages described are epiphyseal widening, ossification of the sesmoid of the thumb, capping of selective epiphyses over their diaphyses and ending with the fusion of selected epiphyses and diaphyses. (Figure 3.)

Figure 3: The 4 discrete stages of ossification described in FMP[1].



This index was developed from data based on the Denver Child Research Council and supplemented with cross sectional data. Fishman found the 11-stage sequence of maturation was stable, with only

3 out of 2000 cases deviating from the pattern. He concluded that his method was valid in both the clinical and research situations and also concluded that variations in height and craniofacial growth velocities were directly related to skeletal maturational development. The stages of relative maturation were correlated to specific amounts of completed growth.[16] The SMI, from cross-sectional data, was supplemented with graphs and tables, from the longitudinal study, to estimate the relative growth rate and the percentage growth completed. This evaluation of potential growth remaining was, and still is in some countries, taken into consideration in orthodontic treatment planning.

Assessment of skeletal maturation from hand-wrist radiographs has been shown to be both a reliable and valid method of assessing an individuals stage of maturation in a systematic review completed by Flores-Mir.[64] This paper reviewed the available literature on hand-wrist radiographic analysis as a predictor of facial growth, and found that there were significant correlations between skeletal maturity and mandibular growth velocity. This group also found growth of the mandibular body length was more closely correlated to skeletal maturity than growth of ramus height, and concluded that the overall horizontal and vertical facial growth velocity was related to the SMI determined by the hand-wrist radiograph. However, individual mandibular and maxillary growth velocity was less robustly related to the SMI of Fishman.[54]

Disadvantages of the hand-wrist method of skeletal maturity assessment, include the need for an additional radiographic exposure for the patient and the need for considerable skill in interpretation the hand-wrist radiograph. This is particularly of concern to the orthodontist as it is an area of anatomy that they may not familiar with.

The validity of the hand-wrist skeletal maturity in the evaluation of craniofacial growth has been confirmed by numerous studies,[2,18,19] however, the additional and/or avoidable radiographic exposure is an important patient safety consideration. The goal of dental radiography is to obtain diagnostic information while keeping the exposure to the patient and dental staff at minimum levels.[76] While some exposure to radiation is acceptable in medical and dental practice, it should be understood that levels of radiation exposure to patients, dental staff, and other nearby occupants should be kept to As Low As Reasonably Achievable (ALARA) to reduce health risks from ionizing radiation.[76] The Ionising Radiation (Medical Exposures) Regulations (IRMER) were produced and state any methods that can reduce patient radiation exposures should be practiced.[77] Practitioners must always balance the risk of patient exposure with the benefit. Therefore, if the orthodontist is not confident at interpreting the radiograph, they should not take it and secondly, if there is an alternative method that reduces the radiation risk to the patient then this, should be viewed as superior.

It should also be noted that there are other limitations to the use of the hand-wrist radiograph for assessment of skeletal maturity, including the polymorphism and sexual diamorphism of the ossification sequence and timing of events.[62] This may reduce the accuracy of this method.

Cervical Vertebrae Maturation as a method of prediction of skeletal maturity

The morphogenic changes of the cervical vertebrae has been investigated and categorized by many authors, as a method of predicting biological maturity and growth potential of an individual.[11,20,22,24,26,28,29,31,34,37,87,88-91] This method is based on assessing the shape of the cervical bodies, seen on routine lateral cephalograms.

Anatomy of the cervical spine

The anatomy of the vertebrae, making up the cervical spine, is not uniform throughout each vertebra. The first and second differ in structure from C3-C7 (Figure 4).[78] The first vertebra is the Atlas. It is the uppermost and, along with C2, the Axis, it forms the connecting joint of skull to the spine. It is formed from three primary ossification centres - the body and two neural arches. The body of C1 is not ossified at birth, but becomes visible as one or two ossification centres during the first year of life. C2, forms from four primary ossification centres, the odontoid, the body and two neural arches. The most distinctive feature of C2 is the strong odontoid process, which rises perpendicularly from the upper surface of the body. The anterior portion of the body is deeper than the posterior and prolonged downward anteriorly. Ossification of the body of C2 begins at approximately 5 months' intrauterine life. C3-C7 are structurally very similar and have the same developmental pattern. (Figure 5).[80] Their body arises from a single ossification centre at approximately 5 months' intrauterine life.[79]

Figure 4: The anatomy of the Cervical Spine[80]

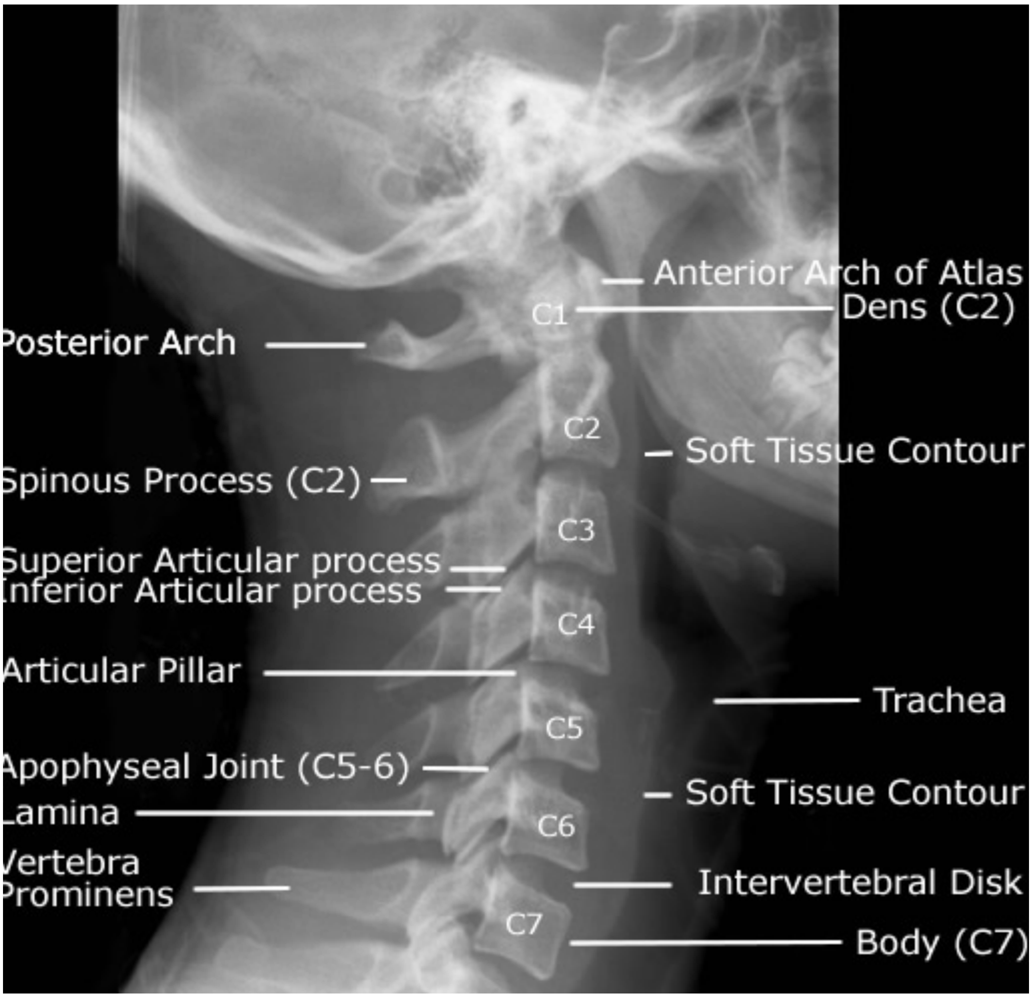
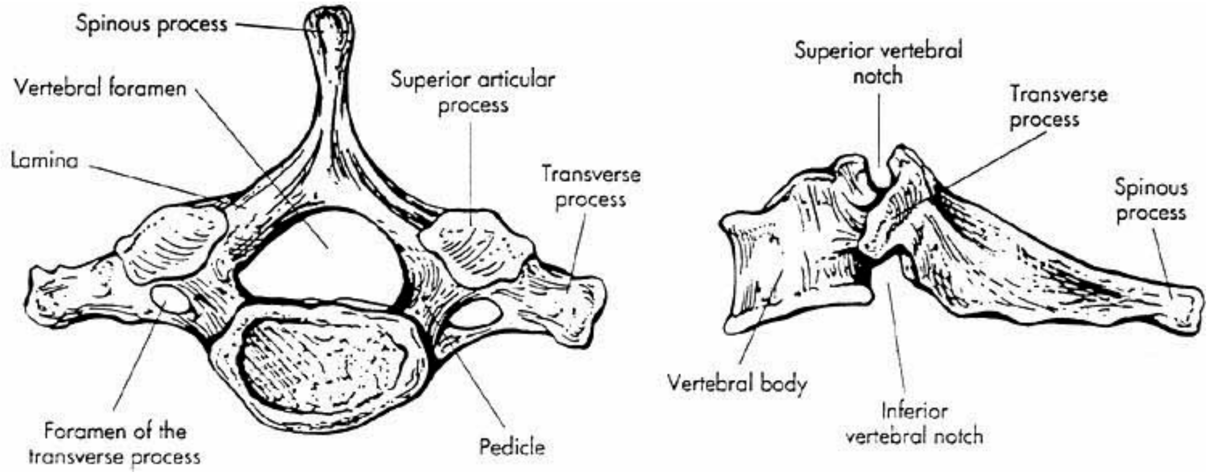


Figure 5. Anatomy of Cervical Vertebra[79]



The study of the growth and development of the cervical vertebrae has revealed both changes in the size and shape of the cervical vertebrae throughout growth.[81-2] Bick et al.,[83] autopsy study of human specimens, at succeeding ages, from a 14 week old foetus to a 25 year old female, documented that ossification events began in fetal life and continued to adulthood.

The longitudinal growth of the body of the vertebrae takes place through endochrondrol development from epiphyseal plates.[83] After completion of endochondral ossification, growth of the vertebral body takes place by periosteal apposition and remodelling. Initially, apposition occurs on the external surfaces in a horizontal direction. During the adolescent growth spurt, the body of the vertebra grows in a vertical direction on the anterior and posterior lower body of the body. Secondary ossification centres appear at puberty along the superior and inferior aspects of the cervical bodies (the superior and inferior epiphyseal rings). Maturational changes can be seen from birth to puberty. These maturational changes in the cervical vertebra, like most bones, are evident radiographically.[81]

2.3 Development of Cervical Vertebrae Maturational Indices

In 1972, Lamparski[8] completed his unpublished thesis exploring "Skeletal Age Assessment Utilizing Cervical Vertebrae". He studied changes in size and shape of cervical vertebrae (C2–C6) to create a defined set of categories for the maturation of cervical vertebrae in males and females. He based his work on previous authors' findings.[80] Lamparski's study only included children aged 10-15 years. This limits the application of the results as it has been shown that the circumpubertal growth spurt can continue beyond this, especially in males.

Hassel and Farman[20] used a sample (N=220) from the Bolton-Brush growth centre, to identify maturational markers in the cervical vertebrae that correlated with Fishman's Skeletal Maturity Index (SMI) using hand-wrist radiographs.[1] The subjects were mainly white and primarily of Northern European descent. They staged the hand-wrist radiographs of the participants according to Fishman's SMI and then studied the lateral cephalogram taken on the same day to identify a pattern. Their work concentrated on the maturation of C2 to C4 as these were the vertebrae that were easily identifiable and not obscured by the use of a thyroid collar. These vertebrae were traced and morphological changes reported. The study provided 6 distinct phases of vertebral maturation for C2, C3 and C4 involving changes in the gross morphology (wedge shaped or square), the vertical and horizontal dimensions and the curvature of the lower border:

Cervical vertebra maturation indicators as described by Hassel and Farman[20]

- **Category 1 was called *Initiation***. This correlated to SMI 1 AND 2. This category represents the beginning and at this stage 80-100% of adolescent growth was still expected. The vertebrae were wedge shaped and the vertebral superior borders were tapered from posterior to anterior.
- **Category 2 was called *Acceleration***. This correlated with SMI stage 3 and 4. Growth acceleration was beginning at this stage with 65-85% of adolescent growth expected.

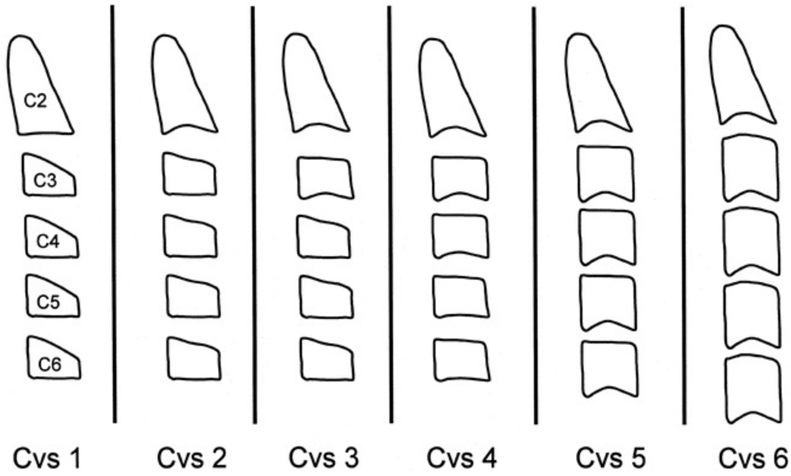
Concavities were developing in the inferior borders of C2 and C3. The inferior border of C4 was flat. Bodies of C3 and C4 were nearly rectangular in shape.

- **Category 3 was called *Transition.*** This correlated to SMI stage 5 and 6. Adolescent growth was approaching the peak growth velocity with 25-65% of growth expected. A concavity was beginning to develop on the inferior surface of C4 with C3 and C4 being rectangular in shape.
- **Category 4 was called *Deceleration.*** This correlated to SMI stage 7 and 8. Only 10-25% of growth was expected at this stage. Concavities were seen in the inferior border of C2-4. C3 and C4 were becoming square in shape.
- **Category 5 was *Maturation.*** This correlated to SMI stage 9 and 10. Only 5-10% of adolescent growth could be expected at this stage. The inferior border of C2-4 had more accentuated concavities whilst C3 and C4 were almost square in shape
- **Category 6 was *Completion.*** This correlated to SMI 11. Growth was considered to be finished. Deep concavities were seen on the lower border of C2-4. C3 and C4 were square or vertically rectangular.

Hassel and Farman[20] reported significant intra-observer and inter-observer reliability. They concluded that improved training in staging would improve the reliability of the staging method, and that it was comparable with Fishman's SMI method for assessment of individual skeletal maturity. Correlation of the Hassel and Farman's method with SMI was supported by Pancherz et al., in his study assessing treatment effects of the Herbst applance.[84] San Roman[24] reported greater validity when the Hassel and Farman[20] method was compared to the Lampriski[8] method due to the greater description of each stage and a more accurate sample from which the results were derived. The Hassel and Farman method has draw backs in that it relies on cephlometric tracing initially which can result in the introduction of error.[85] It describes definite stages of maturation for a continuously changing area of anatomy, and therefore the authors conclude that borderline cases may be difficult to interpret.

Franchi and co-workers[34,37] confirmed the validity of Lamparski's original method as a biologic indicator for both mandibular and somatic skeletal maturation. They initially modified Lamparski's[8] original method by making it applicable to both sexes, to make it easier to use and suitable for the vast majority of patients. (Figure 6)[34]

Figure 6: Schematic morphology of Cervical Vertebral Matuartion Stages as described by Franchi[34]

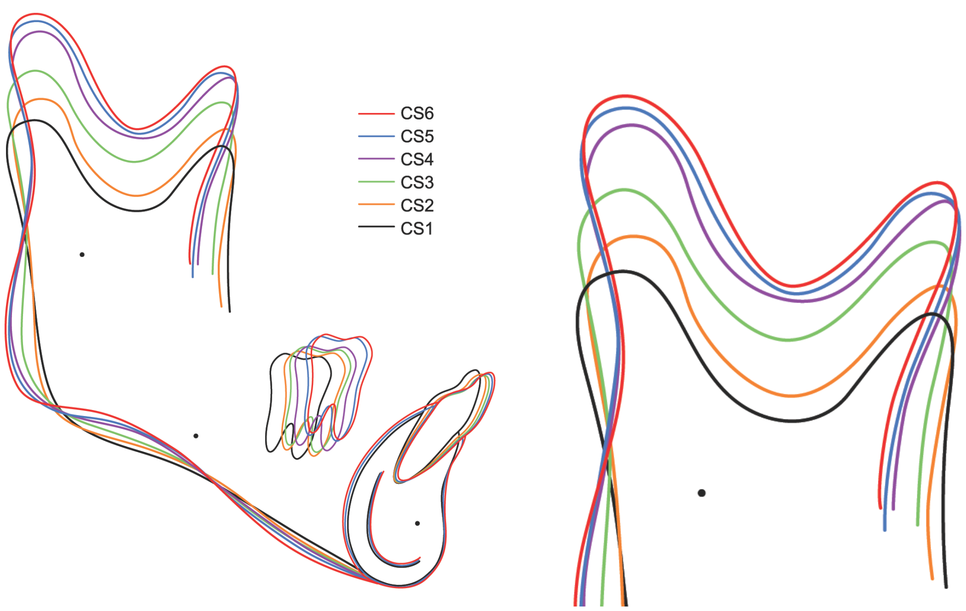


Developmental stages of cervical vertebrae.

- **Stage 1 (Cvs 1):** the inferior borders of the bodies of all cervical vertebrae are flat. The superior borders are tapered from posterior to anterior.
- **Stage 2 (Cvs 2):** a concavity develops in the inferior border of the second vertebra. The anterior vertical height of the bodies increases.
- **Stage 3 (Cvs 3):** a concavity develops in the inferior border of the third vertebra.
- **Stage 4 (Cvs 4):** a concavity develops in the inferior border of the fourth vertebra. Concavities in the lower borders of the fifth and of the sixth vertebrae are beginning to form. The bodies of all cervical vertebrae are rectangular in shape.
- **Stage 5 (Cvs 5):** concavities are well defined in the lower borders of the bodies of all 6 cervical vertebrae. The bodies are nearly square in shape and the spaces between the bodies are reduced.
- **Stage 6 (Cvs 6):** all concavities have deepened. The bodies are now higher than they are wide.

Franchi and his team's initial modification and validation of the CVM staging method, in 2000, retrospectively assessed longitudinal growth records of 24 subjects (15 males and 9 females) from the Michigan Elementary growth study. They concluded that the modified CVM index was able to detect the greatest increment in mandibular and craniofacial growth and that this corresponded to the interval between stage 3 and 4 in 100% of males and 87% of females. This was the period of maximum growth in the overall statural height, as well as the mandible. They also stated that if a cephalogram demonstrated either stage 1 or stage 2, it could be assumed that pubertal peak in growth has not commenced. Gu and McNamara's[86] study analysing retrospective records from the Mathews and Ware implant study, originally conducted at the University of California San Francisco in the 1970s, looked at longitudinal cephalometric records of 20 subjects (13 female, 7 male). They similarly found that the peak mandibular growth was noted during the interval from CS3 to CS4 and forward rotation of the mandible was due to greater mandibular growth posteriorly than anteriorly. (Figure 7). O'Reilly and Yanniello[32] also found similar results when they assessed annual cephalograms of 14 female patients.

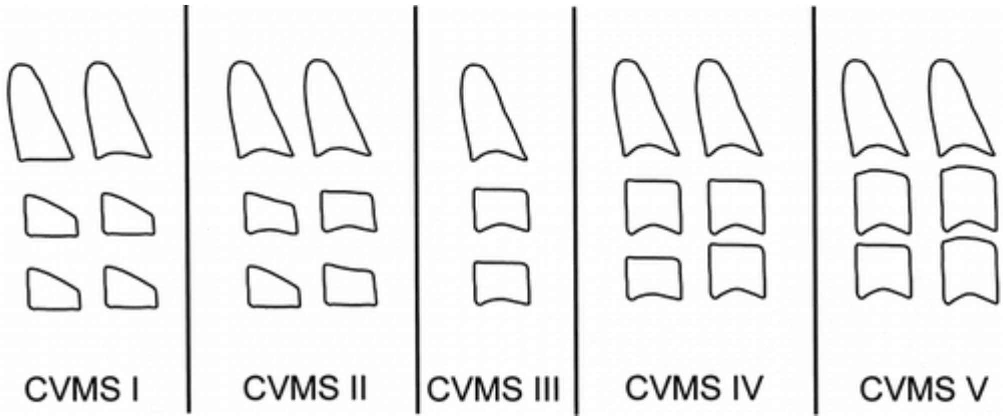
Figure 7. Diagrammatic summary of findings of Gu et al.,[86] demonstrating the amount of mandibular growth between each interval of cervical vertebrae maturation; CS1-6.



Franchi et al., reported 98% reproducibility with this method of cervical vertebrae maturation. However, this method is liable to the same criticisms as the previous work of Hassel, in that the test of reproducibility in both studies, was carried out using observers with increased familiarity with the method in comparison to the general professional body.

Franchi's team further modified their original method in 2002.[37] They followed Hassel and Farman by reducing the number of cervical vertebrae to be included in the staging process, from 5 to 3, C2-C4. They also modified the definitions of the stages, based on comparative assessment of between stage changes, to allow straightforward determination of the stage from a single cephalogram (Figure 8).

Figure 8 : The newly improved CVM Method (five developmental stages, CVMS I through CVMS V). Different combinations of morphological features in the bodies of C2, C3, and C4 are presented for the new method[37]



In this study, they reviewed longitudinal records of 30 subjects, 18 boys and 12 girls, from the Michigan growth study. For each subject the 2 consecutive lateral cephalograms demonstrating peak mandibular growth were used to represent the peak pubertal growth period. The 2 previous and 2 later cephalograms, to these radiographs, formed a total sample of 6 cephalograms for each subject. They carried out visual analysis and digitized tracings, of the morphological features of the vertebrae, C2-C4, on the consecutive radiographs and developed the following staging categories:

Table 1: Cervical Vertebrae Maturation stages[37]

Cervical Vertebral Maturation Stage (CVMS)[37]	FEATURES
I	The lower borders of all the three vertebrae are flat, with the possible exception of a concavity at the lower border of C2 in almost half of the cases. 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 not earlier than one year after this stage
II	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 within one year after this stage.
III	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 one or two years before this stage.
IV	The concavities at the lower borders of C2, C3, and C4 still are present. At least one of the bodies of C3 and C4 are squared in shape. If not squared, the body of the other cervical vertebra still is rectangular horizontal. The peak in mandibular growth has occurred not later than one year before this stage.
V	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 occurred not later than two years before this stage.

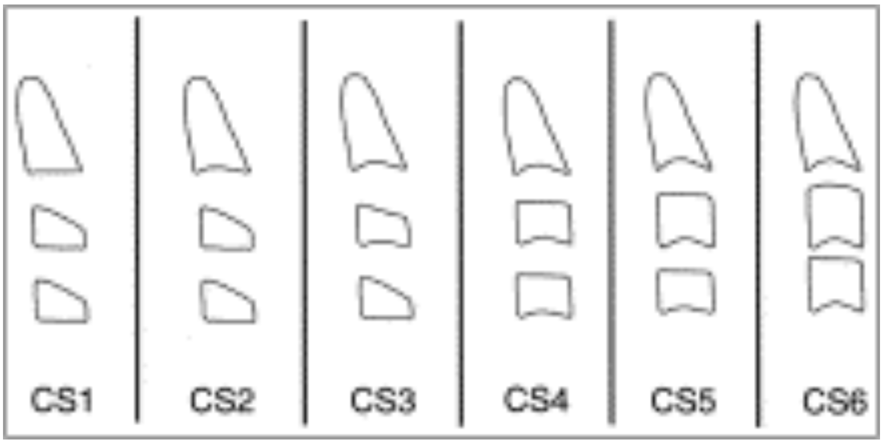
In 2005, the Franchi and Baccetti team again revised their CVM staging method to make it more valid for the appraisal of mandibular skeletal maturity in the individual patient.[38] They reverted back to a 6 stage maturation sequence and refined it to be more practical to apply. They used a description of each stage of maturation that did not rely on what the previous stage definition had been. The new clinically improved CVM method is comprised of six maturational stages (cervical stage 1 to cervical stage 6, i.e., CS1 to CS6). CS1 and CS2 were classified as pre-peak stages; with the peak in mandibular growth occurring between CS3 and CS4. CS6 was described as indicating at least 2 years after the peak. These stages are summarised in Table 2 and Figure 9.

Table 2: Cervical Vertebrae Stages[38]

Cervical Stage (CS) [38]	FEATURES
I	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
II	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
III	Concavities at the lower border of both C2 and C3 are present. The bodies of C3 and C4 may be either trapezoid or rectangular horizontal in shape.
IV	Concavities at the lower bordersof 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.
V	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.
VI	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

Figure 9: Cervical Vertebrae Maturation stages[38]



The validity of the maturational staging, using the cervical vertebrae, has been assessed by many authors, mainly through comparison with the hand-wrist method of skeletal maturation assessment. Mitani and Sato[36] reported that changes in the cervical vertebrae also correlated significantly with increases in mandibular size. Hassel and Farman[20] and Garcia-Fernandez et al.,[21] found a high correlation between cervical vertebrae maturation and the skeletal maturation of the hand-wrist.

Baccetti[37] demonstrated the validity of the method of cervical vertebral maturation for the evaluation of skeletal maturity and for the identification of the pubertal peak in craniofacial growth rate in individual subjects.

It is therefore assumed that cervical vertebral staging is as valid as the well accepted method of skeletal maturity assessment, the hand-wrist radiograph. The principal advantage of the cervical vertebral staging method is that the information is readily available from the lateral cephalogram, which is taken routinely for orthodontic purposes and therefore, does not necessitate additional radiation exposure, as is the case with the hand-wrist method of evaluating maturation. This makes this index very appealing to the orthodontic profession.

2.4 Evaluation of Cervical Vertebrae Maturational Index

The features of any successful diagnostic tool include validity and reliability. Ideally it must do what it is reported to in a quick, easy and reproducible way. The ideal features of a CVM staging method according to Baccetti et al.,[38] also include that it must detect the peak in mandibular growth, in a consistent manner, with inter-examiner error as low as possible. The available literature assessing the reliability of the CVM staging method is conflicting, with intra-observer and inter-observer correlation ranging from perfect agreement to poor agreement. The findings of the main studies addressing reliability of the various methods of CVM staging are summarised in Table 3.

Reliability of Cervical Vertebrae Maturational Index

Most studies have reported high reproducibility results for the various staging methods.[11,20,22,24,26,28,29,31,34,37,38,76-80] The majority of the studies quoting almost perfect correlation have used tracings of the vertebrae rather than the actual radiograph. This may introduce error either by the accidental random tracing error or systematic error, introduced by investigators attempting to define stages more easily. Some of the research studies also used the same observers to trace the cephalogram, again resulting in bias. Another common criticism of many of the reliability tests to date, is the role of the authors of the various CVM staging methods, as observers. The bias introduced by this may be significant, as these individuals will have knowledge, training and familiarity above the average clinician. Reliability tests using experts as observers may result in the reduction of the generalisability of the findings as these results would only be applicable to highly trained individuals. Other problems identified from reviewing the literature include small image sample sizes assessed and small numbers of observers, both reducing the generalisability of the results.

Gabreil et al,.[78] attempted to address these methodological concerns in his study. They used 10 orthodontic specialists as observers, with no known prior training and a large sample of 90 cephalograms. This study displayed the cephalograms to the observers in a cropped form, showing only the cervical vertebrae, and concluded poor reliability of the CVM method and criticised many of the previous reliability studies. The reasoning for this was that he felt it would minimise possible bias from other aspects of the cephalogram. Cropping of the cephalogram, to include only the cervical vertebrae, is an artificial manipulation of the clinical record that can influence results, as clinicians or researchers are likely to have the entire radiograph available in an everyday setting. Cropping reduces the resemblance of the test environment to the normal clinical situation, and is therefore felt to be an unnecessary step that could potentially influence the reliability and reproducibility of the method.

A recent systematic review of the CVM method highlighted the methodological flaws of previous research assessing the reliability of the index and encouraged more robust testing of the index to establish if it was a clinically applicable tool.[94]

			INTRA-OBSERVER RELIABILITY			INTER-OBSERVER RELIABILITY		
Reliability study	**CVM method assessed**	**Sample of cephalograms assessed for reliability**	**No. of observers used for intra-operator reliability**	**Experience of observers**	**Intra-observer reliability**	**No. of observers used for Inter-operator reliability**	**Experience of observers**	**Inter-observer reliability**
Hassel and Farman[20]	Hassel and Farman[20]	N=20 9 male 11 female Broadbent growth study	1	Named author	19/20 CVM stages coincided $r^2 = 1.00$	2	independent observers (A and B)	r^2=0.85, p<0.001 (for observer A and B)
Uysal et al[29]	Hassel and Farman[20]	N=30 Turkish	2	Named authors	0.955-0.987	2	Named authors	0.955-0.987
Ozer et al[28]	Kucukkeles et al[22]	N=150	2	Named authors	99.3%	2	Named authors	98%
San Roman et al[24]	Lamparski[8] Hassel and Farman[20]	N=50	Not stated	Not stated	0.96-0.99 (Pearson's correlation	-	-	-

			INTRA-OBSERVER RELIABILITY			INTER-OBSERVER RELIABILITY		
Reliability study	**CVM method assessed**	**Sample of cephalograms assessed for reliability**	**No. of observers used for intra-operator reliability**	**Experience of observers**	**Intra-observer reliability**	**No. of observers used for Inter-operator reliability**	**Experience of observers**	**Inter-observer reliability**
					coeff.)			
Flores-Mir et al[26]	Baccetti et al[37]	N=10	Not stated	Not stated	0.723-0.968 (intra-class correlation coefficient)	-	-	-
Kucukkles et al[22]	-	_	_	_	45-65%	-	-	-
Franchi et al[34]	Franchi et al, [34]	N=50	1	Named author	100%	2	Named authors	98.6%
Baccetti et al[37]								
Ballrick et al[87]	Baccetti at al[37]	N=15	13	Orthodontic residents	0.82 (weighted kappa)	-	-	-

			INTRA-OBSERVER RELIABILITY			INTER-OBSERVER RELIABILITY		
Reliability study	**CVM method assessed**	**Sample of cephalograms assessed for reliability**	**No. of observers used for intra-operator reliability**	**Experience of observers**	**Intra-observer reliability**	**No. of observers used for Inter-operator reliability**	**Experience of observers**	**Inter-observer reliability**
Alkhal et al. [88]	Baccetti et al[38]	N=25	1	Principal investigator	24/25 were the same	2	Not stated	0.846 ($p<0.001$)
Soegiharto et al. [31]	Baccetti et al[37]	N=300 (200 Indonesian +100 white	-	-	0.85-0.95 (Cohen kappa statistic)	-	-	-
Lai et al. [11]	Baccetti et al[38]	N=30	1	Not stated	90%	3	Not stated	90%
Gabreil et al.,[89]	Baccetti et al[38]	N=90 30+30 pairs	10	Independent from research	0.4-0.8 (Weighted kappa)	10	Independent from research	0.72-0.74 (Kendall's W)

			INTRA-OBSERVER RELIABILITY			INTER-OBSERVER RELIABILITY		
Reliability study	**CVM method assessed**	**Sample of cephalograms assessed for reliability**	**No. of observers used for intra-operator reliability**	**Experience of observers**	**Intra-observer reliability**	**No. of observers used for Inter-operator reliability**	**Experience of observers**	**Inter-observer reliability**
Jaqueira et al.,[90]	Baccetti et al[37]	N=23	1	Radiologist	No figure given but kappa statistic reported to be good (0.61-0.81)	4	1 radiologist 3 orthodontists	0.73 (weighted kappa)
Nestman et al.,[91]	Baccetti et al[38]	N=30	10	Independent from research	44-62%	10	Independent from research	0.45 (Kendall's W)

Table 3: Summary of main Reliability Studies

2.5 Reliability Statistical analysis

Reliability statistical analysis of intra and inter-observer agreement is carried out to determine the level of homogeneity, or consensus between observations. Intra-observer reliability, also referred to as repeatability, is a measure of agreement between observations, when the same observer assesses data, on two separate occasions. Inter-observer reliability is a measure of agreement, between two or more observers, assessing the same data.

Percentage agreement has been widely used as a measure of agreement,[11,22,34,88,91] but may be misleading, as it makes no allowance for the agreements that would occur by chance. Cohen's kappa statistic was introduced as a measure of agreement which avoids the problems of percentage agreement, by adjusting the observed proportional agreement, to take into account the amount of agreement that would be expected by chance.[100] Cohen's kappa is the proportion of agreement adjusted for that expected by chance with values ranging from -1 to +1. It is the amount by which the observed agreement exceeds that expected by chance alone, divided by the maximum that this difference could be. Negative kappa values represent agreement less than chance, values of zero represent exactly chance agreement, and positive results indicated better than chance agreement. The kappa statistic may be used for dichotomous data, such as the presence or absence of disease, or polychotomous data, such as the CVM index. The CVM index, as described by Baccetti et al.,[38] is a categorical, ordinally scaled index.

When the researcher can specify the relative seriousness of each kind of disagreement, they may employ the weighted kappa statistic.[102] The weighted kappa statistic allows for scoring of partial agreements between observations. Unweighted kappa statistic only scores agreement when there is exact agreement between observations. Unweighted kappa is unsuitable for ordinal data. The CVM index describes a continuous biological process of maturation, consequently disagreements may be seen in late subsequent early stages of each category and therefore the use of weighted kappa for assessment of reliability of the CVM index would be most appropriate, allowing credit for complete and partial agreement.[102] Weighting attaches greater emphasis to large differences than small differences. The magnitude of kappa is influenced by the choice of weighting and naturally the larger the number of categories the greater the potential for disagreement.[102] The two most commonly applied weights are linear and quadratic. Linear weightings are proportional to the number of categories apart, whereas quadratic weightings are proportional to the square of the number of categories apart. The weightings for a 6 category scale can be seen in Figure 12 .

Figure 12: a) Linear weightimgs and b) Quadratic weightings

a) Linear weightings

	1	2	3	4	5	6
1	1	0.8	0.6	0.4	0.2	0
2	0.8	1	0.8	0.6	0.4	0.2
3	0.6	0.8	1	0.8	0.6	0.4
4	0.4	0.6	0.8	1	0.8	0.6
5	0.2	0.4	0.6	0.8	1	0.8
6	0	0.2	0.4	0.6	0.8	1

b) Quadratic weights

	1	2	3	4	5	6
1	1	0.96	0.84	0.64	0.36	0
2	0.96	1	0.96	0.84	0.64	0.36
3	0.84	0.96	1	0.96	0.84	0.64
4	0.64	0.84	0.96	1	0.96	0.84
5	0.36	0.64	0.84	0.96	1	0.96
6	0	0.36	0.64	0.84	0.96	1

Landis and Koch[100] described the boundaries for the levels of agreement for Linear weighted kappa statistic . (Figure 13).

Figure 13. Linear weighted kappa statistic parameters.

Weighted kappa	Agreement
<0	Poor
0-0.2	Slight
0.21-0.4	Fair
0.41-0.6	Moderate
0.61-0.8	Substantial
0.81-1	Perfect

Quadratic weighting increase the kappa value with an increase in number of categories particularly from 2 to 5 categories.[103] This is desirable since as the number of categories increase, so does the proportion of variability in the true variable.[104] Quadratic weightings are reported to be used when most disagreements are only one category apart and is usually recommended because it is similar to Intra Class Correlation coefficient.[105] Linear weighted kappa varies much less with the number of categories.

Intra Class Correlation (ICC) was devised to deal with the relationship between variables within classes and has been extended to the comparison of observers, with the condition that the observers are regarded as a random sample of all possible observers, and hence part of the measurement error.[106] It is advocated in situations when observer data are continuous and parametric . The ICC was formulated to be applied to exchangeable measurements. In assessing the agreement among observers, if the same observers rate each component being studied, then systematic differences among observers are likely to exist, which can conflict with the principal of exchangeability. If the ICC is used in a situation where systematic differences exists, the result is an amalgamated measure of intra and inter-observer variability.[107] As a result of this amalgamation, it can be difficult to interpret, when observers are not exchangeable. In these cases the use of kappa statistics are advocated.[108]

Kendall's W, also referred to a Kendall's Coefficient of Concordance, has been used in some studies as a measure of agreement. It can be useful to assess trends between observers however, Kendall et al.,[109] originally described it as a method of assessing agreement between observers ranking the order of n subjects according to some quality. The example in Kendall's paper is the agreement of students ranking quality of pieces of poetry. Therefore, although Kendall's W has evolved as a measure of agreement between multiple observers, it is not how it was intended to be used. Failure to appreciate this can result in inappropriate interpretation of statistics.

Chapter 3: Rationale for research

The reliability of a clinicians' staging of a patients' CVM is an important consideration if a diagnostic index or scale is to be advocated for routine use to aid orthodontic diagnosis and treatment planning. The index must be reliable and sufficiently objective to give similar results for different observers and the same observers on different occassions.[93] Reliability can be defined as two different subtypes (1) agreement between ratings made by two or more clinicians (inter-rater/ inter-operator/ inter-observer reliability) or (2) agreement between ratings made by same clinician on two different occasions, (intra-rater/ -operator/ -observer reliability).

The reliability of the Cervical Vertebrae Maturation staging method suffers conflicting results in the orthodontic literature. Before the CVM staging method can be used routinely it must be shown not only to be valid but also to be reliable. The aim of this study is to address the methodology deficiencies of previous research, highlighted by Santiago et al.,[94] and determine the reliability of the improved method of CVM staging.[38]

Chapter 4: Preliminary audit

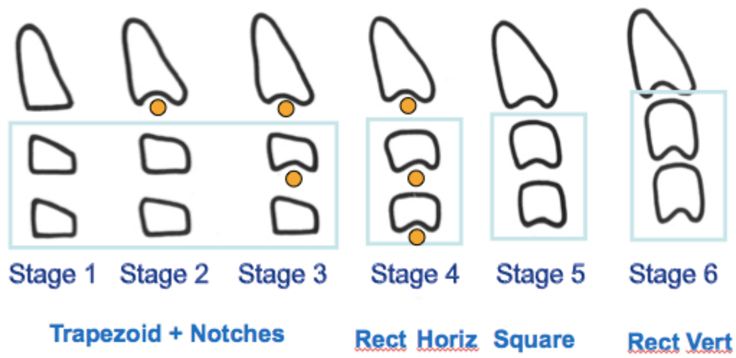
The suggested advantages of CVM staging, to assess growth status, for orthodontic patients are:[38]

1. The cervical vertebrae are routinely visible on a lateral cephalogram therefore no additional images are required.
2. The index is easy to use with good reported reliability.
3. The CVM index has a defined stage that corresponds to the onset of the pubertal growth phase.

4.1 Aim

Application of the CVM index relies on the lateral cephalogram clearly displaying the cervical vertebrae. The aim of this preliminary audit was to evaluate whether lateral cephalograms taken at Liverpool University Dental Hospital routinely included the second, third and fourth cervical vertebrae, allowing application of the Cervical Vertebrae Maturation (CVM) staging index (Figure 4.1).

Figure 4.1 CVM index[38]



4.2 Null Hypothesis

The cervical vertebrae bodies, C1-C4, are not visible on routine standardised lateral cepholgram, taken in the radiology department at Liverpool University Dental Hospital.

4.3 Design and setting:

This was a two phase audit cycle looking at all lateral cephalograms from two consecutive first year Specialty Registrar (StR) patient cohorts at Liverpool University Dental Hospital.

In the initial audit, all lateral cephalograms requested by first year Speciality Registrars (StRs) from the 1st October 2010 to 31st January 2011, were assessed retrospectively for the inclusion of vertebrae CV2,3,4 and the ability to stage the CVM (see Figure 4.2 for examples).

Figure 4.2 : Assessment Criteria: Lateral cephalogram clearly displaying:

A) CV2, CV3 and CV4 **B) CV2 and CV3** **C) Only CV2 .**

4.4 Standards

Ideally all lateral cephalograms should include the cervical vertebrae CV2-CV4[3] and the cervical vertebrae image should be a sufficient diagnostic quality. The standard for the first cycle of the audit was that 90% lateral cephalograms would have a clear representation of the cervical bodies CV2, CV3 and CV4.

Recommendations made following initial audit:

Discussions with the radiology department regarding correct positioning the patient in the cephalostat and training on how to position patient in the natural head posture were made.

After a 9 month wash-out period, the process was re-audited. In the re-audit all lateral cephalograms requested by first year StRs from 1st October 2011 to 31st January 2012 were assessed using same criteria

4.5 Results

Initial audit October 2010-January 2011

264 lateral cephalograms were assessed in the initial audit. All lateral cephalograms assessed clearly displayed CV2, 97% clearly displayed CV3 and 83% of lateral cephalograms displayed CV4. Therefore, overall 83% were suitable for CVM staging using the Baccetti method[38] (Table 4).

In the first audit the target, of 90% of lateral cephalograms having CV2, 3 and 4 displayed, was not achieved. The failure to attain the target appeared to be a result of the patient not being positioned correctly in the cephalostat, with the patient's neck often appearing hyperextended. As a result of this staff in the Radiology department were provided with appropriate training on positioning patients correctly in natural head posture when having a lateral cephalogram taken.[96] After a washout period, the second audit was commenced to assess if there had been an improvement.

Table 4: Results from initial audit (October 2010-January 2011) and re-audit (October 2011- January 2012

	No. of cephalo-grams analysed	No. with c2 visible	Proport-ion with c2 included	No. with c3 visible	Proport-ion with c3 included	No. with c4 visible	Proport-ion with c4 included	No. where cvm possible	Cvm staging possible
INITIAL AUDIT OCT 2010-JAN 2011	264	264	100%	256	97%	219	83%	219	83%

RE-AUDIT OCT 2011-JAN 2012	134	134	100%	134	100%	125	93%	125	93%

Re-audit October 2011- January 2012

In the second cycle, 134 lateral cephalograms were assessed according to the same criteria. Again, all lateral cephalograms in this sample clearly displayed CV2. The proportion clearly displaying CV3 increased to 100% and 93% clearly displayed CV4. Therefore CVM staging was possible in 93% of cephalograms assessed (Table 1).

Overall, the ability to stage lateral cephalograms using the Baccetti[38] method of CVM staging improved from 83%, in the initial audit, to 93% in the re-audit. This improvement was statistically significant [OR 2.72 (1.08, 6.89)].

4.6 Discussion

The initial audit found that 17% of the lateral cephalograms were unsuitable for skeletal maturation staging using cervical vertebrae. This is similar to a clinical trial that reported that it was not possible to stage 16% of lateral cephalograms using the Hassel and Farnam method,[20] as the radiographs were not clear in the cervical vertebrae region.[97]

The results of this audit show that CV2 was present on all cephalograms and it was CV3 and/or CV4 that were not visible. It was identified that this was a result of patient positioning. This happens when patients were positioned with their neck hyper-extended, rather than in a more vertical position, and not in natural head posture, so that the cervical bodies of CV4 and CV3 may be cropped from the radiographic field.

4.7 Conclusions

The target, that 90% of lateral cephalograms had CV2,3,4 clearly visible, was not met in the first cycle but was met in the second cycle. Significantly more lateral cephalograms were suitable for CVM staging in the second cycle which may help clinicians to plan treatment for their patients.

4.8 Recommendations

Failing to display CV2, 3, 4 clearly on a lateral cephalogram is a result of failing to position the patient correctly in the cephalostat. Further training and feed back will be given to staff in the Radiology department.

Chapter 5 : Study Aims and Objectives

5.1 Study Objectives

The primary outcome was to measure:

- The reliability and reproducibility of CVM stage determination amongst orthodontists in training and specialist orthodontists.

The secondary outcomes were to:

- Compare the agreement of specialist orthodontists with orthodontists in training.
- Assess whether the agreement of CVM stage determination, by orthodontists in training and specialist orthodontists, was different when staging an ideal sample of lateral cephalograms.
- Assess whether increased experience with the index improved the agreement between principal investigator (BJR) and Research Supervisor (JEH).
- Determine whether BJR and JEH agree with the experts and co-authors of the index (JMN/LF).

Chapter 6: Null hypothesis:

The null hypothesis was that the Cervical Vertebrae Maturation (CVM) stage determination was not a reproducible or reliable method of assessing the stage of cervical vertebrae maturation, when assessed by orthodontic specialists and orthodontists in training.

The maximum weighted kappa value less than 0.4 (moderate agreement) was established as the level at which the index was not clinically useful.

Chapter 7: Materials and Methods

7.1 Design

The study was conducted as a two phase reliability study, of the diagnostic test CVM staging.[29] The CVM index was applied by a group of orthodontic specialists and orthodontists in training, to a sample of consecutive and contemporary, digital lateral cephalogram radiographs.

7.2 Sponsorship and Ethical Approval

NRES East Midlands Research Ethics Committee (REC) approved the study protocol, initially on the 9th March 2012 and again, after amendment, on the 4th September 2012. The REC reference was 12/EM/0126. Informed consent was obtained from all clinicians who agreed to participate. The REC permitted access to patient radiographs without patient consent, providing images used were fully anonymised.

Ethical approval, was obtained from NRES Committee East Midlands- Northamptom (Appendix 1a and b).

Sponsorship was shared jointly between University of Liverpool and Royal Liverpool and Broadgreen University Hospital Trust. The Royal Liverpool and Broadgreen University Hospital Trust, number 4032 granted co-sponsorship in March 2012 followed immediately by the University of Liverpool, number UOL000840 (Appendices 2 and 3).

7.3 Setting

The study was carried out at the Mersey and North Wales orthodontic audit meetings. The rationale for choosing the Audit meetings as the forum for carrying out the study was that it was a convenient location where all orthodontic clinicians, consultants and orthodontists in training from the Mersey and North Wales region, were at attendance. The meetings were also held in a suitable venue that was well-equipped for presentations.

7.4 The Sample

7.4.1. Participating clinicians as observers.

All orthodontic specialists and orthodontists in training in the Mersey and North Wales region were invited to participate as observers in the reliability study at the preceding audit meeting. Clinician information sheets and consent forms were given to all possible participants (Appendices 3 and 4). A review of the minutes from previous audit meetings revealed that approximately 15-20 clinicians attended each meeting. The aim was to recruit a minimum of 15 clinicians as observers as this would be a larger sample than in any previous stiudies. 20 observers agreed to participate.

7.4.2 Lateral Cephalogram Image Sample

Sample size.

One criticism of previous literature was the use of small samples and the lack of a sample size calculation in the methodology.[94] The sample size in previous research ranged from 10 to 300 lateral cephalograms, often with little justification of sample size.[11,20,22,24,26,28-9,31,34,37,81,88-91] The factors that were considered in the sample size calculation were firstly that there needed to be a large enough sample to evaluate comprehensively the index with minimum bias, while keeping the sample a practical size for observers. The minimum number of cepholgrams in the image sample, that was required for the valid use of weighted kappa, was approximated by the sample size equation, $2k^2$, where k is the number of categories in the rating scale.[98] The rating scale used here has 6 categories, giving a minimum sample size of 72 radiographs.

Sample frame

The sample was taken consecutively from the first 72 lateral cephalograms fulfilling the inclusion/exclusion criteria. The decision to use consecutive images was influenced by convenience and time constraints. Liverpool University Dental Hospital moved to digital radiographs in September

2011 and therefore there was not a large sample of radiographs in digital format. Therefore, using consecutive images permitted expedient sample identification.

The sample of lateral cephalogram for this study was taken from non-surgical, orthodontic patients, being treated by Speciality Registrars (StRs) at Liverpool University Dental Hospital (LUDH). Patients who had had a lateral cephalograms taken were identified from first year StRs patient log books. Lateral cephalograms were selected in the consecutive order based on the date they had the radiograph was carried out. NRES permitted access to patient case notes to determine they satisfied the inclusion and exclusion criteria.

When a suitable lateral cephalogram was identified it was exported from the hospital networked radiograph programme to a fully anonymised image database. Each radiograph was given a unique numerical identification code, in the order it was exported, from 01 to 72.

The sample included only patients who had been assessed and deemed appropriate for treatment under the NHS at LUDH.

The sample of consecutive lateral cephalogram images was then presented in a *Powerpoint*™ presentation. Each radiograph extended to display all structures that are routinely visible on a lateral cephalogram,[99] in order to make the study environment as close to clinical practice as possible.

7.4.3 Radiographic exposure

All participants had undergone radiographic exposure in line with normal clinical practice. There was no additional exposure. The LUDH Consultant Radiologist, Mr Paul Nixon, confirmed the use of the radiographs for the purpose of the study.

7.5 Inclusion Criteria

Lateral cephaolgrams were included if they were of patients who:

- Were below the age of 18 years,
- Were of either gender,
- Had no previous orthodontic treatment,
- Were commencing treatment with first year StRs in the academic year 2010-2011,
- Had complete visualisation of cervical vertebrae CV2,3,4.

7.6 Exclusion criteria

Lateral cephalograms were excluded if they were of patients who:

- Were over the 18 years old at the start records appointment.
- Had previous orthodontic treatment.
- Had been diagnosed with any congenital clefts of the lip or palate, or known or suspected craniofacial syndromes or growth related conditions.
- Required orthognathic surgery.
- Radiograph was of unsuitable quality or did not display clearly the cervical vertebrae required.

7.7 Gold Standard sample

Validation of CVM Training

A supplemental sample of 11 'ideal' radiographs, that had previously had their CVM stage determined by Professor J McNamara, formed the "gold standard" sub-sample of final image sample. These radiographs were presented in a cropped format, extended to include the cervical vertebrae only. They were presented in this format as this was how the authors of the CVM staging method originally described using it. As previously mentioned, this format was not used for the main image sample, because it was thought not to be an accurate depiction of the environment in which the index would be used in clinical practice. The sub-sample, described by Professor McNamara, portrayed clearly the stages of CVM. The purpose of this supplemental, pre-staged sample, was to validate the training provided to the observers and allow comparison of the study sample with the 'gold standard' sample. This gave a total image sample of 83.

Validation of LUDH experts

The PI and supervisor involved in this study (BJR and JEH) staged the sample of radiographs independently under same conditions as clinician observers. Their agreement was assessed separately. Professor J McNamara, co-author of the index also independently staged the complete sample of radiographs. The PI and supervisor's ratings were then compared to the Professor McNamara's to assess whether BJR and JEH could be considered experts and if their determination of the stage of each lateral cephalogram was valid.

7.8 phase I

All clinicians who agreed to participate as observers in the reliability study partook in phase 1 in April 2012. The session commenced with a training presentation, given by BJR, on the CVM staging method described by Baccetti et al.,[38]

Professor McNamara, co-author of the index, provided the material for the training presentation. The format of the training presentation was:

1. Morphological changes the cervical bodies undergo during maturation were diagrammatically described.
2. Morphological changes the cervical bodies undergo during maturation were described radiographically.
3. Each CVM stage was then related to mandibular growth.

4. Observers were Instructed in a easy way to remember each stage, that was developed by McNamara and Baccetti.
5. A review of each stage radiographically again.
6. A short group calibration exercise to ensure all observers were comfortable and confident with the method,
7. Finally the opportunity was given for observers to clarify any queries.

This training exercise was designed to introduce the features of the index to the observers gradually, progressively building on their understanding, and then allow them to apply their knowledge in the calibration exercise.

Observers were provided with hard copies of reference material for consideration throughout the reliability study. (Appendices 6 and 7)

Immediately following the training, the observers were shown the image sample of lateral cephalograms. Random number tables determined the order of the image sample. The gold standard images were inserted into the image presentation at regular intervals. Random number tables also determined the order of these images. Each image was shown for 30 seconds, in which time each observer recorded the CVM stage, they felt best described the image, on score sheet. (Appendix 8). At the mid way point the observers were given a 10 minute break to avoid fatigue.

7.8 phase 2

The second phase of the study was carried out in July 2012. Observers were again trained in the same way and under the same conditions as phase 1. However, the images were displayed in a different random order from the first phase. Random number tables were used to determine the order.

7.10 Study summary

Approching the observer population

- All orthodontic clinicians in Mersey and North Wales regionwere invited to participate as observers in reliability study of CVM Index.

Recruiting radiograph sample

- 72 conseutive lateral cephalograms satisfying the inclusion and exclusion criteria, were identified.
- BJR and JEH completed 2 phase reliability study.
- CVM stage of each radiographs determined by BJR and JEH independently. Disagreements were resolved by discussion.

Presentation of image sampel

- Lateral cephlograms were anonoymised.
- Additional 11 radiographs, pre-staged by JMcN, were interspersed into presentation as Gold standard.
- Radiographs were displayed as a Powerpoint presentation, in random order.

Recruitment of observers

- Orthodontic trainees and specialist orthodontists who had been previously approached, confirmed or declined to partcipate.
- Those willing to participate given a unique identifcation code.

Phase 1 CVM Staging

- Observers trained in CVM staging method.
- Immediately afterwards they were shown images of lateral cephlograms.
- 30 seconds per image to record CVM stage.
- 10min rest at half way mark to avoid observer fatigue.

Phase 2 CVM Staging 3/12 later

- Observers retrained in CVM staging method.
- Immediately afterwards the same images were shown, in a different random order from first presenatation.
- Observers repeated staging under same conditions.

7.11 Statistics

Statistical support was sought from Dr G. Burnside, Liverpool University Dental Hospital.

Statistical Analysis

Intra-observer agreement was determined using percentage agreement and using the weighted kappa co-efficient, to calculate the chance corrected agreement.[100]

In this study, linear weighted coefficients were determined for intra-observer reliability and were calculated by manual construction of a 6x6 comparison tables (see example Table 7.1.) that comprised the intra-observer agreements for each of the 2 samples of radiographs and all clinicians. These data were then entered into the *Vassarstat®* program to calculate the linear weighted kappa.

Table 7.1. Example of 6x6 comparison table for intra-observer staging of the LUDH image sample.

		CVM STAGE AT PHASE I						TOTAL
		1	**2**	**3**	**4**	**5**	**6**	
CVM STAGE AT PHASE II	**1**	95	17	22	5	3	0	**142**
	2	21	100	36	13	5	0	**175**
	3	13	29	110	45	14	1	**212**
	4	2	6	39	206	60	7	**320**
	5	0	1	13	88	217	43	**362**
	6	0	0	2	8	62	157	**229**
TOTAL		**131**	**153**	**222**	**365**	**363**	**208**	**1440**

Inter-observer agreement was calculated using percentage agreement and weighted Fleiss' kappa coeffient using *Agreestat* ® statistic programme.

7.12 Direct Access to Source Data and Documents

The Investigator(s) will permit study-related monitoring, audits, REC review and regulatory inspections (where appropriate) by providing direct access to source data and other documents (ie patients' case sheets, X-ray reports etc).

Quality Assurance

This study was be monitored by the lead researcher to ensure compliance with Good Clinical Practice and scientific integrity was managed and retained by the Co-sponsors (Trust/University).

Data Handling

The Chief Investigator, JEH, acted as custodian for the study data. The following guidelines were strictly adhered to:

- Patient data were anonymised.
- All anonymised data were stored on a password protected computer
- All study data were stored and archived in line with the Medicines for Human Use (Clinical Trials) Amended Regulations 2006 as defined in the Joint Clinical Trials Office Archiving SOP.

Chapter 8: Results

8.1 General Characteristics

The intra-observer and inter-observer agreements for the application of the CVM staging index[38] was explored using a two phase reliability study. Twenty orthodontic clinicians staged a sample of 83 lateral cephalograms presented in two different random orders. The cephalogram sample was made up of 72 consecutive lateral cephalograms taken at Liverpool University Dental Hosital (LUDH), and 11 'ideal' pre-staged images, clearly depicting various CVM stages, which were provided by Professor J McNamara.

- The overall intra- and inter-observer reliability for the LUDH sample was 'substantial', (weighted kappa 0.6-0.8).
- The overall intra-observer weighted kappa statistic was 0.70 (SE 0.01) with percentage agreement of 89%.
- The inter-observer linear weighted kappa statistic on the first occasion was 0.68 (SE 0.03) and 0.66 (SE 0.03) on the second occasion, with an average inter-observer agreement of 88%.

8.2 Intra-observer Reliability of CVM index

The intra-observer agreement, between all 20 clinicians for the consecutive image sample from LUDH (Table 8.1.), was 'substantial', with a κ_W of 0.70, (SE 0.01). The clinicians' intra-observer agreement was 'almost perfect' agreement, κ_W 0.82 (SE 0.02), for the "ideal images" sample. The lower limit of the 95% intra-observer agreement for the ideal image did breach the 'substantial' agreement category, however it was still a significantly greater kappa value than that for the LUDH image sample.

TABLE 8.1: Weighted Intra-observer reliability

	Linear weighting				
INTRA-OBSERVER AGREEMENT		% agree	K_w	S.E.	95%CI
ALL OBSERVERS N=20	LUDH SAMPLE (72)	89	0.70	0.01	0.67-0.73
	IDEAL SAMPLE (11)	95	0.82	0.02	0.77-0.86

TABLE 8.1.2: 6x6 comparison table for intra-observer staging of the LUDH image sample.

	CVM STAGE AT PHASE I							TOTAL
		1	**2**	**3**	**4**	**5**	**6**	
CVM STAGE AT PHASE II	**1**	95	17	22	5	3	0	**142**
	2	21	100	36	13	5	0	**175**
	3	13	29	110	45	14	1	**212**
	4	2	6	39	206	60	7	**320**
	5	0	1	13	88	217	43	**362**
	6	0	0	2	8	62	157	**229**
TOTAL		**131**	**153**	**222**	**365**	**363**	**208**	**1440**

TABLE 8.1.3: 6x6 comparison table for intra-observer staging of the Ideal image sample.

	CVM STAGE AT PHASE I							TOTAL
CVM STAGE AT PHASE II		**1**	**2**	**3**	**4**	**5**	**6**	
	1	20	0	0	0	0	0	**20**
	2	0	38	1	1	1	0	**41**
	3	1	5	29	5	0	0	**40**
	4	0	1	3	14	11	0	**29**
	5	0	0	1	13	54	10	**78**
	6	0	0	0	0	8	4	**12**
TOTAL		**21**	**44**	**34**	**33**	**74**	**14**	**220**

Table 8.1.2 and 8.1.3 demonstrates the frequency of each intra-observer agreement, with the blue cells representing the absolute agreement. When staging the LUDH sample this population of observers achieved absolute agreement in 61.4% of observations and off the disagreements a further 30.0% were only one stage apart. When staging the Ideal image sample, 72.3% of agreements were absolute, and of the disagreements a further 25.4% were only one stage apart.

8.3 Inter-observer reliability of CVM index

The inter-observer agreement between all clinicians, Fleiss' weighted kappa, was calculated for the each of the 2 separate phases of the reliability study, using the raw data and the *Agreestat® for Mac* statistical program.

8.3.1 Phase I

On the first occasion the agreement between clinicians, when they staged the LUDH image sample, was 88% and classified as ‘substantial’, with weighted Fleiss’ kappa K_w 0.68 (SE 0.03). The agreement between observers looking at the ideal image sample, was 93% , K_w 0.83 (SE 0.03). (Table 8.2) This implies ‘almost perfect agreement’ between observers. However, the lower limit of the 95% confidence interval did encroach into the ‘substantial’ agreement classification.

This difference in inter-observer reliability was a statistically significant when looking at the ideal images, compared to the LUDH images.

Table 8.2 .Weighted inter-observer agreement between all observers for phase I of study.

phase I	**Linear weighting**				
INTER-OBSERVER AGREEMENT		% agree	K_w	S.E.	95%CI
ALL OBSERVERS N=20	LUDH SAMPLE (72)	88	0.68	0.03	0.62-0.74
	IDEAL SAMPLE (11)	93	0.83	0.03	0.76-0.90

8.3.2. Phase II

The results of the agreement on phase II of the reliability study were reassuringly very similar to phase I. (Table 8.3) The agreement when clinicians staged the LUDH sample was 88%, K_w 0.66 (SE 0.03).

The agreement when the ideal image sample was staged was 92%, K_w 0.83 (SE 0.04). The comparison of the kappa values for agreement between observers looking at the LUDH and ideal sample again demonstrates a higher level of agreement between observers staging the ideal sample, which is statistically significant.

Table 8.3. Weighted inter-observer agreement between all observers on phase II of study.

phase II	**Linear weighting**				
INTER-OBSERVER AGREEMENT		% agree	K_W	S.E.	95%CI
ALL OBSERVERS N=20	LUDH SAMPLE (72)	88	0.66	0.03	0.61-0.72
	IDEAL SAMPLE (11)	92	0.83	0.04	0.75-0.91

8.4 Comparison of agreement between orthodontists of different grades.

Of the 20 clinicians who acted as observers in this reliability study, 9 were consultant orthodontists, and 11 were trainee orthodontists. The level of agreement between these two groups was looked at separately to evaluate any differences.

8.4.1 Intra- observer agreement between orthodontists of different grades

Table 8.4 summarizes the agreement within the consultant orthodontist and the orthodontists in training programme groups. The intra-observer agreement within the consultant orthodontist group for the LUDH sample was 88% agreement, $\mathbf{K_w}$ 0.69 (SE 0.02). This was a 'substantial agreement' with the 95% confidence limits contained within the margins for this level of agreement. The agreement for the ideal image sample was 'almost perfect ' with 94% agreement, $\mathbf{K_w}$ 0.82 (SE 0.03), although the lower 95% confidence limit did lie in the 'substantial' agreement interval. The difference in agreement between the two image samples, for consultant orthodontists, was significant.

The intra-observer reliability for the trainee orthodontists group were 'substantial' agreement for the LUDH sample and 'almost perfect agreement ' for the ideal image sample, $\mathbf{K_w}$ 0.73 (SE0.01) and $\mathbf{K_w}$ 0.81 (SE0.03) respectively. It was not possible to determine whether this was a statistically significant greater kappa score.

Comparing the two groups' intra–observer reliability, the trainee orthodontists had better kappa scores for the LUDH image sample, than the consultant orthodontists. However, the converse was true for the ideal image sample. It is not possible to conclude if either of these differences were statistically significant.

Table 8.4. Weighted intra-observer agreement of consultant orthodontists and trainee orthodontists

		Linear weighting			
INTRA-OBSERVER AGREEMENT		% agree	**K_w**	S.E.	95%CI
CONSULTANT ORTHODONTISTS N=9	IMAGE SAMPLE (72)	88	0.69	0.02	0.66-0.73
	IDEAL SAMPLE (11)	94	0.82	0.03	0.75-0.88
TRAINEE ORTHODONTISTS N=11	IMAGE SAMPLE (72)	91	0.73	0.01	0.70-0.76
	IDEAL SAMPLE (11)	94	0.81	0.03	0.75-0.87

8.4.2. Inter-observer reliability between orthodontists of different grades

Inter-observer agreement within the consultant orthodontists and trainee orthodontists groups was evaluated by comparing the results from phase I and phase II of the reliability study. (Table 8.5)

In phase I of the reliability study the consultants' inter-observer agreement was 88%, **K_w** 0.66 (SE 0.03) for the LUDH image sample, and improved to 92% **K_w** 0.81 (SE 0.04) when staging the ideal image sample, which means 'substantial' and 'almost perfect' agreement respectively. (Table 8.5a) The agreement was better for the ideal image sample, however, as the confidences intervals overall, this was not significant.

The inter-observer agreement within the orthodontists in training group for phase I was 'substantial' for the LUDH image sample and 'almost perfect agreement' for the ideal image sample group, with agreement of 89%, **Kw** 0.70 (SE 0.03) and 92%, **Kw** 0.81 (SE 0.04) respectively. These results are not statistically significantly different between the groups, for either the ideal image sample or the LUDH image sample, during phase I of the study.

Agreement within the consultant orthodontists' group and trainee orthodontists' group was also evaluated in phase II. (table 8.5b) Overall the results followed a comparable trend to the first session. At the second session of the reliability study, the consultants' inter-observer agreement was 87%, K_w 0.61 (SE 0.03) for the LUDH image sample, and improved to 94% K_w 0.85 (SE 0.03) when staging the ideal image sample. The agreement for these samples was 'substantial' for the LUDH image sample and 'almost perfect' for the ideal image sample. This difference between the image samples was statistically significant for the consultants' inter-observer agreement.

Inter-observer agreement within the trainee orthodontists' group for phase 1 was 'substantial' for the LUDH image sample and 'almost perfect' agreement for the ideal image sample group, 89%, **Kw** 0.70 (SE 0.03) and 92%, **Kw** 0.81 (SE 0.04) respectively. The agreement was greater when staging the LUDH sample compared to the ideal image sample, although it is not statistically significant.

In phase II the inter-observer reliability for the LUDH sample was 90%, **Kw** 0.71 (SE 0.03). The agreement between the orthodontists in training programme improved when staging the ideal image sample to 93% **Kw** 0.85 (SE 0.03).

Table 8.5 Weighted Inter-observer agreement of consultant orthodontists and trainee ortodontists at:

a) Phase I

INTER-OBSERVER AGREEMENT PHASE I	Linear weighting				
		% agree	K_w	S.E.	95%CI
CONSULTANT ORTHODONTISTS N=9	LUDH SAMPLE (72)	88	0.66	0.03	0.60-0.75
	IDEAL SAMPLE (11)	92	0.81	0.04	0.72-0.91
TRAINEE ORTHODONTISTS N=11	LUDH SAMPLE (72)	89	0.70	0.03	0.64-0.76
	IDEAL SAMPLE (11)	92	0.81	0.04	0.72-0.90

b)Phase II

INTER-OBSERVER AGREEMENT PHASE II	Linear weighting				
		% agree	K_w	S.E.	95%CI
CONSULTANT ORTHODONTISTS N=9	IMAGE SAMPLE (72)	87	0.61	0.03	0.55-0.68
	IDEAL SAMPLE (11)	94	0.85	0.03	0.78-0.92
TRAINEE ORTHODONTISTS N=11	IMAGE SAMPLE (72)	90	0.71	0.03	0.65-0.77
	IDEAL SAMPLE (11)	93	0.85	0.03	0.77-0.92

8.5 Validation of the McNamara CVM teaching method

The ideal image sample was provided by the co-author of the index, Dr J McNamara. This image sample had been staged by Dr McNamara and was described as "clearly representing clear CVM stages". Assessing the agreement between all observers, when they applied the index to the ideal image sample of 11 standardized images, was used to validate the CVM teaching method described by McNamara.[38] The agreement with this ideal sample was overall 'almost perfect'. (Table 8.6). The overall intra-observer agreement was 95%, K_w 0.82 (SE 0.02). The inter-observer agreement on the first and second occasions was 93%, K_w 0.83 (SE 0.03) and 92%, K_w 0.83 (SE 0.04) respectively. The high level of agreement in staging ideal image sample suggests that observers understood the teaching method and were able to apply it in the method it was intended, to the ideal sample.

TABLE 8.6 Intra- and Inter- observer agreement using standardized ideal image sample.

OVERALL OBSERVER AGREEMENT (N=20)	Linear weighting				
		% agree	K_w	S.E.	95%CI
INTRA-OBSERVER	IDEAL SAMPLE (11)	95	0.82	0.02	0.77-0.86
INTER-OBSERVER PHASE I	IDEAL SAMPLE (11)	93	0.83	0.03	0.76-0.90
INTER-OBSERVER PHASE II	IDEAL SAMPLE (11)	92	0.83	0.03	0.76-0.91

8.6 Does increased experience with the use of the CVM index improve agreement?

8.6.1. Intra-observer agreement of researchers

The agreement of the PI (BJR) and CI (JEH) was evaluated separately, to determine if they had a different level of agreement from the overall group of observers.

The intra-observer agreement of JEH and BJR, looking at both the LUDH image sample, was 'almost perfect' with K_w 0.83 (SE 0.04) and 0.88 (se 0.03), respectively. (Table 8.7).

Evaluating the intra-observer agreement, for the LUDH image sample, BJR had intra-observer agreement of 96%, K_w 0.88 SE(0.03) and JEH had agreement 90%, K_w0.83SE (0.04). Both of these results suggest 'almost perfect agreement' although JEH's lower 95% confidence limit did fall in the 'substantial agreement' category. BJR's intra-observer agreement was significantly higher than JEH's.

The intra-observer agreement when staging the ideal image sample was for BJR was 100%, K_w1, and for JEH 98%, K_w0.94SE (0.06). It was not possible to determine whether this was statistically significant.

TABLE 8.7: Intra-observer reliability of experienced researchers compared to results of clinicians with no previous experience.

		Linear weighting			
		% agree	K_w	S.E.	95% CI
JEH	LUDH SAMPLE (72)	90	0.83	0.04	0.74-0.92
	IDEAL SAMPLE (11)	98	0.94	0.06	0.81-1
BJR	LUDH SAMPLE (72)	96	0.88	0.03	0.81-0.95
	IDEAL SAMPLE (11)	100	1	-	-
ALL OTHER CLINICIANS N=20	LUDH SAMPLE (72)	89	0.72	0.01	0.67-0.73
	IDEAL SAMPLE (11)	95	0.82	0.02	0.77-0.86

8.6.2. Intra-observer agreement of researchers compared to clinicians with no previous experience

Comparison of the intra-observer agreement of BJR, JEH and all other clinicians, for the LUDH image sample was K_w 0.88 (SE0.03), 0.83 (SE 0.04) and 0.72 (SE 0.01), respectively. BJR and JEH both had statistically significant greater weighted kappa than all other clinicians.

Comparing BJR and JEH's intra–observer agreement, to the intra-observer agreement between all other clinicians, for the ideal image sample, BJR had intra-observer agreement of 100%, $K_w 1$, JEH had agreed 98%, $K_w 0.94$SE (0.06) and the other observers had agreement 95%, K_w 0.82 (SE 0.02).

In summary, BJR's intra-observer agreement was greater than JEH's which is in turn was greater than the clinicians with no previous experience. BJR and JEH's intra-observer agreement for the LUDH image sample, was greater than the other clinicians. This difference was statistically significant. These findings correspond, approximately, to the level of experience. This confirms that the greater the experience with the CVM index, the greater the intra-observer agreement for this sample.

8.6.3 Inter-observer agreement of researchers

The inter-observer agreement between JEH and BJR was compared over the 2 phases of the experiment (table 8.8). Their agreement was 'almost perfect' for the LUDH sample in both phase I and II, with 93%, K_w 0.80 (SE 0.05) and 93%, K_w 0.81 (SE 0.05), agreement respectively. The point kappa score was in the 'almost perfect' agreement category however, the lower limits of the 95% confidence intervals, for both phases, was located approximately midway into the 'substantial" agreement category.

The inter-observer agreement, between BJR and JEH for the ideal image sample, at phase I and II was 'almost perfect', with 99%, K_w0.94SE (0.06) and 100%, K_w1, agreement respectively.

There was an improvement in the agreement between BJR and JEH, from phase I to phase II, for both image samples, although this was not statistically significant.

8.6.4 Inter-observer agreement of researchers compared to clinicians with no previous experience in CVM staging.

Comparison of inter-observer agreement between JEH and BJR, compared to the overall inter-observer agreement between all other clinicians was evaluated. (Table 8.8)

When the inter-observer agreement of the researchers was compared to those with no previous experience, for the LUDH image sample, the kappa score was 'almost perfect' for the researchers and 'substantial' for the other observers for both phase I and II. The weighted kappa ranged from 0.66 to 0.81 with no evidence of a statistically significant difference between the groups.

When the inter-observer kappa for the ideal image sample was investigated the agreement for all groups was classified as 'almost perfect' agreement, ranging from 0.83 to 1. The researchers' inter-observer kappa was greater than the other observers, although the results do not confirm this as statistically significant in phase I. However the inter-observer agreement difference in phase II, for the ideal images, was statistically significant.

TABLE 8.8. Inter-observer reliability of experienced researchers compared to results of clinicians with no previous experience.

IMAGE SAMPLE	phase	Linear weighting			
		% agree	K_w	S.E.	95%CI
JEH + BJR LUDH IMAGE SAMPLE (72)	phase I	93	0.80	0.05	0.71-0.90
	phase II	93	0.81	0.05	0.72-0.90
ALL OBSERVERS LUDH IMAGE SAMPLE (72)	phase I	88	0.68	0.03	0.62-0.74
	phase II	88	0.66	0.03	0.61-0.72
JEH + BJR IDEAL IMAGE SAMPLE (11)	phase I	99	0.94	0.07	0.81-1
	phase II	100	1	-	-
ALL OBSERVERS IDEAL IMAGE SAMPLE (11)	phase I	93	0.83	0.03	0.76-0.90
	phase II	92	0.83	0.04	0.75-0.91

8.7 Validation of LUDH researchers, BJR and JEH, as experts in CVM staging

To establish the researchers as 'experts' in his method of CVM staging, the researchers' mutually agreed CVM stages for each radiograph in the LUDH sample was compared to the CVM stage determined by the authors of the index, Dr. J McNamara (JM) and Dr. L Franchi. (LF) The overall agreement between the 2 results was 'substantial' 93% K_w 0.78 (SE 0.03). The upper limit of the 95% confidence interval fell in the 'perfect agreement' category. This confirms the LUDH researchers had an acceptable level of agreement with the authors of the index to be awarded 'expert' status.

The authors of the index also provided written commentary on each of the images in the sample. Their thoughts will be explored further in the discussion chapter.

TABLE 8.9: Inter–observer reliability between LUDH (JEH and BJR) and Michigan University Experts (JM, LF)

	Linear weighting			
	% agree	K_w	S.E.	95%CI
INTER-OBSERVER AGREEMENT	93	0.78	0.03	0.70-0.84

Chapter 9. Discussion

Orthodontists often take into consideration the skeletal growth potential of an individual patient when formulating a treatment plan. This may influence their decision on the timing at which to start treatment, and indeed even the method of treatment they choose.

Peak growth velocity in standing height has been shown to be one of the most accurate methods of determining the overall skeletal growth rate, however, this is of little benefit to the orthodontist. It is only on retrospective analysis of the growth velocity that the peak velocity is identifiable and therefore it would be difficult for the orthodontist to distinguish and is of little assistance when it comes to assessing the remaining growth potential. It is debatable whether this growth analysis, which is largely representative of long bone growth, is comparable to mandibular growth.

Assessment of the morphology of the cervical vertebrae, using various indices, has been proposed as a method of identifying the timing of the onset of the pubertal peak in skeletal growth, and estimates the proportion of growth remaining.

The Cervical Vertebrae Maturation (CVM) index described by Baccetti et al.,[38] follows the morphological changes in the cervical vertebrae throughout growth and relates this to mandibular growth potential. Many studies have assessed alternative methods of CVM staging, including the Lamparski[8] and the Hassel and Farman[20] methods, and found very positive reliability.[20,24,28,29] These studies will not be compared to the results of the Liverpool study because they were testing a different index. The same principal applies to reliability tests of the earlier version of the CVM index used in this study.[34,37]

Intra-observer and inter-observer agreement between orthodontic clinicians, with no previous experience of CVM staging, was found overall to be 'substantial' (weighted kappa 0.6-0.8) when applying this CVM index.[38] No statistically significant difference in reliability was found when the group of observers was compared according to their hospital grade.

When the results of the observers that had "research level" experience of the CVM index were analysed the intra-observer agreement significantly improved to 'almost perfect' agreement. The inter–observer reliability for this group also improved but it was not statistically significant.

9.1 Limitations of this study

Image sample identification

The LUDH image sample was formed from consecutive lateral cephalograms prescribed by first year registrars from October 2011 until the required number was attained. This method of sample collection was decided upon because it was the quickest way to identify the large sample of lateral cephalograms that would satisfy the inclusion/ exclusion criteria, in the most convenient time frame. Patients treated by new registrars are largely peri-pubertal and therefore cephalograms would display the various stage of cervical maturation. The second reason for choosing consecutive images was that the radiology department LUDH had recently transferred from plain film cephalograms to digital images. This meant that there was not a large bank of current lateral cephalograms in the digital format from which to select such a large sample of radiographs. Digital images were chosen on the basis that they would be more readily transferred to the research presentation, without the need for the scanning, which would be necessary, if plain films had been used.

Random identification of suitable radiographs was recommended by Santiago's systematic review, as a means of improving the generalizability of the results of a reliability.[94] This recommendation was made to overcome the problem of authors pre-selecting ideal images for reliability studies, which would bias findings. However, the LUDH image sample was selected purely on the basis of satisfying the inclusion criteria, regardless of the image quality. This, along with the fact a sample size calculation was performed to determine accurately the number of images necessary to test the index, should have reduced the introduction of selection bias.

Presentation of the images

As stated in the protocol, the images were exported from the PACS radiography database and presented to observers in a random order in a *PowerPoint* presentation. Each image was displayed to include the whole lateral cephalogram and was shown for 30 seconds. The aim was to keep the image in a format that represented the clinical environment. This representation of the image was closer to the clinical environment than the previously described traced cervical vertebrae, however, is not completely comparable. With the advent and use of digital radiography analysis software, it is possible for clinicians to employ a wide variety of aids for image inspection. These include the ability to enhance images, by magnification or contrast,[111] and therefore possibly allow a more accurate stage determination. Another technique possible with image analysis software is the use of calibrated length measures. These would allow the clinician to assess the vertical and horizontal dimensions of the cervical vertebrae more accurately and assign a CVM stage with more confidence and accuracy.

Presentation environment

The study was carried out in a classroom style environment, with a large projector screen at the front, and observers positioned with a direct view of the screen. This environment was deemed the most suitable for providing the presentation to a large group of observers. Clinicians, however, would be more familiar with viewing radiographs in a more intimate environment such as directly on a computer screen. This may have influenced the results. A more suitable environment may be considered as a computer suite, where each observer had an individual monitor, on which the image presentation was shown.

Quality of training

The results of this reliability study may be attributed to the training given to the observers. As discussed previously the author who provided the training in this study had substantial agreement with the authors of the index. Reasons for this have been deliberated above. One way around this difference in interpretation of the index would have been for the authors of the index to provide more explicit classification of what to do when images are between stages, or to expand the categories to define late and early stages more explicitly.

Another limitation of this study is that the reliability is only achievable when training is provided in the same format as described in this study, and cannot be generalized to populations with no training or those who have merely read the scientific paper describing the stages.[38]

Statistical Analysis

The use of the linear weighted kappa statistic may be seen as a limiting factor. The weighting value assigned to each level of agreement is an arbitrary number. We used linear weighting for reasons discussed previously however, it we could be criticised, as we have in fact allowed poor levels of agreement to contribute to our overall kappa value. For example if a cephalogram was scored as a one on the first phase and a 6 on the second, the weighting would allow this poor level of agreement to contribute albeit in a small part to the kappa agreement. In hindsight an individualised weighting scheme may have been developed, only allowing near misses to contribute to the overall kappa.

9.2 Intra-observer Reliability

The overall intra-observer reliability, for the 20 clinicians judging the CVM stage of this sample of 72 LUDH lateral cephalograms, was 89%, with a $\mathbf{K_w}$ 0.70, (95% CI 0.67-0.76), (Table 8.2.1). This corresponds to 'substantial' or 'good' agreement depending on which scale is used to interpret the kappa statistic. This is an acceptable level of intra-observer agreement to recommend the use of this CVM index.

These results are, to some extent, in contrast to earlier studies that assessed the Baccetti improved method of cervical vertebrae staging.[38] There have been many studies reporting very high or almost perfect intra-observer reliability. Alkhal et al., reported 96% intra-observer agreement when they applied the index to a sample of 25 lateral cephalograms, 3 weeks apart.[88] The 25 radiographs were randomly selected from a much larger image sample of 400 Chinese subjects. No details of the randomization process were given or justification of why 25 had been chosen to assess intra-observer reliability. The principal investigator was the only observer tested for the intra-observer reliability. The results of this study therefore may be misleading on account of the observer having 'research level' experience and only staging a small sample. Having had only one observer, the results of this study cannot be generalized to application of the index by other clinicians. The authors also only report the percentage agreement and not the chance corrected kappa statistic, therefore the agreement may be inflated.

Lai et al., also reported intra-observer agreement of this index at 90%, again, using only 1 observer and 30 radiographs randomly selected from a sample of over 600 images.[11] The relative experience of the observer was not stated, and neither was why they decided to assess only 30 radiographs. This study opens itself up to the same criticism of the previous study, in that it reports only the intra-observer agreement for 1 person, therefore has little or no generalizability. The percentage agreement of Lai et al.,[11] does state similar results to this study, however no chance corrected kappa statistic was given to increase the impact of this finding. Obuchowsk et al.,[106] reported that a minimum of 10 observers was needed to be able to generalize results of a diagnostic test and that single observer studies give no opportunity to assess frequency of observer differences as observers often possess different cognitive, visual, and perceptual abilities, ie. be biased.

Neither Lai et al.,[11] or Akhal et al.,[88] specifically state the format of the image sample, however if they followed the instructions of the authors of the index they would have traced the cervical vertebrae prior to staging. This method has been criticized by other researchers[89] because of possible tracing error, and also as it is not how the vertebrae are routinely assessed in the clinical environment.

Therefore, in comparison to Lai[11] and Akhal[88] the results of this study are more generalizable and applicable to the clinical environment.

Other studies have reported lower intra-observer agreement using this index than found in this study. Gabreil et al.,[89] conducted a well-designed reliability study to assess the reliability of the CVM index,[38] using a random sample of 90 lateral cephalograms (30 individual and 30 pairs) and 10 (specialist practitioners) observers with no previous experience. No details of how the random sample was identified were given on the study. All images were displayed in the sample format and quality as digital radiographs. Intra-observer agreement is not given a point score in this paper but the author's state 62% exact intra-observer agreements and with the weighted kappa range from 0.4-0.8 for the 10 individual observers. This corresponds to a moderate-substantial agreement according to Cohen et al.,[101] however, the authors interpreted it as evidence of overall poor reproducibility of the CVM index. When staging the LUDH sample this population of observers achieved absolute agreement in 61.4% of observations and of the disagreements a further 30.0% were only one stage apart. This means 38.6 % of intra-observer scores were disagreements. Of these disagreements 79.0% were one stage apart, 17.2% were 2 stages apart and 3% were more than 2 stages apart, meaning 91% were at the most one stage apart.

Gabreil et al., reported 55% of intra-observer scores were disagreements. Of these disagreements 66.7% were 1 stage apart, 25.7% were 2 stages apart, 5.3% were 3 stages apart, 2.7% were 4 stages apart and 0% were 5 stages apart.[89] This means 81.7% of were, at the most one category apart. These results do not support the title of Gabreil's paper- "*Cervical Vertebrae maturation method: poor reproducibility*"[89]

Possible explanations for the greater variation in kappa found by Gabreil [89] compared to this study, are firstly the training the observers received and how each observer interpreted it, and also the difference in the demographics of the observers. All observers in this study work in hospital environments whereas in the Gabriel study the observers were specialist practitioners.

Jaqueira et al.,[90] reported an intra-observer reliability of the index as 'substantial'. This is similar to this study, however the authors do not report exactly how they collected image sample of 23 lateral cephaolgrams , but suggest they were included because they were of good quality. Three observers, with varying experience, assessed the traced cervical vertebrae of image sample and applied 3 different CVM indices. Of all methods assessed the intra-observer kappa statistic was greatest for the Baccetti method.[38] Although Jaquerira's study does agree with the reliability of the Liverpool study, the robustness of the results of the Liverpool study may be greater due to the greater number of images and observers, the use of digital images rather than traced images of the vertebrae, and the method by which the sample of radiographs was selected.

9.3 Impact of 'research experience' on intra-observer reliability

As discussed previously, studies that use the principal researchers as observers may be biased as observers are thought to have a greater understanding than the wider population. In this study the intra-observer reliabilities of the PI and research supervisor were 96%, $\mathbf{K_w}$ 0.88 (95% CI 0.81-0.95) and 90% $\mathbf{K_w}$ 0.83 (95% CI 0.741-0.92). Both of the researchers' intra-observer weighted kappa statistics were greater than the observers with no experience. This finding provides evidence to support the criticism of Gabreil et al.,[89] that studies using researchers as the observers are biased. This is not a surprising finding as it seems only natural that the more confident an observer is with any diagnostic tool, the less likely they are to hesitate or deliberate when applying it, a therefore the more likely they will agree with themselves.

This finding it not necessarily a negative finding, as it also suggests that the more experience an individual has with the index, the more accurate they may be when applying it.

9.4 Inter-Observer reliability

The inter-observer reliability between all clinicians was assessed over the two phases of this study and was reassuringly very similar on both occasions. The inter-observer reliability, for the LUDH sample on phase I was 88% $\mathbf{K_w}$ 0.68 (95% CI 0.62-0.74) and on phase II was 88% $\mathbf{K_w}$ 0.66 (95% CI 0.61-0.72). These finding both equate to a 'substantial' or 'very good' level of agreement. This result supports the recommendation for the use of the CVM staging method. Other studies assessing inter-observer reliability of this CVM index, report results of varying degrees of agreement with this finding. [11,88]

Alkhal et al.,[88] also evaluated inter-observer agreement between 2 observers, one of whom was the principal author and the other who was described as *"another orthodontist*". They reported inter-observer agreement of kappa 0.846 with 92% agreements between the observers. However, this sample was small and with no justification of sample size. This kappa score corresponds to 'excellent' agreement, however may be biased due to the researcher acting as one of the observers.

Lai et al., assessed inter-observer agreement of the CVM index between 3 observers on one occasion, looking at 30 radiographs.[11] They reported a 90% overall agreement. This is similar to the findings of this study however, they do not give a chance corrected kappa score or any details on the range of the agreement. There was also no comment on the experience of the 2 additional observers (one was the researcher).

Gabreil et al.,[89] looked at the inter-observer agreement between 10 observers looking at 90 radiographs. As mentioned previously, this study had a good methodology. They found inter-observer agreement, using Kendall W statistic, to be 0.74 on the first occasion and 0.72 at the 3-week follow-up study. The researchers classified this a 'moderate' level of inter-observer agreement, but did not give a measure of the range of results for the 10 observers. Gabriel used these findings to conclude that this method of CVM staging was poor, although this does not seem to be the case from analysis of the results in the paper. He bases his "poor reproducibility" on the failure of observers to consistently agree on the exact same stage. Kendall's *W* statistic is a non-parametric statistic and is used for assessing agreement among observers when they have to rank the order of a set of data.[107] This test would be appropriate if observers had to rank the cephalograms in order of their stage of maturation, therefore in the methodology described in Gabreil's study, calculation of Fleiss' weighted kappa for multiple observers would have been more appropriate.[108] This may be therefore interpreted as a inaccurate or misleading conclusion.

It is well recognized that growth and development is a gradual process, albeit interspersed with periods of increased activity, therefore it is intrinsic that some "between stages" will be difficult to differentiate and therefore may result in a variation of staging. In the Liverpool study, the researchers have tried to overcome this difficulty by asking observers to round up if they were unsure of which category to put the radiograph into. This may not necessarily be what a clinician would do in a clinical environment, as there may be other factors, which will influence their clinical judgment, but nonetheless it allowed observers with no experience of the index to classify those radiographs that may have been more difficult to stage.

Jaqueira[90] calculated inter-observer reliability for the same images described previously. They found weighted inter-observer kappa statistic of 0.73. They gave no measure of the error in this point kappa score. Their kappa statistic is within the same parameter of agreement as this study and even though it is higher, creditability of the results in this study may be more robust due to the methodology.

9.5 Impact of 'research experience' on inter-observer reliability

This study has found that the researchers' inter-observer kappa was greater than the other observers for the LUDH image sample. The inter-observer reliability for both phases I and II, for the researchers was in the ‘almost perfect’ agreement category although the lower 95%CI limit was in the ‘substantial’ agreement category. The other clinicians’ inter-observer agreement was in the ‘substantial’ agreement category. However, it was not possible to determine whether this difference was not statistically significant increase. Therefore, it is not possible to support Gabriel’s[89] theory that research level experience makes a significant impact on inter-observer reliability.

9.6 Impact of characteristics of observer population

It has been suggested that observers with researcher level experience are more likely to be consistent with their application of the CVM staging index and demonstrate higher inter- and intra agreement. [94]

The observers in this study were made up of clinicians with different levels of orthodontic experience. The intra- and inter-observer reliability was compared between the consultant orthodontist group and the trainee orthodontists’ group. No statistically significant difference was identified between these two groups therefore the results of this study are comparable to other studies using either population of observers. No previous studies have used either of these observer population exclusively when looking at the agreement using the Baccetti improved method of CVM staging.[38] Ballrick et al., [87] assessed the agreement of between 15 trainee orthodontists, looking at 15 images, 2 weeks apart. Ballrick reported intra–operator weighted kappa of 0.82 and an inter-observer agreement 0.84 (almost perfect agreement). This is greater than the intra- and inter-observer agreement found in this study. This is not surprising as the image sample was small and chosen from the Ohio University historical archive, based on their quality. This implies that bias from the image sample may have resulted in an exaggerated agreement score, even though they used a relatively large sample of observers.

9.7 Impact of characteristics of the image sample

The quality of the image sample may have impact on the observed reliability. Quality may be judged on a number of levels. Firstly, patient-positioning error in the lateral cephalostat is a common problem and structures, such as the lower border of the mandible, are often overlapped on routine lateral cephalograms. Patient positioning errors may have two main effects on the quality of the presentation of the cervical vertebrae. It may lead to the vertebrae not being displayed on the lateral cephalogram, making CVM staging impossible, or overlapping of the vertebrae, which may make CVM staging more difficult.

Along side the main aim of this study, the researchers assessed the lateral cephalogram images that were routinely taken in Liverpool University Dental Hospital, for inclusion of cervical vertebrae 2,3 and 4. (Chapter 4). The initial audit was undertaken in the planning stage of the main study to determine how many of the radiographs, taken for patients commencing treatment, were theoretically suitable for inclusion in the main study. It highlighted a problem that in 17% of lateral cephalograms, taken in the specified time frame, did not include the cervical vertebrae necessary for CVM staging and therefore would not be suitable for inclusion in the study. This result was communicated to the staff in the radiology department who were subsequently given the appropriate training. This resulted in a reduction of 10% in number of cephalograms not being suitable for staging, during the same time period in the following year (2011-2012). This prospective education of the staff in the radiology department was the only effort put in place to improve the image quality of the sample. This is in contrast to other studies. Jaqueira[90] reports that they only included lateral cephalograms where there was no overlapping of the cervical vertebrae. This would improve the image sample quality and therefore may increase the agreement between observers in comparison.

This study included all radiographs satisfying the inclusion and exclusion criteria, of which overlapping was not considered. This resulted in some images with overlapping of the vertebrae being included in the sample so, although this may make staging more difficult, it is more representative of the material an orthodontist may encounter in everyday practice.

Another way of quantifying the quality of the image sample may be in terms of how closely the images represent each of the six defined categories of the index. As discussed previously, growth does not follow six distinct categories, like the index, and therefore it is inevitable that some presentations of the cervical vertebrae will be more difficult to stage and others will be straightforward.

A plausible statement is that those cephalograms with distinct features of a specific category will be easier to stage. This theory is supported by the difference in intra- and inter-observer reliability seen between the ideal image sample and the LUDH image sample. The ideal image sample was comprised of 11 cervical vertebrae images that had been staged by JMcN. They were described as clearly depicting one of the six stages on the index. The inter-observer agreement was greater for the ideal image sample compared to the LUDH image sample in both phases of the study. However, this improvement was statistically significant at the first phase only. This confirms that selecting images in the consecutive order they were taken gave a more generalizable result than using an ideal, standardized image sample.

The results of Ballrick et al., illustrate this point further by demonstrating 'perfect agreement' between 15 orthodontic residents (trainee orthodontists), assessed an image sample selected on the basis of image quality.[87] However, it was not possible to determine what effect the image quality had and how it contributed to the 'prefect agreement' that the authors reported.

In summary studies, that show selection bias, when generating the images for the reliability study may overestimate the intra- and inter-observer reliability.

9.8 Observer training in application of the CVM method for assessment mandibular growth

The level of experience of the observers has previously been discussed as a factor that may influence the results of the reliability study, consequently it is plausible that the quality of the training given to the observers may also influence their ability to stage the image sample accurately.

The training of observers, in the understanding and application of the CVM staging method, was provided at the beginning of each phase of the reliability study by the principal researcher, BJR.

The format of the training began with a description of the learning objectives for the training session. Stating clear learning objectives has been advocated by Cohen et al., as an integral component of the acquisition of new knowledge.[108] If observers do not know what is expected of them, they may miss salient points. The learning outcomes of the training session were:

- To understand the CVM method for mandibular growth[38]
- To know the morphological features of each stage of the CVM method
- To understand McNamara's simple strategy to help remember features of each stage.
- To be able to apply the CVM method to lateral cephalograms.

The overall training was through a lecture type presentation. To maximize the effectiveness of the training the researcher structured the teaching to a satisfy the well-recognized learning model, Visual, Auditory and Kinesthetic (VAK).[109] This model of learning focused on the different ways in which people learn. Visual learners have a preference for seeing, they think in pictures; visual aids such as overhead slides, diagrams and handouts. Auditory learners best learn through listening to lectures or discussions, and kinesthetic learners prefer to learn via experience. In theory, if training or teaching includes all of these elements then the training should satisfy the learning styles of the group of observers.[110]

To introduce the observers to the CVM method, the presentation began with a diagrammatic overview of the six different stages. This was followed by a detailed stage-by-stage description of the index. The method of training was based on gradually increasing the amount of detail conveyed, through oral presentation and diagrammatic and radiographic examples, allowing the observers to familiarize themselves with the index. The final stage of the training was to a group exercise, in which the group applied the index and put into practice the knowledge they had acquired. This was to ensure that everyone understood the index prior to the reliability test. Following the training each observer was provided 2 hand-outs, one with an overall diagrammatic representation of the index, and the other a stage-by-stage description (Appendices 4 and 5). All observers were given the opportunity to ask questions or discuss any queries they had prior to commencing the reliability test.

It is important to recognize that just because knowledge or skills have been taught that it does not mean the knowledge and skills have been learned.

One way of quantifying the success of the teaching was to look at how accurately the observers staged the sample of ideal radiographs. As mentioned previously, this sample of 11 radiographs, provided by McNamara, were '*textbook*' examples corresponding to various stages of the CVM index. In theory, if the observers could stage this sample then they understood the index. The intra– and inter-observer reliability was 'almost perfect' for the ideal image sample. This was an encouraging result and showed that with ideal images, the observers were competent in applying the index.

Many studies, assessing the CVM method of assessing mandibular growth, have used experts or individuals with research level experience with the index who have not required any formal training. Other studies have included a training component, which they described was based on the instruction provided by Baccetti et al.,[38] paper.[89,91] The quality and complexity of the training provided in these studies, and the understanding by the observers, may have influenced the results of the study. A strength of this study was that the training provided was based on a presentation provided by McNamara, co-author of the index.

8.9 Liverpool experts agreement with CVM staging experts

An important aspect of this reliability study was establishing the Liverpool researchers as experts in the method of CVM staging for mandibular growth. This was valuable as the researchers provided the training to the observers therefore, it is paramount to ensure they agree with the authors of the index. It is also central to any further research by the Liverpool team, using the application of the CVM method of staging mandibular growth.

Following completion of the two phase reliability study, the researchers, BJR and JEH, independently staged the image sample and then brought their scores together to determine a definitive stage for each radiograph for the LUDH sample. McNamara and Franchi staged the sample and provided commentary on the sample, individually and overall.

The general commentary provided by McNamara's group was that presenting the images in the full lateral cephalogram format, instead of in cropped format, may be misleading and that the index should be applied to traced cervical vertebrae rather than the radiographic image. The experts also commented that some of the images were difficult to stage. In response to McNamara's points it was decided that the format should remain as the full lateral cephalogram, as this is how the index would be applied in the clinical situation. It was decided not to trace the vertebra as this could have introduced tracing error and again, would reduce the generalizability of the results for use in the clinical environment. Regardless of the level of difficulty in staging, described by McNamara, none of the images were substituted. By maintaining the consecutive image sample, this again increases the generalizability of the results.

The intra-observer agreement between the Liverpool researchers and the Michigan experts was found to be ‘substantial’ with a weighted kappa statistic of 0.78 (SE 0.04). This result indicated ‘substantial agreement’ but was surprisingly lower than expected, considering the higher level of agreement between the Liverpool experts. To explain the difference, each of the Michigan experts comments were compared to the Liverpool experts’ scores and overall two main differences were identified. The main difference was found between a numbers of radiographs that McNamara had staged as CVM Stage 4, but commented that they were late stage 4 in the commentary. The Liverpool researchers had staged these images as cervical stage 5. This difference in classification can be explained by how the Liverpool researchers have interpreted the index. BJR and JEH, recognized that some late stage/ very early stage examples were very similar. In order for observers to have clear idea of what to do with such images, they were advised if they were unsure which category to put the image into they should choose the one they felt it represented most described the image. Therefore, it transpired that what BJR and JEH felt was a stage 5, the experts felt was a late stage 4. This highlights the problem of using an index, which describes a process of continual morphogenesis in six distinct categories. The between stage changes are gradual and therefore naturally difficult to differentiate between.

9.10 Implications for practice

9.10.1 Clinical

The main implication for practice is that this research indicates the method of CVM staging was reliable, when applied by consultant and trainee orthodontists, who have had training in the use of the improved version for cervical vertebrae staging.[38] This may help clinicians to determine the most appropriate treatment timing for their individual patients based on the assessment of the morphology of their cervical vertebrae.

9.10.2 Research

The implication of this research for future orthodontic research, is that cervical vertebrae staging can be included reliably as a variable, in research that assesses treatment effects, or compares characteristics of different treatment groups.

This is particularly important in research that investigates optimal treatment timing for different treatment modalities or investigates treatment efficiency with respect to treatment timing.

Cervical vertebrae maturation is a reliable index that should also be included in any future growth studies.

This study has shown that the application of the index is reliable but further research is required to determine its validity with a contemporary series of cephalograms and children. The teaching method used in this study will also require further validation. This may be determined by comparing the staging ability of observers who have had training versus observers provided with only Baccetti et al., paper.[38]

9.11 Recommendation for future

- To increase further the reliability of the CVM index, digital analysis of the shape and dimensions of the cervical vertebrae for stage determination could be assessed. If this was achievable, software could be developed to determine the stage, which would remove the variability in application and interpretation of the index. This would remove the need for training and interpretation of the index.
- The next logical step is to validate the CVM index for mandibular growth in a contemporary sample. It would be of particular interest to identify how much mandibular and stature growth can be expected at a particular stage.

References

1. Fishman LS. Radiographic evaluation of skeletal maturation. A clinically oriented method based on hand-wrist films. Angle Orthod 1982;52:88-112.
2. Bergersen EO. The male adolescent facial growth spurt: Its prediction and relation to skeletal maturation. Angle Orthod 1972;42:319-338.
3. Mitani H, Sato K. Comparison of mandibular growth with other variables during puberty. Angle Orthod 1992;62:217-222.
4. Hunter CJ. The correlation of facial growth with body height and skeletal maturation at adolescence. Angle Orthod 1966;36:44-54.
5. Van der Beek MC, Hoeksma JB, Prahl-Andersen B. Vertical facial growth and statural growth in girls: A longitudinal comparison. Eur J Orthod 1996;18:549-555.
6. Bjork A, Helm S. Prediction of the age of maximum pubertal growth in body height. Angle Orthod 1967;37:134-143.
7. Hagg U, Taranger J. Menarche and voice change as indicators of the pubertal growth spurt. Acta Odontol Scand 1980;38:179-186.
8. Lamparski DG. Skeletal Age Assessment Utilizing Cervical Vertebrae. Department of Orthodontics. Pittsburgh, Penn: The University of Pittsburgh; 1972.
9. Houston WJ. Relationships between skeletal maturity estimated from hand-wrist radiographs and the timing of the adolescent growth spurt. Eur J Orthod 1980;2:81-93.

10. Hagg U, Taranger J. Skeletal stages of the hand and wrist as indicators of the pubertal growth spurt. Acta Odontol Scand 1980;38:187-200.

11. Lai E, Lui JP, Chang J, Tsai S, Yao C, Chen M, Lin C. Radiographic assessment of skeletal maturational stages for orthodontic patients: Hand-wrist bones or cervical vertebrae? J Formos Med Assoc 2008;107:316-325.

12. Zerin JM, Hernandez RJ. Approach to skeletal maturation. Hand Clin 1991;7:53-62.

13. Greulich WW, Pyle SI. Radiograph atlas as skeletal development of the hand and wrist. 2nd ed Stanford University Press 1959.

14. Tanner JM, Whitehouse RH, Healy MJR. A new system for estimating skeletal maturity from the hand and wirst, with standards derived from a study of 2600 healthy British children. Part II: The scoring system. Paris: International Children's Center;1962.

15. Milner GR, Levick RK, Kay R. Assessment of bone age: a comparison of the Greulich and Pyle, and the Tanner and Whitehouse methods. Clin Radio 1986;37:119-121.

16. Fishman L. Maturational Patterns and Prediction During Adolescence. Angle Orthod 1987;57:178-193.

17. Fishman L. Chronological Versus Skeletal Age, an Evaluation of Craniofacial Growth. Angle Orthod. 1979;49;181-189.

18. Chapman SM. Ossification of adductor sesamoid in adolescent growth spurt. Angle Orthod 1972;42:236-245.

19. Pileski RC, Woodside DG, James GA. Relationship of the ulnar sesamoid bone and maximum mandibular growth velocity. Angle Orthod 1973;43:162–170.

20. Hassel B, Farman AG. Skeletal maturation evaluation using cervical vertebrae. Am J Orthod Dentofacial Orthop 1995;107:58-66.

21. Garcia-Fernandez P, Torre H, Flores L, Rea J. The cervical vertebrae as maturational indicators. J Clin Orthod 1998;32:221-225.

22. Kucukkeles N, Acar A, Biren S, Arun T. Comparison between cervical vertebrae and hand-wrist maturation for assessment of skeletal maturity. J Clin Pediatr Dent 1999;24:47-52.
23. Chang HP, Liao CH, Yang YH, Chang HF, Chen KC. Correlation of cervical vertebrae maturation with hand-wrist maturation in children. Kaohsiung J Med Sci 2001;17:29-35.
24. San Roman P, Palma JC, Oteo MD, Nevado E. Skeletal maturation determined by cervical vertebrae development. Eur J Orthod 2002;24:303-11.
25. Grave K, Townsend G. Hand-wrist and cervical vertebral maturation indicators: how can these events be used to time Class II treatment? Aust Orthod J 2003;19:33-45.
26. Flores-Mir C, Burgess CA, Champney M, Jensen RJ, Pitcher MR, Major PW. Correlation of skeletal maturation stages determined by cervical vertebrae and hand-wrist evaluations. Angle Orthod 2006;76:1-5.
27. Gandini P, Mancini M, Andreani F. A comparison of hand-wrist bone and cervical vertebral analyses in measuring skeletal maturation. Angle Orthod 2006;76:984-989.
28. Ozer T, Kama JD, Ozer SY. A practical method for determining pubertal growth spurt. Am J Orthod Dentofacial Orthop 2006;130:131.e1-6.
29. Uysal T, Ramoglu SI, Basciftci FA, Sari Z. Chronologic age and skeletal maturation of the cervical vertebrae and hand-wrist: Is there a relationship? Am J Orthod Dentofacial Orthop 2006;130:622-8.
30. Chen LL, Xu TM, Jiang JH, Zhang XZ, Lin JX. Quantitative cervical vertebral maturation assessment in adolescents with normalocclusion: a mixed longitudinal study. Am J Orthod Dentofacial Orthop 2008;134:720.e1-7.
31. Soegiharto B, Cunningham S, Moles D. Skeletal maturation in Indonesian and white children assessed with hand-wrist and cervical vertebrae methods. Am J Orthod Dentofacial Orthop 2008;134:217-26.
32. O'Reilly M, Yanniello GJ. Mandibular growth changes and maturation of cervical vertebrae: a longitudinal cephalometric study. Angle Orthod 1988;58:179-84.

33. Huggare J. The cervical vertebrae as an indicator of mandibular growth. Eur J Orthod 1989;11:10-16.
34. Franchi L, Baccetti T, McNamara J. Mandibular growth as related to cervical vertebral maturation and body height. Am J Orthod Dentofacial Orthop 2000;118:335-340.
35. Franchi L, Baccetti T, McNamara JA Jr. Thin-plate spline analysis of mandibular growth. Angle Orthod 2001;71:83-89.
36. Mito T, Sato K, Mitani H. Predicting mandibular growth potential with cervical vertebral bone age. Am J Orthod Dentofacial Orthop 2003;124:173-177.
37. Baccetti T, Franchi L, McNamara J. An improved version of the cervical vertebral maturation (CVM) method for the assessment of mandibular growth. Angle Orthod 2002;72:316-323.
38. Baccetti T, Franchi L, McNamara J. The cervical vertebral maturation method for the assessment of optimal treatment timing in dentofacial orthopedics. Semin Orthod 2005;11:119-129.
39. British Orthodontic Society, 2011. Orthodontics and you. (online) Available at: <http://www.bos.orthodonticssandyou/information> [Accessed 4th June 2011].
40. Bishara SE. Facial and dental changes in adolescents and their clinical implications. Angle Orthod 2000;70:471-484.
41. Mantani H, Sato K. Comparison of mandibular growth with other variables during puberty. Angle Orthod 1993;62:217-222.
42. Sadowsky PL. Craniofacial growth and the timing of treatment. Am J Orthod Dentofacial Orthop 1998;113:19-23.
43. Coburne MT, DiBiase AT. Handbook of Orthodontics 2010. Mosby: Elsevier.
44. Nanda RS. The contributions of craniofacial growth to clinical orthodontics. Am J Orthod Dentofacial Orthop 2000;117:535-555.
45. Demirjian A, Buschang PH, Tanguay R. Interrelationships among measures of somatic, skeletal dental and sexual maturity. Am J Orthod Dentofacial Orthop 1985;88:433-438.

46. Taranger J. Evaluation of biological maturation by means of maturity criteria. Acta Paediatrica Scandinavia supplement 1976;258:77-82.

47. Fishman LS. Chronological versus skeletal age, an evaluation of craniofacial growth. The Angle Orthod 1979;49:181-189.

48. Baccetti T, Franchi L, De Toffel L. The diagnostic performance of chronologic age in the assessment of skeletal maturity. Progress in Orthod 2006;7:176-188.

49. The Royal College of Paediatrics and Child Health, and Scientific Committee of Nutrition. Application of the WHO growth standards in the UK: Expert group on growth standards 2007.

50. Tanner TM, Whitehouse RH, Takaishi M. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British Children, 1965, Part I. Arch Dis Child 1996;41:454-472.

51. Gardiner D, Pearson A. Growth chart for premature and other infants. Arch Dis Child 1971;46:783–787.

52. Buckler J, Tanner J. Revision and update of Tanner–Whitehouse clinical longitudinal charts for height and weight [letter]. Eur J Paediatr 1997;156:248–249.

53. Freeman JV, Cole TJ, Chinn S, Jones PRM, White EM, Preece MA. Cross sectional stature and weight reference curves for the UK, 1990. Arch Dis Child 1995;73:17-24.

54. Wright CM, Booth IW, Buckler JMH, Cameron N, Cole TJ, Healy MJR, Hulse JA, Preece MA, Reilly JJ, Williams AF. Growth reference charts for use in the United Kingdom. Arch Dis Child 2000;86:11-14.

55. de Onis M, Garza C, Victora CG, Bhan MK, Norum KR, and guest editors. The WHO multicentre growth reference study (MGRS). Rationale, planning and implementation. Food and nutrition bulletin 2004;25:Supplement 1.

56. de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J. The WHO multicentre growth reference study (MGRS). Planning, study design and methodology. Food and nutrition bulletin. 2004;25:Supplement 1.

57. Wright CM. Growth charts for babies. BMJ 2005;330:1399-1400.

58. Preece MA, Reilly JJ, Williams AF. Growth reference charts for use in the United Kingdom. Arch Dis Child 2002;86:11-14.

59. Sullivan PG. Prediction of pubertal growth spurt by measurement of standing height. Eur J Orthod 1983;5:198-207.

60. Tanner JM, Whitehouse RH, Marubini E, Resele LF. The adolescent growth spurt of boys and girls in the Harpenden Growth Study. Ann Hum Bio 1976;3:109-126.

61. Nanda RS. Growth changes in skeletal-facial profile and their significance in orthodontic diagnosis. Am J Orthod 1971;59:501.

62. Houston WJB. The current status of facial growth Prediction: a Review. Brit J Orthod 1979;6:11-17.

63. Moore RN, Moyer BA, DuBois LA. Skeletal maturation and craniofacial growth. Am J Orthod Dentofacial Orthop 1990;98:33-40.

64. Flores-Mir C, Nebb B, Major W. Use of skeletal maturation based on hand-wrist radiographic analysis as a predictor of facial growth: A systematic review. Angle Orthod 2004;74:118-124.

65. Tanner JM 1962. Growth at Adolescence, 2nd edn. Blackwell: Oxford.

66. Hagg U, Taranger J. Maturational indicators and pubertal growth spurt. Am J Orthod 1982;82:299-309.

67. Sierra AM. Assessment of dental and skeletal maturity. Angle Orthod 1987;57:194-208.

68. Coutinho S, Buschang PH, Miranda F. Relationships between mandibular canine calcification stages and skeletal maturity. Am J Orthod Dentofacial Orthops 1993;104:262-268.

69. Franchi L, Baccetti T, DeToffol L, Polimeni A, Cozza P. phases of the dentition for the assessment of skeletal maturity. Am J Orthod Dentofacial Orthop 2008;133:395-400.

70. Baccetti T, Franchi L, De Lisa S, Giuntini V. Eruption of maxillary canines in relation to skeletal maturity. Am J Orthod Dentofacial Orthop 2008;133:748-751.

71. Gajoks ZK, Henderson KD, Hirshham JN, Palmert MR. Genetic determinants of pubertal timing in general population. Mol Cell Endocrin 2010;5:21-29.

72. Euling SY, Herman-Giddens ME, Lee PA, Selevan SG, Juul A, Sørensen TI, Dunkel L, Himes JH, Teilmann G, Swan SH. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. Pediatrics 2008;121:172-191.

73. Marshal WA. Interrelationships of skeletal maturation, sexual development and somatic growth in man. Ann of Human Bio 1974;1:29-40.

74. Daniel WA, Hanlan WR, Harland EA, Grill GP. Secondary sex characteristics of girls 12 to 17 years of age: The U.S. Health Examination Survey. J Pediatrics 1980;96:1074-1078.

75. Hanlan WR, Grill GP, Coroni-Huntley J, Leaver P. Secondary sex characteristics of boys 12 to 17 years of age: The U.S. Health Examination Survey. J Pediatrics 1979;95:287-289.

76. Isaacson KG, Thom AR, Horner K, Whaites E. Orthodontic radiographs – guidelines. British Orthodontic Society Third Ed 2008.

77. Department of Health Ionising Radiation (Medical Exposure) Regulations 2000 http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4007957 (last accesssed Dec 2012)

78. X-ray of Normal Spine, 2010. Bone and Spine. (online). Available at <http://www.boneandspin.com/spine/cervicalspine> [Accessed July 2011].

79. Hochman M, Tuli S. Cervical Spondylotic Myelopathy: A Review. The Internet Journal of Neurology 2005; 4:54-57.

80. Spine: Anatomy and Biomechanics, 2011. Orthoteers, the guiding light in orthopaedics. (online). Available at:< http://www.orthoteers.com> [Accessed July 2011].

81. Todd TW, Pyle SI. A quantitative study of the vertebral column by direct and roentgenoscopic method. Am J Phys Anthrop 1928:321-338.

82. Taylor JR. Growth of human intervertebral disk and vertebral bodies. J Anat 1975;120:49-68.

83. Bick E, Copel J. Longitudinal growth of the human vertebrae. J of Bone and Joint Surgery Am Edition 1950;32:803-813.

84. Pancherz H. The mechanism of Class II correction in Herbst appliance treatment. Am Journal Orthod 1982;82:104-113.

85. Houston WJB, Maher RE, McElroy D. Sources of error in measurements from cephalometric radiographs. Eur J Orthod 986;8:149-151.

86. Gu Y, McNamara JA. Mandibular growth changes and cervical vertebral maturation. A cephalometric implant study. Angle Orthod 2007;77:947-953.

87. Ballrick J, Fields H, Vig KW, Beck FM, Germark T, Baccetti T et al. Reliability and validity of CVM and Hand-wrist radiograph. IADR 2005. (abstract available at: http//iadr.confex.com/iadr/2005Balt)

88. Alkhal AH, Wong RWK, Rabie BM. Correlation between chronological age, cervical vertebral maturation and fishman's skeletal maturity indicators in southern chinese. Angle Orthod. 2008;78:591-596.

89. Gabriel DB, Southard KA, Qian F, Marshall SD, Franciscus RG, Southard TE. Cervical vertebrae maturation method:poor reproducibility. Am J Orthod Dentofacial Orthop 2009;136:e1-7.

90. Jaqueria LMF, Armond ML, Pereira LJ, Alcantara CEP, Marques LS. Determining skeletal maturation staging using cervical vertebrae: evaluation of 3 diagnostic methods. Braz oral 2010;24:984-989.

91. Nestman TS, Marshall SD, Qian F, Holton N, Franciscus RG, Southard TE. Cervical vertebrae maturation method morphologic criteria: Poor reproducibility. Am J Orthod Dentofacial Orthop 2011;140:182-8.

92. Wong RWK, Alkhal AH, Rabie BM. Use of cervical vertebral maturation to determine skeletal age. Am J Orthod Dentofacial Orthop 2009;136:484.e1-484.e6.

93. Ball G, Woodside D, Tompson B, Hunter WS. Relationship between cervical vertebral maturation and mandibular growth. Am J Orthod Dentofacial Orthop 2011;139:e455-e461.

94. Santiago R, Vitralb R, Fragac M, Bolognesed A, Maiae L. Cervical vertebral maturation as a biologic indicator of skeletal maturity. A systematic review. Angle Orthod 2012;1123-31.

95. Bland JM, Altman DG. Validating scales and indexes. BMJ 2002;324:607.

96. Moorrees C, Kean M. Natural head position, a basic consideration in the interpretation of cephalometric radiographs. Am J Phys Anthropology 1958;16:213-234.

97. O'Brien K, Wright J, Conboy F, Sanjie YW, Mandall N, Chadwick S, Connolly I, Cook P, Birnie D, Hammond M, Harradine N, Lewis D, McDade C, Mitchell L, Murray A, O'Neill J, Read M, Robinson S, Roberts-Harry D, Sandler J. Effectiveness of treatment for Class II malocclusion with the herbst or twin-block appliances: a randomized, controlled trial. Am J Orthod Dentofacial Orthop 2003;124:128-137.

98. Cicchetti DV, Fleiss JL. Comparison of the null distribution of weighted kappa and the C ordinal statistic. App Psycholo Measurement 1997;1:195-201.

99. Smortree V, Athonasiou AE. Anatomy, radiographic anatomy and cephalometric landmarks of craniofacial skeleton, soft tissue profile, dentition, pharynx and cervical vertebrae. In: Athonasiou, AE Orthodontic Cephalometry. London: Mosby-Wolfe, 1995; p.45-47.

100. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37-46.

101. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

102. Cohen J. Weighted kappa: nominal scale agreement with provision for scale and disagreement or partial credit. Psychol Bull 1968;70:213-20.

103. Schester C. A note on the interpretation of weighted kappa and its relations to other observer agreement statistics for metric scales. Educ Psychol Meas 2004;64:243-253.

104. Sim J, Wright C. The kappa statistic in reliability studies: use, interpretation and sample size requirements. Phys Ther 2005;85:257-268.

105. Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. Epidemiology 1996;7:199-202.

106. Bland JM, Altman DG. A note on use of intra class correlation coefficients in the evaluation of agreement between two methods of measurement. Comm Biol Med 1989;2:337-340.
107. Shrout PE, Fleiss JL. Intraclass correlation: Uses in assessing rater reliability. Psychol Bull 1979;2:420-280.
108. Nickerson CA. A note on a concordance correlation coefficient to evaluate reproducibility. Biometrics 1997;53:1503-1507.
109. Kendall MG, Babington-Smith B. The problem with n rankings. Annals of Math Stat 1939;10:275-287.
110. Fleiss JL, Cohen J. The equivalence of weighted kappa to intraclass correlation coefficient of measures of reliability. Educational Psychol Meas 1973;33:613-619.
111. Van der Stelt P. Better imaging. The advantages of digital radiographs. Jour Am Den Assoc 2008;139:75-35.
112. Obuchowki NA. How many observers are needed in clinical studies of medical imaging. Am Jour Roentg 2004;182:867-869.
113. Legendre P. Species Associations: The Kendall Coefficient of Concordance Revisited. Jour Agricu, Bio and Environ Stat 2005;10:226–245.
114. Cohen SA. Instructional alignment: searching for the magic bullet. Educational researcher 1987;16:16-20.
115. Fleming N. Learning styles again: VARKing up the right tree!! Educat develop 2006;7.4:4-7.
116. Harden RM, Crosby JR. AMEE Education Guide No 20: the good teacher is more than a lecturer – the twelve roles of a teacher. Med teacher 2000;22:334-47.
117. Van der Stelt P. Better imaging. The advantages of digital radiography. Jour Am Den Assoc 2008;139:75-135.

Appendices

Appendix 1

Favourable Ethical Opinion 12/EM/0126

a)

NRES Committee East Midlands - Northampton
The Old Chapel
Royal Standard Place
Nottingham
NG1 6FS

Telephone: 0115 8839425
Facsimile: 0115 8839294

09 March 2012

Dr Jayne Harrison
Consultant Orthodontist
Royal Liverpool and Broadgreen University Hosptials NHS Trust
Orthodontic Department
Liverpool University Dental Hospital
Pembroke Place
L35PS

Dear Dr Harrison,

Study title:	**A Study to assess the reliability and reproducability of the Cervical Vertebrae Maturation Staging Method, amongst orthodontic trainees and specialist orthodontists.**
REC reference:	**12/EM/0126**

Thank you for your application for ethical review, which was received on 09 March 2012. I can confirm that the application is valid and will be reviewed by the Proportionate Review Sub-Committee on 15 March 2012.

One of the REC members is appointed as the lead reviewer for each application reviewed by the sub-committee. I will let you know the name of the lead reviewer for your application as soon as this is known.

Please note that the lead reviewer may wish to contact you by phone or email between 13 March 2012 and 15 March 2012 to clarify any points that might be raised by members and assist the sub-committee in reaching a decision.

If you will not be available between these dates, you are welcome to nominate another key investigator or a representative of the study sponsor who would be able to respond to the lead reviewer's queries on your behalf. If this is your preferred option, please identify this person to us and ensure we have their contact details.

You are not required to attend a meeting of the sub-committee.

Please do not send any further documentation or revised documentation prior to the review unless requested.

Documents received

The documents to be reviewed are as follows:

Document	*Version*	*Date*
Covering Letter		01 March 2012
Investigator CV		12 January 2012
Letter from Sponsor		01 March 2012
Other: Email from Dr Peter Cole		
Other: CV - Billie-Jean Rainey		01 March 2012
Other: Intra/Inter-observer assessment Data Collection Sheet		
Other: Trust Co-Sponsorship Letter	4	21 February 2012
Other: Peer Review Assessment Form		17 January 2012
Other: Peer Review Assessment Form		10 January 2012
Participant Consent Form	2.2	01 January 2012
Participant Information Sheet	2.2	01 January 2012
Protocol	2.20	26 January 2012
REC application	77960/300605/1/668	17 January 2012

No changes may be made to the application before the meeting. If you envisage that changes might be required, you are advised to withdraw the application and re-submit it.

Notification of the sub-committee's decision

We aim to notify the outcome of the sub-committee review to you in writing within 10 working days from the date of receipt of a valid application.

If the sub-committee is unable to give an opinion because the application raises material ethical issues requiring further discussion at a full meeting of a Research Ethics Committee, your application will be referred for review to the next available meeting. We will contact you to explain the arrangements for further review and check they are convenient for you. You will be notified of the final decision within 60 days of the date on which we originally received your application. If the first available meeting date offered to you is not suitable, you may request review by another REC. In this case the 60 day clock would be stopped and restarted from the closing date for applications submitted to that REC.

R&D approval

All researchers and local research collaborators who intend to participate in this study at sites in the National Health Service (NHS) or Health and Social Care (HSC) in Northern Ireland should apply to the R&D office for the relevant care organisation. A copy of the Site-Specific Information (SSI) Form should be included with the application for R&D approval. You should advise researchers and local collaborators accordingly.

The R&D approval process may take place at the same time as the ethical review. Final R&D approval will not be confirmed until after a favourable ethical opinion has been given by this Committee.

For guidance on applying for R&D approval, please contact the NHS R&D office at the lead site in the first instance. Further guidance resources for planning, setting up and conducting research in the NHS are listed at http://www.rdforum.nhs.uk. There is no requirement for separate Site-Specific Assessment as part of the ethical review of this research. The SSI Form should not be submitted to local RECs.

Communication with other bodies

All correspondence from the REC about the application will be copied to the research sponsor and to the R&D office for Royal Liverpool and Broadgreen University Hosptials NHS Trust. It will be your responsibility to ensure that other investigators, research collaborators and NHS care organisation(s) involved in the study are kept informed of the progress of the review, as necessary.

12/EM/0126	**Please quote this number on all correspondence**

Yours sincerely

Miss Jessica Parfrement
Committee Co-ordinator

Email: jessica.parfrement@nottspct.nhs.uk

Copy to:	*Sponsor - Mrs Lindsay Carter* *R&D Contact - Mrs Heather Rodgers* *Student – Mrs Rainey*

b)

NHS
Health Research Authority

NRES Committee East Midlands - Northampton
The Old Chapel
Royal Standard Place
Nottingham
NG1 6FS

Tel: 0115 8839435
Fax: 0115 8839294

26 September 2012

Dr Jayne Harrison
Consultant Orthodontist
Royal Liverpool and Broadgreen University Hosptials NHS Trust
Orthodontic Department
Liverpool University Dental Hospital
PEMBROKE PLACE
L35PS

Dear Dr Harrison

Study title:	**A Study to assess the reliability and reproducability of the Cervical Vertebrae Maturation Staging Method, amongst orthodontic trainees and specialist orthodontists.**
REC reference:	**12/EM/0126**
Amendment number:	
Amendment date:	**04 September 2012**

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

- The Committee request that participant data is transferred via safestick USB and that this be reflected in the protocol and submitted as a minor amendment.

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Protocol	2.22	01 August 2012
Notice of Substantial Amendment (non-CTIMPs)		04 September 2012
Covering Letter		04 September 2012

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

A Research Ethics Committee established by the Health Research Authority

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

12/EM/0126:	**Please quote this number on all correspondence**

Yours sincerely

pp.

Mr John Aldridge
Chair

E-mail: georgia.copeland@nottspct.nhs.uk

Enclosures:	*List of names and professions of members who took part in the review*
Copy to:	*Mrs Heather Rodgers, Research Governance Management*
	Mrs LINDSAY CARTER
	Billie-Jean Rainey

Appendix 2

Evidence of Royal Liverpool and Broadgreen University Hospital Trust Co-sponsorship

The Royal Liverpool and Broadgreen University Hospitals **NHS**
NHS Trust

Royal Liverpool University Hospital
Prescot Street
Liverpool
L7 8XP
Tel: 0151 706 2000
Fax: 0151 706 5806

TRUST APPROVAL LETTER FOR NON-CTIMP STUDIES

Mrs Billie-Jean Rainey
Liverpool University Dental Hospital
Department of Orthodontics
Pembroke Place
Liverpool
L3 5PS

Rec Ref: **12/EM/0126**

19/04/2012

Dear Mrs Billie-Jean Rainey

R&D No: 4302
Reliability of the Cervical Vertebrae Maturation (CMV) method

The above study is a Non-commercial, Other Quantitative Research study, sponsored by University of Liverpool, Royal Liverpool & Broadgreen University Hospitals and funded by University of Liverpool. The Trust is now happy for you to commence work on this study, using the following ethically approved documents.

Document	Version	Dated
Protocol	2.20	26 January 2012
Participant Information Sheet	2.2	1 January 2012
Consent form	2.2	1 January 2012
Intra/inter observer assessment data collection sheet		

May I to take this opportunity to remind you of your responsibilities as PI for this study to:-

- Inform RD&I of any SUSAR's within the Trust as per Trust policy
- Report SAE's as per protocol and Trust policy and record total number on OSIRIS
- Provide copies to RD&I of annual progress and safety reports to Ethics
- Complete and return the RD&I annual report form within the stipulated timeline
- Comply with the Research Governance Framework 2nd Ed 2005 including but not limited to the Human Tissue Act 2004 act plus it's appendices and the Data Protection Act 1998
- Read and disseminate to your research team Trust research SOP announcements

- Inform RD&I of any amendments to, or changes of status in the study
- Ensure any conditions to approval stipulated by the REC have been addressed
- Ensure that all screening and recruitment activity is recorded on OSIRIS
- Maintain the study site file (if not provided by the sponsor a template is available on the Trust intranet)
- Acknowledge to RD&I any updates to the study
- Return signed reply slip provided

Investigators who do not comply with the above will be dealt with in accordance with the Trust Disciplinary policy and/or will have their research stopped.

I wish you every success with your research. Please contact the RD&I Department if you require any advice on the above points.

Yours sincerely

Julia West
Operational Director RD&I

cc Head of Directorate
University of Liverpool, Royal Liverpool & Broadgreen University Hospitals

....

I agree to the terms and conditions of the Trust research approval for RD&I **4302, Reliability of the Cervical Vertebrae Maturation (CMV) method** and am aware of my responsibilities under the Research Governance framework and Trust Research SOP's.

Signed: .... Dated: ....

Please return a copy of this letter to the RD&I Department RLBUHT
Thank you

Appendix 3

Evidence of University of Liverpool Co-sponsorship

UNIVERSITY OF LIVERPOOL

Mrs Lindsay Carter
Clinical Research
Governance Manager.

Research Support Office
University of Liverpool
Waterhouse Buildings
3 Brownlow Street
Liverpool
L69 3GL

Tel: 0151 794 8722
Lindsay.Carter@liv.ac.uk

Sponsor Ref: UoL000840

Dr Jayne Harrison (Billie-Jean Rainey)
School of Dentistry

01 March 2012

Dear Dr Harrison

Re: Intention to Sponsor

"Reliability of the Cervical Vertebrae Maturation (CVM) Staging Method."

I can confirm that the University of Liverpool, in principle, will be willing to act as Sponsor with the Royal Liverpool and Broadgreen University Hospitals NHS Trust for the above research project under the Department of Health's Research Governance Framework. The project is currently being reviewed through the necessary sponsorship approval procedure. It is envisaged that we will be able to confirm sponsorship within the next four weeks.

University professional indemnity and clinical trials insurances will apply to the study as appropriate.

In order that you can proceed with your NRES (National Research Ethics Service) application for NHS ethical approval, please contact me on 0151 794 8722 in order to arrange for the Declaration of Sponsor to be completed and signed by the University.

Yours sincerely,

pp
R Turner

Mrs Lindsay Carter
Clinical Research Governance Manager, Research Support Office

Appendix 4

Clinician Information Sheet

UNIVERSITY OF LIVERPOOL

Department of Orthodontics

January 2010 V2.2

Liverpool University Dental Hospital and School of Dentistry

Pembroke Place, Liverpool L3 5PS

CLINICIAN INFORMATION SHEET

Reliability of Cervical Vertebrae Maturation (CVM) staging method.

What is the purpose of the study?

As Orthodontists we regularly treat patients who are growing and maturing as they grow up from a child, to an adolescent and then to an adult. Predicting how much patients will grow may affect the type of braces we offer them and the result they get from their treatment. Cervical vertebral maturation (CVM) is a measure taken from the neck bones on the X-ray pictures we use routinely. Some research suggests that CVM may provide valuable information about growth, which may then allow us to predict how much growth we can expect. This would then allow us to target treatment better and potentially, reduce the length of orthodontic treatment.

Has the study been approved?

Yes. Liverpool Local Research Ethics Committee has given the approval for this study.

Who is paying for the study?

The School of Dental Sciences of the University of Liverpool is paying for the study. The Royal Liverpool and Broadgreen University Hospital Trust and the University of Liverpool are co-sponsoring the study

111

Who will be conducting the study?

The study is being led by Dr Jayne Harrison (Consultant in Orthodontics) and carried out by Mr Richard Gibson (FTTA in Orthodontics), Mrs Billie-Jean Rainey

Why have you been asked to take part?

We are asking all consultants and trainee orthodontists in the Mersey region to take part. It is our aim to recruit the highest number of observers compared with similar reliability research studies. This will therefore improve the generalisability of our findings.

What will I have to do?

You will have to attend 2 consecutive educational sessions of the Mersey Deanery Audit Meeting and participate in 2 rounds of this study. Round 1 of this study will begin with a 30 minute training presentation, detailing how to use the CVM staging index. You will then be shown a random sample of 72 cephalograms and asked to stage each image appropriately. Round 2 will take place approximately three months later ,at the next educational session training. The training presentation will be repeated and you will be asked to stage the same 72 cephalograms, only in a different random order.

Your individual scores will be analysed for intra-examiner reliability, and compared with your collegues, for inter-examiner reliability.

.

What happens if I don't want to take part?

Please feel free to volunteer or decline participation in this study. You may also withdraw at any time without explanation.

What if I have a question or there is a problem on the trial?

If you have a concern about any aspect of this study, you should ask to speak to a member of the research team on 0151 706 5252. They will do their best to answer your questions.

How will you collect and look after data (information)?

No one will be able to identify any of the data we collect about you. As soon as we have collected the necessary data, we will remove all information that identifies you and replace it by a code number. Only members of the research team will process and analyses your data. The person responsible for security and access to your data is Dr Jayne Harrison, the Co-Chief investigator of the Study. The data will be stored safely for ten years.

What do I do if I want to take part?

If you would like to take part in our study, please sign all the appropriate parts of the consent form that we will give you.

THANK YOU FOR TAKING THE TIME TO READ THIS LEAFLET.

Appendix 5

Consent form

Centre Number:

Study Number:

Patient Identification Number for this trial:

CONSENT FORM

Title of Project: **Reliability of Cervical Vertebrae Maturation (CVM) staging method.**

Name of Researcher: Billie-Jean Rainey

Please initial box

1. I confirm that I have read and understand the information sheet dated January (version 2.2) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.

3. I understand that the data collected during the study will be analysed by the study investigators. I give permission for these individuals to analyse my results for phase I and II of the study

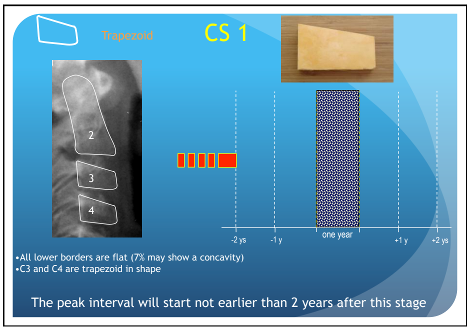
4. I agree to take part in the above study.

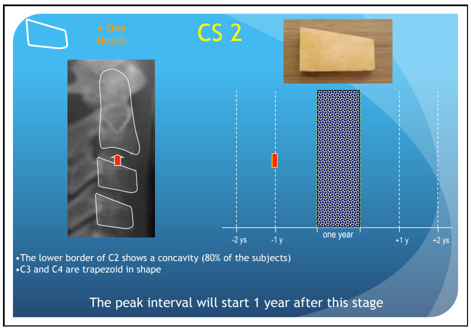
Name of participant	Date	Signature
Name of Person taking consent	Date	Signature

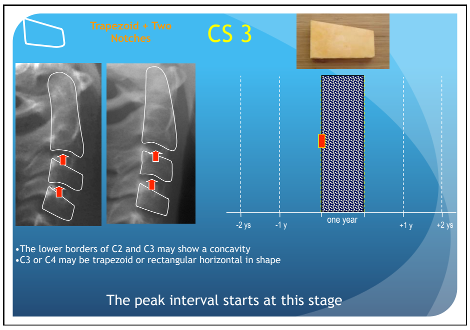
Appendix 6

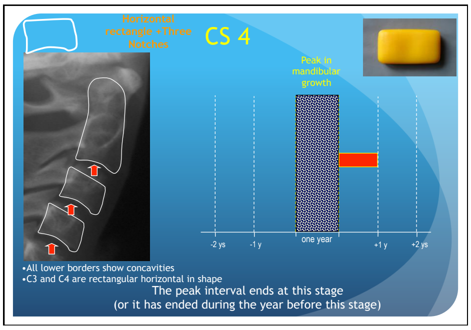
Observer handout I

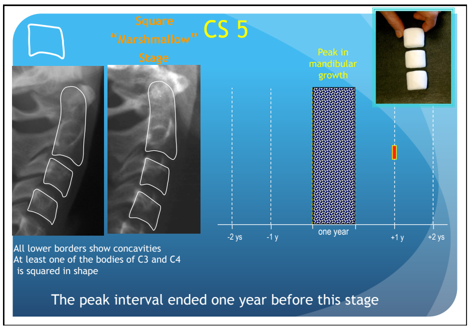
Cheese	Soap	Marshmallow
CS1,2,3	CS4	CS5,6

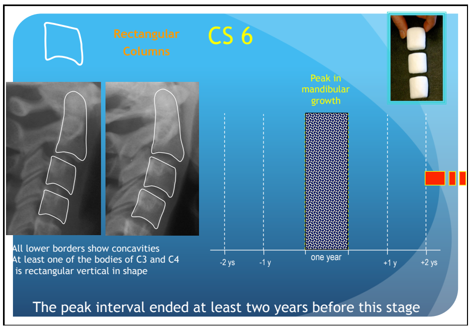






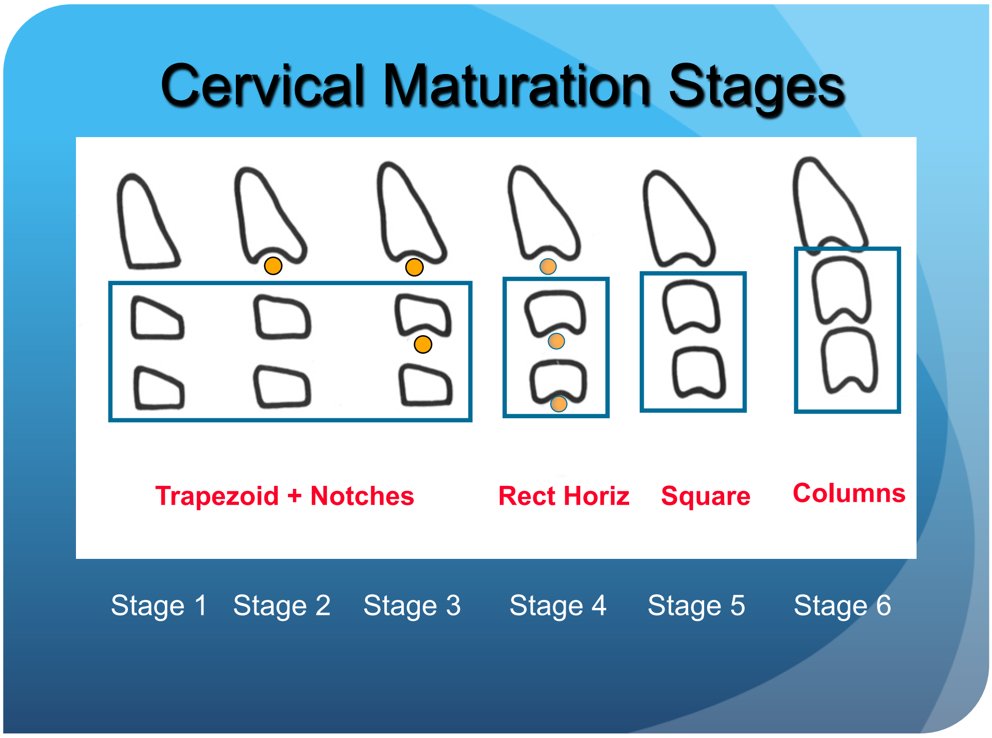






Appendix 7

Observer handout II



Appendix 8

Observer score sheet

Cervical Stage Score Sheet

Name: ______________________________

RADIOGRAPH NO.	CERVICAL STAGE	NOTES
1		
2		
3		
4		
5		
6		
A		
7		
8		
9		
10		
11		
12		
B		
13		
14		
15		
16		
17		
18		

19		
20		
21		
22		
23		
24		
D		
25		
26		
27		
28		
29		
30		
E		
31		
32		
33		
34		
35		
36		
F		
37		
38		
39		
40		
41		

42		
G		
43		
44		
45		
46		
47		
48		
H		
49		
50		
51		
52		
53		
54		
I		
55		
56		
57		
58		
59		
60		
J		
61		
62		
63		

64		
65		
66		
K		
67		
68		
69		
70		
71		
72		

Raw data

PHASE 1

CEPH	C1	C2	C3	C4	C5	C6	C7	C8	C9	T1	T2	T3	T4	T5	T6	T7	T8	T9	T 10	T 11
1	4	4	4	4	4	4	3	4	3	4	4	4	3	4	4	4	3	4	3	4
2	5	4	5	4	5	5	5	5	5	5	4	5	4	5	5	5	5	5	5	5
3	6	6	6	5	6	6	6	6	6	4	6	6	6	6	6	6	6	6	6	6
4	2	2	3	2	3	3	3	2	2	3	2	2	3	2	3	2	3	3	2	2
5	5	6	6	4	5	6	5	6	5	5	5	6	5	5	6	5	6	6	6	5
6	6	6	5	5	6	6	6	6	6	6	4	6	6	4	6	6	6	6	5	6
7	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2
8	5	5	4	4	5	5	5	5	5	5	4	5	5	4	5	4	5	5	5	5
9	1	3	3	1	1	3	1	1	1	1	2	1	2	1	1	1	3	3	1	1
10	4	4	4	4	4	4	4	3	3	4	4	3	3	4	4	3	5	4	4	4
11	5	5	3	3	3	3	4	3	3	4	3	4	4	3	3	3	2	3	3	3
12	1	3	3	1	4	3	1	1	1	1	2	1	3	1	1	3	2	3	1	1
13	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	3	3	2	3
14	6	5	5	5	6	5	5	6	5	6	4	6	6	5	6	5	5	5	5	5
15	4	5	5	4	6	5	4	4	4	4	4	4	5	4	4	4	4	4	4	4
15	5	4	6	5	4	5	5	5	5	5	6	5	6	5	5	5	5	6	5	5
17	4	5	6	4	6	6	5	6	5	5	4	5	6	4	5	4	6	6	6	6
18	4	4	4	4	5	3	4	4	4	4	4	3	3	4	4	4	4	4	4	4
19	5	5	4	4	4	4	4	4	4	4	4	4	4	5	5	4	4	5	5	4
20	5	6	6	5	5	6	5	4	4	6	6	6	5	3	6	4	5	6	4	5
21	4	4	5	5	4	4	5	5	5	3	4	4	3	5	5	4	5	5	5	5
22	2	3	2	2	2	2	3	2	2	3	2	2	2	2	2	2	2	2	2	2
23	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
24	5	3	5	4	5	4	5	5	5	5	5	4	5	4	4	4	4	4	5	5
25	5	3	4	3	3	4	4	3	3	5	3	3	4	5	3	3	3	4	3	3
26	5	5	5	5	5	5	5	3	4	5	4	5	3	5	6	4	5	6	5	5
27	4	4	6	4	5	4	5	5	5	5	5	5	6	4	5	5	5	5	5	6
28	5	6	5	5	5	5	5	6	5	4	4	5	5	6	5	5	5	5	5	6
29	2	1	3	1	1	2	3	1	1	2	1	1	2	2	1	3	1	3	1	2
30	1	2	2	2	2	2	3	2	2	2	2	3	3	3	2	2	2	3	2	2
31	5	4	5	5	6	6	5	5	5	5	5	5	6	5	5	5	5	6	5	5
32	6	6	6	6	6	6	5	6	6	6	6	6	6	6	6	6	5	6	6	6
33	1	3	2	2	3	3	3	3	3	3	1	3	3	3	3	3	2	3	4	1
34	5	5	5	5	4	5	5	4	5	5	2	5	5	5	5	4	5	4	5	5
35	4	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	4	5	4	5
36	3	1	1	1	1	1	1	1	1	2	1	1	3	1	1	1	1	1	1	1
37	3	4	3	3	4	3	4	4	4	4	4	4	4	4	4	3	3	3	4	3

38	5	5	4	4	4	5	4	4	5	4	5	5	6	4	5	4	4	4	5	4
39	6	4	5	5	6	5	5	5	5	6	5	5	5	5	6	5	4	5	5	5
40	5	4	5	5	4	5	3	5	5	3	4	4	4	4	5	4	4	4	4	5
41	4	4	4	3	2	3	3	3	3	3	3	4	3	3	4	3	3	3	4	3
42	2	2	3	3	2	3	3	3	3	3	3	3	4	3	3	3	1	3	3	3
43	5	6	6	5	6	5	6	6	6	6	6	6	6	6	6	6	6	6	5	6
44	2	3	3	3	2	3	4	3	3	3	3	3	3	2	4	3	3	3	3	3
45	5	4	5	5	4	6	5	4	5	4	6	5	5	5	5	6	4	4	6	5
46	4	4	4	4	4	4	4	4	3	5	4	4	5	4	4	4	5	5	4	5
47	5	3	2	2	2	5	4	5	2	6	3	3	3	4	5	5	5	2	5	5
48	4	4	3	2	3	3	4	2	3	3	3	4	3	3	3	3	3	4	3	3
49	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
50	5	6	6	6	5	6	6	5	4	6	6	6	6	6	6	4	5	6	6	6
51	4	4	5	5	4	5	3	4	5	5	5	5	5	5	6	5	4	4	4	4
52	4	5	4	1	4	4	5	2	4	4	4	3	3	2	4	4	4	4	4	4
53	2	3	4	2	2	3	3	2	3	3	4	3	2	3	3	3	3	3	3	3
54	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
55	2	4	4	1	2	4	4	2	1	4	1	4	3	4	2	4	4	4	4	3
56	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
57	5	4	4	4	4	4	4	2	4	4	2	3	3	4	4	4	3	4	4	4
58	6	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	5	6	6	6
59	6	6	5	5	6	6	6	6	6	6	6	6	6	5	6	6	6	6	5	6
60	5	5	6	5	5	5	5	5	5	5	6	5	5	6	5	5	5	6	5	6
61	6	6	6	6	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
52	1	3	3	3	3	3	3	2	3	3	1	3	2	3	3	3	3	3	3	3
63	5	6	5	5	4	5	5	6	5	5	5	5	6	5	5	5	5	5	5	5
64	1	1	4	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
65	1	4	2	2	2	2	2	2	1	4	1	2	3	2	1	2	1	2	2	2
66	4	5	5	4	4	4	4	5	5	5	5	4	5	4	5	4	4	4	4	5
67	1	2	2	2	1	2	1	2	1	2	1	2	3	2	2	2	2	2	2	2
68	6	5	5	5	5	6	5	5	6	6	5	5	6	5	5	5	5	5	5	5
69	5	4	4	2	4	4	4	2	4	4	4	4	4	4	4	4	4	4	4	3
70	4	4	5	4	4	5	5	4	5	5	4	4	5	4	3	4	4	4	4	4
71	3	1	3	1	2	1	4	1	1	2	1	1	1	2	2	1	4	2	4	1
72	5	5	5	2	3	3	3	4	1	3	3	5	5	5	4	5	3	4	5	5

PHASE 2

CEPH	C1	C2	C3	C4	C5	C6	C7	C8	C9	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
2	5	5	5	3	5	5	5	5	5	5	5	5	5	5	5	4	3	5	5	5
3	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
4	3	4	3	2	2	3	1	2	2	2	2	2	1	2	2	2	2	2	2	2
5	6	6	6	5	5	6	6	5	6	5	6	5	6	6	6	5	5	5	5	5
6	6	6	5	6	6	5	6	6	6	6	6	6	6	6	6	4	6	6	5	6
7	2	4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3
8	5	5	4	4	4	3	5	4	4	4	4	4	5	5	5	4	5	4	4	4
9	3	3	3	1	1	3	3	3	1	1	3	3	1	3	3	3	1	1	1	1
10	4	4	4	4	4	3	4	3	4	4	4	3	4	4	4	4	4	4	4	4
11	4	4	3	3	3	4	4	3	3	4	3	3	4	3	3	3	3	3	3	3
12	3	3	5	1	1	3	1	3	1	1	1	3	2	2	1	1	1	1	1	4
13	2	3	3	2	2	2	3	2	1	3	2	2	2	3	2	2	2	3	2	3
14	5	5	5	5	6	5	5	5	5	5	5	5	6	5	5	5	5	5	5	6
15	4	4	4	4	6	4	4	4	5	5	4	4	3	4	4	4	4	4	4	4
15	5	4	5	5	4	5	5	5	5	5	6	5	4	5	5	5	4	5	5	5
17	5	5	6	5	6	5	6	4	6	5	6	4	5	6	5	5	5	4	5	5
18	4	4	4	4	3	3	4	4	3	4	4	4	3	4	4	4	4	4	4	4
19	5	4	4	4	4	5	4	4	4	4	5	4	4	4	4	4	4	4	4	4
20	5	6	6	5	3	5	6	3	4	6	6	4	6	6	5	4	5	3	5	5
21	4	4	5	4	4	4	5	4	5	4	4	4	5	5	5	5	5	4	4	5
22	3	2	1	2	1	2	2	2	1	2	1	2	2	3	2	2	2	2	2	1
23	6	6	6	6	6	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6
24	5	4	4	4	5	4	4	4	4	5	5	4	5	5	4	4	5	4	4	4
25	4	5	4	3	4	3	4	3	4	4	3	3	3	5	3	3	3	3	3	5
26	5	5	5	5	3	4	5	3	5	5	4	4	5	5	5	3	5	3	4	5
27	5	5	5	4	4	4	5	5	4	4	5	5	6	4	5	5	4	4	5	6
28	6	5	5	5	2	5	5	6	5	5	3	5	5	5	6	5	5	6	5	6
29	1	3	3	1	1	2	3	3	2	1	1	2	2	2	1	1	1	1	1	4
30	1	3	2	2	1	2	2	2	1	1	2	2	2	3	3	2	2	3	2	2
31	5	5	6	3	5	5	5	6	5	6	5	5	5	5	6	5	4	5	5	5
32	6	6	5	6	6	4	6	6	6	6	5	6	6	6	6	6	6	6	6	5
33	1	3	1	1	3	3	3	1	1	3	1	3	1	2	3	3	3	3	3	3
34	5	5	5	5	4	3	5	5	5	4	4	5	4	5	5	4	4	4	5	5
35	5	5	4	5	5	5	5	5	5	5	5	4	5	5	4	5	5	4	4	5
36	3	3	5	1	1	3	1	1	1	3	1	1	1	2	1	1	1	1	1	1
37	4	4	5	3	3	4	4	3	3	3	4	3	4	3	4	4	4	3	4	4
38	5	5	4	5	3	4	5	5	4	4	4	4	5	4	5	4	4	4	4	5
39	6	4	6	5	5	4	5	5	4	5	4	5	5	6	5	5	5	5	5	5
40	4	5	5	4	3	5	5	3	3	2	3	3	4	4	4	5	4	5	4	4
41	4	4	4	2	2	4	4	3	4	4	4	3	4	3	4	4	3	3	3	3
42	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3
43	6	6	6	5	6	5	5	6	6	6	6	6	6	6	6	5	6	6	5	5

44	3	3	3	3	2	3	3	2	3	2	2	3	3	2	3	4	3	3	3	4
45	5	5	5	5	4	4	5	4	4	4	4	4	6	4	5	5	5	6	5	6
46	4	5	4	4	4	4	4	4	4	5	5	4	4	4	4	4	5	4	4	5
47	5	4	3	3	3	5	5	5	5	5	5	4	5	4	5	5	4	5	3	5
48	4	4	3	3	3	3	3	3	4	3	3	3	3	3	3	3	3	4	4	4
49	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
50	6	6	6	6	5	5	5	4	5	6	6	6	6	5	6	6	6	5	6	6
51	5	5	4	5	4	5	4	5	4	5	3	5	5	5	5	5	5	5	4	5
52	4	4	5	4	4	5	5	4	4	4	4	4	4	4	4	4	4	4	4	5
53	3	5	4	4	3	3	3	3	3	3	3	3	3	3	3	4	4	3	3	4
54	1	2	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1
55	2	4	5	1	3	3	4	4	4	5	1	4	4	4	1	2	2	2	1	3
56	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
57	4	4	4	3	5	4	3	3	4	3	4	4	4	4	4	4	4	3	4	4
58	6	6	6	6	6	6	5	6	6	6	6	6	6	6	6	6	6	5	6	6
59	6	6	5	6	6	5	6	6	6	6	5	5	6	6	6	6	6	6	5	6
60	6	5	6	5	5	6	5	6	6	5	5	5	6	6	5	5	5	6	5	5
61	6	6	6	6	5	6	5	6	6	6	6	6	6	6	6	6	6	6	6	6
52	3	3	4	3	3	3	4	3	3	4	3	3	3	3	3	3	3	3	3	4
63	5	5	5	5	5	5	6	5	5	5	6	5	5	5	5	5	4	5	4	6
64	1	3	2	1	1	2	2	1	1	2	1	1	1	1	1	1	1	1	2	1
65	2	3	2	2	2	4	3	2	2	4	3	2	2	3	2	2	2	2	2	2
66	4	5	5	4	4	4	5	4	4	5	5	4	5	4	4	4	4	4	4	5
67	2	5	3	2	5	1	4	3	1	1	1	2	2	2	2	2	2	2	2	2
68	5	6	5	5	5	5	5	5	5	6	5	5	6	5	5	4	6	4	5	5
69	4	4	4	3	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4	4
70	4	4	4	4	4	5	5	5	4	5	5	4	5	5	4	4	4	4	4	5
71	2	3	2	1	3	2	4	3	1	4	1	1	1	4	1	1	2	1	2	2
72	5	5	5	5	4	5	5	4	5	5	4	4	5	5	5	4	5	4	4	5