



OPEN

Cross-sectional data accurately model longitudinal growth in the craniofacial skeleton

Kevin M. Middleton¹✉, Dana L. Duren^{2,3}, Kieran P. McNulty⁴, Heesoo Oh⁵, Manish Valiathan⁶ & Richard J. Sherwood^{2,3,6}

Dense, longitudinal sampling represents the ideal for studying biological growth. However, longitudinal samples are not typically possible, due to limits of time, prohibitive cost, or health concerns of repeat radiologic imaging. In contrast, cross-sectional samples have few such drawbacks, but it is not known how well estimates of growth milestones can be obtained from cross-sectional samples. The Craniofacial Growth Consortium Study (CGCS) contains longitudinal growth data for approximately 2000 individuals. Single samples from the CGCS for individuals representing cross-sectional data were used to test the ability to predict growth parameters in linear trait measurements separately by sex. Testing across a range of cross-sectional sample sizes from 5 to the full sample, we found that means from repeated samples were able to approximate growth rates determined from the full longitudinal CGCS sample, with mean absolute differences below 1 mm at cross-sectional sample sizes greater than ~200 individuals. Our results show that growth parameters and milestones can be accurately estimated from cross-sectional data compared to population-level estimates from complete longitudinal data, underscoring the utility of such datasets in growth modeling. This method can be applied to other forms of growth (e.g., stature) and to cases in which repeated radiographs are not feasible (e.g., cone-beam CT).

Longitudinal biological growth data provide tremendous potential for investigating details of growth trajectories and their milestones. Requiring both dedicated participant commitment and long-term financial obligations, the development of longitudinal datasets for humans is logistically complex. As a result, such studies can be limited in the range of biological variation captured. The challenge is greater for studies of skeletal growth where radiographic imaging is ideal for assessment but where repeated exposure to radiation severely limits the maximum number of observations per individual. In contrast, cross-sectional studies are logistically simple, are able to enroll a wider range of participants, and can be completed within a shorter timescale by not requiring participants to age naturally for years or decades. For these reasons, the efficiency of cross-sectional data allows for larger data sets¹. While the benefits of cross-sectional data are clear, the question remains as to what limitations are imposed by cross-sectional analysis of growth.

To answer this question, we use examples from the field of craniofacial growth. Assessment of growth status is important in a number of craniofacial fields from orthodontics to craniofacial and maxillofacial surgery, where optimal treatment timing is critical. Studies into craniofacial growth have regularly used relatively small sample sizes of 100 or fewer individuals^{2–5}. Frequently, those individuals are binned into yearly age classes (e.g., participants between their 12th and 13th birthdays are "12 year-olds") for which means or percentiles are calculated^{6–10}. In an alternate approach, several studies have examined only two timepoints for an individual^{11,12}, typically before and after the adolescent growth spurt. Although technically a longitudinal study, utilizing only two observation timepoints severely limits inferences regarding patterns of growth. Our own work in craniofacial growth modeling and clinical application has often highlighted the benefits of longitudinal data^{13–18}, and we have had similar success modeling human stature and skeletal maturation with longitudinal data^{19–24}.

When analyzed statistically, age groups, facial types, or sexes are most often statistically compared to one another via parametric independent samples *t*-tests or non-parametric *U*-tests tests to answer questions like "do

¹Division of Biological Sciences, University of Missouri, Columbia, MO, USA. ²Department of Orthopaedic Surgery, University of Missouri School of Medicine, Columbia, MO, USA. ³Department of Pathology and Anatomical Sciences, University of Missouri School of Medicine, Columbia, MO, USA. ⁴Department of Anthropology, University of Minnesota, Minneapolis, MN, USA. ⁵Department of Orthodontics, Arthur A. Dugoni School of Dentistry, University of the Pacific, San Francisco, CA, USA. ⁶Department of Orthodontics, School of Dental Medicine, Case Western Reserve University, Cleveland, OH, USA. ✉email: middletonk@missouri.edu

sexes differ significantly in age at peak growth velocity?" or "do facial classes differ in peak growth velocity at age 12?". However, longitudinal data have also been modeled using multilevel polynomial functions^{16,25–28}. In clinical practice, patient-specific measurements from one or more observations are often compared to standards from reference populations^{6,8,10,29–31} or a set of graphic standards such as the Bolton Standards³². In the latter, standards are used as a "target" for comparison between the patient's observed morphology and an ideal or mean configuration (i.e., a "normal" facial configuration)¹⁰. It remains unclear how well cross-sectional, or limited-range longitudinal data, accurately reflect the true underlying longitudinal growth pattern in the craniofacial complex. Thus, determining the agreement between age-specific craniofacial trait values estimated from longitudinal and those from cross-sectional data is of significant importance.

Herein we evaluate the ability of a large cross-sectional sample to replicate growth metrics estimated from well-characterized, similarly large longitudinal samples, and we address the impact of cross-sectional sample size on prediction ability in craniofacial growth milestones. Rather than choosing "ideal" or "normal" morphologies, the present analysis examines the full range of variability present in untreated individuals from historic growth studies. Importantly, this sample includes two sources of variation: naturally occurring phenotypic variation as well as measurement error associated with radiographs and landmark digitization.

To fully utilize the rich dataset described below, which includes a large number of serial observations for an extremely large sample of individuals, we model growth using a multilevel^{28,33,34} double-logistic growth equation³⁵. This model allows for a smooth growth curve that includes two periods of rapid growth, an asymptote at growth cessation, and biologically meaningful parameters for the timing and magnitude of growth milestones. Finally, we assess the performance of these growth models using two primary applications of growth data: (1) estimation of measurement percentile intervals (i.e., growth standards) and (2) growth rates and ages at peak growth velocity. The former is used to assess a patient's current status and future growth potential relative to the entire population (e.g., similar to standards for standing height in children). The latter can be used to determine timing relative to peak growth velocity¹³ or to determine if growth cessation has been achieved¹⁴, which are critical aspects of patient assessment and treatment planning.

Methods

Sample

All data for the present analysis derive from the Craniofacial Growth Consortium Study (CGCS). The CGCS combines data from six separate historic growth studies that span 100 years (1919–2018)¹³. Although this historic sample is primarily composed of self-identified white individuals, it is represented by populations spanning the geographic breadth of North America and has considerable time depth. We have previously shown that the different growth studies that constitute the CGCS do not differ in their overall growth parameters, and thus it is likely that the performance of cross-sectional data relative to longitudinal data is unlikely to vary among population samples¹³. For this study, the sample includes all individuals from the CGCS with at least two cephalographs ($n = 959$ females; $n = 980$ males), yielding a data set with a median count of 9 cephalographs per individual (range: 2–22). The Institutional Review Board of the University of Missouri–Columbia approved all procedures used in this study, and all methods were performed in accordance with the relevant guidelines and regulations. Informed consent was waived by the institutional review board of University of Missouri–Columbia due to the retrospective nature of the study.

Landmark data collection and computation of craniofacial measures

Linear traits are measured between pairs of anatomical locations (e.g., nasion and basion). To capture the two-dimensional (x, y) coordinates, we used a standard landmarking protocol. The full landmarking protocol, including study- and date-specific corrections for radiographic enlargement are provided in Sherwood et al.¹³, and we include only an overview here. Landmarks ($n = 119$) were placed using the eDigit software (Craniofacial Research Instrumentation Lab; Arthur A. Dugoni School of Dentistry, University of the Pacific) by three separate assessors and the coordinates averaged¹³⁶. This software system includes a series of internal checks against large deviations among the three replicate landmark sets by ensuring that landmarks fall within a specified error envelope³⁶.

From the full landmark set, we focused on 12 linear cephalometric traits, defined by pairs of (x, y) coordinates, broadly describing the shape of the basicranium, palate, face, and mandible. Trait values were calculated as inter-landmark distances and corrected for radiographic enlargement prior to analysis¹³. We analyzed growth trajectories and estimated growth rates for these twelve traits separately by sex and compared the performance of modeling these data longitudinally or via cross-sectional subsets.

Statistical methods

We modeled the growth of craniofacial traits as a function of age: $y(\text{age})$, using a double logistic equation proposed by Bock et al.³⁵

$$y(\text{age}) = \frac{a_1}{1 + e^{-b_1(\text{age}-c_1)}} + \frac{f - a_1}{1 + e^{-b_2(\text{age}-c_2)}}$$

in which growth is modeled with two additive phases: prepubertal and adolescent. In the equation above, the f and a_1 parameters represent the asymptotic and prepubertal trait size measurements, b_1 and b_2 rates of growth, and c_1 and c_2 ages at maximal growth rates. These six parameters, when applied across an *age* range, result in a continuously increasing length measurement with two periods of rapid growth (e.g., purple line in Fig. 1a). Although in the original presentation of this growth equation, which focused on modeling stature, asymptotic size at cessation (f) was assumed to be known a priori, here we estimate f from the data. Other modeling

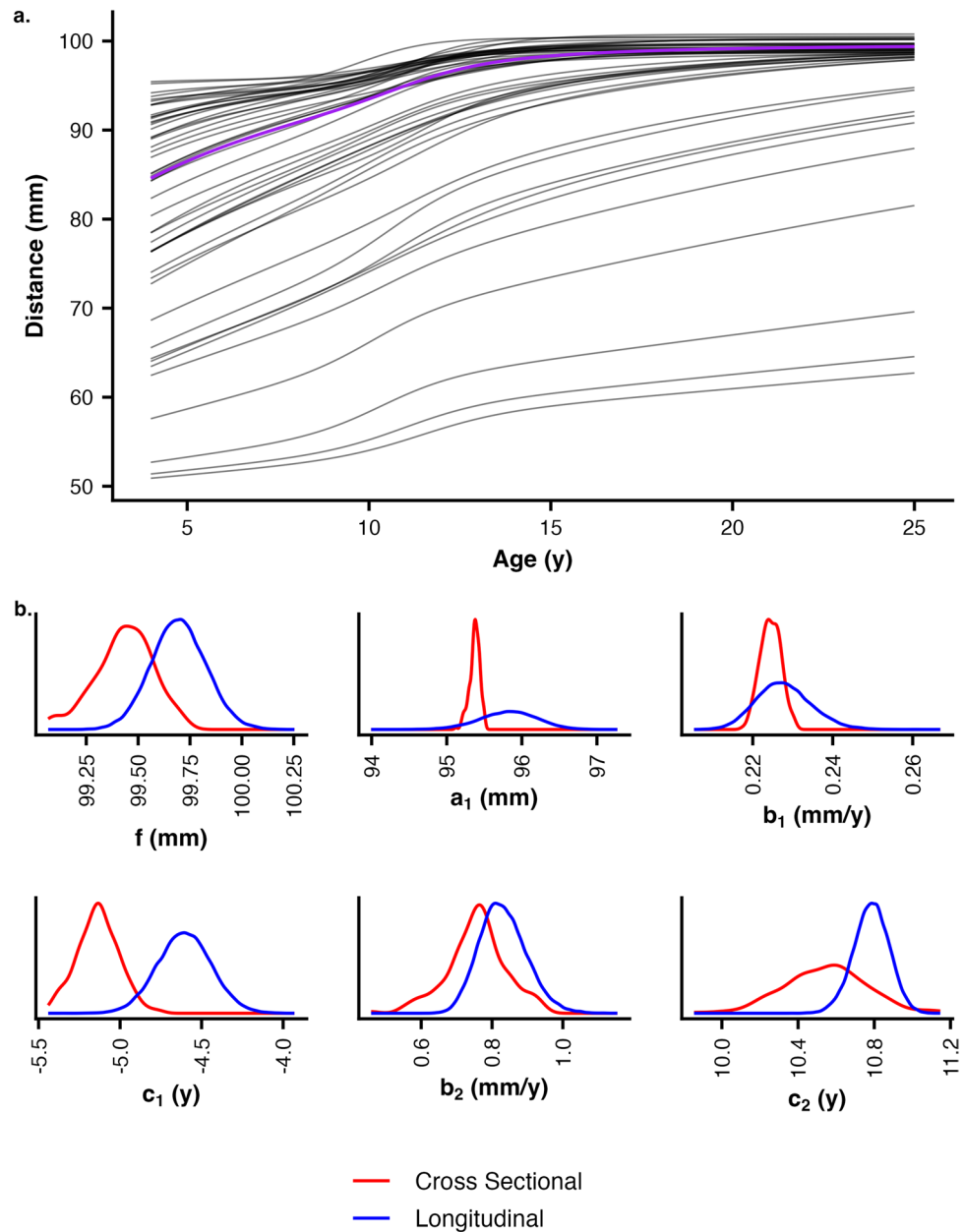


Figure 1. (a) The prior predictive check for female nasion to basion distance. The purple line represents the prediction from the mean prior for each of the six double logistic equation parameters. The gray lines represent 50 predicted growth curves using parameter values randomly sampled from the trait-specific priors. These lines give a representative set of predicted growth patterns before the data has influenced the model. (b) Posterior densities for parameter estimates. Six posterior distributions are compared for the longitudinal model (blue lines) and the median of the full-sample cross-sectional models (red line). f and a_1 represent asymptotic and pre-pubertal lengths, b_1 and b_2 rates of growth, and c_1 and c_2 ages.

approaches have been used to estimate growth parameters, including polynomial and related spline functions, but we prefer the double logistic equation here. While other models are fit more easily, the double logistic equation has several biologically relevant characteristics that make it particularly well suited for studying growth. First, the parameter estimates have coefficients that are directly interpretable, representing distances (f and a_1 in mm), rates (b_1 and b_2 in mm/year), or ages (c_1 and c_2 in years). Second, the f parameter represents the asymptotic size at growth cessation^{13,14}. Finally, by restricting the b_1 and b_2 rate parameters to be positive, growth can be restricted to increase monotonically. The second two benefits are not possible with polynomial modeling.

Long-term growth studies like those that make up the CGCS are increasingly rare and unlikely to be repeated. Thus, we sought to effectively replicate the results of dense longitudinal data modeling using cross-sectional data. Using an identical analysis protocol, we compare the performance of models fit to cross-sectional subsets of the full data set to a longitudinal model fit using the full data set. The cross-sectional models analyzed in this set of

comparisons varied in the number of individuals included. For this comparison, we consider the longitudinal model to represent the best approximation of the growth pattern for a trait and evaluate how well that pattern is approximated by the analysis of cross-sectional samples. The reference model was a Bayesian multilevel model, where we translated the double logistic equation into model syntax with priors for parameter estimates. That model has following specification:

$$\begin{aligned}
 y_i &\sim \text{Normal}(\mu_i, \sigma) \\
 \mu_i &= \frac{a_1}{1 + e^{-b_1(\text{age}_i - c_1)}} + \frac{f - a_1}{1 + e^{-b_2(\text{age}_i - c_2)}} + a_{ID}[ID] \\
 f &\sim \text{Normal}(Pr_f, 2) \\
 a_1 &\sim \text{Normal}(Pr_{a_1}, 2) \\
 b_1 &\sim \text{Normal}(Pr_{b_1}, 0.2) \\
 c_1 &\sim \text{Normal}(Pr_{c_1}, 1) \\
 b_2 &\sim \text{Normal}(Pr_{b_2}, 0.2) \\
 c_2 &\sim \text{Normal}(Pr_{c_2}, 1) \\
 \sigma &\sim \text{Exponential}(0.25) \\
 a_{ID,j} &\sim \text{Normal}(0, \sigma_{ID}), \text{ for } j = 1..n_{ID} \\
 \sigma_{ID} &\sim \text{Exponential}(0.25).
 \end{aligned}$$

In this model description, y_i represents the i th observed value of a trait for an individual with a particular ID . This value follows a normal distribution with a mean (μ_i) defined by the double logistic equation. $a_{ID}[ID]$ represents the random intercept for each individual, which is drawn from a normal distribution with a standard deviation estimated from the data (σ_{ID}) and represents an individual-specific size offset from the population-level growth pattern. The remaining lines f through σ_{ID} represent the priors, which are distributions for the priors for the model parameters, including trait specific values for the means of priors for each parameter (Pr).

In order to determine how well a cross-sectional sample can approximate a complete longitudinal sample for the goal of estimating growth milestones, we first needed to subsample the full dataset into smaller sets for which we could estimate those milestones. We generated 200 random cross-sectional subsets of the data for each of $n = 5, 10, 20, 50, 100, 200, 300, 500$, and ~ 870 – 881 individuals to explore the role of cross-sectional sample size in estimation of growth patterns. Importantly, each of these datasets included only a single observation per individual (i.e., mimicking a cross-sectional data set). Sample sizes for the largest cross-sectional samples varied from 870 to 881 due to differences in the observations of individual trait values across the full data set.

A model was then fit for each of these randomly generated subsets similar to the model above, but without the separate intercept for individual or its associated standard deviation. Aside from only including individuals with two or more observations, we did not impose any additional conditions on the resampling procedure, such as restriction to those with observations within a certain age range or individuals with a minimum number of measurement points. For example, individuals for whom only two observations are present will be disproportionately overrepresented in the larger samples.

Bayesian inference requires a prior for each parameter to be estimated (i.e., f, a_1, b_1 , etc.). These priors define the plausible range of values for a specific parameter, which allow the sampler to efficiently search for posterior values. For example, a prior for f with a mean of 100 mm and standard deviation of 2 mm means that 95% of the prior weight for f falls between 96 and 104 mm. Priors for Bayesian models were set to be broad but mildly regularizing and determined using a genetic algorithm^{13,37–39} with the *rgenoud* package (version 5.9-0.3)⁴⁰. These priors included positive constraints on the b_1 and b_2 rate parameters to ensure monotonically increasing length estimates. Adequacy of priors, neither too restrictive nor too wide, was assessed via prior predictive checks (Fig. 1a). These checks were visually inspected to ensure that the outcome scale of predicted distance approximated the same scale as the observed measures and that the priors were not overly restrictive to exploration of the parameter space.

Bayesian inference via Monte Carlo methods requires a mechanism for sampling from the posterior given the data, the priors, and the model. Models were estimated using Hamiltonian Monte Carlo via the *stan* statistical programming language^{41,42} with the *cmdstanr* package (version 0.5.0)⁴³ in R (version 4.3)⁴⁴. Models were sampled for 10,000 iterations with 50% warmup in four parallel chains. Starting values for parameters were drawn randomly from the priors separately for each chain. Model convergence was assessed by inspection of \hat{R} values and rank histograms⁴⁵. After sampling, model parameter estimates had ~ 2000 – $30,000$ effective samples.

Posteriors, prediction intervals, and growth rates

Bayesian sampling results in posteriors, which include simultaneous estimates for all parameters in the model, which are summarized for presentation and comparison. While the longitudinal model had a single posterior, each of the 200 randomly resampled cross-sectional models had a separate posterior. To combine these posteriors for comparison, we created a distribution of the median parameter estimates from all the cross-sectional models for each sample size (Fig. 1b). To compare posterior predictive ability between longitudinal and cross-sectional models, which represent percentile size intervals, we calculated middle 50%, 80% and 98% quantile intervals from the posteriors (Fig. 2a), using the aggregated samples from the set of 200 random datasets. Second, we calculated the first derivative of the predicted trait size over time, the estimate of growth rate^{13,16,23,46–48} separately for each of the 200 random samples as well as for the median cross-sectional model and the longitudinal model (Fig. 2b).

Effects of sample size

To assess the role of sample size in the ability of a cross-sectional sample to approximate a hierarchical longitudinal analysis, we analyzed subsamples ranging from the full sample size ($n = 959$ females or 980 males) down to $n = 5$ observations (i.e., a model fit to only 5 data points). These models were fit 200 times each using different random samples and posterior predictive mean measurements were calculated (Fig. 3a). Finally, to quantify the

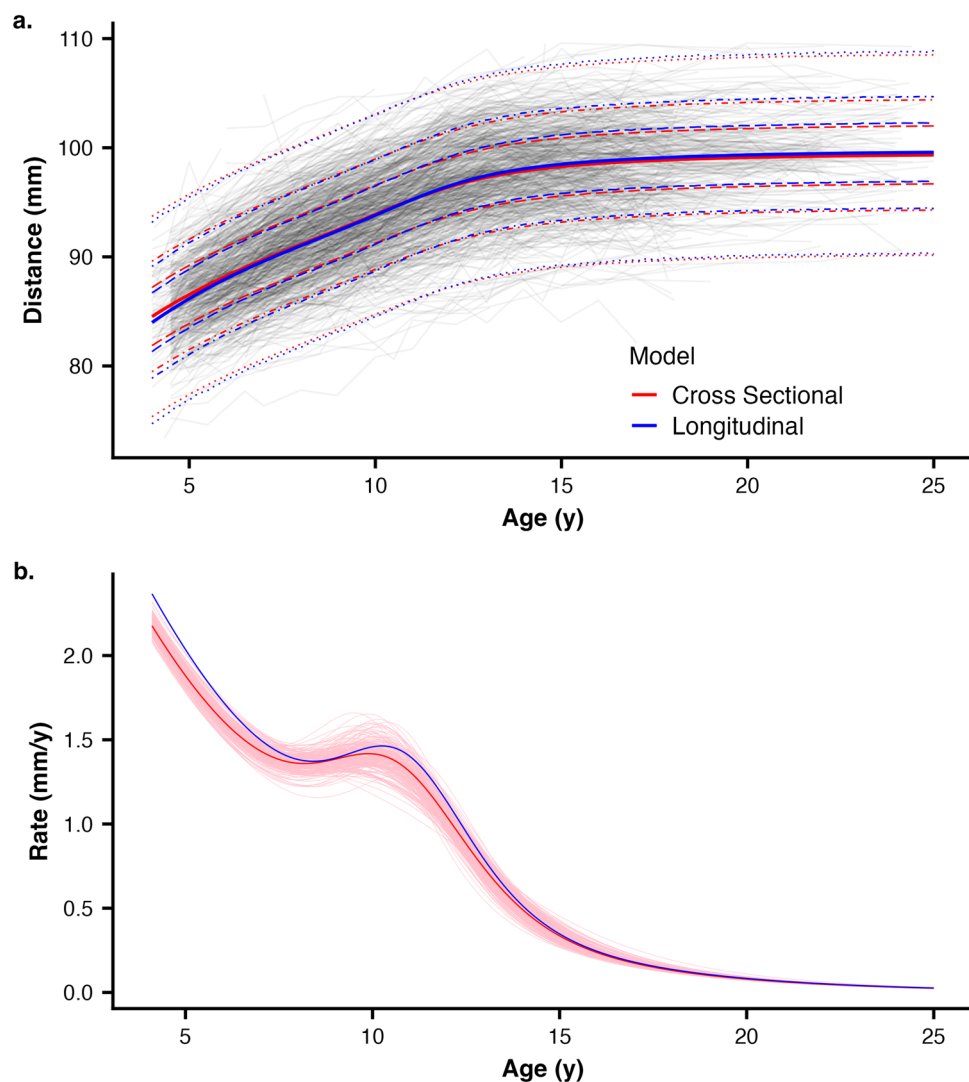


Figure 2. (a) Comparison of size percentiles. The dark red and blue lines represent median cross-sectional and longitudinal model estimates, respectively. The dotted and dashed lines represent 1st, 10th, 25th, 75th, 90th, and 99th percentiles, which are nearly identical between cross-sectional and longitudinal models as indicated by near overlap of percentiles. (b) Comparison of estimated growth rates as the first derivative the growth curve. As in (a), the dark red and blue lines represent median cross-sectional and longitudinal model estimates. All 200 estimated rates for the full-sample cross-sectional analysis are shown in pink lines, indicating general agreement between the two approaches.

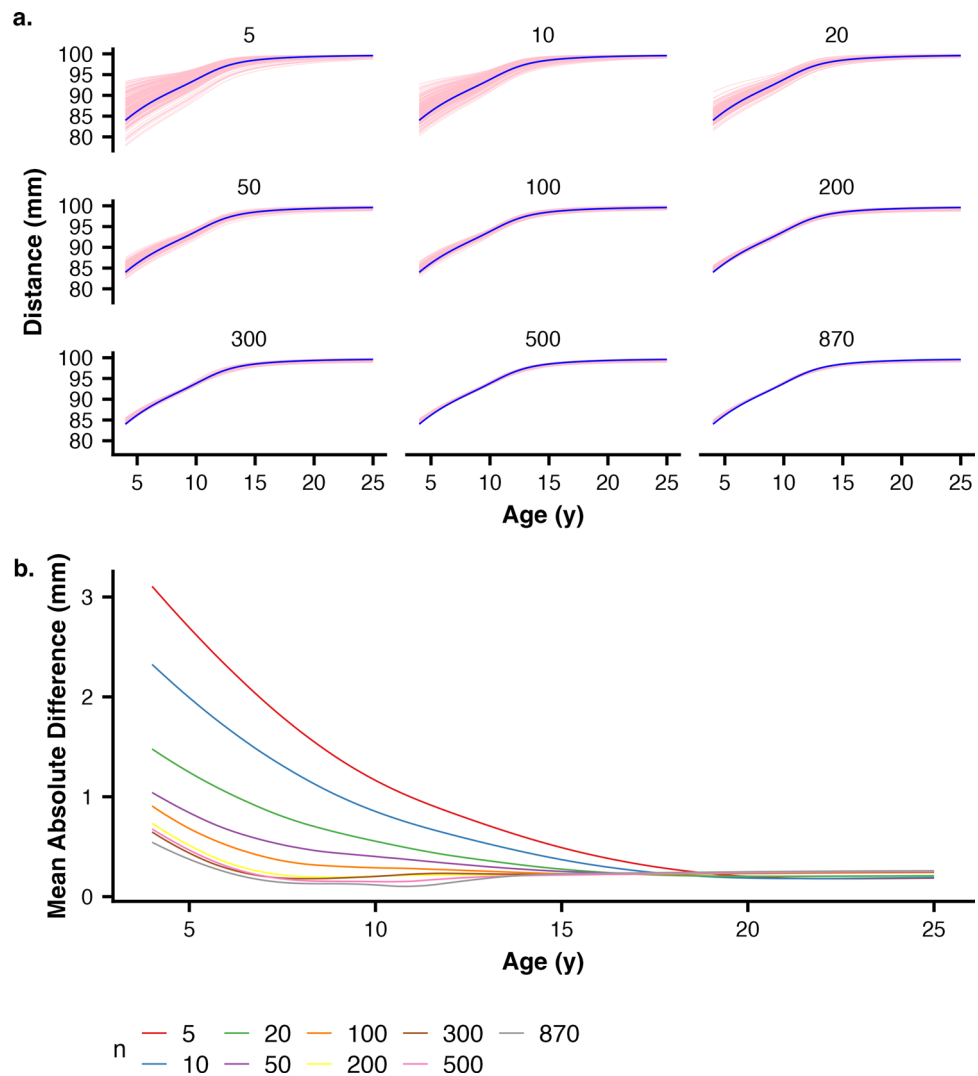


Figure 3. (a) Effect of sample size on nasion to basion distance prediction. The blue line represents the median predicted measurement from the longitudinal model. Each of the pink lines represents one of 200 random subsample models. As sample size increases, the effect of single extreme points has gradually less effect. (b) Differences in predictive ability. Mean absolute difference between the 200 samples and the longitudinal model was calculated across the predictive range. Average difference was usually less than 1 mm except for sample sizes below ~50, even though individual differences vary considerably more as seen in (a).

difference between the predictive ability across sample size, we calculated the mean absolute difference between each of the cross-sectional sample's predicted sizes and predicted measures from the longitudinal model (Fig. 3b).

Results

Overall results were similar across all traits and in both sexes. Thus, for simplicity we present only those results for one trait: the distance from Nasion to Basion in the female CGCS participants (Figs. 1 and 2). Full analysis of all traits in both sexes are included in the supplementary information (<https://figshare.com/s/a0d90329654c2a3d8fcf>).

Posterior distributions

We focus here on the comparison of the full sample cross-sectional models to the longitudinal model. Comparisons of posterior distributions of parameter estimates (Fig. 1b) are similar between the longitudinal model and the median of the 200 cross-sectional models where one sample was drawn from each individual ("Full sample"). Most parameter estimates show almost complete overlap between the two approaches. When the posteriors differ (e.g., f and c_1), the difference is small on the parameter scale: 0.5 mm or 0.5 years.

Posterior predictions and growth rates

Median posterior predictive size was nearly identical between the longitudinal and cross-sectional models (Fig. 2a). Similarly, the 98% prediction intervals (the age-specific range in which 98% of new observations are expected to be located) agree very well, and only small deviations are observable at the earliest ages. Furthermore, both 98% intervals very well encompass the observed data, indicating that the double-logistic model used here is able to model the growth of craniofacial traits. Median predicted growth rates and age at peak growth velocity estimated from the growth curves differ by less than 0.1 mm on average (RMSE = 0.064 mm/y; Fig. 2b).

Effects of sample size

As predicted, models estimated from small sample sizes show greater variation in predicted growth pattern. At young ages, models with 5–20 observations show greater variation than models fit to larger datasets (Fig. 3a), with some deviating by 10 mm or more (i.e., > 10% error). At sample sizes of 200 or more, all of the random-sample cross-sectional models closely approximate the longitudinal model. For the full sample model, the two are largely indistinguishable (Fig. 3a, lower right panel). Mean absolute difference between the predicted measure from the cross-sectional models and the predicted measure from the longitudinal model follows a similar pattern: as sample size increases, deviations decrease (Fig. 3b). Across the age range, all sample sizes of 100 or more show mean absolute deviations of 1 mm or less. We compared the growth milestones peak growth velocity (PGV) and age at peak growth velocity (aPGV) to those estimated from the longitudinal model across the different cross-sectional sample size. We found that PGV was < ~0.5 mm/y and aPGV < ~0.5 y, on average, with larger samples showing lower error in general (Supplemental information).

Discussion

Longitudinal approaches are recognized as the gold standard for growth modeling^{49,50}. As discussed above, however, longitudinal studies incur significant costs in terms of resources and participant commitment. When feasible to carry out, the advantage to this approach comes with the characterization of ontogenetic trajectories at the individual level in addition to the population level. It is clear that individuals, even though they might be similarly sized at the beginning and end of a growth spurt, may experience unique patterns of rate changes and thus differences in the timing of milestone achievements because the pacing of growth differs among individuals. We operate under the paradigm that longitudinal approaches provide a unique opportunity for an increased understanding of the variation in trajectories and of the influences on that variation. That knowledge can then be applied to analysis conducted on cross-sectional datasets to improve those results.

Population modeling

The common goal of growth modeling using both longitudinal and cross-sectional data is to use observed data for individuals to estimate the patterns of trait change over time and the timings and magnitudes of growth milestones. In longitudinal samples such as the CGCS, experimenters can observe variation between different individuals in growth trajectories as well as within individual variation attributable to the combined sources of measurement error. In cross-sectional data, both of these sources of variation—between individuals and measurement—are present, but they are not directly estimable. Thus, an implicit assumption of cross-sectional data is that the observed trait value falls at the true measurement for that individual at that age. Without any sense for measurement error, the observed value must be taken as accurate. That this assumption cannot be true is made clear by observing longitudinal data. Mean trait values increase monotonically, but individual observations do not, a pattern which is exhibited in Fig. 2a: measurements for some individuals appear to increase in size and then decrease, before increasing again. Thus, with longitudinal data, a single individual's trait measurements will increase monotonically on average but will appear to locally increase and decrease. We do not believe that individuals are growing and shrinking in successive measurements but rather that we are observing a "noisy" pattern around a changing trait value representing growth. Nonetheless, without any additional information or a model which explicitly includes measurement error, single observations for any single individual must be assumed to be the true measure. In other words, with cross-sectional data, the observed values are assumed to accurately represent the real measurement.

If interest lies only in the population mean growth pattern or in percentile intervals for observations at a specific age, then the ratio of phenotypic variation to measurement error is of less concern. Indeed, Cock⁵¹ argued that cross-sectional data are useful in the study of growth but only for population patterns. Our results underscore the value of cross-sectional data in elucidating population patterns, and we add that via (1) biologically grounded modeling, (2) priors that mildly constrain parameter estimates, and (3) Bayesian sampling, population growth patterns can be used to inform individual-level prediction. Improved individual prediction is a long-term goal of this research program. With only a single data point, individual trajectories will closely resemble a shifted population curve. However, with two or more observations (e.g., pre- and post-treatment), particularly if these fall at or around the timing of peak growth velocity, the population growth pattern can be used to inform and potentially constrain the range of possible trait values at maturity.

Sources of variation in data influencing parameter estimates

Observed craniofacial growth data can vary as a result of phenotypic variation in a population or via sources of error, each of which can impact parameter estimates and subsequent analyses. In the context of the craniofacial traits studied here, measurement error is introduced from a range of possible sources, including those stemming from collection of the radiograph (e.g., enlargement factor), non-orthogonal positioning of the participant to the radiographic plate¹³, and the scanning, digitizing and landmarking of the radiograph³⁶. Although steps are taken to mitigate each of these sources of error, some of the observed variation in craniofacial traits inevitably stems

from these sources. This experimental error is added to naturally occurring phenotypic variation. Despite these drawbacks, we have shown that the overall measurement error of a very large cross-sectional sample approximates the total variation of a longitudinal sample when all individuals are included in the sample, as indicated by the similarity in ranges of percentile intervals and overlap of the observed data with the 98% posterior interval (Fig. 2a). Although the CGCS is ancestrally homogeneous and we focus here on craniofacial traits, we believe that our results are broadly applicable to other groups and other traits. Our overall results are to be expected, because phenotypic variation and measurement error are indistinguishable from one another in real data measured on individuals such as the CGCS. In contrast, simulated data traditionally model only phenotypic variation either explicitly or implicitly.

Both phenotypic variation and measurement error contribute to observed (i.e., "apparent") variation, but in the absence of additional information, such as repeated sampling of the same individual at the same age, the relative magnitudes of each are unknown. Longitudinal modeling offers one approach to estimate measurement error for a trait. For a monotonically increasing trait value, the observed measurements could be assumed to oscillate around the true trait measurement, with deviations drawn from a normal distribution with a mean of zero. Indeed, the multilevel longitudinal model with random intercepts for each participant used in this study fits exactly this model. Using the multilevel longitudinal model, we estimated the median measurement error across all traits in both sexes at 0.87 mm (~1.2%). However, the percent error was lower for larger traits ($r = -0.64$, $P < 0.001$) where more of the observed variation can be attributed to inter-individual variation.

When considering phenotypes with large absolute magnitude relative to their associated measurement error, such as stature, observed variation can be reliably assigned as phenotypic variation rather than due to measurement error^{23,46–48,52,53}. In stature (or recumbent length in the very young), substantial increase is seen during growth. For example, an infant that is 80 cm at age 1 may increase to 180 cm at age 18. If the measurement error remains constant across the age range at ~0.2 cm, then the percent error decreases from 0.25% at age 1 year to 0.11% at age 18. Comparing this to the craniofacial traits examined here, largest craniofacial trait studied here (Nasion to Menton distance in males) is six times smaller than the stature of even a one-year-old child. The consequence is that models of large magnitude traits with relatively low measurement error (e.g., stature) will be better able to estimate population-level variation, whereas relatively small traits where measurement error can be relatively greater (e.g., nasion-menton) are more challenging to model longitudinally. Further, cross-sectional modeling across age may be more sensitive to such error, whereas longitudinal measurements on individual children can have internal checks on serial measures. Care must be taken, therefore, in craniofacial growth modeling, particularly in cross-sectional population modeling. Comparing PGV and aPGV milestones estimated from cross-sectional samples to the complete longitudinal data (Supplemental Information Fig. 1) for the smallest trait (ANS-PNS), a mid-sized trait (Sella-Gonion), and the largest trait (Nasion-Menton), we find no clear pattern in aPGV. It is slightly overestimated by about 6 months for ANS-PNS and N-M, but very accurately estimated for S-Go. Similarly, PGV is very accurate for ANS-PNS and N-M but slightly underestimated for S-Go by ~0.4 mm/y. Given that the CGCS represents the combined sample from separate growth studies spanning nearly 100 years, each of which used different imaging protocols and technologies, we find this level of accuracy to be acceptable. In the supplemental information, we provide these comparisons for all traits in both sexes.

How large of a cross-sectional sample is adequate?

We found that cross-sectional modeling using increasingly larger subsets of the entire CGCS sample of participants approximates longitudinal models very well for craniofacial traits. Both percentile size intervals and rates of growth from cross-sectional subsamples agree with longitudinal models fit to the full dataset (Fig. 2). Although any single model fit to a relatively small dataset may deviate substantially from the population pattern, an aggregated summary of a large number of models is a good approximation, even if those subsample sizes are relatively small (Fig. 3). It is important to note that the random samples in this study are themselves drawn from a very large sample (>7000 observations per trait for each sex). A single random set of even many hundred individuals might not adequately represent the population-level patterns, but repeated resampling of those data will provide an improved estimate of the overall growth pattern compared to a single model estimate (Fig. 3a). In these data, we found that a sample of at least 200 individuals used with resampling approaches to best estimate population growth patterns from cross-sectional data.

In conclusion, we found that cross-sectional data can provide robust estimates of not only growth in craniofacial traits over time, but also milestones derived from those growth curves, including peak growth velocity and age at peak growth velocity. Importantly, these results are similar to those from a large, densely sampled longitudinal data dataset, provided that the cross-sectional data are sampled repeatedly and at sufficiently large size. Our results provide a method by which to fully utilize cross-sectional data in situations when longitudinal samples are not possible, for example due to time, cost, or health concerns. We envision application of this cross-sectional resampling methodology to the analysis of 3D craniofacial imaging via cone-beam computed tomography (CBCT). Although CBCT involves a relatively low dose of radiation, serial imaging for research purposes is neither warranted nor feasible. The methods described here will prove invaluable to the study of not only craniofacial growth, but other forms of growth.

Data availability

Supporting information for this manuscript is available via Figshare (<https://figshare.com/s/a0d90329654c2a3d8fcf>), which includes R and stan code to reproduce all analyses and figures. Due to the raw data containing personal health information (PHI), we are not able to share the raw data. However, we do share the posterior samples for both longitudinal and cross-sectional analyses, which allow users to carry out all subsequent analyses.

Requests for data should be directed to Richard J. Sherwood (sherwoodrj@health.missouri.edu; Department of Pathology and Anatomical Sciences, University of Missouri School of Medicine, Columbia, Missouri, USA).

Received: 11 May 2023; Accepted: 26 October 2023

Published online: 07 November 2023

References

- German, R. Z. The ontogeny of sexual dimorphism: the implications of longitudinal versus cross-sectional data for studying heterochrony in mammals. in *Shaping Primate Evolution* (eds. Anapol, F., German, R. Z. & Jablonski, N. G.) 11–23 (Cambridge University Press, Cambridge, 2004).
- Scheideman, G. B., Bell, W. H., Legan, H. L., Finn, R. A. & Reisch, J. S. Cephalometric analysis of dentofacial normals. *Am. J. Orthod.* **78**, 404–420 (1980).
- Nanda, S. K. Patterns of vertical growth in the face. *Am. J. Orthod. Dentofacial Orthop.* **93**, 103–116 (1988).
- Arat, M., Köklü, A., Ozdiler, E., Rübendüz, M. & Erdoğan, B. Craniofacial growth and skeletal maturation: A mixed longitudinal study. *Eur. J. Orthod.* **23**, 355–361 (2001).
- Stahl, F., Baccetti, T., Franchi, L. & McNamara, J. A. Jr. Longitudinal growth changes in untreated subjects with class II division 1 malocclusion. *Am. J. Orthod. Dentofac. Orthop.* **134**, 125–137 (2008).
- Riolo, M. L., Moyers, R. E., McNamara, J. A. & Hunter, W. S. *An Atlas of Craniofacial Growth: Cephalometric Standards from the University School Growth Study, the University of Michigan* Vol. 2 (University of Michigan, 1974).
- Lewis, A. B., Roche, A. F. & Wagner, B. Growth of the mandible during pubescence. *Angle Orthod.* **52**, 325–342 (1982).
- McNamara, J. A. Jr. A method of cephalometric evaluation. *Am. J. Orthod.* **86**, 449–469 (1984).
- Nanda, R. S., Meng, H., Kapila, S. & Goorhuis, J. Growth changes in the soft tissue facial profile. *Angle Orthod.* **60**, 177–190 (1990).
- Bhatia, S. N. & Leighton, B. C. *Manual of Facial Growth: A Computer Analysis of Longitudinal Cephalometric Growth Data* (Oxford University Press, Oxford, 1993).
- Auconi, P. *et al.* Understanding interactions among cephalometrics variables during growth in untreated class III subjects. *Eur. J. Orthod.* **39**, 395–401 (2017).
- Barelli, E. *et al.* Exploiting the interplay between cross-sectional and longitudinal data in class III malocclusion patients. *Sci. Rep.* **9**, 6189 (2019).
- Sherwood, R. J. *et al.* Bayesian approach to longitudinal craniofacial growth: The craniofacial growth consortium study. *Anat. Rec.* **304**, 991–1019 (2021).
- Hardin, A. M. *et al.* Estimating craniofacial growth cessation: Comparison of asymptote- and rate-based methods. *Cleft Palate Craniofac. J.* **59**, 230–238 (2022).
- Knigge, R. P. *et al.* Craniofacial growth and morphology among intersecting clinical categories. *Anat. Rec.* **305**, 2175–2206 (2022).
- Nahhas, R. W., Valiathan, M. & Sherwood, R. J. Variation in timing, duration, intensity, and direction of adolescent growth in the mandible, maxilla, and cranial base: The Fels longitudinal study. *Anat. Rec.* **297**, 1195–1207 (2014).
- Oh, H. *et al.* Predicting adult facial type from mandibular landmark data at young ages. *Orthod. Craniofac. Res.* **22**(Suppl 1), 154–162 (2019).
- Hardin, A. M. *et al.* Clinical implications of age-related change of the mandibular plane angle. *Orthod. Craniofac. Res.* **23**, 50–58 (2020).
- Chumlea, W. C. *et al.* The first seriatim study into old age for weight, stature and BMI: The Fels longitudinal study. *J. Nutr. Health Aging* **13**, 3–5 (2009).
- Nahhas, R. W., Sherwood, R. J., Chumlea, W. C. & Duren, D. L. An update of the statistical methods underlying the FELS method of skeletal maturity assessment. *Ann. Hum. Biol.* **40**, 505–514 (2013).
- Duren, D. L., Nahhas, R. W. & Sherwood, R. J. Do secular trends in skeletal maturity occur equally in both sexes?. *Clin. Orthop. Relat. Res.* **473**, 2559–2567 (2015).
- Boeyer, M. E., Sherwood, R. J., Deroche, C. B. & Duren, D. L. Early maturity as the new normal: A century-long study of bone age. *Clin. Orthop. Relat. Res.* **476**, 2112–2122 (2018).
- Boeyer, M. E., Middleton, K. M., Duren, D. L. & Leary, E. V. Estimating peak height velocity in individuals: A comparison of statistical methods. *Ann. Hum. Biol.* **47**, 434–445 (2020).
- Munger, A. M. *et al.* The more the merrier: Integrating multiple models of skeletal maturity improves the accuracy of growth prediction. *J. Pediatr. Orthop.* **41**, e580 (2021).
- Chvatal, B. A., Behrents, R. G., Ceen, R. F. & Buschang, P. H. Development and testing of multilevel models for longitudinal craniofacial growth prediction. *Am. J. Orthod. Dentofac. Orthop.* **128**, 45–56 (2005).
- Jacob, H. B. & Buschang, P. H. Vertical craniofacial growth changes in French–Canadians between 10 and 15 years of age. *Am. J. Orthod. Dentofac. Orthop.* **139**, 797–805 (2011).
- Buschang, P. H., Jacob, H. B. & Demirjian, A. Female adolescent craniofacial growth spurts: Real or fiction?. *Eur. J. Orthod.* **35**, 819–825 (2013).
- Tu, Y.-K., Chiu, Y.-W. & Pandis, N. Analyzing longitudinal orthodontic data. Part 1: multilevel linear and curvilinear models. *Am. J. Orthod. Dentofac. Orthop.* **144**, 481–486 (2013).
- Steiner, C. C. Cephalometrics for you and me. *Am. J. Orthod.* **39**, 729–755 (1953).
- Ricketts, R. M. Cephalometric analysis and synthesis. *Angle Orthod.* **31**, 141–156 (1961).
- Buschang, P. H., Roldan, S. I. & Tadlock, L. P. Guidelines for assessing the growth and development of orthodontic patients. *Semin. Orthod.* **23**, 321–335 (2017).
- Broadbent, B. H. & Golden, W. H. *Bolton Standards of Dentofacial Developmental Growth* (The C.V. Mosby Company, Maryland Heights, 1975).
- Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, Cambridge, 2007).
- Pinheiro, J. C. & Bates, D. M. *Mixed-Effects Models in S and S-PLUS* (Springer, Berlin, 2000).
- Bock, R. D. *et al.* A parameterization for individual human growth curves. *Hum. Biol.* **45**, 63–80 (1973).
- Baumrind, S. & Miller, D. M. Computer-aided head film analysis: the University of California San Francisco method. *Am. J. Orthod.* **78**, 41–65 (1980).
- Sekhon, J. S. & Mebane, W. R. Genetic optimization using derivatives. *Polit. Anal.* **7**, 187–210 (1998).
- Holland, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence* (University of Michigan Press, Ann Arbor, 1975).
- Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Boston, 1989).
- Mebane, W. Jr. & Sekhon, J. Genetic optimization using derivatives: The rgenoud package for R. *J. Stat. Softw.* **42**, 1–26 (2011).
- Hoffman, M. D. & Gelman, A. The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
- Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).

43. Stan Development Team. *RStan: the R interface to Stan*. (2020).
44. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2022).
45. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-normalization, folding, and localization: An improved rhat for assessing convergence of MCMC. *Bayesian Anal.* **16**, 667–718 (2021).
46. Preece, M. A. & Baines, M. J. A new family of mathematical models describing the human growth curve. *Ann. Hum. Biol.* **5**, 1–24 (1978).
47. Jolicoeur, P., Pontier, J., Pernin, M. O. & Sempé, M. A lifetime asymptotic growth curve for human height. *Biometrics* **44**, 995–1003 (1988).
48. Sanders, J. O. *et al.* The uniform pattern of growth and skeletal maturation during the human adolescent growth spurt. *Sci. Rep.* **7**, 16705 (2017).
49. Singer, J. D. & Willett, J. B. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* Vol. 644 (Oxford University Press, Oxford, 2003).
50. Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. *Applied Longitudinal Analysis* (Wiley, New Jersey, 2011).
51. Cock, A. G. Genetical aspects of metrical growth and form in animals. *Q. Rev. Biol.* **41**, 131–190 (1966).
52. Cole, T. J., Donaldson, M. D. C. & Ben-Shlomo, Y. SITAR—a useful instrument for growth curve analysis. *Int. J. Epidemiol.* **39**, 1558–1566 (2010).
53. Simpkin, A. J., Sayers, A., Gilthorpe, M. S., Heron, J. & Tilling, K. Modelling height in adolescence: A comparison of methods for estimating the age at peak height velocity. *Ann. Hum. Biol.* **44**, 715–722 (2017).

Acknowledgements

This work was supported by National Institutes of Health, National Institute of Dental and Craniofacial Research Grant Nos. R01 DE024732, R01 DE024732-06S1, R03 DE021435. Additionally, creation of the CGCS was supported by the American Association of Orthodontists Foundation. We are grateful to the participants in all the studies that constitute the CGCS. This work would not have been possible without their sustained and generous contributions. We acknowledge the thoughtful comments of three anonymous reviewers, whose suggestions greatly improved this manuscript.

Author contributions

All authors conceived and designed the work. D.L.D., K.P.M., H.O., M.V., and R.J.S. acquired data. K.M.M. analyzed the data, drafted the manuscript, and produced the figures. All authors reviewed and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-46018-x>.

Correspondence and requests for materials should be addressed to K.M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023