The Use of Readily Obtained Patient and Parent Data in Mandibular Growth Prediction

Aaron B. Christopher

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Dentistry

University of Washington

2013

Committee:

Anne-Marie Bollen

Charles F. Spiekerman

Susan W. Herring

Program Authorized to Offer Degree:

Orthodontics

The Use of Readily Obtained Patient and Parent Data in Mandibular Growth Prediction

Aaron Christopher, DMD

Abstract

The ability to predict a patient's mandibular growth would be a powerful tool for the orthodontic practitioner. Despite several past attempts, a simple, accurate and reliable method still does not exist for mandibular growth prediction. Using only data readily obtained in a routine orthodontic clinical exam, I created several unique predictive models over an average T1-T2 interval of 2.77 years for 447 patients between the ages of 9 and 16 at T1 using a best-fit least squares linear method. Evaluating all possible combinations of 14 different non-age-related variables and three different age-related variables, 65,536 unique models were constructed to predict either the change in articulare-pogonion (ΔArPg) or the difference between the actual ΔArPg and the expected ΔArPg as predicted by a historical population average from T1 to T2 for three groups: males, females, and a subset of the females whose menarchal status is known. I assessed the mean absolute error (MAE) of each model using a leave-one-out cross-validation approach on half of the data (model-building sets) for each group. I then fit the most accurate models to the model-building sets, tested their accuracies (MAE) at predicting ΔArPg on the other half of the data (validation sets), and compared these accuracies to those obtained when using historical population averages to predict ΔArPg. The population averages were represented by polynomial growth curves (PGCs) which were fit to the pooled data of three published longitudinal growth studies (Michigan, London, and Philadelphia). Only three models proved to be more accurate than the PGCs when applied to the validation sets, but not to a clinically significant degree, being merely 0.07, 0.06, and 0.01mm more accurate. The variables used in this study provide no clinically significant insight into a patient's mandibular growth rate over a typical orthodontic treatment interval beyond what can be obtained from a historical population average.

Introduction

The size and position of a patient's mandible relative to the maxilla and to the cranial base are fundamental to consider for the orthodontic practitioner striving to produce an ideal occlusion and an esthetic facial profile. Given this reality, one of the most difficult challenges faced in orthodontic practice is the treatment of growing patients, whose mandibular size can change dramatically over the course of treatment.

This challenge has been recognized from the outset of the orthodontic profession [1, 2] and thus attempts have been made over the past six decades to predict the amount of mandibular growth in patients – prior to treatment, during the course of treatment, and thereafter [3-26]. Such prediction could theoretically guide an orthodontist to know the appropriate timing of and approach to treatment and whether or not an ideal result can be obtained non-surgically.

Some of the initial attempts to predict mandibular growth involved the addition of population averages to a patient's starting mandibular size as the basis for prediction. Ricketts [3] proposed an arcial method which sought to identify a center of rotation within the ramus that would aid prediction of the vector of growth. The magnitude of growth, extracted from population averages, was then used to predict final mandibular position. Ricketts [4] later employed a more sophisticated computerized system to improve the accuracy of his technique. Johnston [5] proposed a system based on the addition of mean increments by direct superimposition on a printed grid. Later, Popovich and Thompson [6] characterized the average change in cephalometric landmarks relative to sella-nasion to create cephalometric templates to be used in growth prediction. Despite such efforts, these methods lack simplicity. Also, their reliance on population means disallows prediction tailored to the individual, and thus reduces their accuracy on the very subset of patients most difficult to treat—those whose growth patterns are farthest from the mean [7-10].

Others have hypothesized that mandibular growth patterns can be predicted by analyzing a patient's starting craniofacial morphology as obtained in an initial lateral cephalometric radiogroaph. Johnston [11] used multivariate analysis of various linear, angular, area, and proportional measurements from the lateral cephalogram as a tool in developing a growth prediction model for various elements of the facial skeleton, including the mandible. This approach attempted to tailor prediction to the individual, but still lacked in accuracy. Balbach [12] attempted a similar approach directed solely at the prediction of

mandibular growth and eventual occlusal position between the ages of 6 and 16, but did not find any strong associations.

Additional methods have been employed to predict the onset, peak and end of the adolescent growth spurt, and to estimate the proportion of facial growth remaining based on radiographic analysis of skeletal maturation. The earliest of these methods is based on the assessment of hand-wrist radiographs. A system originated by Greulich and Pyle [13] and later modified by Tanner [27], involved the comparison of an individual to standards of normal skeletal maturity to predict a patient's skeletal age. The most recent, and perhaps widely used, hand-wrist method was developed by Fishman, known as the "skeletal maturity index" (SMI) [28-30]. This index uses 11 discrete skeletal maturity indicators on six anatomical sites to relate skeletal maturity directly to the pubertal growth curve, independent of chronological age. Some controversy exists regarding the effectiveness of the SMI technique at determining facial growth potential, with some studies showing a significant association between SMI and facial growth [14-16] and others showing little association [31-33]. In either case, the use of hand-wrist radiographs can be controversial due to the added amount of radiation exposure to the patient.

The second major radiographic method for facial growth prediction involves the analysis of cervical vertebrae. Lamparski [34] was the first to develop a system for prediction of biologic age based on the maturation of cervical vertebrae visible in lateral cephalometric radiographs. Since then, several cervical vertebral maturation (CVM) methods have been proposed and improved upon [17, 18, 35, 36]. The most widely used of these methods was proposed by Baccetti et al [18] in 2005. Until then CVM methods relied on cross-sectional data and were correlated to hand-wrist skeletal age. Baccetti's method was the first to be established using longitudinal data from radiographs and the first to correlate changes in cervical vertebrae to changes in mandibular growth. Though somewhat widely used, some question the reliability of this method [37, 38]. Reliability aside, neither CVM nor SMI are

simple enough for routine clinical use and both are too subjective to keep interobserver reliability to a minimum [38].

Many consider the use of serial cephalometric radiographs to be the gold standard in evaluating the onset, peak and termination of the adolescent growth spurt as it pertains to the facial skeleton [19, 20, 39]. It can be somewhat challenging, however, to implement the knowledge gained from analyzing serial cephalometric radiographs due to the retrospective nature of the analysis. By the time one has determined either the onset or the peak of an adolescent's growth spurt, this critical time point has already passed. While being a great retrospective assessment of where a patient was along his/her pubertal growth curve, the serial cephalometric method has not been successfully used alone to predict future mandibular growth rate. This method is also somewhat controversial due to the added degree of radiation exposure to the patient, and its implementation requires a nontrivial amount of time and resources.

Perhaps the simplest and most widely used method employed in clinical practice to predict mandibular growth is the assessment of patient data readily obtained in an initial patient history, such as: patient's gender, age, height, weight, and menarchal status; parents' and siblings' heights; and any recent changes in clothing or shoe sizes. Such data are so efficiently and non-invasively obtained that it has become a nearly ubiquitous practice to record at least some of these patient variables during the initial orthodontic exam. It is unclear how orthodontic practitioners apply this information to their treatment planning and treatment of patients, but a search of the orthodontic literature reveals that to-date no predictive models have been developed and tested with the exclusive use of these variables. It can only be assumed therefore that such variables merely guide an orthodontist's intuition as to how much mandibular growth can be expected. Without a predictive model that makes use of these data, one might ask whether their recording and assessment is of any benefit to the orthodontist.

The medical profession has developed and tested several models over the years to predict adult statural height from childhood height [40-47]. The most accurate have incorporated multiple variables into the prediction model such as: skeletal age as assessed on hand-wrist radiographs and growth velocity obtained from serial height measurements, along with many of the same variables routinely recorded by orthodontists—parents' heights, patient's gender, age, height and menarchal status. In regards to parents' heights, the medical prediction models incorporate what is referred to as midparental stature [43, 44]. This calculation is made by simply taking the mean of two values: 1) the height of the same gender parent, and 2) a "gender-adjusted" height of the opposite gender parent. The gender adjustment is made by altering the height of the opposite gender parent by 5.12 inches (adding 5.12 inches to a female parent when used for male subjects, or subtracting 5.12 inches from a male parent for female subjects). A Pubmed search for, "midparental height," or "midparental stature," limited to the dental (including orthodontic) literature did not yield any results, suggesting that this calculation may not be employed by orthodontists despite the (assumed) widespread recording and usage of parental heights in treatment planning. The midparental height calculation, which was introduced by Tanner [43, 44], has shown to predict adult stature with a correlation coefficient of 0.70 and its incorporation into predictive models for statural height has greatly increased their accuracy. For example, Bayley and Pinneau [42] created a predictive model based on statural height and skeletal age alone, which was shown to lack enough accuracy to be widely used in practice. Later, Bayley [48] added midparental height to his prediction model and greatly increased its accuracy. The Roche-Wainer-Thissen (RWT) method [45, 46] which was developed later likewise incorporated midparental height and has likewise proven to be relatively accurate. Such models have been shown to predict adult stature with a prediction accuracy ranging from 84% to 96% [42, 45, 46].

Given the relative success of these multiple-variable medical models at predicting statural height, is seems possible that similar models could be created to predict mandibular growth. Moreover, given

that mandibular growth is closely associated to statural growth [15, 16, 21-24, 49, 50], it seems likewise plausible that an effective mandibular growth prediction model could be created using many of the same variables used in the prediction of statural height.

A simple, yet accurate and reliable method of mandibular growth prediction would be a powerful clinical tool for the orthodontic practitioner. Simplicity demands variables that are efficiently and non-invasively obtained. Accuracy requires a model that can effectively stray from populations averages, tailoring prediction to the individual. Reliability necessitates clearly defined, reproducible, and non-subjective techniques for obtaining the values of the variables in the model. Despite several decades of attempts at mandibular growth prediction, a growth prediction model that is simultaneously simple, accurate, and reliable still does not exist.

The purpose of this study is to evaluate the accuracy of several simple mandibular growth prediction models which use only that information readily and reliably obtained in the routine workup of an orthodontic patient: patient's gender, age, height, weight, menarchal status, maxillomandibular relationship, and mandibular length; parents' heights; and published mandibular growth curves. Evaluating the accuracy of these prediction models provides further insights into the predictive value of each individual variable and the clinical utility of collecting such data for the purposes of predicting a patient's mandibular growth.

Materials and Methods

The protocol for this retrospective study was reviewed and approved by the Institutional Review Board at the University of Washington, Seattle.

Patient Sample

All data were obtained from patient charts at the University of Washington, Department of Orthodontics. All charts that had not been moved to an offsite storage facility were eligible for inclusion, primarily consisting of patients treated within the last 10 years. Inclusion in this study was based on age and availability of data. For inclusion, patients had to be between 9 years 0 months old and 16 years 11 months old at the time of the initial lateral cephalometric radiograph (T1). I roughly estimate that the average adolescent orthodontic patient presents between the ages of 12 and 13. My sample therefore seeks to encompass those of average age at presentation ± 3 years. Additionally, patients had to have all of the following data available in their charts for inclusion: parental heights, patient's height within a month of T1, patient's weight within a month of T1, patient's date of birth, patient's gender, and initial and final lateral cephalometric radiographs of good quality. Patients with a known craniofacial anomaly were excluded from participation in this study.

Data Collection

At the initial records appointment, height was determined for each patient by using either a wall-mounted ruler or stadiometer and recorded to the nearest 0.5 inches. Weight was determined by means of a standard mechanical scale. The patient's date of birth, menarchal status, and the height of both of his/her biological parents were obtained by verbal report from the patient, parent(s), or guardian(s) present at the initial records appointment. Both height and weight were recorded by the treating resident in the patient's chart. Lateral cephalometric radiographs were also obtained at the initial records appointment.

For each patient meeting the inclusion and exclusion criteria, the following data were collected from his/her chart: patient's gender, date of birth (month and year), height, weight and menarchal status (premanarchal, postmenarchal, or not specified); parents' heights. The following cephalometric

measurements were also calculated and recorded for each patient: A-point – nasion – B-point (ANB) angle (Fig 1) using a cephalometric protractor to the nearest 0.25 degree at T1, articulare-pogonion (Ar-Pg) length (Fig 1) using digital calipers to nearest hundredth of a millimeter at T1 and T2, and other measurements related to ArPg (see 'Magnification Correction') at T1 and at T2. T1 was defined as the date of the first cephalometric radiograph taken after the age of 9 years 0 months for which there was also corresponding patient height and weight data recorded within one month of this date. T2 was defined as the date of the cephalometric radiograph closest to 33 months after T1. For all surgical cases, the T2 cephalometric radiograph had to be a pre-surgical one. The 33 months value was derived from evaluating the intervals of all radiographic timepoints for all patients, including initial, progress, and final radiographic timepoints. The median value of these intervals was 33 months, which also roughly corresponded to the center of the bell curve distribution of intervals. The decision to keep the T1-T2 intervals as close to 33 months as possible was made for two reasons. First, improving the uniformity of the interval lengths improves the accuracy of the prediction models. Second, 33 months represents an interval within which an orthodontist is very likely to complete patient care, making my prediction model useful within this clinical range. Data were recorded by six different evaluators; the principal investigator and five dental students who were trained by the principal investigator. To evaluate the inter-rater reliability, all T1 and T2 cephalometric measurements were repeated by all evaluators on a random subset of 20 patients and intra-cluster correlation coefficients calculated.

Magnification Correction

All lateral cephalometric radiographs used in this study were taken using a 1986 cephalometric radiograph unit, model 6520.607, from the Continental X-ray Corporation, Chicago, IL, USA. For all radiographs, this machine had a standard source to object distance. The object to film distance was not standardized during this study and ranged from 11.6 cm to 20.8cm. This machine also lacks a ruler in

the nosepiece or elsewhere to be used in adjusting for the magnification differences arising from the different object-film distances. I therefore developed a novel magnification-correction technique which utilizes the wooden arms that hold the ear rods of the cephalometric machine to standardize the magnification of the radiographs.

In the range from 35mm to 55mm above the center of the ear rods, the width of the wooden arms that hold the rods is exactly 34mm, serving as the "absolute arm width." For each radiograph, the width of both the near and far arms were measured 45mm above the ear rods (mechanical porion – Po; see Fig 1) and the mean value recorded, yielding a "radiographic arm width". The "absolute arm width" was then divided by the "radiographic arm width" to provide a "magnification factor." To standardize the ArPg measurements, each was multiplied by its corresponding "magnification factor" to yield an "absolute ArPg length." All ArPg lengths used in this study are the absolute lengths and therefore represent the patients' actual (unmagnified) mandibular lengths.

Validation of Magnification Correction Technique

In order to evaluate the accuracy and reliability of the above magnification correction technique, I placed midline radio-opaque markers on nine dry human skulls obtained from the Department of Biologic Structure, University of Washington. Each radio-opaque marker consisted of a paperclip bent 90 degrees from the skull surface at an arbitrary location along the midline of the frontal bone (Fig 2). Lateral cephalometric radiographs were taken of each dry skull at three different object-film distances (13 cm, 16.5 cm, 20 cm) on the same machine that was used for all the radiographs in this retrospective study. For each radiograph, all six evaluators independently measured the widths of both the wooden arms 45mm up from the ear rods to obtain the "radiographic arm width" and to calculate the "magnification factor." The evaluators also measured the distance from anterior nasal spine (ANS) to the tip of the paperclip (radio-opaque marker) for each radiograph. The radiographic ANS-paperclip

lengths were then multiplied by the corresponding "magnification factors" to obtain the "estimated absolute ANS-paperclip lengths". The absolute distance from ANS to paperclip was also measured directly on each skull, three times on three separate days. The mean value of the three measurements was calculated to obtain the "true absolute ANS-paperclip length" for each skull. The absolute value of the difference between the "estimated absolute ANS-paperclip lengths" and the "true absolute ANS-paperclip lengths" yielded absolute errors. The mean absolute errors (MAEs) were used to evaluate the accuracy of this method.

Mandibular Growth Curve

A mandibular growth curve representing a historical population average is needed for two reasons. First, it serves as a standard against which I can compare the accuracy of my prediction models. Second, it is included in some of the prediction models as a starting point, producing models that predict how much a patient will differ from the population average based on certain predictor variables. To be appropriate the historical population used to create the growth curve must match my study population as much as possible, particularly in terms of ethnicity, and enrollment date. The data from the historical population also have to use the same mandibular measurement, ArPg, provide a way to account for radiographic magnification, and include a sufficient number of patients to be a reasonable representation of the general population. With these parameters in mind, and an assumption that patients treated at the University of Washington are primarily of Caucasian origin, I selected and pooled the mandibular data from three longitudinal growth studies: the Michigan Growth Study, the London Growth Study, and the Philadelphia Growth Study. The data from these studies are published in their respective atlases [50-53]. The Michigan Growth Study took place between 1953 and 1966 at the University of Michigan Dental School. The atlas presents the data of 47 males and 36 females documented from 6 to 16 years of age. These subjects represent an "average school population" from

within this geographic region. The ArPg lengths are not corrected for linear enlargement and represent a 12.9% radiographic magnification. The London Growth Study took place between 1952 and 1993 at King's College School of Medicine and Dentistry in London, UK. Following a rigorous selection process based on the quality and number of records, 58 males and 63 females were chosen from among the 736 original participants "of Caucasian origin" for inclusion in the atlas. These individuals were documented from 4 to 20 years of age. The ArPg lengths are presented as natural (unmagnified) dimensions. The Philadelphia Growth Study was conducted between 1948 and 1968 at the Philadelphia Center for Research in Child Growth. Participants were selected from European ancestry with "good medical and dental health". While there are conflicting reports of the data and methods it is assumed that roughly 50 males and 50 females took part in the study. These subjects were documented from 6 to 25 years of age. The ArPg lengths are not corrected for linear enlargement and represent a 6.0% radiographic magnification.

The ArPg data from the Michigan and Philadelphia studies were converted to the natural (unmagnified) dimensions and then pooled with the London study for ages 9 through 20 for each gender, an age-range that encompasses all T1 and T2 ages in the present study. To allow for calculation of average ArPg at any precise age between annual intervals a mathematically defined polynomial curve was constructed to represent the data for each gender. To accomplish this, five polynomial curves, ranging from second- to sixth-degree polynomials were each fit to the mean male and female datapoints respectively using a best fit "least squares" technique. "Least squares" means that the overall solution minimizes the sum of the squares of the errors made, where the errors are the absolute values of the differences between the actual data at each age and the estimates made by the polynomial curve for each age. I then assessed the accuracy of the fit of the various curves to the ArPg data by calculating the mean absolute errors (MAEs). The polynomial curve with the least MAE was used to represent the ArPg data for each gender. Higher degree polynomials were not assessed for two reasons: 1) to avoid over-fitting the curves to the

data, 2) to eliminate Runge's phenomenon, the highly oscillatory nature of high order polynomials at the edges of an interval [54]. In addition to serving as a standard against which I can evaluate the accuracy of my prediction models, the ArPg polynomial growth curve (ArPg PGC) for each gender served to generate "expected" changes in ArPg length (Exp ΔArPg) from T1 to T2 for each patient. Exp ΔArPg was used as a starting point in some of the prediction models.

Model-Building

Overview**.** Models to predict 1) ΔArPg between T1 and T2, and 2) the difference between ΔArPg and Exp ΔArPg (ΔArPg - Exp ΔArPg) were constructed using a best fit least squares method. Separate prediction models were constructed for males and for females. For females separate prediction models were constructed using a subset of the data for which menarchal status was known. For each gender the data were randomly divided into two roughly equal-sized subsets by assigning each patient either a "1" or a "2" using randomly generated strings of 1s and 2s from the website random.org. The 1s were assigned to the model building set, while the 2s were assigned to the validation set. Prediction models were constructed using data exclusively from patients in the model-building sets. The best models were then chosen and the accuracy of these models assessed using the validation sets.

Predictor variables. The predictor variables used to construct the prediction models were:

- T1 Age (years): Calculated as the difference between first day of the patient's birth month (exact birth dates not recorded), and the exact T1 date, and converted to a decimal.
- T1-T2 interval (years): Calculated as the difference between exact T1 and T2 dates, converted to a decimal.
- T1 ArPg (mm): Calculated using the above magnification correction technique, thus representing the absolute (unmagnified) linear measurements.
- T1 ANB (degrees)

- T1 height (inches)
- T1 weight (lbs)
- T1 BMI percentile (%): Raw BMI scores were calculated using height and weight data. The raw scores were then converted to BMI percentiles using the gender-specific growth charts from the Centers for Disease Control (CDC) [55].
- Mother's height (inches)
- Father's height (inches)
- Midparental height (inches): Using the method outlined by Tanner et al [43], the mean of two values was calculated for each patient: 1) the height of the same gender parent, and 2) a "gender-adjusted" height of the opposite gender parent. The gender adjustment was made by altering the height of the opposite gender parent by 5.12 inches (adding 5.12 inches to a female parent for male patients, or subtracting 5.12 inches from a male parent for female patients).
- Height remaining (inches): Estimated by subtracting the patient's T1 height from the midparental height.
- Menarchal status: A binary variable represented by a 0 (premenarchal) or a 1 (postmenarchal) and used only on the subset of female patients for which menarchal status was known.
- Overweight: A categorical variable, which applies a coefficient in the prediction models to all patients with BMI percentiles >85% and <95%, thus following the convention established by the CDC [55].
- Obese: A categorical variable, which likewise applies a coefficient in the prediction models to all patients with BMI percentiles ≥95%, thus following the convention established by the CDC [55].
- Prognathic: A categorical variable, which applies a coefficient in the prediction models to all patients with an ANB angle $\leq 1.5^{o}$.

- Retrognathic: A categorical variable, which applies a coefficient in the prediction models to all patients with an ANB angle ≥5°.

Prediction outcome. Foundational to constructing the prediction models was deciding what aspect of mandibular growth to predict. The options considered were: T2 ArPg, ΔArPg, and (ΔArPg - Exp ΔArPg). Of these options, ΔArPg seemed the most appealing since a clinician could make immediate use of the prediction without measuring a cephalometric radiograph or using a growth curve as a starting point. Predicting (ΔArPg - Exp ΔArPg) was also appealing given that it incorporates the non-linear growth curve into an otherwise linear prediction model. After a preliminary evaluation of various models, it became clear that: 1) models predicting T2 ArPg are extremely inaccurate without T1 ArPg as a covariate, and 2) models predicting T2 ArPg with T1 ArPg as a covariate have the exact same MAEs as models predicting ΔArPg. Given these findings, I chose to eliminate T2 ArPg as a prediction outcome and to focus on predicting ΔArPg and (ΔArPg - Exp ΔArPg).

Age input. Age can be input into the prediction models in various ways. Three inputs were explored, inputting age: as a linear covariate, as a set of categorical variables, and as a set of polynomial covariates. Since ArPg changes with age in a non-linear manner, inputting age as a linear covariate may produce models with reduced accuracy. Inputting age as a set of categorical variables could potentially overcome this deficiency by creating a distinct coefficient for each distinct age category. The seven categories used were the following ages at T1: 9≤age<10, 10≤age<11, 11≤age<12, 12≤age<13, 13≤age<14, 14≤age<15. A category for 15≤age was not needed since this is the default when the other categories are not true, and adding it would be a redundancy that would be rejected by the statistical software used in this study, “R”. Inputting age as a set of polynomial covariates is another non-linear input created by fitting a mathematically defined polynomial curve to the data. Five polynomial functions were explored for age at T1: second-, third-, fourth-, fifth-, and sixth-degree polynomials.

For models predicting ΔArPg, all three types of age input were assessed. When predicting (ΔArPg - Exp ΔArPg), age was input as a linear covariate only since the non-linear polynomial-defined growth curve (the ArPg PGCs) for each gender was already being used as a starting point. In addition to having one of three inputs for T1 age, T1-T2 interval length was also a necessary input due to the heterogeneity of interval lengths in this study. T1-T2 interval length was thus input into every model with the assumption that a clinician would use the model to predict the change in mandibular length over a predetermined interval.

Model-building: The model-building process was completed by constructing and evaluating as many unique models as possible. Every possible combination of the remaining 14 predictor variables was used to construct unique models for each of the age inputs and predictor outcomes described above for both males and females from within the model-building sets. Models were similarly constructed on the subset of females for whom menarchal status was known. Using the statistical software, "R", each model was evaluated for predictive value using a "leave one out" cross-validation approach. In this approach, using a best-fit least squares linear method, each model is fit to the data from all patients within the model-building data set except one, who is left out. The resulting model is then applied to the patient who is left out and the prediction error is calculated. The prediction error is equal to the absolute value of the millimetric difference between the value predicted by the model and the true value for that patient. This process is repeated, leaving one patient out at a time until every patient has been left out once, with the prediction error of the model calculated for each patient. The mean of the prediction errors, which is equivalent to the mean absolute error (MAE), is then calculated and used to evaluate the accuracy of the model from within the model-building sets.

Model analysis: The models were grouped by prediction outcome for each gender: ΔArPg or (ΔArPg -Exp ΔArPg). The ΔArPg group was further subdivided into groups based on which age input was used: linear,

categorical, or polynomial. The models within these 12 groups (four for each gender, and four for the female subset for whom menarchal status is known) were then ranked by their MAEs. The models with the lowest MAE in each group (the "top" models) then had their MAEs compared to the MAEs obtained when using the ArPg PGCs as prediction models. Those groups whose top models did not have MAEs lower than those obtained from the ArPg PGCs were not analyzed further. In the remaining groups, the three models with the lowest MAEs were selected for final analysis. The final analysis consisted of: 1) plotting absolute errors against fitted values to assess the linearity of the fit, and 2) an assessment simplicity, where the models comprised of the fewest covariates are considered superior.

Model validation: After analysis, one model was selected from each group for each gender and for the subset of females for whom menarchal status was known. The best fit least squares method was used to fit all of the model-building data to each model (ie- not leaving anything out). The coefficients for the variables comprising each model were then applied to the validation sets of data for each model, and the MAE calculated for each. The male and female ArPG PGCs were also applied to the validation sets and MAEs calculated. The MAEs of the prediction models and those of the ArPg PGCs were finally compared to determine which prediction models are more accurate than a historical population average (the ArPg PGCs).

Female Substratum Based on Menarchal Status: As stated previously, for each of the female groups, a substratum was created by pooling all those for whom menarchal status is known (84% of females). Menarche was used as a binary covariate in all the models in this substratum (premenarchal = 0, postmenarchal = 1).

Results

Validation of Magnification Correction Technique

The absolute ANS-paperclip distances used ranged from 85.85 to 108.29mm, roughly corresponding to the range of ArPg values measured in this study. The MAEs of my method for magnification correction are summarized in Table 1. The accuracy of this method is represented by the overall MAE of 0.60mm. To assess possible differences in accuracy between skulls, the MAEs of the 9 skulls were calculated, with MAEs ranging from 0.47mm to 0.71mm. MAEs were calculated for the three object-film distances to determine if the accuracy of this method differs with differing object-film distances. The MAE was 0.55mm for the near distance, 0.66mm for the middle distance, and 0.59mm for the far distance. Comparing the MAEs of the six judges, which ranged from 0.49mm to 0.66mm, gives a sense for the reliability of this method.

Patient Data

There were 447 patients who qualified for inclusion in this study, 218 males and 229 females. The means and standard deviations of the data collected are summarized in Table 2. The mean age at T1 was 12.58 for males and 12.63 for females with standard deviations of 1.77 and 1.82 years respectively. These ages are fairly representative of the age of adolescent patients beginning orthodontic treatment. The T1-T2 intervals for males and females of 2.87 and 2.68 years likewise approximate typical treatment intervals. Figure 3 details the distribution of T1-T2 intervals for both genders combined. The vast majority of these intervals lie between two and four years, with the data normally distributed. The change in ArPg lengths over these intervals differed greatly between males (6.25 mm) and females (3.52 mm), with the change in ArPg being 78% larger for males (Table 2). This difference highlights the much higher rate of mandibular growth among males over the age range of this study. Several patients

differed from the mean, however, as illustrated by the very large standard deviations for both males (3.6mm) and females (3.0mm). The ANB angles at T1 were very similar for males and females, with the averages for both groups similar to that reported by other studies [51] at 3.7° with standard deviations of roughly 2.4°. Patient's height, weight, and BMI percentiles were pretty similar for males and females, as were parents' heights. It is interesting to note that the BMI percentiles for both males (63%) and females (61%) are higher than the average established by the CDC in 2000 (50%). Males had mid-parental heights and heights remaining that were roughly 4 inches greater than that of the females on average.

Figure 4 shows the temporal distribution of T1 dates for both males and females. All patient charts that had not been moved to an offsite storage facility were eligible for inclusion, which consisted mostly of patients who presented after 2000. There were, however, a quarter of the patients who presented between 1995 and 2000; and ~1% who presented before 1995.

Inter-Rater Reliability

The determinations of ΔArPg involved six measurements: ArPg at T1 and T2, and the widths of the near and far ear rod arms at both T1 and T2. The inter-rater reliability of the ΔArPg determinations as calculated by an intra-cluster correlation coefficient (ICC) was 0.96 (0.93-0.98, 95% CI). The ANB measurements involved only one measurement per patient. The inter-rater reliability study for this measurement revealed an ICC of 0.98 (0.96-0.99, 95% CI).

Model Building

Following randomization, 113 males were allocated to the model-building set and 105 to the validation set (Fig 5). For females, 113 were allocated to the model-building set, while 116 were allocated to the validation set. Females were further subdivided into model-building and validation subsets comprising

just those patients for whom menarchal status is known. These subsets consisted of 94 females in the model-building subset and 99 patients in the validation subset.

Table 3 shows the mean and standard deviations of the model-building and validation sets for both males and females. Randomization led to very similar groups. There was, however, a statistically significant difference between model-building and validation sets for males for three of the patient variables: T1 age, T2 age and T2 ArPg.

Mandibular Growth Curves

The ArPg data which were pooled from the Michigan Growth Study, London Growth Study, and Philadelphia Growth Study are shown in Figure 6. Also shown are the associated ArPg PGCs, with their respective polynomial expressions shown below the figure. Sixth-degree polynomials were used for both males and females as these had the best fit (least MAE) of the different order polynomials tested. Each ArPg PGC mathematically represents the historical population average.

Accuracy of Prediction Models

Table 4 shows the number of unique models possible for each prediction outcome and age input tested. Whether predicting ΔArPg or (ΔArPg – Exp ΔArPg), the number of combinations for each age input is 8192 due to the fact that there are 13 different covariates used in all possible combinations to create the various models (age and T1-T2 interval being common to all models, and menarche being common to all models in that subset). The exception is the group of models for which age was input as a polynomial. When constructing all possible models for this age input, all polynomials were evaluated from 2^{nd} degree to 6^{th} degree polynomials, making the number of unique combinations five times greater at 40960. Of these unique combinations, the MAEs of the most accurate models from the leave-one-out cross-validation approach on the model-building sets are summarized in Table 4. Also shown in

the table, are the MAEs of the ArPg PGCs when applied to the model-building sets. The male ArPg PGC (M ArPg PGC) had an MAE of 2.05mm, while the female ArPg PGC (F ArPg PGC) had an MAE of 1.55mm when applied to the entire female model-building set and an MAE of 1.47 when applied to the female model-building subset for which I have menarche data. The MAEs marked by asterisks are those corresponding to models more accurate than the ArPg PGCs. For the male and female model-building sets, even the most accurate models failed to be more accurate than the growth curves when predicting ΔArPg. When predicting (ΔArPg – Exp ΔArPg), on the other hand, the top models were slightly more accurate than the growth curves. It is important to note, however, that these models used the ArPg PGCs as a starting point (Exp ΔArPg), and predicted how the actual ΔArPg would differ from the expected values obtained from the ArPg PGCs. When applying the models to the subset of females for which there are menarche data, the models with the lowest MAEs are each more accurate than the ArPg PGCs for all prediction outcomes and age inputs tested.

The top three models chosen from within the model-building sets with linear age inputs predicting (ΔArPg – Exp ΔArPg) are summarized in Table 5. For males the most accurate models, organized in ascending order of MAE, are labeled “M1”, “M2”, and “M3”. When taken to the nearest hundredth of a millimeter, each has a MAE of 1.94mm. M1 consists of four covariates: linear age, father’s height, patient’s height, and patient’s weight. M2 also utilizes linear age and father’s height, but adds obese as a categorical variable. M3 comprises the same covariates as the first with the exception of father’s height. All three models have similar linearity (data not shown). Of these three models, M3 was chosen for testing on the validation set of data as it was comprised of the fewest covariates. For females the most accurate models, organized in ascending order of MAE, are labeled “F1”, “F2”, and “F3”. Each has an MAE of 1.49. Obese is an additional covariate in F1 and weight is an additional covariate in F2. All three models have similar linearity (data not shown). Of these three, F3 was chosen for testing on the validation set because it has the fewest covariates. For the subset of females with menarche data, the

three most accurate models are labeled, “FM1”, “FM2”, and “FM3”, organized by MAE in ascending order. Their MAEs are 1.31mm, 1.32mm, and 1.32mm respectively. All utilize linear age and menarche as covariates, and obese as a categorical variable. FM2 uses mother’s height as a covariate, while FM3 uses prognathic (ANB≤1.5) as a categorical variable. All three models have similar linearity (data not shown). FM1 was chosen for testing on the validation set because it has the lowest MAE and the fewest covariates.

The top three models with ΔArPg as the prediction outcome chosen from within the model-building subset for which I have menarche data are summarized in Table 6. All of the top three models for each of the age inputs utilize menarche as a covariate. The top three models using age as a linear covariate are labeled “FML1”, “FML2”, and “FML3”. These are arranged in ascending order based on MAE. Rounded to the nearest hundredth of a millimeter, each has a MAE of 1.38mm. Each uses BMI percentile as a covariate and obese as a categorical variable. FML1 and FML3 also utilize ANB, while FML3 adds father’s height as a covariate. All three models have similar linearity (data not shown). FML2 was chosen for testing on the validation set, given that it has the fewest covariates. The top three models which use age as a set of categorical variables are labeled “FMC1”, “FMC2”, and “FMC3”, arranged in ascending order by MAE. These have MAEs of 1.37mm, 1.38mm, and 1.38mm respectively. Each uses BMI percentile, ANB and height as covariates, and obese as a categorical variable. FMC2 uses prognathic (ANB ≤ 1.5) as a categorical variable, while FMC3 uses ArPg as a covariate. All three models have similar linearity (data not shown). FMC1 was chosen for testing on the validation set because it had the lowest MAE and the fewest covariates. The top three models which input age as a polynomial are labeled “FMP1”, “FMP2”, and “FMP3”, arranged in ascending order of MAE. When rounded to the nearest hundredth of a millimeter, each has a MAE of 1.33mm. All of these models have similar linearity (data not shown). Each utilizes 5^{th} degree polynomials as covariates. FMP1 was chosen for testing on the validation set as it was comprised of the fewest covariates.

The models chosen using the model-building sets are shown in Tables 7 and 8. The MAEs of these models when applied to the validation sets are shown. Also shown for comparison are: 1)the MAEs and 25th and 75th percentiles of the absolute errors (AEs) calculated from within the model-building sets using the leave-one-out cross-validation as described above, and 2) the MAEs and the 25th and 75th percentiles of the AEs calculated when applying the ArPg PGCs to both the model-building and validation sets. When predicting (ΔArPg – Exp ΔArPg) and inputting age as a linear covariate (Table 7), the chosen models either fail to be more accurate on the validation data sets than the historical population averages (ArPg PGCs) or are insignificantly more accurate. The male model has an MAE of 1.73mm on the validation set compared to the MAE of the male ArPg PGC at 1.80mm. For the female model, the MAE is 1.61mm when applied to the validation set compared to 1.60mm for the female ArPg PGC. The model chosen for the female subset for which I have menarche data has an MAE of 1.57 when applied to the validation set, compared to 1.58mm for the F ArPg PGC.

When predicting ΔArPg on the female subset for which I have menarche data (Table 8), only FMP1 has a lower MAE (1.52mm) than the female ArPg PGC (1.58mm) when applied to the validation set. In comparison, FML1 has an MAE of 1.60mm, while FMC1 has an MAE of 1.64mm when applied to the validation set.

Discussion

Articulare (Ar) is defined as the point of intersection of the posterior border of the mandibular ramus and the inferior border of the basilar part of the occipital bone [56] (Fig 1). Strictly speaking, therefore, Ar is not a purely mandibular landmark, but rather a constructed point based on the relative position of

the ramus to the occipital bone. For this reason, some have disputed its use in measuring mandibular length [57, 58], arguing that Ar is affected by mandibular posture. It is undisputed that condylion-pogonion (CoPg) is a more true measure of mandibular length than ArPg. Nonetheless, there is a precedent for the use of ArPg as a surrogate for mandibular length in the orthodontic literature [25, 59]. It is an attractive surrogate due to the ease of its identification relative to condylion (Co). Co is often obscured by the superimposition of various dense radiopaque structures in the cranial base and middle cranial fossa [60]. Several studies have therefore shown that Co cannot be accurately and consistently located on routine (closed-mouth) lateral cephalometric radiographs [61, 62]. Moreover, the validity of using ArPg as a measure of mandibular length has been supported in prior studies. One study in particular reports a correlation coefficient between CoPg and ArPg of 0.93 [59]. Given that Ar is more reliably identified than Co, and with prior reports showing ArPg to be a good proxy measurement for CoPg, I felt comfortable using ArPg to represent mandibular length in this study. It is important to note, however, that many of my T2 radiographs are of patients with treatment in progress. Orthodontic appliances can often rotate the mandibular plane clockwise, even if for a transient period of time during the orthodontic treatment. Such increases in mandibular plane angle could potentially increase the ArPg measurement, since Ar would be higher up the posterior border of the condylar neck. Any mandibular posture-related increases in ArPg are assumed to be trivial.

One of the major challenges to gathering retrospective data from lateral cephalometric radiographs is the magnification differences between radiographs. Most new machines have a ruler fixed in the nosepiece to allow for standardization with other machines that may have different source-object or object-film distances. Many older machines, however, lack this feature. Furthermore, some older machines allow for variable object-film distances like the one used for the radiographs in this study. If this distance varies between two radiographs, their linear measurements cannot be compared without correcting for the magnification differences. My method to standardize the magnification of the

radiographs in this study proved to be both accurate and reliable. The accuracy was demonstrated by comparing my technique to absolute measurements on dry skulls, the MAE being 0.60mm. The reliability is perhaps even more important and is shown by an intra-class correlation coefficient of 0.96 for ΔArPg determinations in this study. As stated previously, these determinations were made as a result of six measurements. Taken together, these results indicate this method to be a viable magnification-correction technique, which may be useful for institutions facing similar challenges when gathering retrospective data from lateral cephalometric radiographs.

This study relied on growth curves which represent the population average. These growth curves were essential for evaluating the accuracy of the various models and were used as a starting point in some of the models. Finding appropriate data for this purpose is particularly challenging. For a growth curve to be appropriate for comparison with a study population, three conditions must be met: 1) the landmarks that define mandibular length must be the same between the growth curve data and the study population data, 2) the radiographic magnification must be standardized between the growth curve data and the study population data, 3) the population used for the growth curve must be similar to the study population, particularly in terms of ethnicity, enrollment date, and geographic location. Another challenge is compiling a large enough population of individuals with longitudinal data so as to be truly representative of the population average. This is difficult for several reasons. First, the same three conditions must be met when pooling data from different longitudinal growth studies, however, mandibular length can be reported as ArPg, CoPg, Co-gnathion, sella-gnathion, or other measurements; radiographic magnification is typically not reported; and there is a wide temporal distribution of enrollment among the various longitudinal growth studies. Second, there is a scarcity of publicly available longitudinal cephalometric data. Longitudinal growth studies which expose patients to annual cephalometric radiographs are no longer possible given the ethics of exposing them to radiation now known to increase cancer risk, so the limited data we now have are all that are and will be available.

Several longitudinal growth studies have been done since the 1920s, including 12 in North America, but much work needs to be done to preserve these data, determine their radiographic magnification, and to make them more publicly accessible. The first step to both preservation and public accessibility is to digitize the radiographs. Secondly, these data need to be made available online for researchers to have ready access. There has been an effort by the American Association of Orthodontists Foundation (AAOF) to accomplish this. The AAOF has compiled data from nine different longitudinal growth studies from North America: the Bolton-Brush, Burlington, Denver, Fels, Forsyth, Iowa, Mathews, Michigan, and Oregon Longitudinal Growth Studies. The digitizing and compiling of these data, which make up the AAOF Craniofacial Growth Legacy Collection, is an important step toward preservation and public accessibility, but much still needs to be done. For example, only a limited subset of subjects from each of the growth studies is available on the AAOF Legacy Collection website (www.aaoflegacycollection.org). Also, nowhere on the website is there a report of the radiographic magnification of the cephalometric radiographs in each study. These challenges, along with the fact that mandibular length is only reported as CoPg, made these data useless for the present study.

After an exhaustive PubMed and internet search, I identified three longitudinal growth studies with data deemed appropriate for this study—the Michigan, Philadelphia, and London growth studies. The data from these studies are reported in their respective atlases as well as by Dibbets et al [63]. These studies: 1) use ArPg as the measure for mandibular length, 2) have a known radiographic magnification, 3) comprise individuals from North America or Europe of Caucasian origin, and 4) have enrollment dates after 1948. I chose to use data from the London Growth Study despite the differing geographic location because its growth curve has been shown to be extremely similar to that of the Michigan Growth Study [63]. It was important for the date of enrollment to be as late as possible given the secular trend to human growth, which shows that men and women are reaching puberty at increasingly earlier ages [64, 65]. This trend can impact the slope and shape of the mandibular growth curve for a population, with

average subjects after 1960 reaching the onset and peak of the pubertal growth spurt at an earlier age than average subjects before 1960 [66, 67]. The pooled data from these three longitudinal growth studies likely reflect this secular trend to some degree when compared to my study data, which has enrollment dates from 1986 to 2010. I chose to accept this limitation as there are no longitudinal cephalometric growth studies with enrollment dates coinciding with the temporal distribution of my study population, and these three studies are some of the later growth studies done among the various studies.

After identifying appropriate longitudinal data to use in creating population average growth curves, I chose to fit the data to mathematically-defined polynomials. It is extremely helpful to have mathematical expressions to define these curves as it allows for easy calculation of the average change in mandibular length between any two precise ages. However, forcing the data into a polynomial curve may lead to inaccuracies. Lower order polynomials tend to over-smooth growth curve data, minimizing the fluctuations in the slope of the curve. On the other hand higher order polynomials can over-fit the data, producing a near-exact match of all fluctuations of slope. Higher order polynomials also tend to be highly oscillatory near the edges of the curve, leading to inaccuracies (Runge's phenomenon) [54]. I hoped to limit such inaccuracies by finding the best-fit polynomial between second- and sixth-degree polynomials. If the curve was over-smoothed by the lower order polynomials this would be reflected in an increased MAE. Also, limiting the higher degree polynomials to sixth-order avoids both over-fitting and Runge's phenomenon. Despite the shortcomings of fitting the data to mathematically-defined polynomials, the resulting ArPg PGCs proved to be fairly accurate predictors of ΔArPg. These curves in fact did better at predicting ΔArPg than half of the final prediction models chosen for validation. And even the prediction models that were more accurate the ArPg PGCs were insignificantly more accurate, being 0.07mm (M3), 0.01mm (FM1), and 0.06mm (FMP1) more accurate. The relative accuracy of these polynomials therefore serves to validate: 1) this method for expressing mandibular growth curves, and

2) the appropriateness of using data from the Michigan, Philadelphia, and London growth studies to estimate the growth of more recent populations.

In addition to predicting ΔArPg, I also constructed models which sought to predict (ΔArPg – Exp ΔArPg), where Exp ΔArPg is the prediction based on the ArPg PGCs. These models can be viewed as predicting the magnitude and direction an individual will stray from the population average based on the values of certain individual characteristics—the covariates in the models. It is interesting how poorly these models performed. None of the covariates tested could improve upon the population-average-based prediction by more than 0.07mm. This finding suggests that collecting and analyzing such data as found among the covariates used in this study is a particularly ineffective approach to discerning how an individual's growth will stray from the population average.

Collecting data regarding the patient's and parents' heights is a very wide-spread practice among orthodontists trying to establish the growth potential or current growth rate of a given patient. In this study I used five covariates relating to height—patient's height, father's height, mother's height, mid-parental height, and height remaining. The latter was formulated to most directly encompass what an orthodontist seeks to evaluate when comparing a patient's height to that of the parents. It is interesting that height remaining was not used as a covariate in any of the top models for either gender, whether predicting ΔArPg or (ΔArPg – Exp ΔArPg), suggesting that a direct comparison between a patient's height and that of both parents is not a useful tool in assessing a patient's mandibular growth rate. Mid-parental height was not used by any of the top models, suggesting its lack of utility as well. In fact, of the three models which proved to be more accurate than the ArPg PGCs when applied to the validation sets (M3, FM1, and FMP1), a height-related variable was used by only one (M3, patient's height). These findings suggest that such height information is not significantly useful in mandibular growth prediction.

Growth velocity for statural height has been shown to be correlated strongly to growth velocity for mandibular length [15, 16, 21-24, 49, 50]. Most recently, Mellion et al. 2013 [50] used 100 patients from the Bolton-Brush Longitudinal Growth Study to show that the *onset* of pubertal growth spurts in statural height and mandibular length occurred in females at ages 9.3 and 9.5 years respectively, and in males at age 11.9 years. For females this age difference was statistically significant. They also showed that the *peak* of the growth spurt in statural height and mandibular length occurred in females at ages 10.9 and 11.5 years respectively, and in males at ages 14.0 and 14.3 years respectively. For both males and females, this difference was statistically significant. Based on these findings, for males the onset and peak of mandibular growth is nearly coincident between statural height and mandibular length, while for females the onset and peak of statural growth spurt occurs a few months before that of mandibular length. With such a strong correlation between growth velocities in statural height and mandibular length as shown in these subjects, the use of longitudinal statural height data may have strengthened the accuracy of my prediction models. I did not have longitudinal data on statural height to use in my models, but used instead a comparison between a patient's statural height and that of his/her parents. This comparison proved to be an inaccurate predictor of mandibular growth velocity.

Various studies have been done regarding the influence of a patient's weight on his/her individual growth curve, onset of puberty, and onset of menarche. One major theory is that for puberty to occur, the body must reach a certain body weight or fat mass [68]. Due to this phenomenon, diet and exercise are extremely influential in determining an individual's pubertal maturation [69, 70]. It is well-established now that overweight females mature earlier than lean females [71-73]. In addition, obesity, as measured by BMI percentile, has been shown to be significantly associated with early puberty in females [73]. Recently it has also been shown that skeletal maturity determined by cervical vertebral maturation (CVM) staging is more advanced in subjects with increased BMI percentiles [74]. In addition to its impact on puberty and skeletal maturation, it has been shown recently that obesity may impact

orthodontic treatment outcome. Von Bremen et al, 2013 suggest that obese patients may have slightly longer treatment lengths and require more appointments than their normal-weight peers [75].

In this study I used two covariates and two categorical variables relating to weight—patient's weight, patient's BMI percentile, overweight, and obese. Patient's weight was used in some of the top models in this study and was used in one of the three models with MAEs lower than that of the ArPg PGCs when applied to the validation sets, M3. BMI percentile was used in all of the top models from within the dataset for which I have menarche data, when predicting ΔArPg and inputting age as either a linear or categorical variable. None of these models, however, proved more accurate than the female ArPg PGC when applied to the validation set. Overweight did not appear as a categorical variable in any of the top models, but obese appeared in several, including nearly all of the female models with menarche as a covariate. One of the female models which were more accurate than the female ArPg PGC when applied to the validation set had obese as a covariate (FM1). These findings suggest that a patient's weight may play a role in his/her mandibular growth rate, albeit an extremely minor one.

Obesity rates in the United States have shown a dramatic increase in recent years among both children and adults [76]. This is a multifactorial epidemic, including a decrease in physical activity, an increase in caloric consumption, and a change in the societal perception of obesity. For US children, participation in structured physical activities at school is on the decline [77]. Other studies show that obese and overweight individuals in the US are increasingly less likely to perceive themselves as overweight or obese, thus reducing motivation for changes in diet or activity level [78]. Socioeconomic status also may play a role as reported by a North Carolina study showing a five-fold increase in obesity rates over the general population among Medicaid patients aged 13-16 [79]. Among the 447 subjects in my study, 141 (32%) were either overweight (BMI percentile > 85) or obese (BMI percentile >95) at T1 (64 females, 77 males), which is over double the portion of overweight and obese individuals in the general population

as reported by the CDC [55]. Of these, 61 (14% of study population), were obese at T1 (33 females, 28 males), nearly triple the obesity rate of the average population. In contrast, only 14 subjects (3%) were underweight (BMI percentile < 5) at T1 (8 males, 6 females), almost half the portion of underweight individuals in the general population. For many patients of low socioeconomic status, the University of Washington (UW) Orthodontics Clinic has been an attractive alternative to traditional private practices which typically quote much higher treatment fees. This has led to a greater representation of patients of low socioeconomic status at the UW clinic. It stands to reason therefore that the obesity rate among those in this study would be elevated relative to that of the general population, similar to the Medicaid population in the North Carolina study. The rate of 14% in the present study is in fact mid-way between the rate of the North Carolina study (25%) and that of the general population (5%). This finding may limit the generalizability of my study.

Several studies have suggested that there is a specific pattern of growth in patients who are maxillary retrognathic and/or mandibular prognathic, the so-called “Class III growth pattern” [80, 81]. Other studies suggest that there is likewise a distinct “Class II growth pattern” in patients who are maxillary prognathic and/or mandibular retrognathic [60]. Due to the characteristic growth patterns at the two extremes of jaw relationship, I chose to use one covariate and two categorical variables relating to antero-posterior (AP) maxillo-mandibular relationship—ANB angle, retrognathic (ANB≥5), and prognathic (ANB≤1.5). Given that ANB angle (Fig 1) is the most widely used method to assess the AP skeletal relationship, it was used in creating these covariates. ANB was used by most of the top female models for which there were menarche data for predicting ΔArPg. ANB was also used by one of the three models (FMP1) which was more accurate than the ArPg PGCs when applied to the validation sets. Despite the suggestion of a characteristic Class II or Class III growth pattern, the categorical variables, retrognathic and prognathic, were not used in most of the top models. Neither of these categorical variables was used in any of the final models applied to the validation sets. These results suggest that

knowing an individual's ANB angle is of limited to no value in determining his/her growth rate. It is also possible that my cut-offs for retrognathic and prognathic, which were somewhat arbitrarily chosen, could have been misplaced, therefore not truly isolating those with deviant growth rates based on maxillo-mandibular antero-posterior relationship.

The relationship of the female growth spurt to menarche has long been studied [83-89]. Several studies are in agreement, showing that growth velocity in statural height peaks one year prior to the onset of menarche, slows down thereafter, and stops completely within one year following menarchal onset [83, 84, 90, 91]. Although there are conflicting reports, it has also been suggested that females have been reaching menarche at progressively earlier ages from 1960 to the present, a secular trend attributed in part, but not completely to the rising obesity rates in developed countries [67, 68]. Given that 1) growth velocity in statural height is closely related to age at menarche, and 2) the onset and peak of statural height and mandibular length growth spurts are closely correlated [50], it would stand to reason that menarchal status would be an extremely powerful predictor of mandibular growth rate.

In this study, menarche was used as a covariate in a subset of the female population for which there were menarche data. Only these models were accurate enough among the model-building sets to predict ΔArPg more accurately than the ArPg PGCs. When applied to the validation set, however, only one of these models was more accurate than the ArPg PGC (FMP1), and only by 0.06mm. It is surprising that menarche did not prove to be an even stronger predictor of mandibular growth in this study, given its strong correlation to growth shown in so many other studies. Since mandibular growth has been shown to cease one year following onset of menarche, information regarding the exact time since menarchal onset may have improved the effectiveness of the menarchal covariate. In my study, the menarchal covariate was merely given a value of 0 for premenarchal females and a value of 1 for post-menarchal females, which likely limited its effectiveness as a covariate. Despite this limitation, it is clear

that orthodontic practitioners should take caution when interpreting a female's growth based on her menarchal status. Since the most accurate predictive model utilizing menarche status is only less than a tenth of a millimeter more accurate than the population growth curve, one should not assume that a female patient's mandibular growth will significantly stray from the age-dependent population average based on such status.

Apart from those described previously, there are some additional limitations to this study that could limit the weight of its conclusions. One limitation is the sample size. The models were built using samples of just over 100 for both males and females. For the female subset for which there were menarchal data, the model-building set was slightly smaller at 94. The fit of the top models was then tested on validation sets of roughly equal sizes. The smaller the size of a model-building set, the more unique that set becomes, and as a result the greater the tendency to over fit the data to that set. Likewise, the smaller the size of a validation set, the more unique it becomes. When a model is chosen and fit to the model-building set, and then tested against a unique validation set, the MAE will be larger than and not representative of the accuracy of the model on the population as a whole. With my sample size, it is possible that my model-building sets were too small to avoid over-fitting and that my validation sets were too small to avoid inflating the inaccuracies of the models.

The measurement error potentially introduced by the magnification correction method used in this study is another limitation. Haas et al, 2001 [59] showed an inter-examiner intraclass correlation coefficient (ICC) of 0.97 with two examiners measuring ArPg on 60 patients. This study showed an inter-examiner ICC of 0.96 with six examiners determining ΔArPg on 20 patients. It is important to note that determining ΔArPg in this study involved two ArPg measurements and four measurements of the width of the arms that hold the ear rods. Despite the additional measurements, the ICC of 0.96 in this study is

not much lower than the 0.97 obtained by Haas. I assume the potentially higher measurement error introduced by the additional measurements is of trivial significance.

Another limitation is the variable T1-T2 intervals used in this study. An attempt was made to keep the intervals as close to 33 months as possible. Despite this attempt, there were patients in this study with T1-T2 intervals as low as 7 months and as high as 5 years, 6 months. The smaller the interval, the more ΔArPg approaches the measurement error. On the other hand, the higher the interval, the less information it provides since growth rates can fluctuate widely from year to year. If the T1-T2 intervals could have been standardized, the prediction models may have proven to be even more accurate.

The fact that my prediction model is generated from individuals within a treated population can be seen as a limitation, since mandibular length can be influenced by certain treatment modalities. The influence of treatment on mandibular length may be most pronounced in individuals who receive functional appliance therapy. My prediction model may therefore not be applicable to untreated populations. One could argue that this limitation could actually be a strength given the fact that the data obtained from my prediction model will be applied to a very similar population—patients seeking and subsequently receiving orthodontic treatment.

There is a limitation in my patient population which results from the fact that I pool patients from different ethnicities in creating my prediction models, despite obvious differences in growth patterns between ethnicities. Given the lack of ethnic identifiers in the patient charts used in my study and the fact that the majority are assumed to be Caucasian, this is a limitation I have chosen to accept.

Another limitation involves potential error arising from self-reporting. Throughout the T1 time-points in this study it was standard practice at the UW clinic to use either a wall-mounted ruler or stadiometer to record height with the patients' shoes off, and to record weight by means of a standard mechanical scale. Despite these standard practices, there is no way of knowing whether the data recorded in

patient charts for patients' weights, and patients' heights were obtained verbally or assessed directly. Additionally, I do not know whether patients' heights were measured with their shoes off. My assumption is that the vast majority of patient's heights and weights were measured directly with their shoes off and that parental heights were obtained verbally. For those cases in which patient's heights were obtained verbally or with shoes on, data from these patients will confound my results. Moreover, given that likely all of the parents' heights were reported verbally by self or by a family member, there is a possibility for reporting bias that would further confound my results.

The final limitation I mention is the clinical utility of knowing the growth potential or growth rate of just one jaw. Prediction of ArPg length may not accurately indicate the eventual antero-posterior relationship of the mandible relative to the maxilla. The amount of maxillary growth is an important consideration not explored in this study that can impact this relationship. Another important parameter impacting this relationship not explored in this study is mandibular plane angle. If the mandibular plane angle is flat, the patient will have full expression of his or her mandibular length in the antero-posterior dimension, but in cases of steep mandibular plane angles a patient could still remain mandibular retrognathic relative to the cranial base and/or maxilla despite considerable increases in ArPg length.

It is common practice among orthodontists to collect the data used as covariates in the prediction models of this study-- patient's age, height, weight, menarchal status, ANB angle, mandibular length, and parents' heights. My findings clearly show, however, that my ability to predict a patient's mandibular growth rate over a typical orthodontic treatment interval using these data is extremely limited. In fact my data suggest that analysis of these variables provides little to no insight into a patient's mandibular growth rate over a typical orthodontic treatment interval beyond what can be obtained from a mere historic population average. Furthermore, without an adequate mathematical model, the analysis of these variables has served to merely guide an orthodontist's intuition as to how

much mandibular growth can be expected. Such intuition relies on one's recollection of past experiences and is therefore limited and biased. Mathematical modeling overcomes those deficiencies through the unbiased and simultaneous analysis of data from hundreds of patients. If such modeling fails to yield significantly greater predictive accuracy than the population average, clinical intuition will certainly fail as well. In light of these findings, orthodontic practitioners should reevaluate the utility of collecting and analyzing this information for the purposes of mandibular growth prediction.

Conclusions

1. The method of standardizing the magnification of lateral cephalometric radiographs outlined in this study is an accurate and reliable method
2. The ArPg polynomial growth curves (ArPg PGCs) produced from the Michigan, Philadelphia, and London Growth Studies provide relatively accurate estimates of mandibular growth and can be applied to more contemporary populations
3. An analysis of the following variables alone provides no clinically significant value in predicting mandibular growth over a typical orthodontic treatment interval beyond what can be derived from a historical population average—patient's age, height, weight, ANB angle, mandibular length, menarchal status, and parents' heights

Bibliography

1. Marschner, J.F. and J.E. Harris, *Mandibular growth and class II treatment.* Angle Orthod, 1966. **36**(1): p. 89-93.
2. Dudas, M. and V. Sassouni, *The hereditary components of mandibular growth, a longitudinal twin study.* Angle Orthod, 1973. **43**(3): p. 314-22.
3. Ricketts, R.M., *A principle of arcial growth of the mandible.* Angle Orthod, 1972. **42**(4): p. 368-86.
4. Ricketts, R.M., *The value of cephalometrics and computerized technology.* Angle Orthod, 1972. **42**(3): p. 179-99.
5. Johnston, L.E., *A simplified approach to prediction.* Am J Orthod, 1975. **67**(3): p. 253-7.
6. Popovich, F. and G.W. Thompson, *Craniofacial templates for orthodontic case analysis.* Am J Orthod, 1977. **71**(4): p. 406-20.
7. Greenberg, L.Z. and L.E. Johnston, *Computerized prediction: the accuracy of a contemporary long-range forecast.* Am J Orthod, 1975. **67**(3): p. 243-52.
8. Schulhof, R.J. and L. Bagha, *A statistical evaluation of the Ricketts and Johnston growth-forecasting methods.* Am J Orthod, 1975. **67**(3): p. 258-76.
9. Mitchell, D.L., J.F. Jordan, and R.M. Ricketts, *Arcial growth with metallic implants in mandibular growth prediction.* Am J Orthod, 1975. **68**(6): p. 655-9.
10. Thames, T.L., P.M. Sinclair, and R.G. Alexander, *The accuracy of computerized growth prediction in Class II high-angle cases.* Am J Orthod, 1985. **87**(5): p. 398-405.
11. Johnston, L.E., *A statistical evaluation of cephalometric prediction.* Angle Orthod, 1968. **38**(4): p. 284-304.
12. Balbach, D., *The cephalometric relationship between the morphology of the mandible and its future occlusal position. .* The Angle Orthodontist, 1969. **39**: p. 29-41.
13. Greulich, W.W. and S.I. Pyle, *Radiographic atlas of skeletal development of the hand and wrist*. 2nd ed. 1959, Stanford, CA: Stanford University Press. xvi, 256 p.
14. Bambha, J.K. and P. Vannatta, *Longitudinal Study of Facial Growth in Relation to Skeletal Maturation during Adolescence.* American Journal of Orthodontics and Dentofacial Orthopedics, 1963. **49**(7): p. 481-&.
15. Nanda, R., *The rates of growth of several facial components measured from serial cephalometric roentgenograms.* Am J Orthod 1955. **41**: p. 658-673.
16. Rose, G., *A cross-sectional study of the relationship of facial areas with several body dimensions.* Angle Orthod, 1960. **30**: p. 6-13.
17. Baccetti, T., L. Franchi, and J.A. McNamara, *An improved version of the cervical vertebral maturation (CVM) method for the assessment of mandibular growth.* Angle Orthodontist, 2002. **72**(4): p. 316-323.
18. Baccetti T, F.L., McNamara JA Jr. , *The cervical vertebral maturation (CVM) method for the assessment of optimal treatment timing in dentofacial orthopedics.* Semin Orthod, 2005. **11**: p. 119-129.
19. Bergersen, E.O., *The directions of facial growth from infancy to adulthood.* Angle Orthod, 1966. **36**(1): p. 18-43.

20. Hirschfeld, W.J. and R.E. Moyers, *Prediction of craniofacial growth: the state of the art.* Am J Orthod, 1971. **60**(5): p. 435-44.
21. Hunter, C., *The correlation of facial growth with body height and skeletal maturation at adolescence.* Angle Orthod, 1966. **36**: p. 44-54.
22. Pike, J.B., *A serial investigation of facial and statural growth in seven to twelve year old children.* Angle Orthod, 1968. **38**(1): p. 63-73.
23. Baughan B, D.A., Levesque GY, LaPalme-Chaput L, *The pattern of facial growth before and during puberty, as shown by French-Canadian girls.* Ann Hum Biol, 1979. **6**: p. 59-76.
24. Baume, R.M., P.H. Buschang, and S. Weinstein, *Stature, head height, and growth of the vertical face.* Am J Orthod, 1983. **83**(6): p. 477-84.
25. Bishara, S.E., L.C. Peterson, and E.C. Bishara, *Changes in Facial Dimensions and Relationships between the Ages of 5 and 25 Years.* American Journal of Orthodontics and Dentofacial Orthopedics, 1984. **85**(3): p. 238-252.
26. Skieller, V., A. Bjork, and T. Linde-Hansen, *Prediction of mandibular growth rotation evaluated from a longitudinal implant sample.* Am J Orthod, 1984. **86**(5): p. 359-70.
27. Tanner, J.M., *Assessment of skeletal maturity and prediction of adult height (TW2 method), 2nd Edition*. 1983, London ; New York: Academic Press. vii, 99 p.
28. Fishman, L.S., *Radiographic evaluation of skeletal maturation. A clinically oriented method based on hand-wrist films.* Angle Orthod, 1982. **52**(2): p. 88-112.
29. Subtelny, J.D., *Early orthodontic treatment*. 2000, Chicago: Quintessence Pub. Co. viii, 312 p.
30. Fishman, L.S., *Maturational Patterns and Prediction during Adolescence.* Angle Orthodontist, 1987. **57**(3): p. 178-193.
31. Houston, W.J.B., *Relationships between Skeletal Maturity Estimated from Hand-Wrist Radiographs and the Timing of the Adolescent Growth Spurt.* European Journal of Orthodontics, 1980. **2**(2): p. 81-93.
32. Ruf, S. and H. Pancherz, *Development of the frontal sinus in relation to somatic and skeletal maturity. A cephalometric roentgenographic study at puberty.* European Journal of Orthodontics, 1996. **18**(5): p. 491-497.
33. Meredith, H.V., *Childhood Interrelations of Anatomic Growth Rates.* Growth, 1962. **26**(1): p. 23-&.
34. Lamparski, D.G., *Skeletal age assessment utilizing cervical vertebrae*, 1972, University of Pittsburgh: Pittsburgh,. p. 164 p.
35. Hassel, B. and A.G. Farman, *Skeletal Maturation Evaluation Using Cervical-Vertebrae (Vol 107, Pg 58, 1995).* American Journal of Orthodontics and Dentofacial Orthopedics, 1995. **107**(6): p. A19-A19.
36. San Roman, P., et al., *Skeletal maturation determined by cervical vertebrae development.* European Journal of Orthodontics, 2002. **24**(3): p. 303-311.
37. Gabriel, D.B., et al., *Cervical vertebrae maturation method: Poor reproducibility.* American Journal of Orthodontics and Dentofacial Orthopedics, 2009. **136**(4).
38. Zhao, X.G., et al., *Validity and reliability of a method for assessment of cervical vertebral maturation.* Angle Orthod, 2011.

39. Tweed, C.H., *The Frankfort-mandibular plane angle in orthodontic diagnosis, classification, treatment planning, and prognosis.* Am J Orthod Oral Surg, 1946. **32**: p. 175-230.
40. Harris, E.F., et al., *Predicting adult stature: a comparison of methodologies.* Ann Hum Biol, 1980. **7**(3): p. 225-34.
41. Bramswig, J.H., et al., *Adult height in boys and girls with untreated short stature and constitutional delay of growth and puberty: accuracy of five different methods of height prediction.* J Pediatr, 1990. **117**(6): p. 886-91.
42. Bayley, N. and S.R. Pinneau, *Tables for Predicting Adult Height from Skeletal Age - Revised for Use with the Greulich-Pyle Hand Standards.* Journal of Pediatrics, 1952. **40**(4): p. 423-441.
43. Tanner JM, G.H., Whitehouse RH. , *Standards for children's height at ages 2-9 years allowing for height of parents.* Arch Dis Child, 1970. **45**: p. 755-62.
44. Tanner JM, W.R., Marshall WA, , *Prediction of adult height from height, bone age and occurrence of menarche, at ages 4 to 16 with allowance for midparent height.* Arch Dis Child, 1975. **50**: p. 14-26.
45. Roche, A.F., H. Wainer, and D. Thissen, *The RWT method for the prediction of adult stature.* Pediatrics, 1975. **56**(6): p. 1027-33.
46. Roche, A.F., H. Wainer, and D. Thissen, *Predicting adult stature for individuals.* Monogr Paediatr, 1975. **3**: p. 1-114.
47. Walker, R.N., *Standards for somatotyping children: I. The prediction of young adult height from children's growth data.* Ann Hum Biol, 1974. **1**(2): p. 149-58.
48. Bayley, N., *The accurate prediction of growth and adult height.* Mod Probl Paediatr 1962. **7**: p. 234.
49. Bambha, J.K., *Longitudinal cephalometric roentgenographic study of face and cranium in relation to body height.* Journal of the American Dental Association, 1961. **63**: p. 776-99.
50. Mellion ZJ, Behrents RG, Johnston LE Jr. *The pattern of facial skeletal growth and its relationship to various common indexes of maturation.* Am J Orthod Dentofacial Orthop. 2013 Jun;143(6):845-54.
51. Riolo, M.L., *An Atlas of craniofacial growth : cephalometric standards from the University school growth study, the University of Michigan*. Craniofacial growth series. 1974, Ann Arbor: Center for Human Growth and Development, University of Michigan. 379 p.
52. Bhatia SN, Leighton BC. A Manual of Facial Growth. *A Computer Analysis of Longitudinal Cephalometric Data.* Oxford: Oxford University Press, 1993.
53. Saksena SS, Walker GF, Bixler D, YuP. *A Clinical Atlas of Rontgenocephalometry in Norma Lateralis.* New York: Liss, 1987.
54. Runge, Carl (1901), *"Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten"*, Zeitschrift für Mathematik und Physik 46: 224–243. available at www.archive.org
55. Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, et al. *2000 CDC growth charts for the United States: methods and development.* Vital Health Stat 11 2002;246:1-190.

56. Athanasiou, A.E., *Orthodontic cephalometry*. 1995, London ; Baltimore: Mosby-Wolfe. 296 p.
57. Stickel, A. and H. Pancherz, *Can Articulare Be Used in the Cephalometric Analysis of Mandibular Length - a Methodologic Study.* European Journal of Orthodontics, 1988. **10**(4): p. 362-368.
58. Nelson, C., M. Harkness, and P. Herbison, *Mandibular Changes during Functional Appliance Treatment.* American Journal of Orthodontics and Dentofacial Orthopedics, 1993. **104**(2): p. 153-161.
59. Haas, D.W., et al., *Measurements of mandibular length: A comparison of articulare vs condylion.* Angle Orthodontist, 2001. **71**(3): p. 210-215.
60. Aelbers, C.M. and L.R. Dermaut, *Orthopedics in orthodontics: Part I, Fiction or reality--a review of the literature.* Am J Orthod Dentofacial Orthop, 1996. **110**(5): p. 513-9.
61. Adenwalla, S.T., J.H. Kronman, and F. Attarzadeh, *Porion and Condyle as Cephalometric Landmarks - an Error Study.* American Journal of Orthodontics and Dentofacial Orthopedics, 1988. **94**(5): p. 411-415.
62. Moore, R.N., et al., *The Accuracy of Measuring Condylion Location.* American Journal of Orthodontics and Dentofacial Orthopedics, 1989. **95**(4): p. 344-347.
63. Dibbets JMH, Nolte K. *Regional size differences in four commonly used cephalometric atlases: the Ann Arbor, Cleveland (Bolton), London (UK), and Philadelphia atlases compared*.Orthod. Craniofacial Res. 5, 2002; 51-58
64. Karlberg J. *Secular trends in pubertal development.* Horm Res. 2002;57 Suppl 2:19-30.
65. Papadimitriou A. *Sex differences in the secular changes in pubertal maturation.* Pediatrics. 2001 Oct;108(4):E65.
66. Kaplowitz P. *Pubertal development in girls: secular trends*. Curr Opin Obstet Gynecol. 2006 Oct;18(5):487-91.
67. Demerath EW, Li J, Sun SS, et al. *Fifty-year trends in serial body mass index during adolescence in girls: the Fels Longitudinal Study*. Am J Clin Nutr 2004; 80:441–446.
68. Frisch RE. *Pubertal adipose tissue: is it necessary for normal sexual maturation? Evidence from the rat and human female.* Fed Proc 1980; 39: 2395–2400.
69. Frisch RE, Wyshak G, Vincent L. *Delayed menarche and amenorrhea in ballet dancers.* N Engl J Med 1980; 303: 17–19.
70. Warren MP. *The effects of exercise on pubertal progression and reproductive functions in girls.* J Clin Endocrinol Metab 1980; 51: 1150–1157.
71. Garn SM, Haskell JA. *Fat and growth during childhood.* Science 1959; 130: 1711–1712.
72. Frisch RE, McArthur JW. *Menstrual cycles: fatness as a determinant of minimum weight for height necessary for their maintenance and onset.* Science 1974; 185: 949–951.
73. Kaplowitz PB, Slora EJ, Wasserman RC, Pedlow SE, Heman-Giddens ME. *Earlier onset of puberty in girls: relation to increased body mass index and race.* Pediatrics 2001; 108: 347–353.
74. Mack KB, Phillips C, Jain N, Koroluk LD, *Relationship between body mass index percentile and skeletal maturation and dental development in orthodontic patients.* Am J Orthod Dentofacial Orthop 2013;143:228-34.
75. Von Bremen J, Wagner J, Ruf S, *Correlation between body mass index and orthodontic treatment outcome.* Angle Orthod 2013;83:371–375.

76. Flegal KM. *Epidemiologic aspects of overweight and obesity in the United States.* Physiol Behav 2005;86:599-602.
77. Strong WB, Malina RM, Blimkie CJ, Daniels SR, Dishman RK, Gutin B, et al. *Evidence based physical activity for school-age youth.* J Pediatr 2005;146:732-7.
78. Burke MA, Heiland FW, Nadler CM. *From "overweight" to "about right": evidence of a generational shift in body weight norms.* Obesity (Silver Spring) 2010;18:1226-34
79. Lazorick S, Peaker B, Perrin EM, Schmid D, Pennington T, Yow A, DuBard CA. *Prevention and treatment of childhood obesity: care received by a state medicaid population.* Clin Pediatr (Phila). 2011 Sep;50(9):816-26.
80. Shanker, S., et al., *Cephalometric A point changes during and after maxillary protraction and expansion.* Am J Orthod Dentofacial Orthop, 1996. **110**(4): p. 423-30.
81. Chong, Y.H., J.C. Ive, and J. Artun, *Changes following the use of protraction headgear for early correction of Class III malocclusion.* Angle Orthodontist, 1996. **66**(5): p. 351-362.
82. Downs, W., *Quantitative and Qualitative Variations in Facial Growth*, in *Vistas in Orthodontics*, R.R. BS Kraus, Editor. 1962, Lea & Febiger: Philadelphia.
83. Tanner JM *Growth at Adolescence*, 2nd ed. Oxford, England Blackwell Scientific Publications1962
84. Buckler JMH *A Longitudinal Study of Adolescent Growth*. New York, NY Springer-Verlag New York Inc1989
85. Boas F *Studies in growth*. Hum Biol. 1932;4307- 3508
86. Richey HG *The relation of accelerated, normal and retarded puberty to the height and weight of school children*. Monogr Soc Res Child Dev. 1937;2 ((serial No. 8)) 1- 67
87. Shuttleworth FK *The physical and mental growth of girls and boys age 6 to 19 in relation to age at maximum growth*. Monogr Soc Res Child Dev. 1939;4 ((serial No. 22)) 1- 29
88. Simmons K, Greulich W *Menarcheal age and the height, weight and skeletal age of girls aged 7 to 17 years.* J Pediatrics. 1943;22518- 548
89. Marshall WA,Tanner JM *Growth and physiological development during adolescence.* Ann Rev Med. 1968;19283- 300
90. Tanner JM, Davis PSW *Clinical longitudinal standards for height and height velocity for North American children*. J Pediatr. 1985;107317- 329
91. Forbes GB *Body size and composition of premenarchal girls*. AJDC. 1992;14663- 66

Tables and Figures

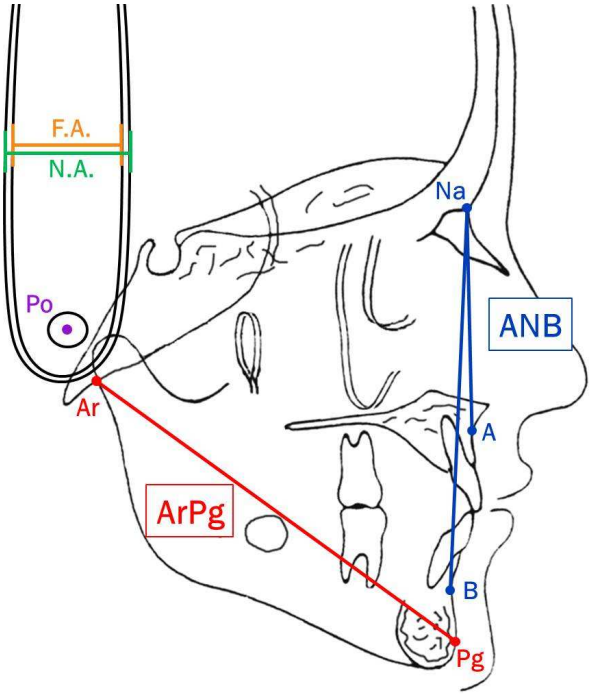

Figure 1 - Cephalometric Landmarks and Measurements
Articulare (Ar) – the point of intersection of the posterior border of the mandibular ramus and the inferior border of the basilar part of the occipital bone. Pogonion (Pg)—the most anterior point on the contour of the bony chin. A-point (A)—an arbitrary point on the innermost curvature from the maxillary anterior nasal spine to the crest of the maxillary alveolar process. B-point (B)— An arbitrary point on the anterior bony curvature of the mandible corresponding to the innermost curvature from chin to alveolar junction. Nasion (Na)—the junction of the nasal and frontal bones at the most posterior point on the curvature of the bridge of the nose. ArPg—linear distance between Ar and Pg. ANB- angle formed between A, Na, and B. Mechanical Porion (Po) – the center of the ear rods. Far ear rod arm (F.A.) – the width of the arm which holds the ear rod farthest from the x-ray source (narrowest dimension radiographically) measured 45mm above Po. Near ear rod arm (N.A.) –the width of the arm which holds the ear rod nearest to the x-ray source (widest dimension radiographically) measured 45mm above Po.

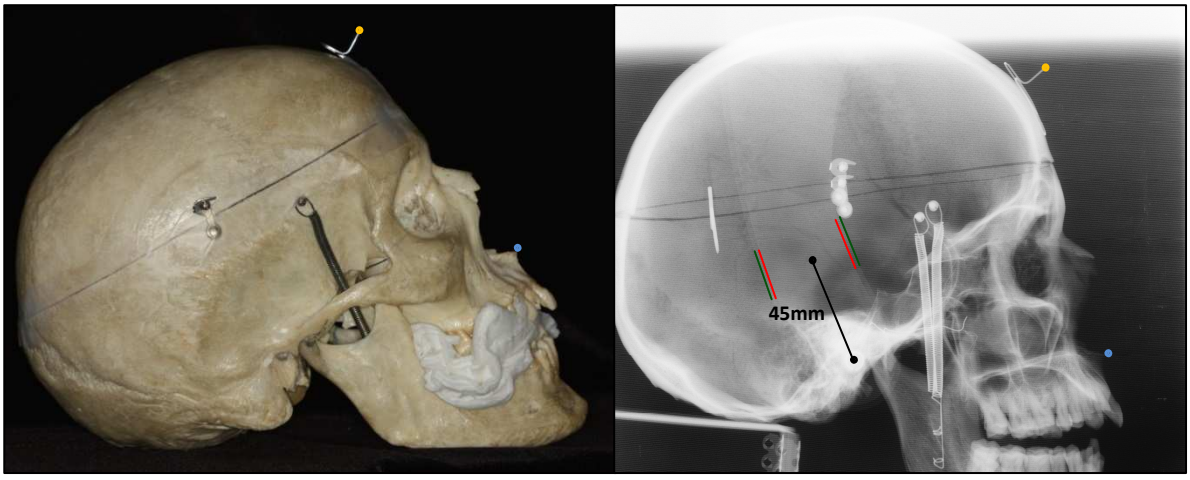

Figure 2 - Dry Skull and Radiograph
Paperclip tip (orange dot), ANS (blue dot), near ear-rod arm (green lines), far ear-rod arm (red lines), width of ear rod arms measured at 45mm above the center of Po.

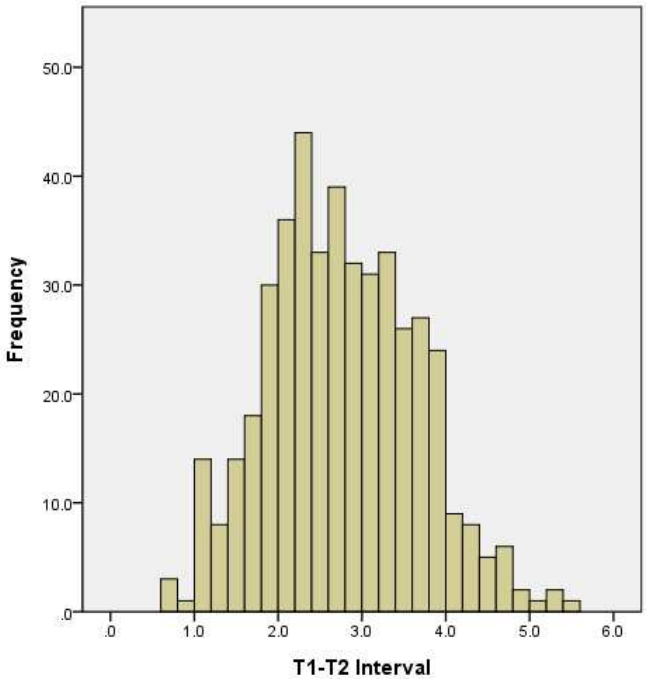

Figure 3 - Distribution of T1-T2 Intervals
N = 447; Mean interval = 2.77 years; Std. Dev = 0.89 years

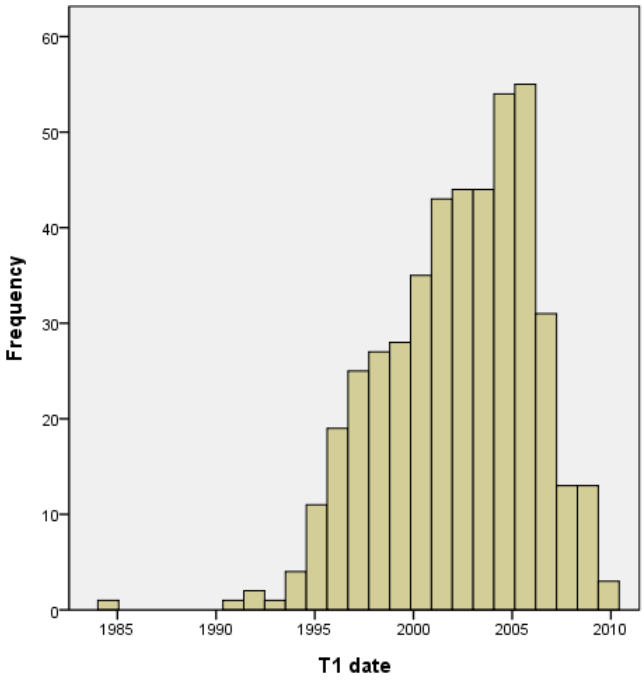

Figure 4 - Temporal Distribution
N=447; Mean T1 date = 03/09/2002; Std. Dev = 3.78 years

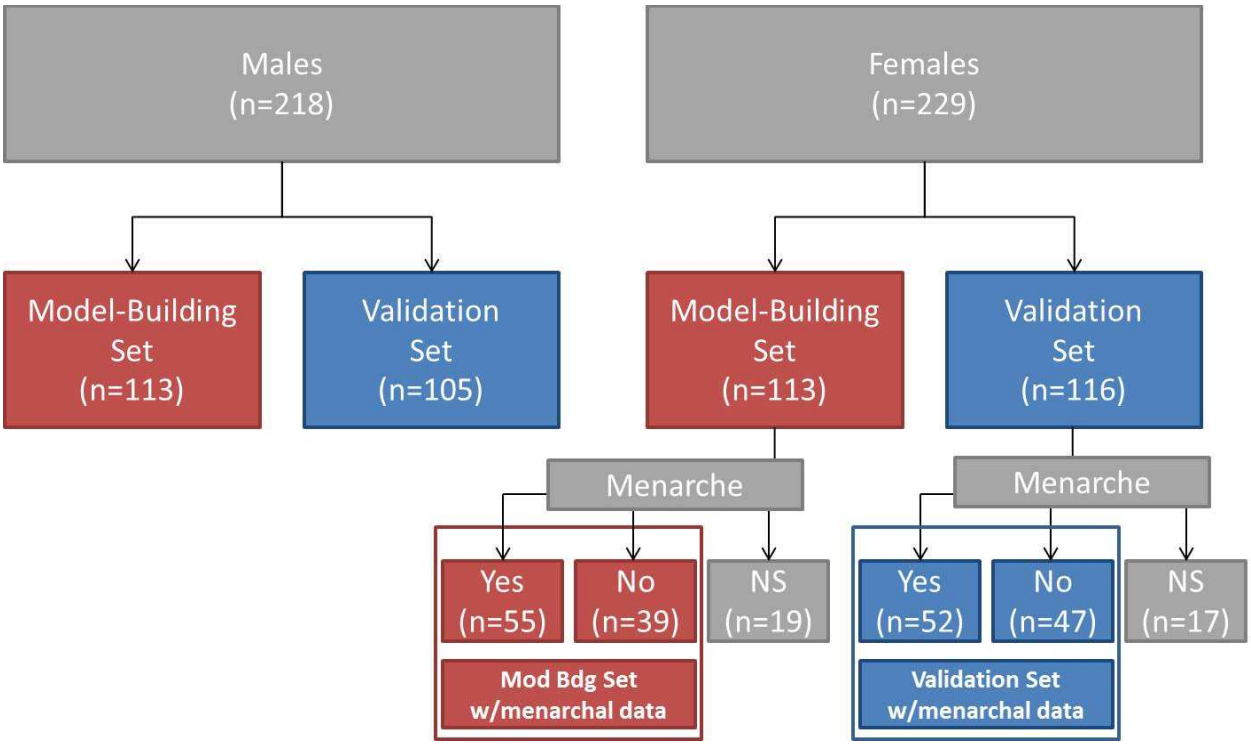

Figure 5 – Allocation of Model-Building and Validation Datasets

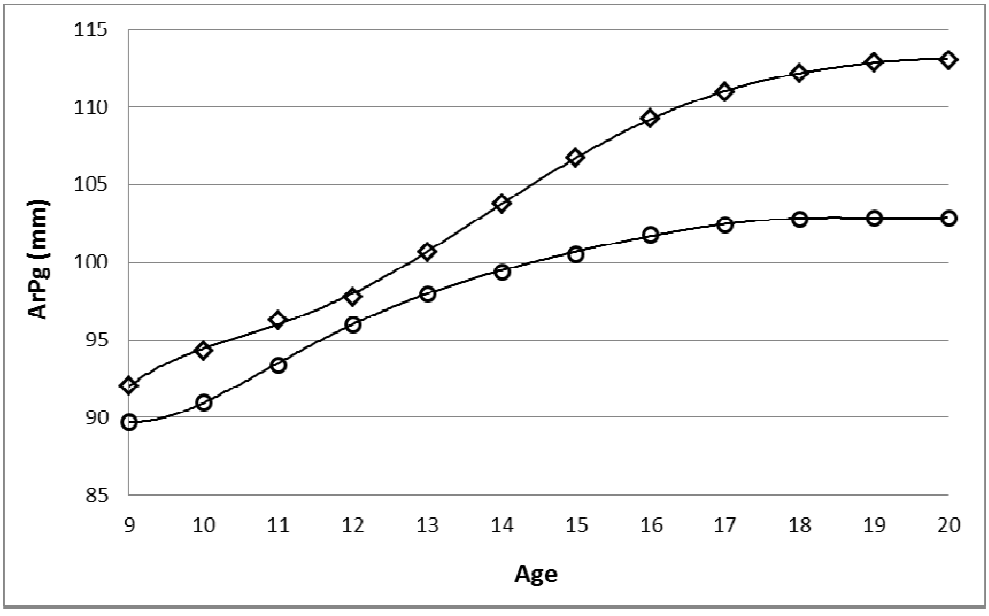

Figure 6 – ArPg Data From Longitudinal Growth Studies with Resulting Polynomial Growth Curves (ArPg PGCs)
Diamonds = mean male ArPg lengths at each age; associated PGC defined as ArPg = -0.0001635*Age^6 + 0.01536*Age^5 - 0.5910*Age^4 + 11.87*Age^3 - 131.0*Age^2 + 754.7*Age - 1684; Circles = mean female ArPg lengths at each age; associated PGC defined as ArPg = 0.0001890*Age^6 - 0.01705*Age^5 + 0.6340*Age^4 - 12.42*Age^3 + 135.0*Age^2 - 769.1*Age + 1873

MAE (mm)								
Skull		Examiner		O-F Dist		Overall		
1)	0.71	1)	0.64	Near	0.55	0.60		
2)	0.65	2)	0.66	Middle	0.66			
3)	0.68	3)	0.57	Far	0.59			
4)	0.51	4)	0.49					
5)	0.47	5)	0.60					
6)	0.59	6)	0.64					
7)	0.71							
8)	0.50							
9)	0.57							

Table 1- Mean absolute error (MAE) of Magnification Correction Technique

Characteristic	Males (n=218)	Females (n=229)
T1 Age (yr)	12.58 ± 1.77	12.63 ± 1.82
T2 Age (yr)	15.44 ± 1.87	15.31 ± 1.87
T1-T2 Interval (yr)	2.87 ± 0.95	2.68 ± 0.81
T1 ArPg (mm)	97.86 ± 6.70	95.48 ± 5.84
T2 ArPg (mm)	104.11 ± 6.83	99.00 ± 5.52
ΔArPg (mm)	6.25 ± 3.22	3.52 ± 2.69
T1 ANB (°)	3.67 ± 2.41	3.72 ± 2.46
T1 Height (in)	61.73 ± 5.30	60.98 ± 4.35
T1 Weight (lbs)	112.37 ± 33.71	111.26 ± 33.65
T1 BMI Percentile (%)	63.36 ± 29.59	61.33 ± 29.32
Mother's Height (in)	64.55 ± 2.97	64.92 ± 2.74
Father's Height (in)	70.25 ± 3.58	70.18 ± 3.22
Midparental Height (in)	69.96 ± 2.73	64.99 ± 2.43
Height Remaining (in)	8.23 ± 5.19	4.01 ± 4.48

Table 2 - Descriptive Statistics

	Males			Females		
Characteristic	Model-Building Set (n=113)	Validation Set (n=105)	*P*	Model-Building Set (n=113)	Validation Set (n=116)	*P*
T1 Age (yr)	12.83 ± 1.73	12.31 ± 1.79	0.032	12.75 ± 1.84	12.51 ± 1.81	0.325
T2 Age (yr)	15.80 ± 1.82	15.06 ± 1.86	0.003	15.37 ± 1.89	15.26 ± 1.86	0.659
T1-T2 Interval (yr)	2.98 ± 0.88	2.75 ± 1.01	0.075	2.62 ± 0.78	2.75 ± 0.84	0.231
T1 ArPg (mm)	98.54 ± 6.83	97.13 ± 6.52	0.123	95.68 ± 6.02	95.29 ± 5.69	0.610
T2 ArPg (mm)	105.04 ± 6.73	103.11 ± 6.83	0.037	98.94 ± 5.57	99.06 ± 5.50	0.875
ΔArPg (mm)	6.50 ± 3.30	5.97 ± 3.12	0.225	3.26 ± 2.58	3.77 ± 2.78	0.151
T1 ANB (°)	3.91 ± 2.19	3.42 ± 2.61	0.136	3.61 ± 2.65	3.83 ± 2.27	0.505
T1 Height (in)	62.19 ± 5.11	61.24 ± 5.48	0.184	61.09 ± 4.48	60.87 ± 4.24	0.708
T1 Weight (lbs)	114.18 ± 33.07	110.42 ± 34.44	0.412	113.48 ± 31.33	109.10 ± 35.77	0.326
T1 BMI Percentile (%)	20.43 ± 3.86	20.30 ± 4.03	0.818	21.16 ± 4.59	20.31 ± 4.65	0.166
Mother's Height (in)	64.41 ± 2.81	64.69 ± 3.14	0.497	64.85 ± 2.78	64.98 ± 2.71	0.710
Father's Height (in)	70.11 ± 3.46	70.41 ± 3.71	0.542	69.88 ± 3.31	70.46 ± 3.13	0.177
Midparental Height (in)	69.82 ± 2.67	70.11 ± 2.80	0.442	64.81 ± 2.53	65.16 ± 2.33	0.269
Height Remaining (in)	7.63 ± 5.19	8.87 ± 5.15	0.078	3.72 ± 4.62	4.29 ± 4.33	0.335

Table 3 – Descriptive Statistics of Model Building Datasets vs Validation Datasets

Model-Building Set	Predicting	Age Input	# Unique Models	Lowest MAE (mm)	MAE (mm) of ArPg PGC
Male	ΔArPg	Linear	8192	2.19	2.05
	ΔArPg	Categorical	8192	2.18	
	ΔArPg	Polynomial	40960	2.14	
	(ΔArPg - Exp ΔArPg)	Linear	8192	1.94*	
Female	ΔArPg	Linear	8192	1.57	1.55
	ΔArPg	Categorical	8192	1.64	
	ΔArPg	Polynomial	40960	1.58	
	(ΔArPg - Exp ΔArPg)	Linear	8192	1.49*	
Female with Menarche Data	ΔArPg	Linear	8192	1.38*	1.47
	ΔArPg	Categorical	8192	1.37*	
	ΔArPg	Polynomial	40960	1.33*	
	(ΔArPg - Exp ΔArPg)	Linear	8192	1.31*	

Table 4 – Lowest MAEs of all Models Using Leave-One-Out Cross-Validation Within Model-Building Sets

MAE of ArPg PGCs (when applied to model-building sets) also listed for comparison

(*) indicates prediction model had lower MAE within model-building set than ArPg PGC

PREDICTING (ΔArPg - EXP ΔArPg), LINEAR AGE MODELS			
Model	*Covariates*	*MAE (mm)*	*AE (mm) (25%, 75%)*
M1	age (linear), father's height, height, weight	1.94	(0.79, 2.60)
M2	age (linear), father's height, obese	1.94	(0.65, 2.61)
M3*	age (linear), height, weight	1.94	(0.82, 2.68)
M ArPg PGC	----------	2.05	(0.90, 2.65)
F1	age (linear), father's height, obese	1.49	(0.46, 2.04)
F2	age (linear), father's height, weight	1.49	(0.46, 2.07)
F3*	age (linear), father's height	1.49	(0.45, 2.00)
F ArPg PGC	----------	1.55	(0.64, 2.14)
FM1*	age (linear), menarche, obese	1.31	(0.52, 1.95)
FM2	age (linear), menarche, obese, mother's height	1.32	(0.45, 1.90)
FM3	age (linear), menarche, obese, prognathic	1.32	(0.42, 1.90)
F ArPg PGC	----------	1.47	(0.65, 2.00)

Table 5 - Top Three Models from Within Model-Building Sets, Predicting (ΔArPg – Exp ΔArPg), Using Linear Age
M1-M3 = top 3 models chosen from within male model-building set; F1-F3 = top 3 models chosen from within female model-building set; FM1-FM3 = top 3 models chosen from within female model-building subset for which I have menarche data; MAEs and absolute errors (AEs) calculated using leave-one-out cross-validation; MAEs and AEs of ArPg PGCs (when applied to model-building sets) also listed for comparison, where M = male and F = female
(*) indicates the models chosen for validation on the validation data sets

PREDICTING ΔArPg, VARIOUS AGE MODELS			
Model	*Covariates*	*MAE (mm)*	*AE (mm) (25%, 75%)*
FML1	age (linear), menarche, obese, BMI %, ANB	1.38	(0.41, 2.09)
FML2*	age (linear), menarche, obese, BMI %	1.38	(0.48, 2.05)
FML3	age (linear), menarche, obese, BMI %, ANB, father's height	1.38	(0.49, 2.07)
FMC1*	age (categorical), menarche, obese, BMI %, ANB, height	1.37	(0.50, 2.00)
FMC2	age (categorical), menarche, obese, BMI %, ANB, height, prognathic	1.38	(0.46, 1.93)
FMC3	age (categorical), menarche, obese, BMI %, ANB, height, ArPg	1.38	(0.44, 2.05)
FMP1*	age (polynomial, n=5), menarche, ANB	1.33	(0.55, 1.85)
FMP2	age (polynomial, n=5), menarche, ANB, height	1.33	(0.45, 1.94)
FMP3	age (polynomial, n=5), menarche, obese, ANB, weight	1.33	(0.46, 1.85)
F ArPg PGC	----------	1.47	(0.65, 2.00)

Table 6 - Top Three Models from Within Model-Building Sets for Females with Menarche Data, Predicting ΔArPg
FML1-FML3 = top 3 linear-age models; FMC1-FMC3 = top 3 categorical-age models; FMP1-FMP3 = top 3 polynomial-age models; MAEs and absolute errors (AEs) calculated using leave-one-out cross-validation; MAEs and AEs of F ArPg PGC (when applied to model-building sets) also listed for comparison, where F = female
(*) indicates the models chosen for validation on the validation data sets

PREDICTING (ΔArPg - EXP ΔArPg), LINEAR AGE MODELS					
		Model-Building Sets †		*Validation Sets* ‡	
Model	*Covariates*	*MAE (mm)*	*AE (mm) (25%, 75%)*	*MAE (mm)*	*AE (mm) (25%, 75%)*
M3	age (linear), height, weight	1.94	(0.82, 2.68)	1.73*	(0.49, 2.39)
M ArPg PGC	----------	2.05	(0.90, 2.65)	1.80	(0.63, 2.59)
F3	age (linear), father's height	1.49	(0.45, 2.00)	1.61	(0.69, 2.25)
F ArPg PGC	----------	1.55	(0.64, 2.14)	1.60	(0.61, 2.15)
FM1	age (linear), menarche, obese	1.31	(0.52, 1.95)	1.57*	(0.69, 2.11)
F ArPg PGC	----------	1.47	(0.65, 2.00)	1.58	(0.59, 2.25)

Table 7 –Final Models Applied to Validation Sets, Predicting (ΔArPg – Exp ΔArPg), Using Linear Age

M3, F3, and FM1 are the models chosen from within the male, female and female with menarche data sets respectively for validation on the validation sets; MAEs and absolute errors (AEs) of ArPg PGCs also listed for comparison, where M = male and F = female

(†) MAE and AE calculated via leave-one-out cross-validation.

(‡) MAE and AE calculated after fitting model to model-building sets, then using resulting coefficients on validation sets

(*) indicates models with lower MAEs than ArPg PGCs when applied to validation sets

PREDICTING ΔArPg, VARIOUS AGE MODELS					
		Model-Building Sets †		*Validation Sets* ‡	
Model	*Covariates*	*MAE (mm)*	*AE (mm) (25%, 75%)*	*MAE (mm)*	*AE (mm) (25%, 75%)*
FML2	age (linear), menarche, obese, BMI %	1.38	(0.48, 2.05)	1.60	(0.61, 2.33)
FMC1	age (categorical), menarche, obese, BMI %, ANB, height	1.37	(0.50, 2.00)	1.64	(0.63, 2.37)
FMP1	age (polynomial, n=5), menarche, ANB	1.33	(0.55, 1.85)	1.52*	(0.60, 2.01)
F ArPg PGC	----------	1.47	(0.65, 2.00)	1.58	(0.59, 2.25)

Table 8- Final Models Applied to Validation Sets, Predicting ΔArPg, Using Various Age Inputs

FML2, FMC1, and FMP1 are the top linear-, categorical-, and polynomial-age models respectively chosen from within the female model-building subset for which I have menarche data; MAEs and absolute errors (AEs) of F ArPg PGC also listed for comparison, where F = female

(†) MAE and AE calculated via leave-one-out cross-validation.

(‡) MAE and AE calculated after fitting model to model-building sets, then using resulting coefficients on validation sets

(*) indicates models with lower MAEs than ArPg PGCs when applied to validation sets