Author:
Sandhu, Satpal S

Title:
Analysis of longitudinal jaw growth data to study sex differences in timing and intensity of the adolescent growth spurt for normal growth and skeletal discrepancies

Analysis of longitudinal jaw growth data to study sex differences in timing and intensity of the adolescent growth spurt for normal growth and skeletal discrepancies

Satpal Singh Sandhu

School of Education

University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Social Sciences and Law.

November 2019

Word count: 79,240

Abstract

The upper jaw (maxilla) and the lower jaw (mandible) grow in a downward and forward direction resulting in normal skeletal jaw relationship (Class I) with normal anteroposterior (upper and lower jaw length) and vertical (total face height) growth changes. An excessive or deficient growth in either jaw or both jaws result in skeletal malocclusions (Class II and Class III) characterized by anteroposterior and vertical skeletal discrepancies. Studying timing and intensity of the adolescent growth spurt is essential for the successful correction of skeletal malocclusions. A natural approach to estimation of timing (age at peak growth velocity, APGV) and intensity (peak growth velocity, PGV) involves fitting growth curve models (GCMs) and estimating derivatives.

Different linear mixed effects (LME) and nonlinear mixed effects (NLME) GCMs have been successfully applied to height data for estimating timing and intensity of the adolescent growth spurt. However, a systematic review of the literature (database searched until December 31, 2016) showed that studies applying GCMs to longitudinal jaw growth data focused exclusively on conventional polynomial based linear GCM. Furthermore, none of the previous studies simultaneously compared anteroposterior and vertical growth changes between normal skeletal jaw relationship and skeletal malocclusions.

In this thesis, I explored the potential of three LME and two NLME GCMs for studying jaw growth data available from the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection. Data comprised of repeated growth measurements of upper and lower jaw length and total face height on 128 males (mean age 11.67 years, standard deviation 2.92) and 139 females (mean age 11.60 years, standard deviation 2.88) between seven and 18 years of age. The LME models included were the conventional polynomial (CP), fractional polynomial (FP), and restricted cubic spline (RCS). The NLME models studied were the super imposition by translation and rotation (SITAR) and Preece-Baines (PB). The research goal was to first evaluate

and compare the fit of LME and NLME GCMs and then apply the best fitting linear or nonlinear GCM to the jaw growth data for studying class differences in the timing and the intensity of adolescent growth spurt between normal skeletal jaw relationship and skeletal malocclusions (i.e., Class I vs Class II, Class I vs Class III, and Class II vs Class III) for males and females.

In the first of the three research studies which make up this thesis, a simulation study was conducted to evaluate and compare the performance of popular information criteria (Akaike information criterion, AIC; Bayesian information criterion, BIC) and prediction criteria (measure of variance explained, R^2; concordance correlation coefficient, CCC) for selecting the optimal functional form for GCMs. I restricted attention to CP GCM in this study. Balanced and unbalanced data were simulated and analysed for different sample sizes and varying model complexity. Different versions of the restricted maximum likelihood (REML) based AIC and BIC were calculated to study the effect of different penalty adjustments on their performance. The AIC and BIC which included the total number of model parameters in their penalty terms performed at least as well and often better than their counterparts which included only the number of variance-covariance parameters. Both AIC and BIC performed consistently better than the prediction criteria in selecting the true model. Amongst the two information criteria, AIC performed better than BIC especially when sample size was small, and the model involved a complex variance covariance structure.

In the second research study, the AIC was then used to compare the fit of covariate adjusted CP, FP, RCS, SITAR and PB GCMs fitted to the upper jaw length, lower jaw length and total face height measurements (hereafter referred as outcomes). Data were analysed separately for males and females. Each GCM was fitted by including all possible individual-specific random effects. In addition to fit to the data, I also compared GCMs in terms of their ability to estimate covariate adjusted growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (APGV and PGV). The PB model failed to converge for any of the three outcomes for

both sexes. Results showed that unlike RCS and the SITAR GCMs, both CP and FP GCMs estimate biologically implausible growth trajectories (negative growth velocity). The RCS GCM fitted best to the data (as measured by the AIC) and therefore was selected for answering the clinical research questions in the final research study.

In the final research study, the RCS GCM was then used to estimate class differences in growth trajectories and the adolescent growth spurt parameters for males and females. Results showed sex differences in the timing and the intensity of adolescent growth spurt for normal growth and skeletal malocclusions. Females, on average, experience a less intense adolescent growth spurt which occurs almost one and half year earlier than males. Results indicated that an early but less intense growth spurt in the upper jaw length and the lower jaw length is mainly responsible for the development of anteroposterior (upper and lower jaw length) and vertical (total face height) skeletal discrepancies for Class II and Class III skeletal malocclusions. The clinical implications of the research findings are discussed.

Dedication

To my son, Ranbir Singh Sandhu

Acknowledgements

I sincerely thank my supervisors Prof. George Leckie, Prof. Kate Tilling and Dr Rachael Hughes for their continued support and guidance. Thank you for the great discussions we had in our fortnightly meetings which helped me immensely throughout my thesis. Any errors and omissions in this thesis are my own.

I extend my deepest gratitude to my family for their love and support. Thank you from the bottom of my heart. This work belongs to you as much as it belongs to me.

I would also like to acknowledge the financial support from Economic and Social Research Council, UK Research and Innovation [grant number S139511-120]. The views and opinions expressed in this thesis are my own and do not represent the views of the Economic and Social Research Council or the UK Research and Innovation.

Table of contents

List of Tables

List of Figures

List of Abbreviations

AIC	Akaike information criterion
APGV	age at peak growth velocity
BIC	Bayesian information criterion
CCC	concordance correlation coefficient
COPAD	condyle to point-A distance
COPOD	condyle to pogonion distance
CP	conventional polynomial
FP	fractional polynomial
GCM	growth curve model
LME	linear mixed effects model
ML	maximum likelihood
NLME	nonlinear mixed effects model
PB	Preece–Baines
PGV	peak growth velocity
RCS	restricted cubic spline
REML	restricted maximum likelihood
RIGLS	restricted iterative generalised least squares
SITAR	superimposition by translation and rotation
TFHNP	total face height nasion perpendicular

Glossary of terms

Term	Definition
Acceleration	A term describing the rate of change in velocity as a function of age.
Adolescent growth spurt parameters	A term referring to timing (age at peak growth velocity) and the intensity (peak growth velocity) of the adolescent growth spurt.
Class differences	Differences in growth parameters for the upper jaw length, lower jaw length and total face height parameters between normal skeletal jaw relationship (Class I) and skeletal malocclusions (Class II and Class III).
Class I	Normal skeletal jaw relationship characterized by harmonious length of the upper and lower jaws with a well-balanced total face height
Class II	Skeletal malocclusion defined by short lower jaw length, large upper jaw length, or a combination of both (i.e., an overbite). The skeletal malocclusion is often accompanied by a short total face height (deep bite).
Class III	Skeletal malocclusion defined by long lower jaw length, short upper jaw length, or a combination of both (i.e., an underbite). The skeletal malocclusion is often accompanied by an increase in the total face height (open bite).
Distance	A term used to describe increase in the upper jaw length, lower jaw length and the total face height as a function of age.
Growth trajectories	A collective term referring to distance, velocity and acceleration.
Intensity of the adolescent growth spurt	Peak growth velocity, which is a biological maturity indicator, reflects the maximum skeletal growth velocity during the adolescent growth spurt.
Lateral cephalogram	A standardized radiographic image of the head and face region taken from the profile (side) view. Age-related jaw growth changes are assessed from serial lateral cephalograms by superimposing them on a stable reference frame (such as Sella–Nasion plane) and recording horizontal and vertical anatomic landmarks.
Linear growth curve model	Synonym for linear mixed-effects model.
Lower jaw (Mandible)	The largest bone in the human skull. The horseshoe-shaped bone, which is composed of the body and the ramus, is located inferior to the maxilla. The alveolar process, which is the superior border, contains the lower teeth.
Lower jaw length	Length of the lower jaw measured on the lateral cephalogram as a linear distance between anatomic landmarks Condyle and Pogonion.

Term	Definition
Nonlinear growth curve model	Synonym for nonlinear mixed-effects model.
Sexual dimorphism	A term that refers to differences between males and females.
Skeletal malocclusion	A set of human craniofacial morphologic characteristics defined by excessive or deficient growth of craniofacial structures such as the upper jaw (maxilla) and the lower jaw (mandible).
Timing of the adolescent growth spurt	The age at peak growth velocity which denotes the age corresponding to the peak growth velocity during the adolescent growth spurt.
Total face height	Total face height measured on the lateral cephalogram as a linear distance between anatomic landmarks Nasion and Menton. The Nasion and Menton distance is recorded along a perpendicular dropped on the Frankfurt horizontal plane from the Nasion.
Upper jaw (Maxilla)	A pyramid shaped bone of the midface with its base adjacent to the nasal cavity and bounded by the alveolar process which serves as an anchor for the teeth of the upper denture.
Upper jaw length	Length of the upper jaw measured on the lateral cephalogram as a linear distance between anatomic landmarks Condyle and Point-A.
Velocity	A term denoting the rate of change in distance as a function of age.

CHAPTER 1. INTRODUCTION

1.1. Background

Human physical growth is not a smooth progression through time, but is inherently dynamic and proceeds in a series of spurts which reflect an increase in the growth velocity (Bogin, 2010; Cameron & Bogin, 2012; Stulp & Barrett, 2016). The adolescent growth spurt (also known as the pubertal growth spurt) is experienced by all individuals and is the most readily recognised aspect of adolescence. Compared to males, females experience a less intense adolescent growth spurt, which occurs almost two years earlier than males (Bernstein, 2018; Gurri, 2018; Hauspie & Roelants, 2012; Rogol et al., 2000; Sanders et al., 2017; Tanner & Cameron, 1980). Skeletal growth changes during adolescence are typically characterised by an increase in bone size (such as height and jaw length) and acquisition of bone mass (Duren et al., 2013; Stagi et al., 2013).

The timing of the peak growth velocity during adolescence is essential to consider when treating children with a wide variety of skeletal disorders, such as scoliosis in which the spine has a sideways curve (Busscher et al., 2012; Sanders et al., 2007). Similarly, the timing of the peak growth velocity spurt is the key factor in treatment planning to correct skeletal malocclusions with discrepancies in the jaw size and face height (Dean et al., 2016; Proffit, 2006; Proffit, 2014; Turpin, 2000).

Skeletal malocclusions result from excessive or deficient growth of the upper jaw (maxilla) and/or the lower jaw (mandible). Interaction between the anteroposterior (horizontal) and vertical vectors of jaw growth determines the final relationship between upper and lower jaw (De Clerck & Proffit, 2015; Dean et al., 2016; Mao & Nah, 2004; Nanda, 1983; Nanda, 1992; Ong, 2004; Proffit et al., 2014). While normal horizontal and vertical growth results in proportional length of the upper and lower jaws with a well-balanced total face height (Class I skeletal jaw relationship), an altered growth results in Class II and Class III skeletal malocclusions.

Depending on how upper and lower jaws are related to each other in the anteroposterior plane, the skeletal malocclusion is classified as Class II or Class III. Class II skeletal malocclusion results from a short lower jaw length, large upper jaw length, or a combination of both (i.e., an increase in overjet). In contrast, Class III skeletal malocclusion results from a short upper jaw length, large lower jaw length, or a combination of both (i.e., a decrease in overjet) (De Clerck & Proffit, 2015; Dean et al., 2016; Proffit et al., 2014).

For both Class II and Class III skeletal malocclusions, anteroposterior skeletal jaw discrepancies are often associated with vertical skeletal discrepancies resulting in skeletal deep bite (short face height) or open bite (long face height) (De Clerck & Proffit, 2015; Dean et al., 2016; Sassouni, 1969). While Class II skeletal malocclusion is often accompanied by a skeletal deep bite (i.e., an increase in overbite), the Class III skeletal malocclusion is typically characterised by a skeletal open bite (i.e., a decrease in overbite). Thus, there are "*two basic types of vertical disproportions (the skeletal deep bite and open bite) and two types of anteroposterior disproportions (the skeletal Class II and Class III)"* (Sassouni, 1969).

Prevalence of skeletal malocclusion is high, affecting both sexes across all ethnicities (Joshi et al., 2014). Class II skeletal malocclusion (global prevalence 20.9%) is more common than Class III skeletal malocclusion (global prevalence 7.3%), as reviewed by Joshi et al. (2014). Despite its lower prevalence, the impact of Class III skeletal malocclusion on quality of life is greater than Class II skeletal malocclusion (Bernabe et al., 2008).

Skeletal malocclusions adversely affect oral and general health, facial aesthetics, and psychosocial wellbeing (Baskaradoss et al., 2013; Bollhalder et al., 2013; Ghafournia & Hajenourozali Tehrani, 2012; Joshi et al., 2014; Martins-Junior et al., 2012; Masood et al., 2013; Proffit et al., 2014). Adolescents are particularly sensitive to the facial aesthetics, and skeletal jaw discrepancies can have a significant negative effect on their emotional and social wellbeing

(Dimberg et al., 2015; Koroluk, 2016). If left untreated, skeletal jaw discrepancies become more profound due to accelerated growth during adolescence (Koroluk, 2016).

Treatment of skeletal malocclusions is a priority for clinical practices worldwide (Bernabe et al., 2008; Joshi et al., 2014). Three treatment options available to correct anteroposterior and vertical skeletal jaw discrepancies are (Proffit et al., 2014): growth modification (modifying jaw growth to correct skeletal jaw discrepancies), camouflage (moving teeth to mask underlying skeletal jaw discrepancies), and orthognathic surgery (surgical correction of skeletal jaw discrepancies).

Treatment of skeletal malocclusions using growth modification procedures is more desirable than camouflage treatments. This is because growth modification improves the overall facial aesthetics by correcting the underlying skeletal jaw discrepancies, and therefore treatment results are more stable (Proffit et al., 2014). Considering the risk to benefit ratio, growth modification procedures are also recommended where possible over orthognathic surgery for correction of skeletal jaw discrepancies. As growth modification utilises the natural growth potential and treatment is possible only during active jaw growth, orthognathic surgery is used to correct skeletal jaw discrepancies in nongrowing adults (Kluemper & Spalding, 2001; Proffit et al., 2014).

Growth modification is thus the treatment of choice to correct skeletal jaw discrepancies when children are growing actively. The desired growth modification to correct Class II skeletal malocclusion is to stimulate lower jaw growth but restrain the upper jaw growth (De Clerck & Proffit, 2015). As Class III skeletal malocclusion is the reverse of Class II skeletal malocclusion, correction involves stimulation of the upper jaw growth and inhibition of the lower jaw growth (De Clerck & Proffit, 2015).

The orthodontic literature reports hundreds of appliances to modify jaw growth (De Clerck & Proffit, 2015). Briefly, functional appliances (such as activators) that position the lower jaw forward are the mainstays of treatment to correct Class II skeletal malocclusion (De Clerck & Proffit, 2015).

In addition to stimulating the growth of the lower jaw, functional appliances also inhibit the upper jaw growth which is a desirable feature when correcting Class II skeletal malocclusion (De Clerck & Proffit, 2015; Proffit et al., 2014).

For Class III skeletal malocclusion, which is typically characterized by deficient upper jaw growth, the facemask is the appliance of choice (De Clerck & Proffit, 2015; Proffit et al., 2014). Since the facemask takes support from the chin, it also restricts the excessive forward growth of the lower jaw which is often associated with Class III skeletal malocclusion.

As growth modification appliances stimulate / inhibit the bone growth, these appliances are typically used when children are still growing and have more pliable bone tissue. Growth modification procedures are most effective when the treatment timing is carefully coordinated with the timing of the adolescent growth spurt.

To correct skeletal jaw discrepancies, growth modification procedures initiated around the age at peak growth velocity provides the best results (Castaldo & Cerritelli, 2015; De Clerck & Proffit, 2015; Fleming & Lee, 2016; Proffit, 2006; Proffit et al., 2014; Turpin, 2000). This to harness the skeletal growth potential of the adolescent growth spurt. Failure to initiate growth modification procedures at the right age can result in a lost opportunity to correct skeletal jaw discrepancies. This then necessitates a surgical procedure, which carries a greater physical risk and a higher financial cost (DiBiase et al., 2015).

Researchers have used different methods to determine the timing of the adolescent growth spurt for jaw growth. These methods include assessment of sexual maturation indicators such as menarche onset in females and secondary sex characteristics in both sexes (such as menarche or genital development) (Bjork & Helm, 1967; Cole, 2003; Fishman, 1979; Hagg & Taranger, 1980; Hägg & Taranger, 1982; Valeggia & Núñez-de la Mora, 2015), chronological age (Hägg & Taranger, 1982), dental age (Bjork & Helm, 1967; Demirjian & Goldstein, 1976; Hägg & Matsson, 1985; San

Roman et al., 2002), and skeletal maturation using radiographs (Fishman, 1979; Grave & Brown, 1976; Hassel & Farman, 1995; Lamparski, 1975; San Roman et al., 2002).

Each of the above-mentioned methods used to assess the skeletal maturation has certain limitations. For example, assessment of sexual maturation based on secondary sex characteristics lacks objectivity because stages of development are somewhat arbitrary and discrete (Beunen et al., 2006). Besides radiation exposure, radiographic methods present difficulty in differentiating contiguous stages of skeletal maturity in adolescents (Shim et al., 2012). Dental age (tooth development and eruption) is not a reliable indicator of growth because dental maturation is significantly influenced by local factors (such as space available for a tooth to erupt) that are not related to the biology of skeletal growth (Buschang et al., 2017; Flores-Mir et al., 2004).

Methods based on longitudinal information pertaining to jaw growth velocity are more accurate than assessment of sexual maturation or radiographic methods (Buschang et al., 2017; Flores-Mir et al., 2004). The peak growth velocity (PGV) is a somatic biological maturity indicator (somatic tissues include musculoskeletal and general body tissues), which reflects the maximum skeletal growth velocity denoting intensity of the adolescent growth spurt (Hauspie et al., 2004). The age at peak growth velocity (APGV) denotes the timing of the adolescent growth spurt (Hauspie et al., 2004) and is considered the gold standard to objectively assess the skeletal maturity of jaw bones (Flores-Mir et al., 2004). Unlike secondary sex characteristic–based assessment of skeletal maturity, which cannot be used to compare males and females, APGV can be used for sex comparisons (Sherar et al., 2004).

1.2. Purpose and scope

The purpose of this thesis is to enhance academic and clinical understanding of jaw growth during adolescence. It focuses on comparing growth trajectories (distance, velocity and acceleration) and the adolescent growth spurt parameters (timing and intensity) for the upper jaw length, lower jaw length and total face height between the normal skeletal jaw relationship (Class I) and skeletal malocclusions (Class II and Class III). Studying adolescent jaw growth trajectories and adolescent growth spurt parameters will help in better understanding the development of skeletal jaw discrepancies and their correction through growth modification procedures.

The adolescent growth period follows immediately after the juvenile growth period (age seven to 10 years for females, and seven to 12 years for males) and ends at approximately 18 years of age (American Psychological Association, 2002; Curtis, 2015). For studying jaw growth during adolescence, data should ideally include jaw growth measurements from at least age eight for females, and age 10 for males (Mellion et al., 2013). The present study includes jaw growth data between seven and 18 years for both sexes.

Data comprised of repeated growth measurements (also referred to as longitudinal growth data) on individuals is central to studying human growth (Cameron & Bogin, 2012; Hauspie et al., 2004). The American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection has recently made available longitudinal jaw growth data for orthodontic research and education purposes (Baumrind & Curry, 2015).

The AAOF Craniofacial Growth Legacy Collection pools longitudinal growth data from nine historic growth studies conducted in the United States and Canada between 1929 and 1982 (Baumrind & Curry, 2015). Data comprised of repeated growth measurements of craniofacial structure such as upper and lower jaw length and total face height for Class I, Class II and Class III skeletal jaw relationships is a valuable source for studying jaw growth (Baumrind & Curry, 2015).

Analysis of longitudinal growth data using an appropriate statistical method allows descriptions of growth trajectories which include distance, velocity and acceleration (Molinari & Gasser, 2004). Distance denotes increase in size with age (such as height and jaw length), whereas velocity (the first derivative of distance with respect to age) describes the rate of increase in size over time. Acceleration (the second derivative of distance or the first derivative of velocity) represents change in growth velocity as a function of age (Molinari & Gasser, 2004). Derivatives (velocity and acceleration) are used to estimate the timing (APGV) and intensity (PGV) of the adolescent growth spurt (Hauspie et al., 2004).

The term growth curve modelling refers to a wide array of statistical methods used for analysing repeated measurement data (Curran et al., 2010). Different approaches have been developed to analyse repeated measurement data. Data can be analysed separately for each child, or simultaneously for all children (Johnson, 2015). Analysing data simultaneously for all children is more efficient and allows for modelling between-individual (inter-individual) variation in individual-specific (intra-individual) growth trajectories, which is not possible with individual growth curve analysis (Ghisletta et al., 2015; Goldstein, 2011b; Johnson, 2015; Snijders & Bosker, 2012). Due to their ability to answer broader research questions, methods that analyse data simultaneously for all children are collectively called growth curve models (GCMs).

"*The contemporary use of the term growth curve model typically refers to statistical methods that allow for the estimation of inter-individual variability in intra-individual patterns of change over time"* (Curran et al., 2010).

GCMs can be broadly grouped into latent growth curve models and mixed-effect models (Johnson, 2015; McNeish & Matta, 2018). This thesis used mixed-effects models. The term mixed-effects implies that a model estimates both fixed (population-average) and random (individual-specific) effects which collectively describe the growth trajectory of each individual in the sample

(Goldstein, 2011b; Johnson, 2015; Johnson et al., 2013). Mixed-effects models allow modelling and quantifying systematic differences in growth trajectories due to independent variables (such as class of skeletal jaw relationship) (Curran et al., 2010; Goldstein, 2011b; Johnson, 2015).

Based on the underlying process of defining growth trajectories, mixed-effects models can be developed as linear mixed effects (LME) models (Bryk & Raudenbush, 1987; Goldstein, 1986a, 1986b; Goldstein, 1989a, 2011b; Laird & Ware, 1982) or nonlinear mixed-effects (NLME) models (Davidian & Gallant, 1992; Lindstrom & Bates, 1990; Mallet et al., 1988; Pinheiro & Bates, 1995a; Sheiner & Beal, 1980; Vonesh & Carter, 1992). Both Bayesian and frequentist approaches are available for estimating GCMs (Goldstein, 2011b; Stegmueller, 2013). Although Bayesian modelling has some advantages over the frequentist approach, especially when sample size is small, the frequentist approach is more popular for fitting GCMs (de Valpine, 2012; Stegmueller, 2013). This thesis used a frequentist approach.

In this thesis, the terms GCM and mixed-effect model were used synonymously. The LME and NLME models were referred to as linear and nonlinear GCMs, respectively. As the longitudinal growth data follows a hierarchical (multilevel) structure in which repeated growth measurements (at level 1 in the data hierarchy) are nested in individuals (level 2), the term multilevel model (MLM) is often used to denote GCM (Goldstein, 2011b; Johnson, 2015).

Use of nonparametric methods such as kernel estimation (Gasser et al., 2004; Gasser et al., 1984) and functional data analysis (Muller & Yao, 2010; Yao et al., 2005) were not considered as regression methods based on a parametric function are the most widely applied methods to analyse longitudinal growth data and in most situations are considered preferable to nonparametric methods (Goldstein, 1986a; Hauspie et al., 2004).

1.3. Aim and research questions

Estimation of timing (APGV) and intensity (PGV) of the adolescent growth spurt requires fitting a GCM and estimating derivatives. As stated earlier (Section 1.2), linear and nonlinear GCMs are methods of choice for analysing longitudinal jaw growth data as they allow modelling and quantifying individual specific variability in jaw growth.

A systematic review of the literature (Chapter 2, Section 2.2) showed that only a few studies (a total nine studies) have applied GCMs to longitudinal jaw growth data. All these studies exclusively focused on a linear GCM using conventional polynomials for modelling jaw growth trajectories. Furthermore, none of the previous studies directly compared the anteroposterior (upper and lower jaw length) and the vertical (total face height) growth changes between normal skeletal jaw relationship (Class I) and skeletal malocclusions (Class II and Class III).

The aim of this thesis was to first evaluate and compare the fit and performance of different linear and nonlinear models and then apply the best fitting GCM to the jaw growth data for studying sex differences in the timing and the intensity of adolescent growth spurt for the upper and lower jaw length and total face height between normal skeletal jaw relationship and skeletal malocclusions. I focused on three linear and two nonlinear GCMs that are popular for modelling longitudinal height data. The three linear GCMs included are the conventional polynomial (CP), fractional polynomial (FP), and restricted cubic spline (RCS). The two nonlinear GCMs are the superimposition by translation and rotation (SITAR) and the Preece–Baines (PB). To achieve the research aim, a systematic approach was adopted, which involved answering three specific research questions.

1.3.1. Research Question 1 (Chapter 5)

How well do different information criteria and prediction criteria perform when choosing between competing GCMs to select a best-fitting GCM?

1.3.2. Research Question 2 (Chapter 6)

Which GCM fits best to jaw growth data and performs well in modelling jaw growth trajectories and estimating adolescent growth spurt parameters?

1.3.3. Research Question 3 (Chapter 7)

How do growth trajectories and the adolescent growth spurt parameters for the upper jaw length, lower jaw length and total face height differ between the normal skeletal jaw relationship (Class I) and skeletal malocclusions (Class II and Class III) for males and females?

The first research question (Chapter 5) addresses the core issue of GCM selection. This is central to answering the second research question (Chapter 6) which addresses the fundamental issue of selecting a GCM that best fits to jaw growth data. The third research question (Chapter 7) addresses the key clinical issues relating to class differences in jaw growth trajectories and the adolescent growth spurt parameters. I further elaborate on these three research questions in Chapter 2 (Section 2.3) and Chapter 4 (Section 4.5).

1.4. Structure and organization

This thesis is organised into the following nine chapters. Chapters 5, 6 and 7 are research chapters addressing the research questions 1, 2 and 3, respectively. Each research chapter is comprehensive by itself and is structured as follow: background and aim, methodology, results, discussion, limitations and conclusion.

In Chapter 2, I first briefly review the concept of skeletal jaw growth during adolescence and then present a systematic review of the literature on jaw growth during adolescence. The focus of the systematic review is to critically review methodology and findings reported in earlier studies investigating timing and intensity of the adolescent growth spurt for the upper jaw length, lower jaw length and total face height for males and females. The chapter concludes by summarizing the key gaps in the clinical literature on jaw growth during adolescence and relating them to three specific research questions posed in this thesis.

Chapter 3 describes the longitudinal growth data available from the AAOF Craniofacial Growth Legacy Collection project which is analysed in this thesis. After providing a detailed description of historic growth studies which contributed to the AAOF Craniofacial Growth Legacy Collection database, the Chapter then presents the key limitations of the data.

Chapter 4 reviews linear (CP, FP and RCS) and nonlinear (SITAR and PB) GCMs and methodology of estimating growth trajectories (distance velocity and acceleration) and adolescent growth spurt parameters (APGV and PGV) using derivatives. The chapter concludes by summarizing the key gaps in the methodological literature on modelling growth data and outlines the methodological and clinical research work which I do in this thesis while addressing each of the three specific research questions.

Chapter 5, which addresses the first research question, presents a simulation study conducted to evaluate and compare the performance of popular information criteria (Akaike information criteria, AIC; Bayesian information criteria, BIC) and prediction criteria (measure of variance explained, R2; concordance correlation coefficient, CCC) for selecting the optimal functional form for GCMs. The chapter concludes by identifying the criterion of choice which is then used to evaluate and compare the fit of different GCMs in the next study (Chapter 6).

Chapter 6 answers the second research question by comparing the fit of linear (CP, FP and RCS) and nonlinear (SITAR and PB) GCMs applied to the upper jaw length, lower jaw length and total face data for males and females. Each GCM is fitted by including class of the skeletal jaw relationship (Class I, Class II and Class III) as a covariate. To compare fit to the data, the *a priori* criterion selected in the previous study (Chapter 5) is used. In addition to the model fit, I also compare GCMs in terms of their ability in modelling covariate adjusted growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (APGV and PGV). The chapter concludes by identifying the best fitting GCM which is then used in the next study (Chapter 7) for answering the clinical research questions.

Chapter 7 applies the best fitting GCM to jaw growth data to quantifying sex differences in growth trajectories and adolescent growth spurt parameters for upper jaw length, lower jaw length and total face height between normal skeletal jaw relationship (Class I) and skeletal malocclusions (Class II and Class III). Class differences in growth trajectories and adolescent growth spurt parameters are appropriately adjusted for potential study effects (historic growth studies which contribute to the data). The chapter discusses the clinical implications of research findings.

Chapter 8 provides a general discussion of methodological (Chapters 5 and 6) and clinical (Chapter 5) research findings and outlines contributions this thesis makes to the methodological and clinical literature on modelling jaw growth data during adolescence. The methodological and

clinical limitations of my work are discussed and specific topics for the future research work are suggested.

Chapter 9 concludes the thesis by summarising the key findings and highlighting the main points.

CHAPTER 2. REVIEW OF THE LITERATURE

This chapter first provides an overview of jaw growth during adolescence, and then presents a systematic review of the literature. The objective of the systematic review of the literature is to synthesise evidence on jaw growth during adolescence with an emphasis on the adolescent growth spurt. The chapter concludes by summarising the review findings and identifying the research gaps.

2.1. An overview of jaw growth during adolescence

The concept that bone grows with age is one of the fundamental paradigms in research on human skeletal growth changes over time (Duren et al., 2013). Growth is more readily understood when a physical pattern is used to describe how bones grow with age (Dean et al., 2016). Scammon (1930) proposed that growth of different body tissues can be summarised as four distinct growth patterns: neural (brain), lymphoid (organs of the immune system), genital (organs of the reproductive system) and somatic (e.g., musculoskeletal tissues comprised of muscles and bones).

Postnatal jaw growth follows a pattern that is intermediate between neural and somatic growth patterns (Mitchell, 2013; Proffit, 2014). I focus here on upper jaw length, lower jaw length and total face height, the three dimensions of jaw growth explored in this thesis and which are considered most important for correction of horizontal and vertical skeletal jaw discrepancies observed in Class II and Class III skeletal malocclusions. The growth pattern of the lower jaw more closely resembles the somatic growth pattern than the upper jaw. Figure 2-1 shows upper and lower jaw growth curves, along with Scammon's growth curves for neural, lymphoid, genital and somatic tissues (Scammon, 1930).

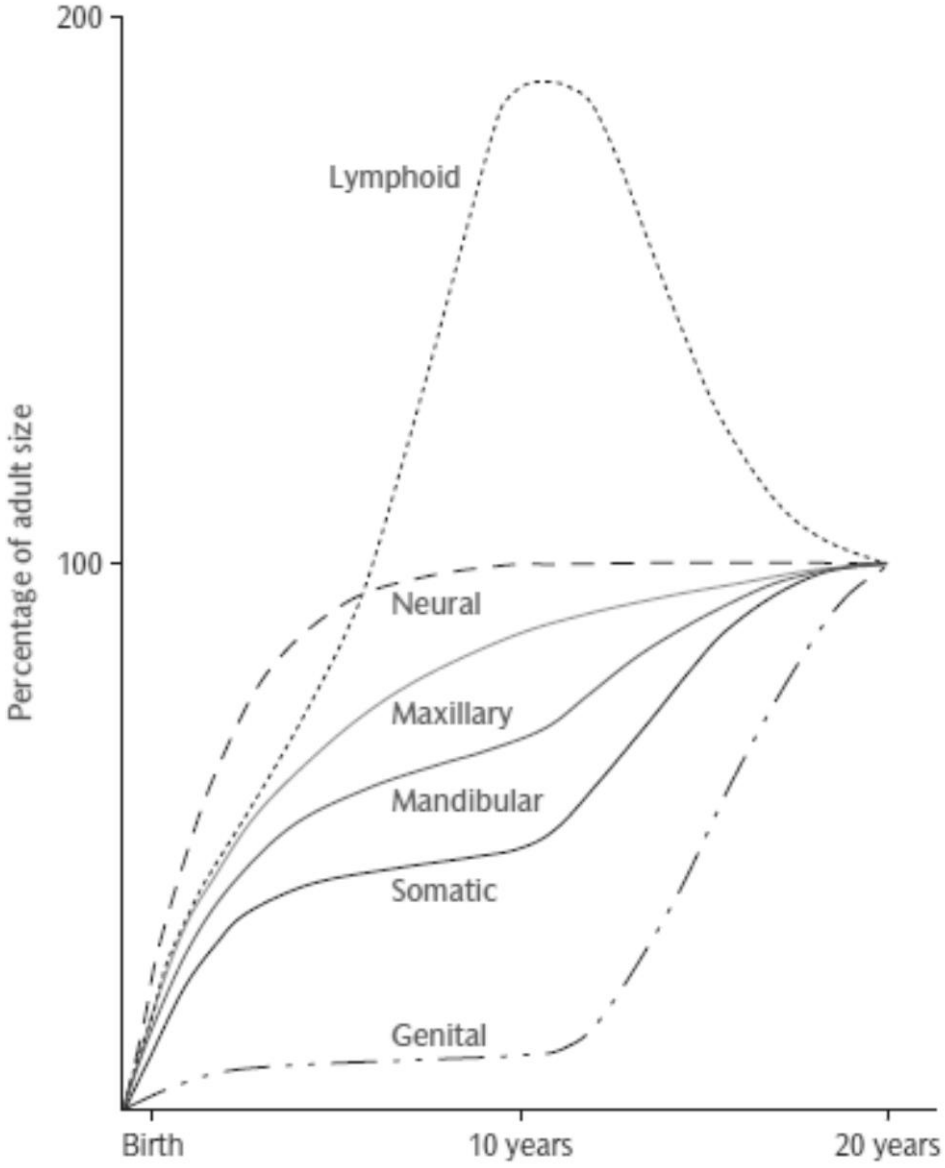


Figure 2-1. Postnatal growth curves for upper and lower jaws. Jaw growth curves are shown against the background of neural, lymphoid, somatic and genital tissue growth curves as percentages of total increase. The patterns for somatic growth, upper jaw growth and lower jaw growth are shown as solid lines. (From Mitchell, L. (2013). An introduction to orthodontics (4th ed.). *Oxford University Press*. Reproduced with permission of *Oxford University Press* through PLSclear. Copyright Laura Mitchell).

During adolescence, both upper and lower jaws (the focus of the current research) grow primarily in a downward (vertical) and forward (horizontal) direction. This is because transverse (width) growth ceases at an earlier age than vertical and horizontal growth (Cobourne & DiBiase, 2010; Mitchell, 2013; Proffit et al., 2014; Thilander, 1995). For both sexes, transverse growth declines to adult levels at around nine years of age for the lower jaw (inter-canine width), and 12 years for the upper jaw (Mitchell, 2013).

In relation to the cranial base (skull base), which completes its entire growth (95% of the adulthood size) by age seven (Cameron, 2008; Cobourne & DiBiase, 2010; Proffit, 2014; Thiesen et al., 2013), the upper jaw continues to grows in a downward and forward direction during adolescence (Cobourne & DiBiase, 2010; Mitchell, 2013; Proffit, 2014), as shown in Figure 2-2. To maintain its structural and functional harmony with the upper jaw, the lower jaw also grows in a downwards and forwards direction (Cobourne & DiBiase, 2010; Proffit et al., 2014), as shown in Figure 2-3.

Since both upper and lower jaws grow in a downward and forward direction, the overall face height also increases (Figure 2-4). Due to the cephalocaudal gradient of growth (see below), the lower jaw matures later than the upper jaw, and shows greater growth (Proffit, 2014). Because face height increases with overall jaw growth, the growth pattern of the total face height matches closely to that of the lower jaw growth pattern.

According to the cephalocaudal gradient of growth, body parts that are farther away from the head, such as legs, mature later but grow more than those that are closer to the head, such as arms (Hermanussen, 2016; Humphrey, 1998; Kingsbury, 1924; Smith & Buschang, 2005). Likewise, the lower jaw, which is away from head, matures later and grows more than the upper jaw, which is closer to the head (Buschang et al., 1983; Proffit, 2014).

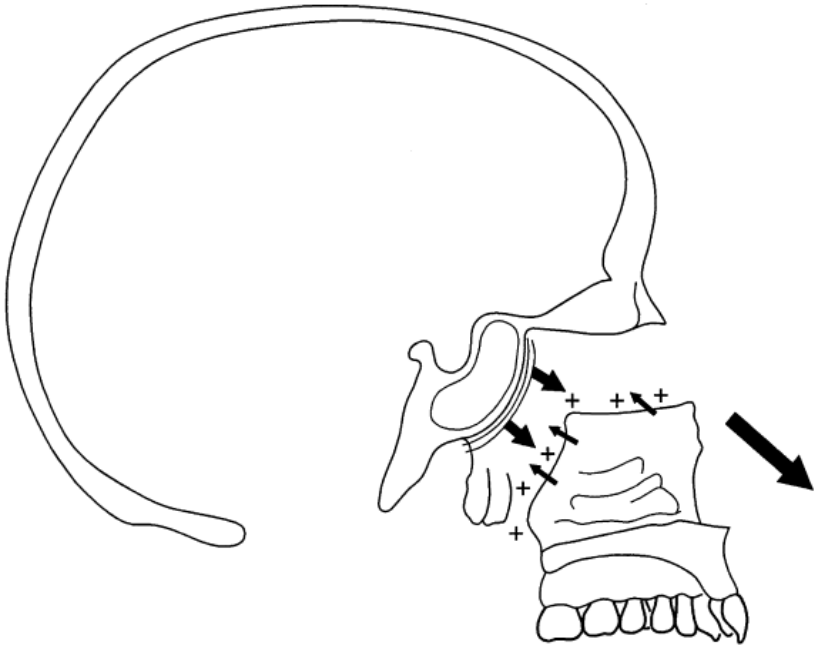


Figure 2-2. Downward and forward growth of the upper jaw. (From Mitchell, L. (2013). An introduction to orthodontics (4th ed.). *Oxford University Press*. Reproduced with permission of *Oxford University Press* through PLSclear. Copyright Laura Mitchell).

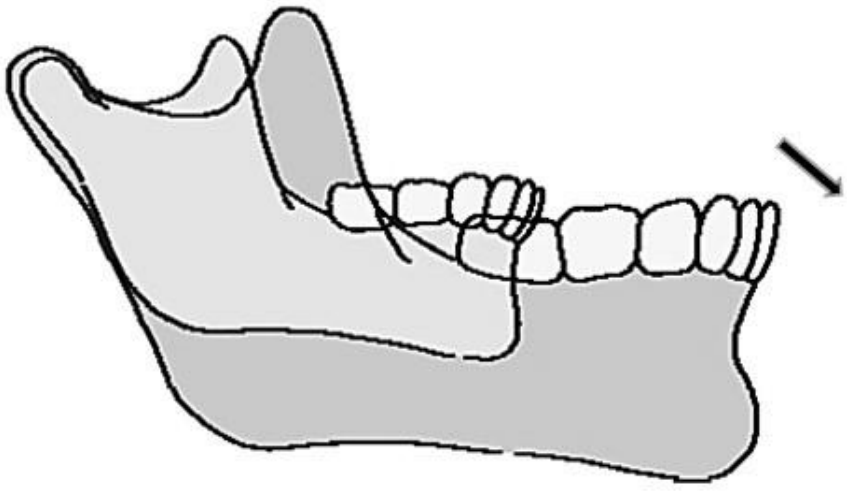


Figure 2-3. Downward and forward growth of the lower jaw. (From Mitchell, L. (2013). An introduction to orthodontics (4th ed.). *Oxford University Press*. Reproduced with permission of *Oxford University Press* through PLSclear. Copyright Laura Mitchell).

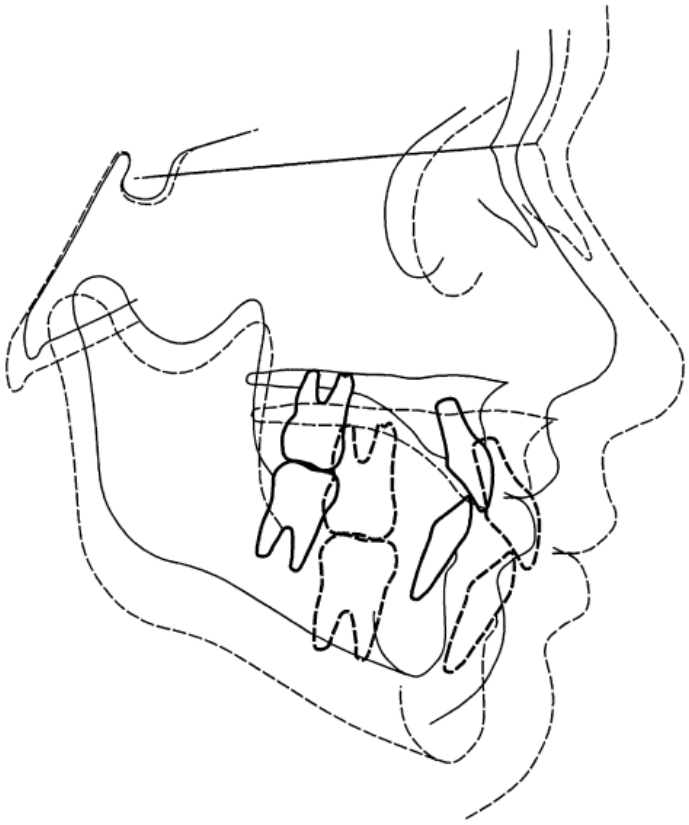


Figure 2-4. An overall downward and forward growth of the upper and lower jaw results in an increase in the total face height. Solid line eight years; broken line 18 years of age. (From Mitchell, L. (2013). An introduction to orthodontics (4th ed.). *Oxford University Press*. Reproduced with permission of *Oxford University Press* through PLSclear. Copyright Laura Mitchell).

2.2. A systematic review of the literature on jaw growth during adolescence

2.2.1. Background and objective

Adolescence is a period of rapid jaw growth. Hellman (1927) and Goldstein (1936) were the first to report "spurts" in the growth of facial bones. However, these early studies, which were based on excavated human skulls (Hellman, 1927) or direct anthropometric measurements on the face (Goldstein, 1936), did not provide a detailed and precise picture of jaw growth dynamics of the adolescent growth spurt.

The feasibility and precision of studying jaw growth were greatly enhanced by the introduction of the radiographic lateral cephalometric technique. This technique, which allows direct horizontal and vertical measurement of facial bones, founded the basis of studying jaw growth in historic studies (see Chapter 3).

Over the decades, attempts have been made to enhance clinical understanding of jaw growth by studying longitudinal lateral cephalometric growth data. In this systematic review, the available evidence on growth of the upper and lower jaw length and total face height for males and females is appraised and summarised, with a focus on the adolescent growth spurt.

2.2.2. Methodology

A systematic approach was adopted to identify and retrieve studies evaluating growth of the upper jaw length, lower jaw and total face height during adolescence. I followed the Centre for Reviews and Dissemination (2009) guidelines for undertaking systematic reviews in health care.

Eligibility criteria

Inclusion criteria:

- Longitudinal (i.e., a single cohort is followed over time) or mixed longitudinal (i.e., different cohorts enter study at different point of time and then followed longitudinally) growth studies using lateral cephalograms to evaluate skeletal growth of at least one of the three growth measures (upper jaw length, lower jaw length and face height) in normal healthy individuals (seven to 18 years of age).
- Studies evaluating jaw growth in orthodontically untreated individuals.
- Studies reporting horizontal (sagittal/antero-posterior) and/or vertical skeletal growth assessment.
- Studies with a minimum of three lateral cephalogram images (repeated growth measures) per individual.

Exclusion criteria:

- Studies not including adolescence and focusing on growth assessment in children (under seven years of age) or adults (over 18 years of age).
- Studies evaluating skeletal jaw growth in orthodontically treated individuals or those with craniofacial malformations such as cleft lip and palate.
- Studies exclusively focusing on transverse skeletal growth.
- Studies published in language other than English that could not be translated appropriately into English[1].

Google Translate (https://translate.google.co.in/).

Search strategy

The Medline and Embase electronic databases were searched to identify and retrieve relevant studies. For reviews of health care, these are the most commonly used databases (Centre for Reviews and Dissemination, 2009). Both databases were searched without language restrictions, from their inception time until 31 December 2016. No restrictions were applied for sex, race or ethnicity.

Details of electronic search strategy are provided in Appendix E. Briefly, the electronic search strategy was based on the search filters available for identification and retrieval of longitudinal epidemiological, case series and cohort studies (Fraser et al., 2006; Marcano Belisario et al., 2013; University of Texas School of Public Health, 2015). After running the search filter, the search results were combined with the results obtained by running the search terms specifically developed to identify and retrieve review-specific studies. In addition to the Medline and Embase electronic databases, grey literature was searched by hand searching the reference lists and through Google Scholar by using "cephalogram" and "growth" as search terms.

Study selection and data extraction

Titles and abstracts of all retrieved studies were screened to remove duplicate records and eliminate studies that failed to meet the review objective. Based on the eligibility criteria, full texts of remaining studies were scrutinised for potential inclusion. For all studies that met the inclusion criteria, data were extracted for the following items: i) study identification: first author's name and year of publication, ii) objective: research focus i.e., growth measure(s), iii) participants characteristics: sample size, age range, number of males and females, and class of skeletal jaw relationship (Class I, Class II or Class III), iv) number of growth measures per individual as well as the total number of measurements, and v) statistical method used for data analysis.

Most of the studies did not report the class of skeletal jaw relationship, even though they stated that both normal skeletal occlusion and skeletal malocclusion cases were included in the

sample. The number of growth measures per individual were not clearly reported in some studies. Some studies vaguely reported that an appropriate statistical approach was used for analysis, such as regression or analysis of variance, without providing further details on the statistical method.

2.2.3. Results

Search results

The search flow diagram is shown in Figure 2-5. A total of 560 studies were retrieved by running the electronic search strategy (until 31 December 2016). After removing 272 duplicate records, 288 studies remained. An additional 13 studies were located by electronically searching Google Scholar and hand searching reference lists. Out of a total 301 studies that were screened for titles and abstracts, 167 were eliminated for reasons outlined in the flow diagram (Figure 2-5). Scrutinising the full text of the remaining 134 articles, based on the eligibility criteria, resulted in further exclusion of 97 studies (see Figure 2-5 for reasons).

Thus, a total of 37 studies, published between 1955 and 2014, were included in this systematic review (Baccetti et al., 2011; Ball et al., 2011; Bambha, 1961; Bambha & Van Natta, 1963; Baughan et al., 1979b; Baume et al., 1983; Bishara, 1981; Bishara et al., 1981; Björk, 1963; Buschang et al., 1983; Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Buschang et al., 1986; Chvatal et al., 2005; Jamison et al., 1982; Lewis et al., 1982; Lewis et al., 1985; Mellion et al., 2013; Mitani, 1977; Moore et al., 1990; Nahhas et al., 2014; Nanda, 1955, 1971; Nanda, 1988; Nanda, 1992; Nanda & Rowe, 1989; Ochoa & Nanda, 2004; Pileski et al., 1973; Savara & Tracy, 1967; Singh & Savara, 1966; Thilander et al., 2005; Thompson et al., 1976; Tracy & Savara, 1966; van der Beek et al., 1996).

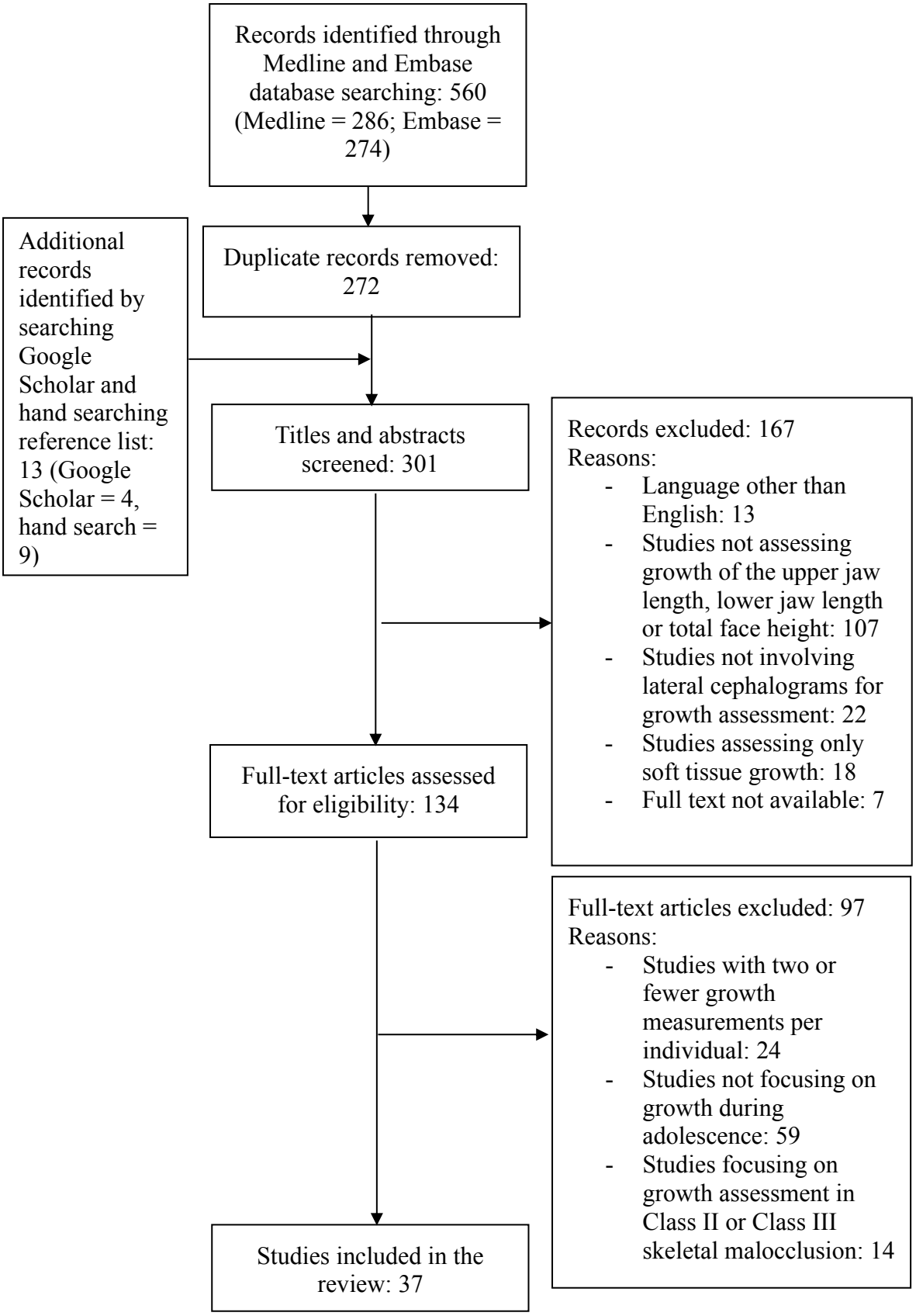


Figure 2-5. Flow diagram showing the different phases of the systematic review.

Characteristics of studies included in the systematic review

Out of the total 37 studies included in the systematic review, eight studies included only females (Baughan et al., 1979b; Buschang et al., 2013; Buschang et al., 1989; Nanda, 1992; Singh & Savara, 1966; Thompson et al., 1976; Tracy & Savara, 1966; van der Beek et al., 1996) whereas five studies included only males (Ball et al., 2011; Björk, 1963; Buschang et al., 1983; Buschang et al., 1986; Savara & Tracy, 1967). The remaining 22 studies evaluated jaw growth for both sexes. The sample sizes across all 37 studies ranged from 15 individuals, 10 males and 5 females (Nanda, 1955) to 459 individuals, 215 males and 244 females (Thilander et al., 2005). The median sample size across all studies was 50.

Out of 37 studies, 29 studies (78.4%) used simple statistical methods (such as ordinary least square regression) with growth velocity estimated as the incremental increase in growth over consecutive measures (Baccetti et al., 2011; Ball et al., 2011; Bambha, 1961; Bambha & Van Natta, 1963; Baughan et al., 1979b; Baume et al., 1983; Bishara, 1981; Bishara et al., 1981; Björk, 1963; Buschang et al., 1983; Buschang et al., 1986; Jamison et al., 1982; Lewis et al., 1982; Lewis et al., 1985; Mellion et al., 2013; Mitani, 1977; Moore et al., 1990; Nanda, 1955, 1971; Nanda, 1988; Nanda, 1992; Nanda & Rowe, 1989; Ochoa & Nanda, 2004; Pileski et al., 1973; Savara & Tracy, 1967; Singh & Savara, 1966; Thompson et al., 1976; Tracy & Savara, 1966).

Eight studies (21.6%) used growth curve models (GCMs) for data analysis (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996). All eight studies used conventional polynomial (CP) GCM for data analysis (see Table 2-1). The authors used conventional polynomials ranging from a second-degree polynomial up to a sixth-degree polynomial for modelling jaw growth trajectories. I discuss model fitting in full detail in Chapter 6 (Section 6.4.6). Here I focus on summarizing the key findings reported in these eight studies.

Table 2-1. Characteristics of studies applying growth curve models for longitudinal jaw growth data analysis.

ID	Author(s)	Year	Growth measure	Number of individuals	Sex (M/F)	Age range	Class of skeletal jaw relationship [a]	Number of growth measurements	
								Per individual	Total
1	Buschang et al.	1988	Lower jaw length	113	71 / 42	6 to 15	Class I: 71 Class II: 42 Class III: 0	8 to 10	918
2	Buschang et al.	1988	Lower jaw length	209	104 / 105	10 to 15	Not reported	4 to 6	1156
3	Buschang et al.	1989	Lower jaw length	105	0 / 105	6 to 15	Not reported	8 to 10	772
4	van der Beek et al.	1996	Total face height	134	0 / 134	7 to 14	Not reported	5 to 10	1071
5	Buschang et al.	1999	Lower jaw length	221	113 / 108	6 to 16	Not reported	8 to 10	1647
6	Chvatal et al.	2005	Lower jaw length	287	128 / 159	6 to 15	Not reported	5 to 10	1941
7	Buschang et al.	2013	Upper jaw length, lower jaw length and total face height	111	0 / 111	10 to 15	Not reported	4 to 6	625
8	Nahhas et al.	2014	Upper jaw length and lower jaw length	193	148 /145	4 to 24	Not reported	4 to 19	3106

[a] Normal skeletal jaw relationship (Class I) or skeletal malocclusion (Class II or Class III).

Considering the sexual dimorphism in jaw growth, each study fit CP GCM separately for males and females. One study analysed data with an age range of four to 24 years (Nahhas et al., 2014). Another study included individuals between six and 17 years of age (Buschang et al., 1999). The remaining five studies analysed data with an age range of seven to 15 years. Four studies evaluated jaw growth in only females (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1989; van der Beek et al., 1996) whereas the remaining four included data for both sexes.

One study analysed the upper and lower jaw length data for both sexes (Nahhas et al., 2014) whereas another study (Buschang et al., 2013) applied CP GCM to the upper and lower jaw length and total face height data for females. The remaining six studies focused only on analysing the lower jaw length data (Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; van der Beek et al., 1996).

With the exception of one study, which compared lower jaw length growth between normal skeletal jaw relationship (Class I) and Class II skeletal malocclusion (Buschang et al., 1988b), all other studies failed to report class-specific distribution of the sample (Class I, Class II and Class III). Four studies (Buschang et al., 2013; Buschang et al., 1999; Chvatal et al., 2005; Nahhas et al., 2014) reported that individuals with normal skeletal jaw relationship and skeletal malocclusions (but did not specify Class II and/or Class III) were included to model normal variation. However, no information was reported regarding the class-specific distribution of the sample or whether and how class was included into model fitting. The remaining three studies (Buschang et al., 1988a, 1989; van der Beek et al., 1996) did not provide any information on inclusion of normal skeletal jaw relationship or skeletal malocclusions.

Below I summarise the findings from this review. I focus on describing the timing and the intensity of the adolescent growth spurt. A few studies did not report numeric data for the timing (APGV) and the intensity (PGV) of the adolescent growth spurt but only displayed growth velocity

plots. When data were reported only graphically, I extracted the numeric data using the Windows-based digitizing computer program UnGraph (version 5.0; Biosoft, Cambridge, United Kingdom). Data extraction with UnGraph has good reliability and validity (Shadish et al., 2009) and has been used in dental research (Sandhu et al., 2016).

Review findings

This section first reports the findings of eight studies which used CP GCMs for data analysis. Results of the other 29 studies, which used statistical methods other than GCMs for data analysis, are summarised towards the end of this section.

Studies using growth curve models

Buschang et al. (1988a) were the first to apply GCMs to jaw growth data. The authors used conventional polynomial function for modelling growth trajectories of the lower jaw length between 10 and 15 years of age. Data were analysed separately for 104 males and 105 females. While a cubic polynomial best described the growth for females, the optimal growth curve for males followed a quartic polynomial. Between-individual variability was higher for males than females. The mean APGV was 14.10 years for males, and 12.10 years for females. The PGV was approximately 1.6 times greater for males (3.80 mm/year) than females (2.35 mm/year). The lower jaw length was larger for males than females across the entire age range of 10 to 15 years.

Buschang et al. (1988b) evaluated and compared growth of lower jaw length in individuals with a normal skeletal jaw relationship i.e., Class I (48 males and 23 females) with those having Class II skeletal malocclusion (23 males and 19 females). Data were analysed separately for males and females with an age range of six to 15 years. For both Class I and Class II, sixth- degree and fifth-degree age polynomials best described the growth for males and females, respectively. The APGV for Class I and Class II was 14.10 years for males and 12.90 years for females. The PGV estimated showed that females (Class I 2.51 mm/year; Class II 2.42 mm/year) experience a less

intense (lower PGV) adolescent growth spurt than males (Class I 4.08 mm/year; Class II 3.62 mm/year). The difference in PGV between Class I and Class II was greater for males than females. The cumulative differences in growth resulted in significantly shorter lower jaw length in Class II skeletal malocclusion than Class I, with difference in lower jaw length greater for males than females. In addition to the adolescent growth spurt, the authors also identified a pre-adolescent spurt for both Class I and Class II, with a common peak growth velocity at 8.7 years for males and 7.7 years for females.

Buschang et al. (1989) analysed the lower jaw length data for 105 females. The best-fitting model involved a fifth-degree polynomial. The PGV for the adolescent growth spurt was 2.25 mm/year. The age corresponding to the peak growth velocity (APGV) was 12.70 years. Between-individual variability was higher for males than females.

van der Beek et al. (1996) studied growth of the total face height for 134 females between seven and 14 years of age and explored its relationship with their height. The authors fitted a multivariate polynomial model to simultaneously describe the total face height growth trajectories and their relationship with stature height. For both the growth measures (total face height and stature height), fourth-degree polynomial provided the best-fitting model. Results showed that total face height growth curve and stature height growth curve were parallel, suggesting a strong correlation between stature height and total face height. The average APGV was 12.30 years. The authors did not report intensity of the adolescent growth spurt.

Buschang et al. (1999) evaluated growth in the lower jaw length for 221 individuals (113 males and 108 females), following them from six to 16 years of age. Data were analysed separately for males and females. The CP GCM using fourth-degree polynomial for males and fifth-degree polynomial for females fitted best to the data. For males, the APGV was 14.30 years. Females experienced the adolescent growth spurt earlier than males, at 12.20 years. The PGV for males and

females were 3.10 mm/year and 2.30 mm/year, respectively. The between-individual variability was higher for males than females.

Chvatal et al. (2005) applied GCM to study growth of the lower jaw length for 159 females and 128 males between six and 15 years of age. Data were analysed separately for males and females. Polynomials ranging from second-degree to fifth degree were used for modelling growth trajectories. The best fitted model involved a fifth-degree polynomial. The lower jaw length showed the characteristic features of an adolescent growth spurt. However, the authors did not report estimates for APGV or PGV or presented growth velocity plots.

Buschang et al. (2013) analysed the upper jaw length, lower jaw length and total face height data for 111 females aged 10 to 15 years. Although length of the upper jaw and the lower jaw increased with age, they did not show the change in growth velocity that characterises the adolescent growth spurt. The total face height showed an adolescent growth spurt with PGV (2.50 mm/year) occurring at 12.30 years. Thus, while females experienced the adolescent growth spurt for the total face height, no evidence was found of the adolescent growth for the upper jaw length or the lower jaw length.

Nahhas et al. (2014) studied timing and intensity of the adolescent growth spurt for the upper jaw length and the lower jaw length for 293 individuals (148 males and 145 females). Data were analysed separately for males and females. The authors included data with an age range of four to 24 years. The best fitted CP GCM for upper and lower jaw length data involved fifth-degree polynomial for both sexes. Both upper and lower jaw length showed adolescent growth spurts. The intensity of the adolescent growth spurt i.e., the PGV was greater for males (upper jaw length 0.96 mm/year; lower jaw length 2.71 mm/year) than females (upper jaw length 0.85 years; lower jaw length 2.29 mm/year). The APGV for the upper jaw length (13.29 years) and the lower jaw length

(13.41 years) was later for males when compared with the timing of the adolescent growth spurt for females (upper jaw length 10.98 years; lower jaw length 10.84 years).

In summary, studies using CP GCM for data analysis report a wide range of the APGV for lower jaw length which varied between 13.41 to 14.30 years for males, and 10.84 to 12.70 years for females. For the upper jaw length, one study analysing the female data did not find an evidence for the adolescent growth spurt (Buschang et al., 2013) whereas another study reported peak growth velocity at 13.29 years for males, and 10.98 years for females (Nahhas et al., 2014). Both studies analysing the total face height data for females reported peak growth velocity at 12.30 years of age (Buschang et al., 2013; van der Beek et al., 1996). No previous study analysed the total face height data for males.

The range of PGV for the lower jaw length reported in previous studies using CP GCM for data analysis vary from 2.71 to 4.08 mm/year for males, and 2.20 to 2.51 mm/year for females. For the upper jaw length, a study analysing the female data (Buschang et al., 2013) did not find any evidence for the adolescent growth spurt. Another study, which evaluated growth in the upper growth length for both sexes, (Nahhas et al., 2014) reported adolescent growth spurt with PGV as 0.96 mm/year for males, and 0.85 mm/year for females. Out of two studies which analysed the total face height data for females (Buschang et al., 2013; van der Beek et al., 1996), one study reported PGV as 2.50 mm/ year (Buschang et al., 2013) whereas another study did not report estimates for the PGV (van der Beek et al., 1996). As stated above, no previous study applied GCM to the total face height data for males.

Out of a total eight studies which applied GCMs to the jaw growth data, only one study investigated class difference in adolescent growth spurt parameters (Buschang et al., 1988b). However, even this study (see below) focused only on comparing adolescent growth spurt parameters between normal skeletal jaw relationship (Class I) and Class II skeletal malocclusion for

a single outcome (the lower jaw length). Thus, no previous study investigated class differences (i.e., Class I vs Class II, Class I vs Class III, or Class II vs Class III) in adolescent growth spurt parameters for the upper jaw length and total face height. Also, no evidence is available on the differences in adolescent growth spurt parameters for the lower jaw length for Class I vs Class III, or Class II vs Class III.

Buschang et al. (1988b) compared timing and the intensity of adolescent growth spurt for the lower jaw length between normal skeletal jaw relationship (Class I) and Class II skeletal malocclusion. The authors report no difference in the timing (APGV) of the adolescent growth spurt between Class I and Class II for both males (Class I 14.10 years, Class II 14.10 years) and females (Class I 12.90 years, Class II 12.90 years). However, the intensity of the adolescent growth spurt (PGV) was weaker for Class II than Class I for both males (Class I 4.08 mm/year; Class II 3.62 mm/year) and females (Class I 2.51 mm/year; Class II 2.42 mm/year).

Studies using simple statistical methods

Besides the above eight studies that used CP GCM for data analysis, many other studies evaluated jaw growth during adolescence by using traditional statistical methods such as simple linear regression.

Evidence suggests that for both sexes, the adolescent growth spurt for the upper jaw length occurs at an earlier age than for the lower jaw length and the total face height (Baccetti et al., 2011; Ball et al., 2011; Bambha & Van Natta, 1963; Baughan et al., 1979b; Baume et al., 1983; Buschang et al., 1999; Buschang et al., 1986; Lewis et al., 1982; Lewis et al., 1985; Mellion et al., 2013; Moore et al., 1990; Nanda, 1955, 1971; Nanda, 1988; Nanda, 1992; Nanda & Rowe, 1989; Ochoa & Nanda, 2004; Thilander et al., 2005; Thompson et al., 1976). The upper jaw completes its growth earlier than the lower jaw due to its proximity to the head (Baughan et al., 1979a; Buschang et al., 1983).

Evidence supports sexual dimorphism in the adolescent jaw growth spurt. The intensity of the adolescent growth spurt for the upper jaw length, lower jaw length and total face height is weaker for females than males, and the timing of the spurt for all three growth measures is approximately 1.5 to 2 years earlier for females than males (Bambha, 1961; Mitani, 1977; Nanda, 1955; Savara & Tracy, 1967; Singh & Savara, 1966; Tracy & Savara, 1966).

The APGV and the PGV for the upper jaw length, lower jaw length and total face height varied across studies (Baughan et al., 1979b; Bishara, 1981; Bishara et al., 1981; Björk, 1963; Buschang et al., 1999; Buschang et al., 1988b; Jamison et al., 1982; Lewis et al., 1985; Ochoa & Nanda, 2004; Pileski et al., 1973; Savara & Tracy, 1967; Thilander et al., 2005; Tracy et al., 1965).

The range of APGV reported is as follows: (i) upper jaw length: males 11.40 to 14.60 years; females 10.50 to 12.50 years, (ii) lower jaw length: males 13.10 to 14.40 years; females 9.60 to 12.90 years, and (iii) total face height: males 13.20 to 14.50 years; females 9.80 to 12.90 years.

The range of PGV reported is as follows: (i) upper jaw length: males 0.90 to 2.60 mm/year; females 0.80 to 1.98 mm/year, (ii) lower jaw length: males 2.10 to 5.40 mm/year; females 1.42 to 2.96 years, and (iii) total face height: males 2.20 to 5.50 mm/year; females 1.50 to 2.90 years.

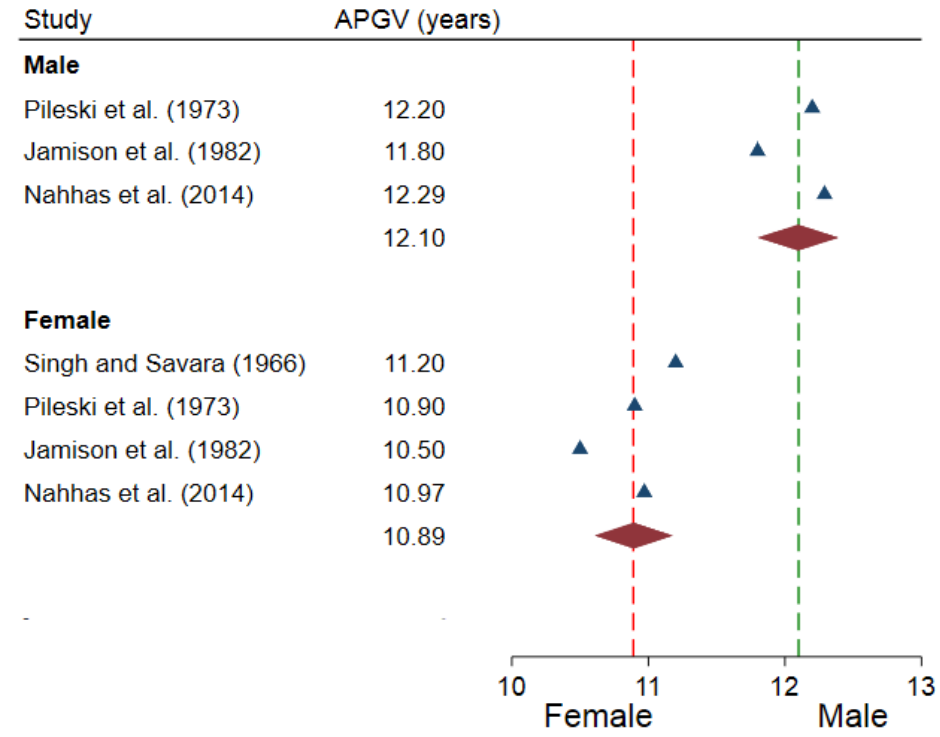
Combining evidence across studies

A logical step to follow a descriptive review of the literature is to conduct a meta-analysis. The meta-analysis estimates an overall summary statistic and its precision (confidence interval), quantifies between-study heterogeneity, and explores potential publication bias.

While all studies included in the systematic review reported the APGV, less than one-third of the studies reported the PGV. That is, most studies focused only on the timing of the adolescent growth spurt, not the strength of this spurt. Therefore, I decided to consider evidence synthesis only for the timing of the adolescent growth spurt.

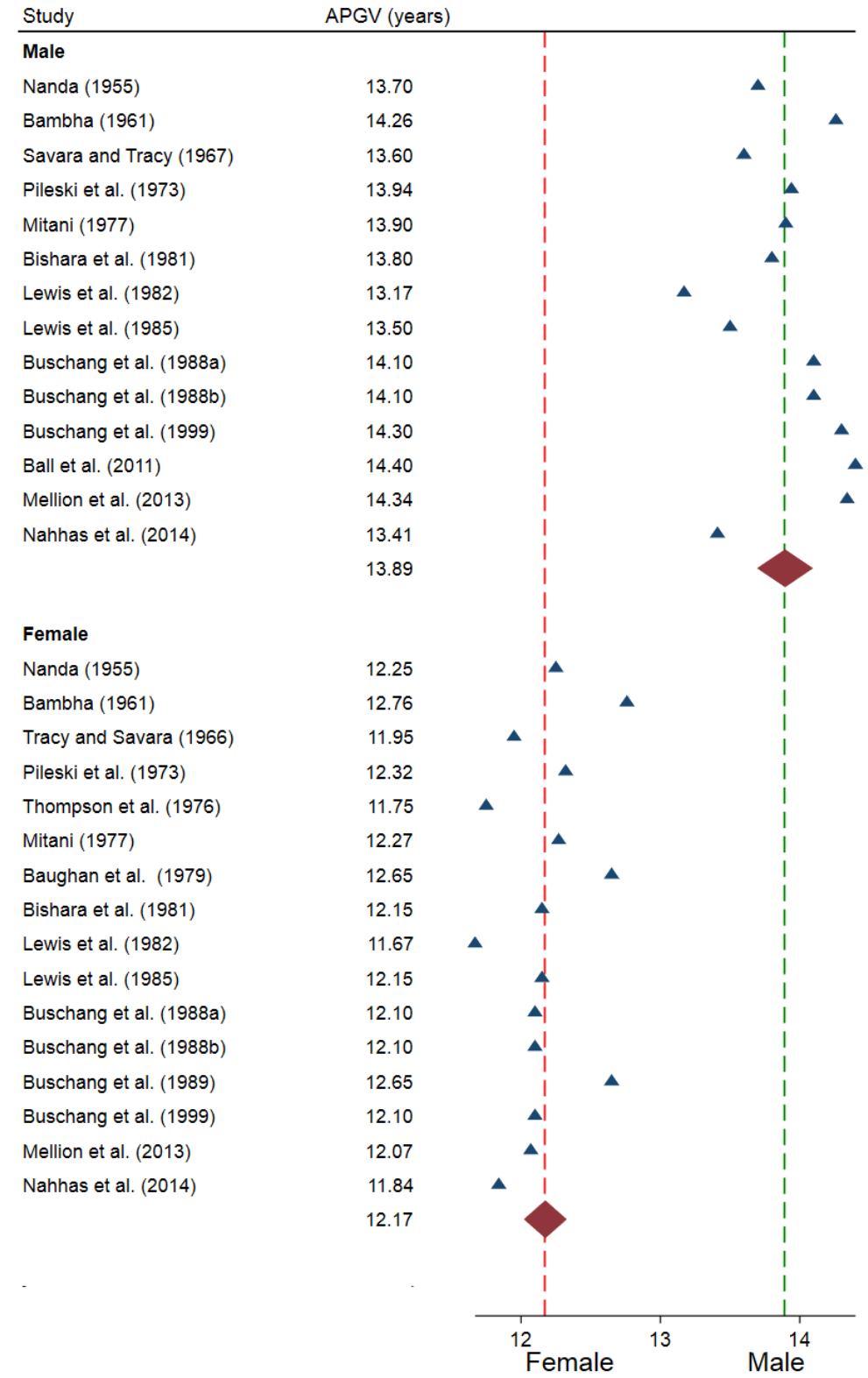
A pre-requisite to conduct a meta-analysis is that studies report effect size as well as the standard deviation. Almost all studies (approximately 94%) failed to report the standard deviation of the APGV. While methods have been developed to compute missing standard deviations by using alternative summary statistics such as confidence intervals and range (Weir et al., 2018), such alternative summary statistics were also missing from the majority of studies (approximately 80%). As a result, meta-analysis could not be conducted.

To provide visualization of study specific estimates of APGV and an average estimate across studies, a customized Stata command was written to draw forest plots without confidence intervals. Results for upper jaw length, lower jaw length and total face height are shown as Figure 2-6, Figure 2-7 and Figure 2-8, respectively. The average difference in the mean average APGV between males and females is 1.21 years for the upper jaw length, 1.72 years for the lower jaw length and 1.69 years for the total face height.



Note:
The green vertical dash line shows mean average APGV (12.10 years) for males
The red vertical dash line shows mean average APGV (10.89 years) for females

Figure 2-6. Age at peak growth velocity (APGV) for upper jaw length (COPAD) reported in previous studies.



Note:
The green vertical dash line shows mean average APGV (13.89 years) for males
The red vertical dash line shows mean average APGV (12.17 years) for females

Figure 2-7. Age at peak growth velocity (APGV) for lower jaw length (COPOD) reported in previous studies.

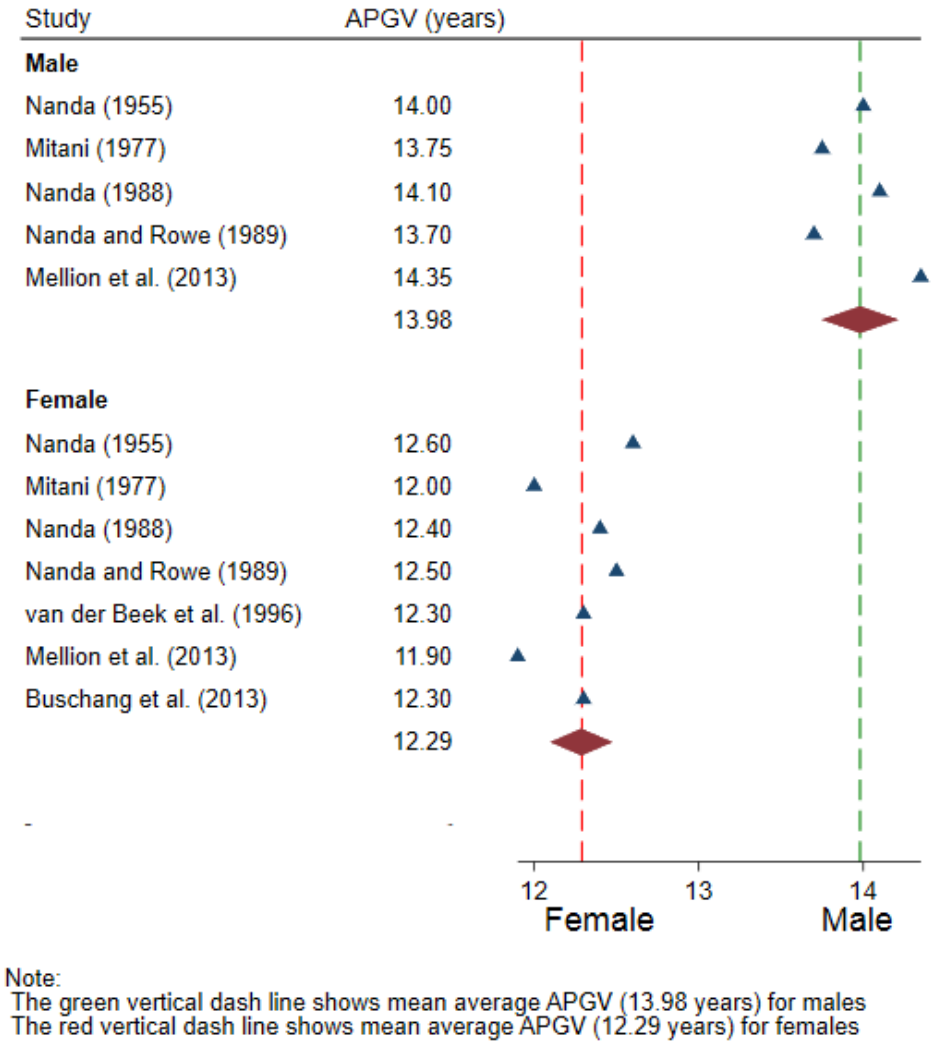


Figure 2-8. Age at peak growth velocity (APGV) for total face height (TFHNP) reported in previous studies.

2.3. Conclusion and research gaps

A substantial amount of work has already been done towards better understanding jaw growth during adolescence. However, the systematic review of the literature shows that most previous studies (78.6%) used simple analytical tools for modelling longitudinal growth data. Such methods are limited as they ignore between individual variation in growth and the serial dependencies that arises in longitudinal data. A few studies (21.6%) did incorporate and study these features of longitudinal jaw growth data by using GCMs. However, here researchers have only explored modelling growth using conventional polynomials.

The systematic review of the literature shows that no previous study compared adolescent growth spurt parameters for the upper jaw length, lower jaw length and total face height between Class I, Class II and Class III skeletal jaw relationships for either males or females. Since an understanding of class differences in adolescent growth spurt parameters is essential for the successful correction of anteroposterior (upper and lower jaw length) and vertical (total face height) skeletal discrepancies (see Chapter 1, Section 1.1 for details), it is vitally important to study class differences in adolescent growth spurt parameters. Considering the sexual dimorphism in jaw growth, it is clinically important to know sex differences in adolescent growth spurt parameters for upper jaw length, lower jaw length and total face height between normal skeletal growth (Class I) and skeletal malocclusions (Class II and Class III).

Therefore, a need exists to compare different GCM approaches for modelling longitudinal growth trajectories and estimating sex differences in adolescent growth parameters for upper jaw length, lower jaw length and total face height between normal skeletal growth and skeletal malocclusions To achieve this research goal, the first objective was to first select a GCM which fits best to the jaw growth data and then apply this GCM of choice for answering clinical research

questions. This thesis addresses these issues by answering the specific research questions outlined in Chapter 1 (Section 1.3). Here I briefly elaborate the research questions posed in this thesis.

2.3.1. Research Question 1 (Chapter 5)

In applications of statistical models, including GCMs, the key aspect of analysis is model selection. My literature review findings show that studies use varying degrees of conventional polynomials for modelling jaw growth but did not focus on model selection when selecting the best fitting GCM. The starting point for the present research work was therefore the question of how to select the best-fitting model from competing GCMs. A simulation study was conducted to answer this question.

2.3.2. Research Question 2 (Chapter 6)

Statistical methods for modelling longitudinal growth data have become increasingly sophisticated over time. In addition to conventional polynomials, fractional polynomials and splines are often used for modelling growth trajectories (Tilling et al., 2014). In addition to linear GCMs, nonlinear GCMs have been applied to study human physical growth such as the stature height. Nonlinear GCMs that follow a specified nonlinear function in age enable modelling complex growth trajectories with parameters that are easily interpretable. My literature review shows that such alternative approaches have not been used for modelling jaw growth. This study attempted to bridge the gap by fitting different linear and nonlinear GCMs to the jaw growth data for males and females. It aimed to assess and compare the performance of different linear and nonlinear GCMs in estimating growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing and intensity).

2.3.3. Research Question 3 (Chapter 7)

Both upper and lower jaws grow in a downward and forward direction during adolescence. However, the upper and the lower jaws do not share a common growth trajectory. The upper jaw

completes its growth earlier than the lower jaw, which continues to grow for a longer duration. This raises the following question: How do the upper and lower jaws grow with age to achieve a normal skeletal jaw relationship (Class I)? An equally important question is: How do skeletal malocclusions (Class II and Class III), which are characterised by skeletal discrepancies in the upper and lower jaw length and total face height, develop with age. My literature review shows that no previous study has compared growth of the upper jaw length, lower jaw length or total face height between Class I, Class II and Class III skeletal jaw relations for either males or females.

A comparison of jaw growth trajectories between normal skeletal jaw relationship (Class I) and skeletal malocclusions (Class II and Class III) was undertaken to answer these questions. A better understanding of class differences in the timing and intensity of the adolescent growth spurt would be helpful in coordinating the timing of treatment to correct skeletal jaw discrepancies in the upper and lower jaw length and total face height seen in Class II and Class III skeletal malocclusions.

CHAPTER 3. DATA

To answer research questions posed in this thesis, longitudinal growth data is essential. Data comprised of repeated growth measurements on individuals allows studying growth changes over time and comparing these changes between individuals (Goldstein, 1968; Hauspie et al., 2004). However, as the collection of longitudinal growth data is resource-intensive and time consuming exercise, the limited availability of such a data is one of the main challenges in studying human physical growth (Hauspie et al., 2004).

For studying skeletal jaw growth which requires radiographic measurements of craniofacial structure such as upper and lower jaw length and total face height, an additional factor to be considered is the radiation exposure. Typically, serial lateral cephalograms (radiographs of skull and face region from sideview) taken repeatedly on young individuals are required to study jaw growth during adolescence. Considering the repeated radiation exposure, it is unethical to conduct a longitudinal jaw growth study routinely.

The primary resource available to the orthodontists and craniofacial researchers for studying craniofacial growth is the collection of longitudinal growth records that were accumulated by dedicated investigators during the early and middle years of the 20th century (1929 to 1982). The collection represents thousands of records gathered from various longitudinal growth studies. The materials available from these historic growth studies is rare and irreplaceable (Boyd et al., 1980; J. M. Tanner, 1981).

Recently in 2014, the American Association of Orthodontists Foundation (AAOF), the charitable arm of the American Association of Orthodontists (AAO) which is responsible for advancement of orthodontic research and education, pooled longitudinal growth data from nine historic growth studies conducted in the United States and Canada between 1929 and 1982 (Baumrind & Curry, 2015). The data available from this project provide longitudinal growth records of craniofacial structures in normally growing and healthy children (Baumrind & Curry, 2015).

The AAOF Craniofacial Growth Legacy Collection project uses the Internet and cloud computing to collect and share data (Baumrind & Curry, 2015). The whole database has been made freely available to orthodontists for research and education purposes. The Craniofacial Research Instrumentation Laboratory (CRIL) at the Department of Orthodontics, Arthur A. Dugoni School of Dentistry, University of the Pacific, USA[2], is responsible for running and maintaining the official website of the AAOF Craniofacial Growth Legacy Collection[3].

In this chapter, I first briefly describe the AAOF Craniofacial Growth Legacy Collection project and then the data available from the official website of the AAOF Craniofacial Growth Legacy Collection project. This is followed by details on the data analysed in this thesis and a summary of data limitations.

[2] See http://www.cril.org/
[3] See http://www.aaoflegacycollection.org/aaof_home.html

3.1. Database

Appendix D provides a detailed review of historical background of the origin of the AAOF Craniofacial Growth Legacy Collection project (Section D.1.1), longitudinal growth studies contributing to the database available from the AAOF Craniofacial Growth Legacy Collection project (Section D.1.2), and longitudinal growth studies which did not participate in the AAOF Craniofacial Growth Legacy Collection project (Section D.1.3). Below I briefly summarise the main points.

3.1.1. Historical background

Out of concern for the preservation of the important collection of longitudinal growth records, the National Institute of Dental Research (NIDR) in 1988 sponsored a survey of existing records available in the United States of America and Canada (Baumrind & Curry, 2015). In his survey, Hunter et al. (1993) identified 12 relevant longitudinal growth studies providing radiographic images (lateral and frontal cephalograms; hand-wrist) of craniofacial structures for children and adolescents with varied growth patterns such as Class I, Class II and Class III skeletal jaw relationships (Baumrind & Curry, 2015; Hunter et al., 1993). Appendix D (Section D.1.1) provides a detailed review of historical background of the AAOF Craniofacial Growth Legacy Collection project.

Each growth study was independent and pursued its own sampling and data collection strategies. Taken together, these different and complementary strategies have produced a rich longitudinal record of growth changes in craniofacial structures such as upper and lower jaw length and total face height (Cavanaugh, 2009; Hunter et al., 1993). The contributing collections, working individually, have produced most of the information that is available in the contemporary literature on the craniofacial growth in untreated children (Baumrind & Curry, 2015).

3.1.2. Longitudinal growth studies contributing to the database

Nine of the 12 longitudinal growth studies identified by Hunter et al. (1993) collaborated and participated in the AAOF Craniofacial Growth Legacy Collection project to share their databases. These studies are (Table 3-1): i) Bolton-Brush growth study, ii) Burlington growth study, iii) Denver growth study, iv) Fels growth study, v) Forsyth growth study, vi) Iowa growth study, vii) Mathews growth study, viii) Michigan growth study, and ix) Oregon growth study. Appendix D (Section D.1.2) provides a detailed description of each individual growth study.

The data collection and pooling from each individual growth study lasted for three years between 2011 and 2014 (Baumrind & Curry, 2015). It was decided to contribute longitudinal growth records of around 100 individuals from each growth studies (except the Mathews growth study which enrolled only 36 individuals).

Although each growth study collected multiple data such as radiographs (lateral cephalograms, frontal cephalograms and hand and wrist images), dental models (replica of dentition) and anthropometric measurements like stature height, the priority was given to include lateral cephalograms in this initial phase of the AAOF Craniofacial Growth Legacy Collection project (2011 to 2014). The growth records included serial lateral cephalograph images for all individuals and the available numeric data from a subset of lateral cephalograms. The numeric data was derived from lateral cephalograms available from individual growth studies for research purposes (before 2011) and later added as part of the project. Table 3-2 summarises the characteristics of participants enrolled in each individual growth study.

Table 3-1. Longitudinal growth studies included in the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection database.

Growth study (enrolment period)	Database source (Department/University/Institute)	Location	Country
Bolton-Brush (1929 to 1959)	Bolton Brush Growth Study Centre, School of Dental Medicine, Case Western Reserve University	Cleveland, Ohio	United States
Burlington (1952 to 1971)	Faculty of Dentistry, Dental School, University of Toronto	Toronto, Ontario	Canada
Denver (1932 to 1967)	Department of Orthodontics, Dental School, University of Oklahoma	Oklahoma City, Oklahoma	United States
Fels (1931 to 1982)	Division of Epidemiology & Biostatistics, Department of Population and Public Health Sciences, Boonshoft School of Medicine, Wright State University	Dayton, Ohio	United States
Forsyth (1959 to 1970)	The Forsyth Institute, Harvard School of Dental Medicine Affiliate	Cambridge, Massachusetts	United States
Iowa (1946 to 1960)	University of Iowa	Iowa City, Iowa	United States
Mathews (1967 to 1979)	Craniofacial Research Instrumentation Laboratory, Department of Orthodontics, Arthur A. Dugoni School of Dentistry, University of the Pacific	San Francisco, California	United States
Michigan (1953 to 1968)	Department of Orthodontics and Paediatric Dentistry, School of Dentistry, University of Michigan	Ann Arbor, Michigan	United States
Oregon (1950 to mid-1970s)	Department of Orthodontics, School of Dentistry, Oregon Health and Science University	Portland, Oregon	United States

Table 3-2. Characteristics of longitudinal growth studies included in the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection database.

Growth study	Characteristics of growth studies				AAOF Craniofacial Growth Legacy Collection		
	Number of participants	Sex	Ethnicity	Enrolment period	Number of participants	Males (%)	Age in years; standard deviation (range)
Bolton-Brush	4,309	Nearly equal number of males and females	American-born Anglo-Saxon (few black children)	1929 to 1959	102	58 (56.86)	12.1 SD 5.6 (0.3–60)
Burlington	1,258	Nearly equal number of males and females	Predominantly European Caucasians	1952 to 1971	100	51 (51)	10.8 SD 4.8 (2–21)
Denver	313	155 males, 158 females	All European Caucasians	1932 to 1967	94	50 (53.19)	12.3 SD 4.43 (4–26)
Fels	1,200	Unclear	Predominantly European Caucasians	1931 to 1982	102	54 (52.94)	11.7 SD 4.5 (0.1–26)
Forsyth	414	226 males, 178 females	All European Caucasians	1959 to 1970	10	6 (60)	11.4 SD 3.9 (4–18)
Iowa	183	92 males, 91 females	All European Caucasians	1946 to 1960	100	50 (50)	8.9 SD 5.1 (0–49)
Mathews	36	13 males, 23 females	Predominantly European Caucasians	1967 to 1979	35	13 (37.14)	11.5 SD 3.38 (4–20)
Michigan	721	Nearly equal number of males and females	Predominantly European Caucasians	1953 to 1968	102	58 (56.86)	9.0 SD 3.5 (3–20)
Oregon	357	153 males, 204 females	All European Caucasians	1954 to 1975	107	49 (45.79)	10.9 SD 4.6 (2–31)

3.1.3. Longitudinal growth studies not part of the database

Out of 12 longitudinal growth studies identified by Hunter et al. (1993), three studies did not participate in the AAOF Craniofacial Growth Legacy Collection project. These studies are (see Appendix D, Section D.1.3 for details): i) Krogman Philadelphia Growth Study (University of Pennsylvania, Pennsylvania USA), ii) Meharry Growth Study (Meharry University, Nashville USA), and iii) Montreal Growth Study (University of Montreal, Montreal Canada). There were no specific reasons why these three growth studies did not participate in the project[4].

In addition to the above mentioned three longitudinal growth studies, two other longitudinal growth studies are not part of the AAOF Craniofacial Growth Legacy Collection project. These studies are (see Appendix D, Section D.1.3 for details): i) the Belfast growth study (Queen's University Belfast, Belfast Northern Ireland), and ii) the Nijmegen growth study (Radboud University Nijmegen, Nijmegen Netherland). Both these studies were not included in the survey report of Hunter et al. (1993) as they were conducted outside the USA and Canada. Therefore, these two growth studies were not invited to participate in the AAOF Craniofacial Growth Legacy Collection project.

[4] Personal communication, Dr Sheldon Baumrind (25/10/2016), Director of the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection project.

3.2. Data availability

Data available from the AAOF Craniofacial Growth Legacy Collection website[5] includes demographics, lateral cephalogram images and the numeric data derived from lateral cephalogram images (see Appendix D, Section D.2 for details). Data were downloaded from the website. As reported earlier (Section 3.1), the AAOF Craniofacial Growth Legacy Collection website allows orthodontists to freely access, download, and use all the available data for research and education purposes.

To answer the research questions posed in this thesis, the data accessed included linear measurements for the upper and lower length and total face height. The upper jaw length (see Appendix D, Figure D-3) is denoted by the Condyle–Point-A Distance (COPAD) which is measured as a linear distance (in millimetres) between anatomic landmarks Condyle and Point-A. The lower jaw length (see Appendix D, Figure D-4) is recorded as Condyle–Pogonion Distance (COPOD) measured in millimetres between anatomic landmarks Condyle and Pogonion. The total face height (see Appendix D, Figure D-5) is measured as a distance (in millimetres) between anatomic landmarks Nasion and Menton along a perpendicular dropped on the Frankfurt horizontal plane from the Nasion. The measurement is recorded as the Total Face Height Nasion Perpendicular (TFHNP).

Appendix D (Section D.2) provides a detailed description of steps involved in downloading and scrutinising the data, removing duplicate records and selecting the final numeric data included for analysis. Briefly, the numeric data on the upper jaw length, lower jaw length and total face height were available for 270 individuals (130 males and 140 females). Of these, data for one male and one female were excluded as they did not provide any growth measurement between seven and 18

[5] https://www.aaoflegacycollection.org/aaof_home.html

years of age. Also, as the Fels growth study contributed growth data for only one individual (a male), it was excluded from the data. Thus, the final data comprised measurements of the upper and lower jaw length and total face height from 2,060 lateral cephalograms of 267 individuals (128 males and 139 females; mean age 11.60 years, SD 2.90, range 7–18 years).

Table 3-3 summarises the characteristics of data available from the AAOF Craniofacial Growth Legacy Collection website and numeric data included in this thesis. Due to the large size of Table 3-3, proportions of males only are reported (number of males out of the total sample size for each growth study). The class-specific sample sizes show combined numbers of males and females with Class I, Class II and Class III skeletal jaw relationships.

Table 3-4 summarises the availability of jaw growth data from all eight growth studies. While all eight growth studies provide data for normal skeletal jaw relationship (Class I), data for Class II and Class III jaw relationships is not available from some of the growth studies. Two growth studies, the Iowa growth study and Forsyth growth study, do not provide data for Class II jaw relationship. Data for Class III jaw relationship is not available from five growth studies which include the Denver growth study, Forsyth growth study, Iowa growth study, Mathews growth study and Michigan growth study.

The nonuniform distribution of the sample is because of following two reasons. First, the prevalence of Class I skeletal jaw relationship is greater than the Class II and Class III skeletal jaw relationships. The Class III skeletal jaw relationship is least common in the population (see Chapter 1, Section 1.1 for details). Second, most of the previous research using data from individual growth studies housed at different academic institutes (see Table 3-1) focused on investigating jaw growth for normal skeletal jaw relationship (see Chapter 2, Section 2.2.3 for details). As a result, the pooled numeric data available from the AAOF Craniofacial Growth Legacy Collection project is greater for normal skeletal jaw relationship than skeletal malocclusions.

Table 3-3. Summary of participant characteristics and data available for analysis.

Growth study	Total number of individuals' data available from the AAOF Collection: 750 [a]	Number of individuals with cephalometric measurements calculated using the Cartesian coordinates data available from the AAOF Collection: 236 [b]	Number of individuals with additional cephalometric measurements data directly available from the AAOF Collection: 34 [c]	Total number of individuals with cephalometric measurements data included: 267 [d]
Bolton-Brush	Number of individuals:102	Number of individuals: 48	-	Number of individuals: 47
	Age: 12.1 SD 5.6 (0.3–60)	Age: 11.6 SD 3.4 (4–25)	-	Age: 11.5 SD 2.9 (4–25)
	Male: 58 (56.86%)	Male: 31 (64.58%)	-	Male: 31 (65.96%)
	Class I: 25 (24.51%)	Class I: 9 (18.75%)	-	Class I: 9 (19.15%)
	Class II: 55 (53.92%)	Class II: 27 (56.25%)	-	Class II: 27 (57.45%)
	Class III: 22 (21.57%)	Class III: 12 (25%)	-	Class III: 11 (23.4%)
	Class Unknown: 0 (0%)	Class Unknown: 0 (0%)	-	Class Unknown: 0 (0%)
Burlington	Number of individuals: 100	Number of individuals: 49	Number of individuals: 01	Number of individuals: 49
	Age: 10.8 SD 4.8 (2–21)	Age: 11.6 SD 3.4 (7–20)	Age: 5.4 SD 0.8 (4–6)	Age: 11.5 SD 2.8 (7–18)
	Male: 51 (51%)	Male: 24 (48.98%)	Male: 1 (100%)	Male: 24 (48.98%)
	Class I: 38 (38%)	Class I: 14 (28.57%)	Class I: 1 (100%)	Class I: 14 (28.57%)
	Class II: 56 (56%)	Class II: 31 (63.27%)	Class II: 5 (15.15%)	Class II: 31 (63.27%)
	Class III: 6 (6%)	Class III: 4 (8.16%)	Class III: 0 (0%)	Class III: 4 (8.16%)
	Class Unknown: 0 (0%)	Class Unknown: 0 (0%)	Class Unknown: 0 (0%)	Class Unknown: 0 (0%)
Denver	Number of individuals: 94	Number of individuals: 50	-	Number of individuals: 50
	Age: 12.3 SD 4.43 (4–26)	Age: 12.4 SD 3.6 (7–23)	-	Age: 11.8 SD 2.9 (7–18)
	Male: 50 (53.19%)	Male: 12 (24%)	-	Male: 12 (24%)
	Class I: 58 (61.7%)	Class I: 30 (60%)	-	Class I: 30 (60%)
	Class II: 22 (23.4%)	Class II: 15 (30%)	-	Class II: 15 (30%)
	Class III: 1 (1.06%)	Class III: 5 (10%)	-	Class III: 5 (10%)
	Class Unknown: 13 (13.83%)	Class Unknown: 0 (0%)	-	Class Unknown: 0 (0%)
Fels	Number of individuals:102	Number of individuals: 01	-	-
	Age: 11.7 SD 4.5 (.1 - 26)	Age: 12.2 SD 4.4 (7 - 18)	-	-
	Male: 54 (52.94%)	Male: 1 (100%)	-	-
	Class I: 44 (43.14%)	Class I: 1 (100%)	-	-
	Class II: 3 (2.94%)	Class II: 0 (0%)	-	-
	Class III: 1 (0.98%)	Class III: 0 (0%)	-	-
	Class Unknown: 54 (52.94%)	Class Unknown: 0 (0%)	-	-
Forsyth	Number of individuals: 10	Number of individuals: 09	-	Number of individuals: 09
	Age: 11.4 SD 3.9 (4–18)	Age: 11.3 SD 3.9 (4–18)	-	Age: 12.4 SD 3.3 (7–18)
	Male: 6 (60%)	Male: 5 (55.56%)	-	Male: 5 (55.56%)
	Class I: 6 (60%)	Class I: 6 (66.67%)	-	Class I: 6 (66.67%)
	Class II: 4 (40%)	Class II: 3 (33.33%)	-	Class II: 3 (33.33%)
	Class III: 0 (0%)	Class III: 0 (0%)	-	Class III: 0 (0%)
	Class Unknown: 0 (0%)	Class Unknown: 0 (0%)	-	Class Unknown: 0 (0%)
Iowa	Number of individuals: 100	Number of individuals: 08	-	Number of individuals: 08
	Age: 8.9 SD 5.1 (0–49)	Age: 9.7 SD 3.5 (4–17)	-	Age: 11.1 SD 2.7 (7–17)
	Male: 50 (50%)	Male: 4 (50%)	-	Male: 4 (50%)
	Class I: 80 (80%)	Class I: 8 (100%)	-	Class I: 8 (100%)
	Class II: 16 (16%)	Class II: 0 (0%)	-	Class II: 0 (0%)
	Class III: 2 (2%)	Class III: 0 (0%)	-	Class III: 0 (0%)
	Class Unknown: 02 (2%)	Class Unknown: 0 (0%)	-	Class Unknown: 0 (0%)
Mathews	Number of individuals: 33	Number of individuals: 33	-	Number of individuals: 33
	Age: 11.5 SD 3.3 (4–20)	Age: 11.5 SD 3.3 (4–20)	-	Age: 11.8 SD 2.9 (7–18)
	Male: 12 (36.36%)	Male: 12 (36.36%)	-	Male: 12 (36.36%)
	Class I: 17 (51.51%)	Class I: 17 (51.52%)	-	Class I: 17 (51.52%)
	Class II: 16 (48.48%)	Class II: 16 (48.48%)	-	Class II: 16 (48.48%)
	Class III: 0 (0%)	Class III: 0 (0%)	-	Class III: 0 (0%)
	Class Unknown: 0 (0%)	Class Unknown: 0 (0%)	-	Class Unknown: 0 (0%)

Growth study	Total number of individuals' data available from the AAOF Collection: 750 [a]	Number of individuals with cephalometric measurements calculated using the Cartesian coordinates data available from the AAOF Collection: 236 [b]	Number of individuals with additional cephalometric measurements data directly available from the AAOF Collection: 34 [c]	Total number of individuals with cephalometric measurements data included: 267 [d]
Michigan	Number of individuals: 102	Number of individuals: 09	Number of individuals: 33	Number of individuals: 42
	Age: 9.0 SD 3.5 (3–20)	Age: 11.1 SD 3.6 (4–18)	Age: 10.2 SD 3.6 (3–20)	Age: 11.6 SD 2.9 (7–18)
	Male: 58 (56.86%)	Male: 6 (66.67%)	Male: 19 (57.58%)	Male: 25 (59.52%)
	Class I: 80 (78.43%)	Class I: 3 (33.33%)	Class I: 28 (84.85%)	Class I: 31 (73.81%)
	Class II: 20 (19.61%)	Class II: 6 (66.67%)	Class II: 5 (15.15%)	Class II: 11 (26.19%)
	Class III: 1 (0.98%)	Class III: 0 (0%)	Class III: 0 (0%)	Class III: 0 (0%)
	Class Unknown: 01 (0.98%)	Class Unknown: 0 (0%)	Class Unknown: 0 (0%)	Class Unknown: 0 (0%)
Oregon	Number of individuals: 107	Number of individuals: 29	-	Number of individuals: 29
	Age: 10.9 SD 4.6 (2–31)	Age: 12.4 SD 4.7 (3–31)	-	Age: 11.7 SD 2.9 (7–18)
	Male: 49 (45.79%)	Male: 15 (51.72%)	-	Male: 15 (51.72%)
	Class I: 67 (62.62%)	Class I: 6 (20.69%)	-	Class I: 6 (20.69%)
	Class II: 34 (31.78%)	Class II: 19 (65.52%)	-	Class II: 19 (65.52%)
	Class III: 4 (3.74%)	Class III: 4 (13.79%)	-	Class III: 4 (13.79%)
	Class Unknown: 02 (1.87%)	Class Unknown: 0 (0%)	-	Class Unknown: 0 (0%)

Note: Age in years, standard deviation (range).
[a] Demographic data available from the AAOF Craniofacial Growth Legacy Collection.
[b] Cephalometric measurements calculated using the Cartesian coordinates data available from the AAOF Craniofacial Growth Legacy Collection.
[c] Additional cephalometric measurements directly available from the AAOF Craniofacial Growth Legacy Collection.
[d] Data included (age range 7–18 years). Data excluded for one individual each from the Bolton-Brush and Burlington growth studies as no measurement is available with the age range of 7–18 years. Fels growth study excluded as data available for only one individual.

Table 3-4. Summary of data available from each of the eight growth studies.

	Growth study							
	Bolton-Brush	Burlington	Denver	Forsyth	Iowa	Mathews	Michigan	Oregon
Male								
Class I	✓	✓	✓	✓	✓	✓	✓	✓
Class II	✓	✓	✓		×	✓	✓	✓
Class III	✓	✓	×	×	×	×	×	✓
Female								
Class I	✓	✓	✓	✓	✓	✓	✓	✓
Class II	✓	✓	✓	×	×	✓	✓	✓
Class III	✓	✓	✓	×	×	×	×	✓

✓ Included × Not included

3.3. Data analysed

Data comprise repeated measurements of the upper jaw length, lower jaw length and total face height for 128 males (mean age 11.67 years, standard deviation 2.92) and 139 females (mean age 11.60 years, standard deviation 2.88) who participated in eight growth studies (see Section 3.2 for details). The total number of observations is 2,060 (1,005 for males and 1,055 for females).

All three growth measures (outcomes) i.e., upper jaw length, lower jaw length and total face height are recorded in millimetres (mm) between seven and 18 years of age. Collectively, these three growth measures describe the anteroposterior (horizontal) and vertical growth-related changes in the face (Jacobson & Jacobson, 2006; McNamara, 1984; Proffit et al., 2014).

Table 3-5 summarises the sex- and class-specific distribution of the sample size across all eight growth studies. For both males and females, the mean age is comparable across all three classes (Class I, Class II and Class III). The number of individuals with Class III skeletal malocclusion is lowest for both sexes (11 males and 13 females). For males, the sample size is largest for Class II skeletal malocclusion (68 individuals) followed by Class I skeletal jaw relationship (49 individuals). For females, the sample size is largest for Class I skeletal jaw relationship (72 individuals) followed by Class II skeletal malocclusion (54 individuals). For both sexes, the sample size is smallest for Class III skeletal jaw relationship (see Section 3.2 for differences in the sample size for Class I, Class II and Class III skeletal jaw relationships). For both sexes, the number of repeated measurements per individual is lowest for Class III skeletal malocclusion (males: median 6, IQR 5–9; females: median 7, IQR 6–9).

Table 3-5. Summary of data analysed.

	Number of individuals	Age (years)				Repeated measurements per individual			
		Mean	SD [a]	Min [b]	Max [c]	Median	IQR [d]	Min [b]	Max [c]
Male									
Total	128	11.67	2.92	7.0	18.0	9	7 to 10	2	12
Class I	49	11.76	2.93	7.0	18.0	9	8 to 10	3	12
Class II	68	11.57	2.89	7.0	18.0	8	7 to 9	4	11
Class III	11	11.90	3.22	7.0	18.0	6	5 to 9	2	10
Female									
Total	139	11.60	2.88	7.0	18.0	8	7 to 9	3	12
Class I	72	11.63	2.89	7.0	18.0	8	7 to 9	3	12
Class II	54	11.58	2.86	7.0	18.0	8	7 to 10	4	12
Class III	13	11.52	2.92	7.0	18.0	7	6 to 9	3	10

Note: Total number of individuals is 267. [a] Standard deviation [b] Minimum [c] Maximum [d] Interquartile range

For both sexes, the median age gap between two consecutive measurements is one year (male IQR: 1.00 to 1.10 years; female IQR: 1.00 to 1.15 years). This is because most of the growth studies focused on collecting data around participants' birthdays. Table 3-6 shows the total number of observations and the missing data (number of missing responses out of the total number of observations) for the upper jaw length, lower jaw length and total face height. For both sexes, missing data is less than 4% for any of the three outcomes. No data were missing for covariate (class of skeletal jaw relationship).

Table 3-6. Number of repeated measurements and missing data for each outcome.

	Number of individuals	Observations	Missing data (%)
Male			
Upper jaw length (COPAD)	128	1005	12 (1.19)
Lower jaw length (COPOD)	128	1005	15 (1.49)
Total face height (TFHNP)	128	1005	38 (3.78)
Female			
Upper jaw length (COPAD)	139	1055	11 (1.04)
Lower jaw length (COPOD)	139	1055	12 (1.14)
Total face height (TFHNP)	139	1055	18 (1.71)

Note. Total number of individuals is 267. Total number of repeated measurements (observations) is 2,060.

COPAD: condyle–point A measurement in millimetres; COPOD: condyle–pogonion measurement in millimetres; TFHNP: total face height measurement in millimetres.

Figure 3-1, Figure 3-2 and Figure 3-3 show individual growth trajectories (raw data) plotted between seven and 18 years of age for upper jaw length, lower jaw length and total face height. The plots show individual growth trajectories of upper jaw length, lower jaw length and total face height appear higher for males than females. These gender differences are more pronounced for lower jaw length and total face height than they are for upper jaw length.

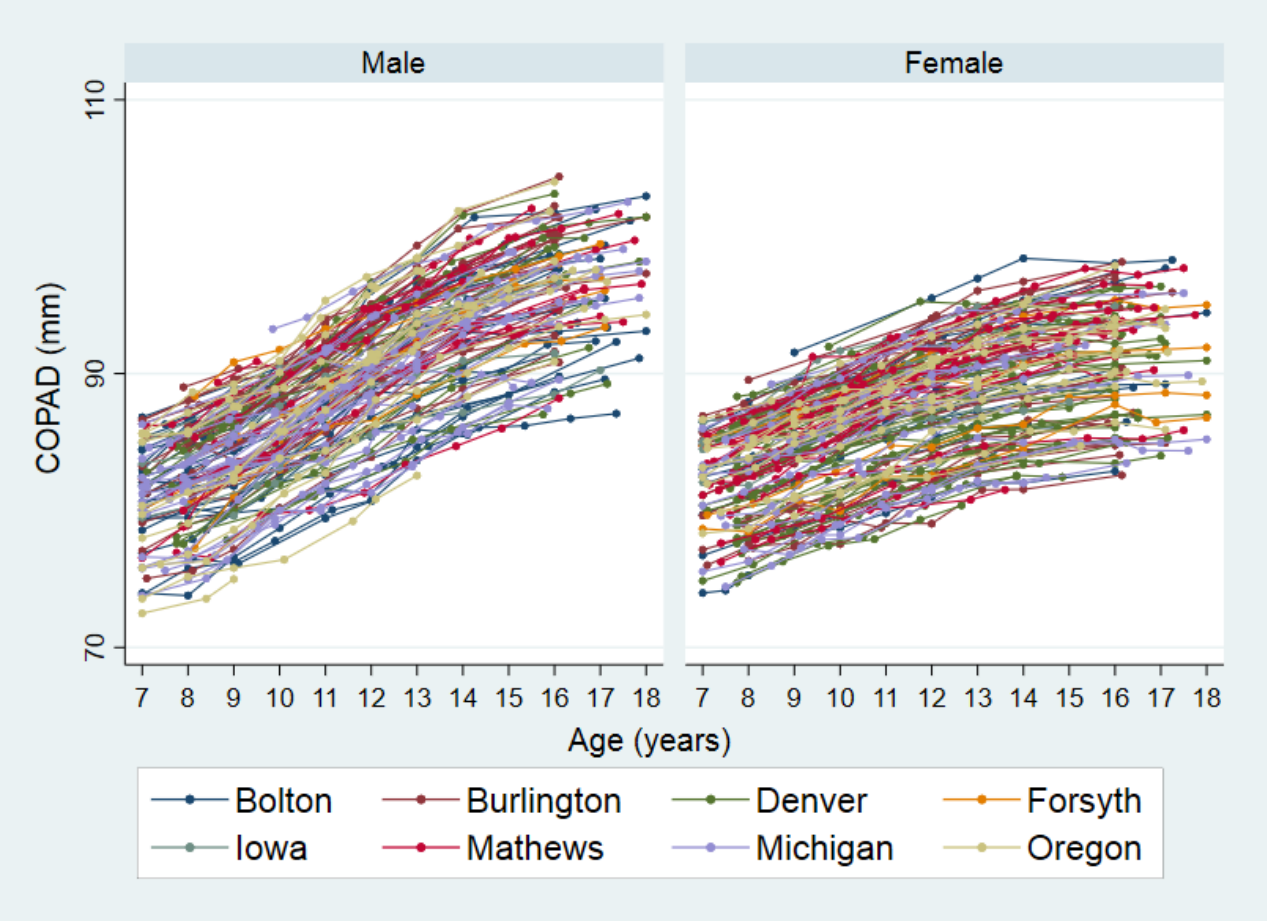


Figure 3-1 Upper jaw length (COPAD) growth trajectories for males and females.

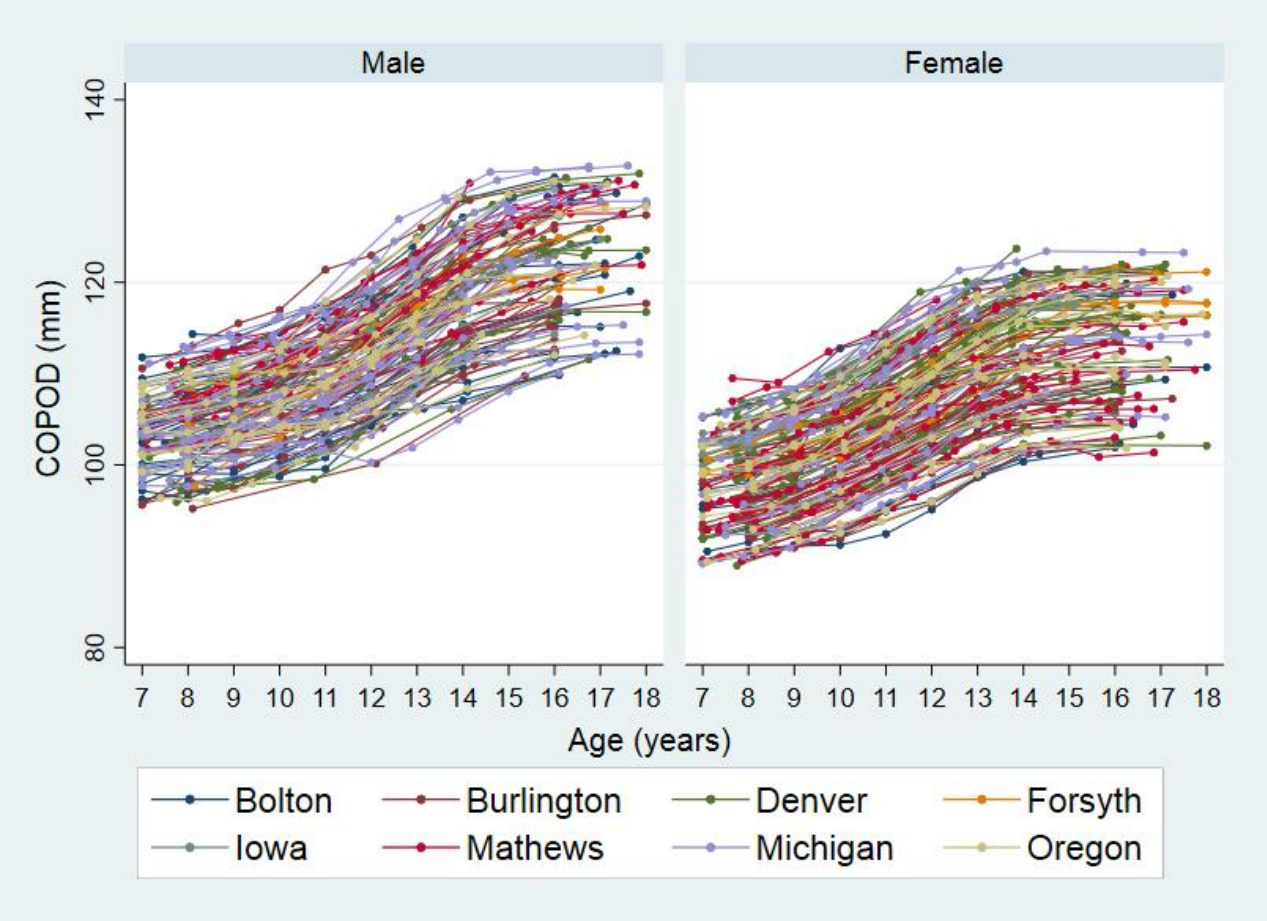


Figure 3-2 Lower jaw length (COPOD) growth trajectories for males and females.

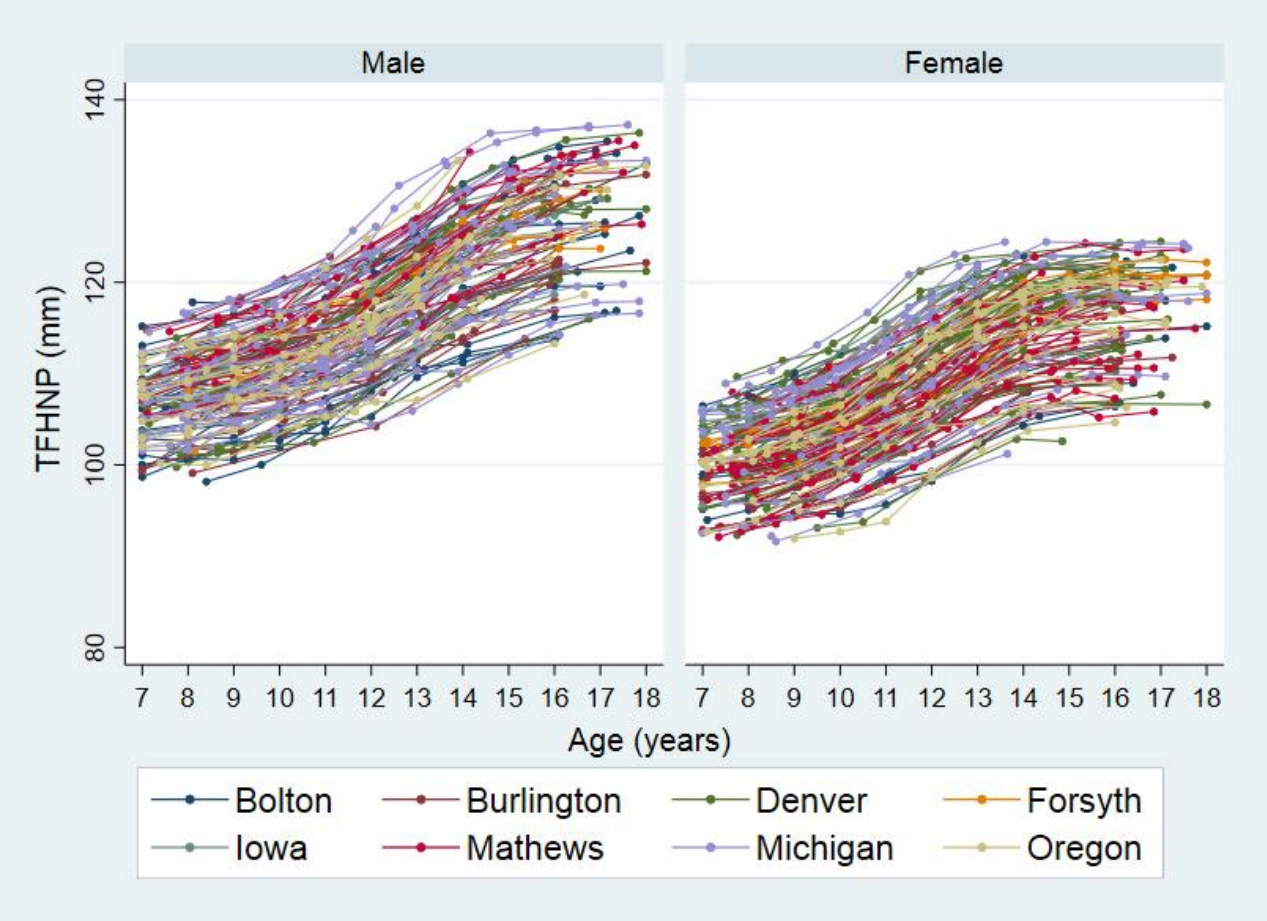


Figure 3-3 Total face height (TFHNP) growth trajectories for males and females.

3.4. Data limitations

Data comprise of repeated growth measurement on individuals who participated in eight different growth studies. Thus, the data follows a three-level structure where repeated measurements (level 1) on individuals (level 2) are further nested in growth studies (level 3). Since the number of growth studies is small, and some provided data for a few individuals (Table 3-3), estimating between-study variance would be difficult. Therefore, instead of fitting three-level growth curve models, an alternative analytical strategy needs to be followed which involves fitting a two-level model and including studies into the fixed effects part of the model to adjust for potential study effects.

Individuals are classified by normal (Class I) or abnormal (Class II and Class III) skeletal jaw relationship. However, the sample size is small and distribution of data is not uniform across Class I, Class II and Class III skeletal jaw relationships (Table 3-3). For example, out of a total 267 individuals, Class III comprised only 24 individuals (11 males and 13 females). This would potentially result in low precision of estimates for class differences in growth trajectories and adolescent growth spurt parameters.

As data were collected in the 20th century over a wide time period (between 1929 and 1982), it would have been interesting to explore secular trends (change in jaw growth parameters over the time period) in jaw growth for Class I, Class II and Class III skeletal jaw relationships. As some growth studies provide no data for Class II and Class III skeletal jaw relationships (Table 3-4), modelling interaction effects between study and class variables is not possible. A statistical model without such interactions assumes that class-specific growth trajectories are same across all growth studies.

Lastly, data are not available for stature height, which otherwise could have been used to establish a relationship between jaw growth and stature height.

3.5. Data summary

Data analysed in this thesis is available from the AAOF Craniofacial Growth Legacy Collection project. Data include repeated measurements of the upper jaw length (COPAD), lower jaw length (COPOD) and the total face height (TFHNP) on 128 males and 139 females who participated in eight growth studies. The age range considered for analysis is seven to 18 years for both sexes. Males and females are classified into Class I, Class II and Class III depending on their skeletal jaw relationship. For both sexes, i) data are unbalanced, ii) missing data (number of missing observations) is less than 4 per cent, iii) the sample size (number of individuals) is smallest for Class III skeletal jaw relationship, iv) the median number of repeated measurements per child is 6 to 8, and v) the median age gap between two consecutive measurements is one year (male IQR: 1.00 to 1.10 years; female IQR: 1.00 to 1.15 years). Data characteristics are further discussed in Chapter 4. (Section 4.1).

CHAPTER 4. GROWTH CURVE MODELS FOR ESTIMATING TIMING AND INTENSITY OF THE ADOLESCENT GROWTH SPURT

4.1. Introduction

As the notion of change is central in studying growth, a correct view of the growth dynamics and the relationship between different growth phases can only be obtained from longitudinal growth data (Cameron & Bogin, 2012; Hauspie et al., 2004). Analysis of longitudinal growth data helps in answering questions related to a change in the outcome(s) measured repeatedly over growth periods, and comparisons of these changes between individuals (Goldstein, 1968). Analysis of longitudinal growth data using an appropriate statistical method allows description of growth trajectories (distance and derivatives) and estimation of the timing and intensity of the adolescent growth spurt (see Chapter 1, Section 1.2).

However, analysing longitudinal growth data, which is essential for understanding dynamics of jaw growth during adolescence, poses analytical challenges. First, growth does not follow a simple linear trajectory, but rather a complex pattern involving periods of acceleration and deceleration in growth velocity (see Chapter 1, Section 1.1). Second, there is a large variability in growth, which results from individual-specific variations in genetic factors (Jelenkovic et al., 2016), overall health status and levels of sex hormones during adolescence (American Psychological Association, 2002; Curtis, 2015; Preedy, 2011), environmental factors (Berman et al., 2016; Cameron & Bogin, 2012; Duren et al., 2013; Preedy, 2011; Ulijaszek, 2010; Wells, 2017), and the way an individual responds to environmental stimuli (Berman et al., 2016; D'Aloisio et al., 2013; Wells, 2017).

GCMs can handle the complexity of longitudinal growth, and allow for modelling and estimation of between-individual variability in the within-individual patterns of change over time (see Chapter 1, Section 1.2). The GCMs can be broadly grouped into latent growth curve models and mixed effect models (McNeish & Matta, 2018). Latent growth curve models treat repeated

measures as multivariate and generally fit within the structural equation modelling (SEM) framework (Bollen & Curran, 2006; Meredith & Tisak, 1990). Mixed effect models are applied to repeated measures set as univariate and fit within the multilevel modelling (MLM) regression framework (Goldstein, 1986a; Laird & Ware, 1982). Although both approaches are analytically *almost* identical (Curran, 2003), Ghisletta and Lindenberger (2004) and Ghisletta et al. (2015) have suggested general practical guidelines (notably not definite rules) described below to aid researchers in selecting one approach over another, depending on the data at hand and the research questions of interest.

Mixed effect models are preferred over latent growth curve models when data are highly unbalanced and the measurement error covariance matrix structure can be represented by a common matrix (Ghisletta & Lindenberger, 2004; Ghisletta et al., 2015). Unbalanced longitudinal data was analysed for the present study (see Chapter 3, Section 3.3), with empirical evidence suggesting that jaw growth trajectories can be represented by a common measurement error matrix i.e., a constant variance across occasions (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996). Therefore, mixed effect models were chosen over latent growth curve models for this thesis.

Mixed effect models can efficiently handle balanced and unbalanced data (Allison, 2012; Goldstein, 2011b; Hox et al., 2018; Singer & Willett, 2003; Snijders & Bosker, 2012). Therefore, no special considerations are required when fitting mixed effects models to the longitudinal data that is time unstructured and unbalanced (i.e., individuals measured at different time points).

For incomplete longitudinal data such as missing jaw measurements, mixed effect models use all available measurements (no listwise deletion) for maximum likelihood (ML) estimation, assuming that data are missing at random (Allison, 2012; Black et al., 2011; Coertjens et al., 2017; Liu, 2016). In other words, for missing at random (MAR) data, any systematic differences between

the observed and missing data can be explained by associations in the observed data (see Chapter 6, Section 6.2.2 for further details).

Unlike in early years when linear (linear mixed effect) GCMs based on conventional polynomials were extensively used for studying human growth (McArdle, 2015), linear GCMs based on alternative transformation of age (such as splines) and nonlinear (nonlinear mixed effect) GCMs particularly developed for modelling growth data are now gaining popularity in modelling longitudinal skeletal growth data such as height (Hauspie et al., 2004; McArdle, 2015). However, findings of the systematic review (Chapter 2, Section 2.2.3) show that studies analysing longitudinal jaw growth data have exclusively focused on conventional polynomial models (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996).

In this thesis, three linear and two nonlinear GCMs were studied. The three linear GCMs included in the thesis were the conventional polynomial (CP), fractional polynomial (FP), and restricted cubic spline (RCS). The two nonlinear GCMs studied are the superimposition by translation and rotation (SITAR) and the Preece–Baines (PB).

The objective of this chapter is to introduce linear and nonlinear GCMs and review the concepts of derivatives estimation and their use in estimating timing (age at peak growth velocity, APGV) and intensity (peak growth velocity, PGV) of the adolescent growth spurt. Towards the end of this chapter, I outlined the methodological and the clinical work done in this thesis while addressing the three specific research questions.

4.2. Growth curve models

4.2.1. Notation

Data comprising repeated measurements on a group of individuals represents a two-level hierarchical structure where measurements (level 1) are nested within individuals (level 2). As repeated measurements obtained from all individuals are analysed simultaneously, GCMs include both level 1 and level 2 information. This thesis follows a commonly used system in multilevel modelling, which denotes level 1 predictors by using subscript *i* and level 2 units by using subscript *j* (Goldstein, 2011a; Kreft & Leeuw, 1998; Rasbash et al., 2016; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). The term measurement occasion (or simply occasion) is used to denote the actual age at which repeated measurements were obtained from an individual (Goldstein, 2011a).

The outcome (also referred to as the dependent variable or response), such as jaw length, measured at the *i*-th occasion for the *j*-th individual is denoted by y_{ij}, and age_{ij} denotes the actual age of individual j at measurement occasion i. The total number of individuals is denoted by n (where n is 128 for males and 139 for females) and the total number of repeated jaw growth measurements made on an individual j are indicated by m_j. Design matrices for fixed effects and random effects are denoted by X_{ij} and Z_{ij}. Fixed effects (population-average regression coefficients) and random effects (individual-specific estimates) are denoted by β and u_j, respectively. Residuals are represented by e_{ij}. Level 2 and level 1 variance-covariances matrices are respectively shown as $\mathbf{\Omega_2}$ and $\mathbf{\Omega_1}$.

4.2.2. Linear growth curve models

The term linear denotes that the function linking the outcome to the predictors is linear in its parameters (Fitzmaurice et al., 2008; Ghisletta et al., 2015; Goldstein, 2011b; Laird & Ware, 1982). In other words, the function is a linear combination of the parameters even when the predictor (such as age) has been transformed (e.g., polynomials and splines) for modelling nonlinear growth trajectories (Fitzmaurice et al., 2008; Ghisletta et al., 2015; Goldstein, 2011b; Laird & Ware, 1982).

Conventional polynomial (CP)

The conventional polynomial functions are popular in modelling growth trajectories (Curran et al., 2010; Goldstein, 1979; Goldstein, 1986a; Hox, 2010; Wishart, 1938) and have been extensively used for analysing jaw growth data (See Chapter 2, Section 2.3). A CP GCM can be written as (Goldstein, 2011b):

$$y_{ij} = \beta_0 + \sum_{r=1}^{p} \beta_r age_{ij}^{r} + u_{0j} + \sum_{s=1}^{q} u_{sj} age_{ij}^{s} + e_{ij} \quad (4\text{-}1)$$

where $\beta_0 + \sum_{r=1}^{p} \beta_r age_{ij}^{r}$ specifies a fixed intercept (β_0) and a p-th order polynomial (positive integer) trend for the population-averaged relationship between y_{ij} and age_{ij}, which is estimated by the fixed effects regression coefficients $\beta_0, \dots, \beta_p$. The $u_{0j} + \sum_{s=1}^{q} u_{sj} age_{ij}^{s}$ defines the random part of the model at the individual level and includes a random intercept (u_{0j}) and random polynomial terms (positive integer) up to q-th order (where $q \leq p$). This allows the population-average trajectory to vary between individuals by entering individual-specific random effects $u_{0j}, \dots, u_{qj}$. The residual e_{ij} captures the variation in the observed measurements around the individual-specific trajectories.

The random effects ($u_{0j}, \dots, u_{qj}$.) and residuals (e_{ij}) are assumed to be mutually independent and to follow normal distributions.

$$\mathrm{cov}\binom{\boldsymbol{u}}{e} = \begin{bmatrix} \boldsymbol{\Omega_2} & 0 \\ 0 & \boldsymbol{\Omega_1} \end{bmatrix}. \tag{4-2}$$

Here $\mathrm{cov}(.)$ is the covariance operator and $\boldsymbol{\Omega_2}$ and $\boldsymbol{\Omega_1}$ are positive-definite, symmetric between-individual and within-individual covariance matrices, respectively. The random effects follow a multivariate normal distribution with mean 0 and a $(q+1) \times (q+1)$ dimensional covariance matrix $\boldsymbol{\Omega_2}$ comprising $[(q+1)(q+2)/2\,]$ unique variance-covariance parameters:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ \vdots \\ u_{qj} \end{pmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{\Omega_2} = \begin{bmatrix} \sigma^2_{u_0} & & & \\ \sigma_{u_{0j}u_{1j}} & \sigma^2_{u_{1j}} & & \\ \vdots & \vdots & \ddots & \\ \sigma_{u_{0j}u_{qj}} & \sigma_{u_{1j}u_{qj}} & \cdots & \sigma^2_{u_{qj}} \end{bmatrix} \right) \tag{4-3}$$

The residuals are assumed to be independent and normally distributed with 0 mean and a $n_j \times n_j$ dimensional identity covariance matrix $\boldsymbol{\Omega_1}$ with a diagonal constant variance parameter, σ^2_e. However, the assumption of homoscedasticity of residuals (constant level 1 variance) can be relaxed. Therefore, the covariance matrix $\boldsymbol{\Omega_1}$ does not necessarily need to be a matrix with zero off-diagonal elements.

$$e_{ij} \sim N(0, \quad \boldsymbol{\Omega_1} = \boldsymbol{I}\sigma^2_e) \tag{4-4}$$

Here $\boldsymbol{I}$ is the identity matrix (diagonal matrix of 1s), and σ^2_e is the residual variance.

Depending on the polynomial power, which is always a positive integer ($1,2,3 \ldots\ldots, p;\ 1,2,3 \ldots\ldots, q$), different CP GCMs can be fitted to the data. The degree of the model refers to the power(s) used to transform the predictor (age). For example, the first degree (degree 1) model involves linear transformation of age (age^1) and the second degree (degree 2) model uses linear (age^1) and quadratic (age^2) transformations of age. Similarly, the third degree (degree 3) model is based on the linear (age^1), quadratic (age^2) and cubic (age^3) transformations of age, and so on. Figure 4-1 shows population-average growth trajectories estimated by different degrees of CP GCM applied to a hypothetical data.

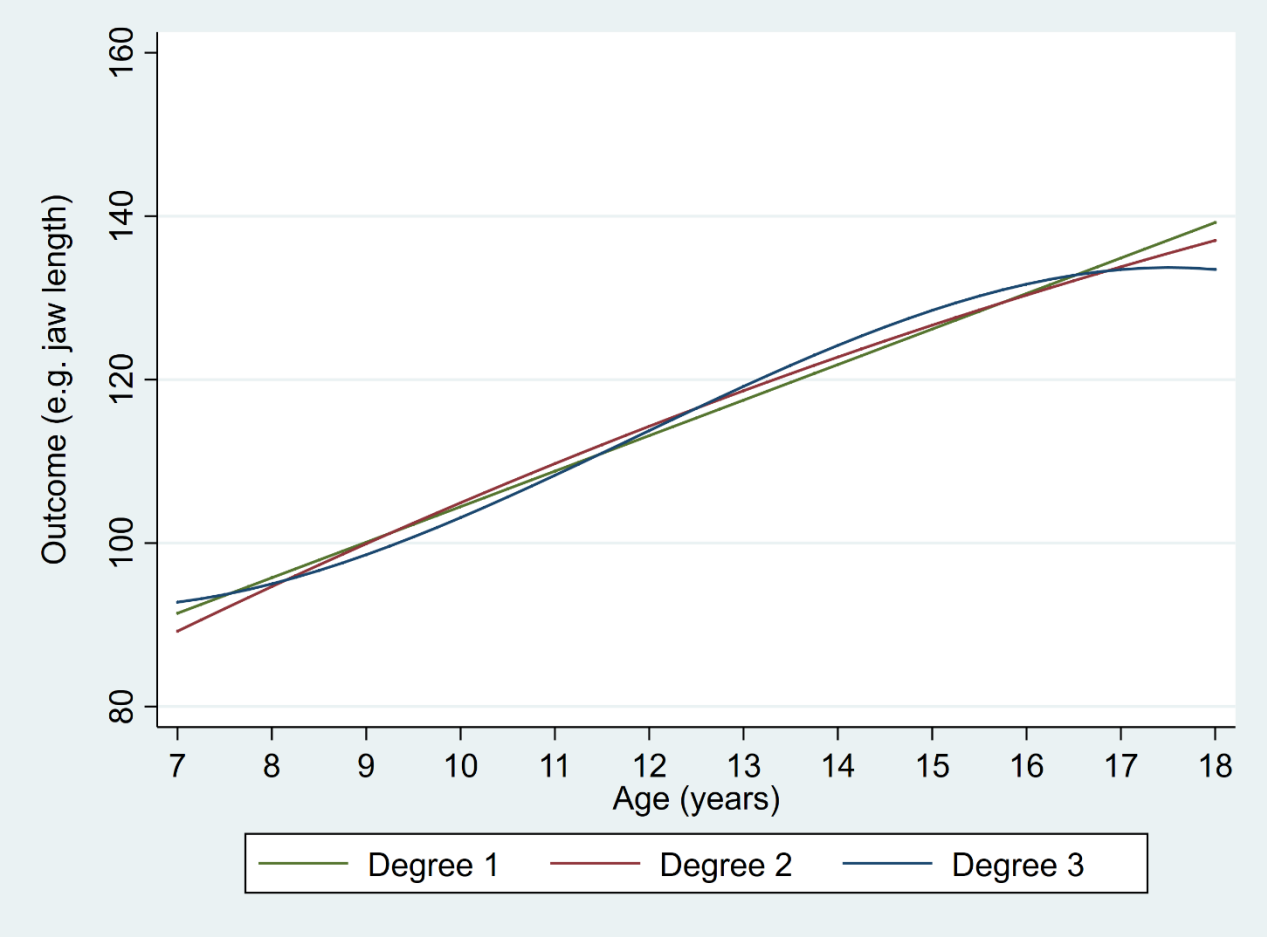


Figure 4-1. An illustrative example showing population-average growth trajectories estimated by conventional polynomial (CP) growth curve model.

Fractional polynomial (FP)

The FP approach offers greater flexibility in modelling growth trajectories than conventional polynomials (Royston & Altman, 1994). Unlike a CP GCM, the FP GCM allows repetition of powers which can be either zero, negative, fractional or a combination of these. When the power is zero, the natural log of age is used instead of age to the power 0, which is simply the constant 1. The log transformation of age represents the Box and Tidwell (1962) transformation of the predictor (Royston & Altman, 1994). Following Royston and Altman (1994) my study considered the following set of eight powers [-2, -1, -0.5, 0, 0.5, 1, 2, 3].

For any degree of FP model (m, which denotes the number of fractional polynomial terms), there is a total of $(m + p - 1)! / m! (p - 1)$ possible FP GCMs [6], where p denotes the number of powers and m is the degree of the model. For example, the total number of possible models resulting from a combination of eight powers for a FP of degree one ($m = 1$), two ($m = 2$), and three ($m = 3$) models is 8, 36, and 120, respectively.[7] The FP approach selects the best-fitting model by selecting the best-fitting model for each degree m being considered (e.g. the best-fitting model of degree 1, the best-fitting of degree 2, etc) and then using the likelihood ratio test to choose amongst these *m* best-fitting models (Royston, 2017; Royston & Altman, 1994). Figure 4-2 shows use of zero, negative and repeated powers for second- and third-degree FP GCMs applied to a hypothetical data.

[6] See https://www.cs.sfu.ca/~ggbaker/zju/math/perm-comb-more.html

[7] [(8! / (1) x (7!) = (8 x 7 x 6 x 5 x 4 x 3 x 2 x 1) / ((1) x (7 x 6 x 5 x 4 x 3 x 2 x 1)) = 8]
[(9! / (2!) x (7!) = (9 x 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1) / ((2 x 1) x (7 x 6 x 5 x 4 x 3 x 2 x 1)) = 36]
[(10! / (3!) x (7!) = (10 x 9 x 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1) / ((3 x 2 x 1) x (7 x 6 x 5 x 4 x 3 x 2 x 1)) = 120]

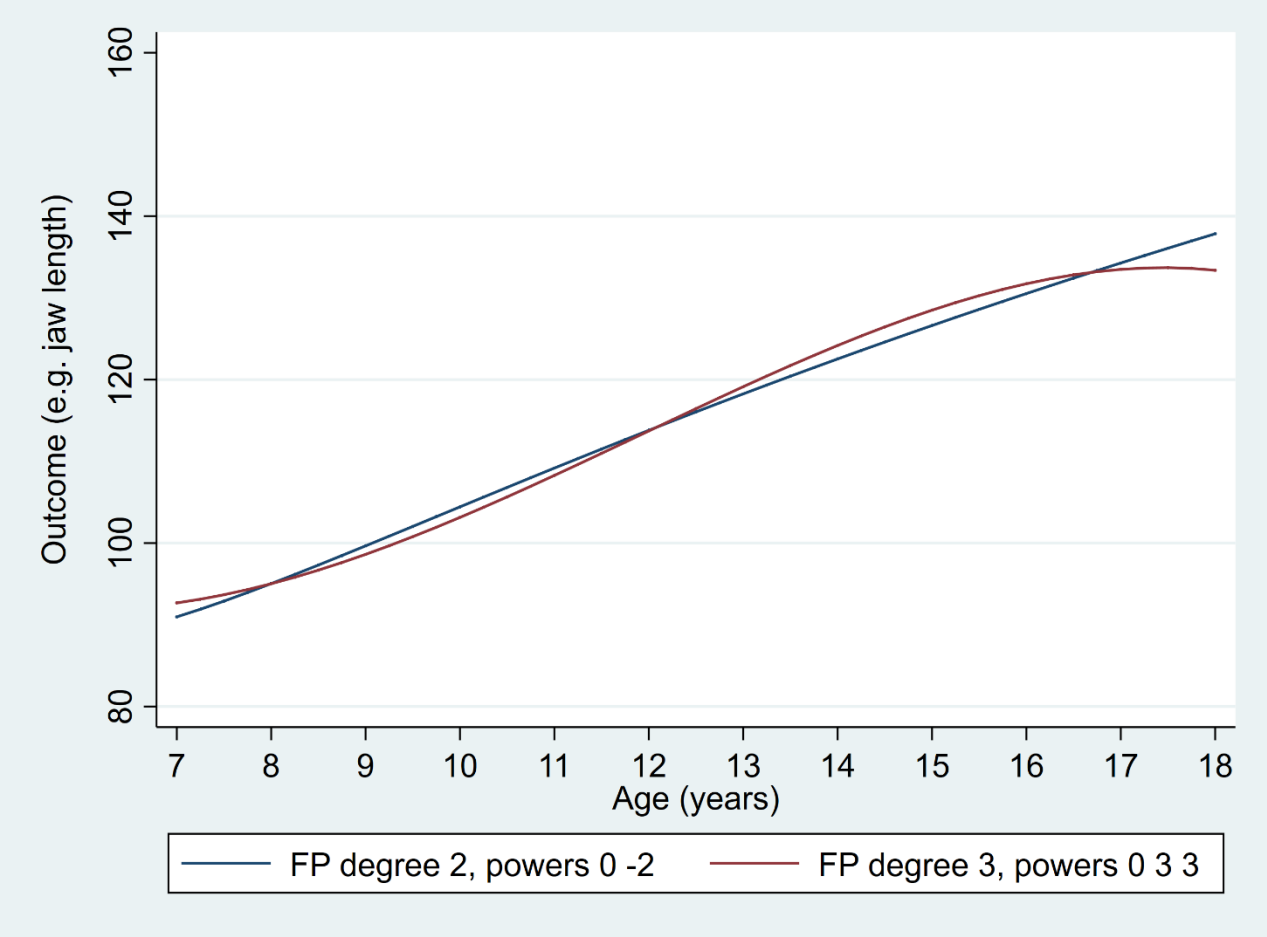


Figure 4-2. An illustration of population-average growth trajectories estimated by fractional polynomial (FP) growth curve model.

Considering the flexibility provided by the fractional polynomials, authors proposing the FP GCM suggested that *"models with degree higher than 2 are rarely required in practice*" (Royston & Altman, 1994). As a result, most studies using fractional polynomials restrict fitting FP GCMs up to a second degree (Binder et al., 2013; Long & Ryoo, 2010; Royston & Altman, 1994; Tilling et al., 2014). Although up to second-degree FP GCMs are routinely used, third- or higher-degree polynomials may be needed for more complex trajectories (Binder et al., 2013; Tilling et al., 2014). Following a recent study that used third-degree FP GCM for analysing height data (Simpkin et al., 2017), it was decided to fit up to third-degree FP models to the jaw growth data. The general form of an FP GCM can be written as (Simpkin et al., 2017):

$$y_{ij} = \beta_0 + \sum_{r=1}^{m_p} \beta_r age_{ij}^{p_r} + u_{0j} + \sum_{s=1}^{m_q} u_{sj} age_{ij}^{p_s} + e_{ij} \tag{4-5}$$

where m_p represents the degree of the fixed effects part of the model, m_q represents the degree of the random effects part of the model, and p_r and p_s denote the power(s) of fixed and random parts of the model, respectively. As powers p_r and p_s can be repeated ($m_p > 1;\ m_q > 1$), the model (4-5) with a mixture of zero and non-zero powers with and without repetitions can be written as:

$$\begin{aligned} y_{ij} = \beta_0 + \sum_{r=1}^{m_p} \beta_r \begin{cases} age_{ij}^{p_r} & if\ p_r \neq 0; c = 1, \\ age_{ij}^{p_r}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_r \neq 0; c > 1, \\ \log(age_{ij})^{p_r} & if\ p_r = 0; c = 1, \\ \log(age_{ij})^{(r-1)} & if\ p_r = 0; c > 1, \end{cases} \\ + u_{0j} + \sum_{s=1}^{m_q} u_{sj} \begin{cases} age_{ij}^{p_s} & if\ p_s \neq 0; c = 1, \\ age_{ij}^{p_s}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_s \neq 0; c > 1, \\ \log(age_{ij})^{p_s} & if\ p_s = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ p_s = 0; c > 1, \end{cases} \\ + e_{ij} \end{aligned} \tag{4-6}$$

where c denotes the count (occurrences) of the same powers. When a power appears for the first time (not repeated), $c = 1$, and when the power is repeated for the first time (repeated once), then $c = 2$, and so on. As an example, consider a third-degree FP GCM with repeated power 2 (powers 2 2 2). The first FP term (power not repeated: c = 1) is created as age_{ij}^{2}, the second FP term (power is repeated for the first time: c = 2) is created as $age_{ij}^{2} * log(age_{ij})^{(2-1)} = age_{ij}^{2} * log(age_{ij})^{1}$, and the third FP term (power is repeated for the second time: c = 3) is created as $age_{ij}^{2} * log(age_{ij})^{(3-1)} = age_{ij}^{2} * log(age_{ij})^{2}$.

The fixed effects and individual-specific random effects are indicated by $\beta_0, \dots, \beta_{m_p}$ and $u_{0j}, \dots, u_{m_q j}$, respectively. The e_{ij} are the residuals. Model assumptions are same as described earlier for CP GCM. The $\mathbf{\Omega_2}$ is a $(m_q + 1) \times (m_q + 1)$ dimensional matrix $\mathbf{\Omega_2}$ comprising $\left[(m_q + 1)\left((m_q + 1) + 2\right)/2\right]$ unique variance-covariance parameters. The $\mathbf{\Omega_1}$ is a $n_j \times n_j$ dimensional identity covariance matrix with a diagonal constant variance parameter, σ_e^2.

Restricted cubic spline (RCS)

Spline functions are an alternative to conventional and fractional polynomial functions for modelling growth data (Howe et al., 2016; Tilling et al., 2014). Unlike global polynomial basis functions, the spline functions are locally flexible (Wold, 1974). "*Global polynomial basis means that local changes in the data have global effects"* (Simpkin et al., 2018). A spline regression model includes piecewise linear or polynomial functions defined over adjacent intervals joined by 'knots' (Desquilbet & Mariotti, 2010; Greenland, 1995; Harrell, 2015; Wold, 1974). Commonly used spline functions include linear, quadratic, restricted quadratic, cubic and restricted cubic spline functions (Desquilbet & Mariotti, 2010; Durrleman & Simon, 1989; Greenland, 1995).

Cubic spline functions are characterised by having continuous first and second derivatives at the knots (Harrell, 2015). As cubic splines can behave poorly in the two tails (before the first knot and after the last knot), where data is often sparse (Stone & Koo, 1985), Stone and Koo (1985) suggested using restricted cubic splines; that is, cubic splines constrained to be linear in the tails. An advantage of using an RCS function (also called natural cubic splines) over a cubic spline function is that a model based on a restricted spline function is more parsimonious than one based on the cubic spline function (Harrell, 2015). For k number of knots, a cubic spline function estimates $k + 4$ parameters (including an intercept), whereas an RCS function estimates k parameters, including the intercept, (Harrell, 2015).

An RCS GCM with k_p knots for the fixed part of the model and k_q knots for the random part of the model can be written as (Harrell, 2015; Korn & Graubard, 2011):

$$y_{ij} = \beta_0 + \beta_1 age_{ij} + \sum_{p=2}^{k_p-2} \beta_p S(age_{ij}) + u_{0j} + u_{1j} age_{ij} + \sum_{q=2}^{k_q-2} u_{qj} S(age_{ij}) + e_{ij} \quad (4\text{-}7)$$

where $\sum_{p=2}^{k_p-2} \beta_p . S(age_{ij})$ and $\sum_{q=2}^{k_q-2} u_{qj} . S(age_{ij})$ parts of the model use a cubic spline function, $S(.)$, to generate RCS design matrix comprising $k_p - 2$ spline terms for the fixed effects parts of the model and $k_q - 2$ spline terms for the random effects parts of the model (see below). The fixed effects ($\beta_0, \beta_1 \ldots \beta_p$) and the random effects ($u_{0j}, u_{1j} \ldots u_{qj}$) describe population average and individual-specific growth trajectories, respectively. The e_{ij} are the residuals. Model assumptions are same as described earlier for CP GCM.

The RCS terms are constructed by using an appropriate spline basis function. There are two main ways to construct RCS terms: the B-splines basis (de Boor, 1978) and the truncated power basis (Smith, 1979). Although B-splines basis are numerically more stable than the truncated power basis, these are more complex and do not allow for extrapolation beyond the outer knots (Harrell, 2015). As modern computation methods of matrix handling are very efficient, the truncated power basis seldom presents estimation problems, and is widely used for the construction of the cubic spline design matrix (Harrell, 2015). To put the truncated power basis on the same scale and to improve the numerical stability, Harrell (2015) recommended dividing each truncated power basis by the square of the difference between the last (t_{k_p}) and the first (t_1) knots: $(t_{k_p} - t_1)^2$.

Using the truncated power basis "+" function, $k_p - 2$ spline terms for the fixed part of a model with k_p knots ($t_p, \ldots, t_{k_p}$ where $p = 1, \ldots, k_p - 2$) are created as (Harrell, 2015):

$$\frac{\left(age_{ij} - t_p\right)_+^3}{\left(t_{k_p} - t_1\right)^2} - \frac{\left(age_{ij} - t_{k_p-1}\right)_+^3 \left(t_{k_p} - t_p\right)}{\left(t_{k_p} - t_{k_p-1}\right)\left(t_{k_p} - t_1\right)^2} + \frac{\left(age_{ij} - t_{k_p}\right)_+^3 \left(t_{k_p-1} - t_p\right)}{\left(t_{k_p} - t_{k_p-1}\right)\left(t_{k_p} - t_1\right)^2} \qquad (4\text{-}8)$$

where "+" function is defined as $S_+ = S \; if \; S > 0; = 0 \; if \; S \leq 0$.

Similarly, $k_q - 2$ new variables (spline terms) are created for the random part of the model:

$$\frac{\left(age_{ij}-t_q\right)_+^3}{\left(t_{k_q}-t_1\right)^2}-\frac{\left(age_{ij}-t_{k_q-1}\right)_+^3\left(t_{k_q}-t_q\right)}{\left(t_{k_q}-t_{k_q-1}\right)\left(t_{k_q}-t_1\right)^2}+\frac{\left(age_{ij}-t_{k_q}\right)_+^3\left(t_{k_q-1}-t_q\right)}{\left(t_{k_q}-t_{k_q-1}\right)\left(t_{k_q}-t_1\right)^2} \quad (4\text{-}9)$$

The number of knots and their locations are usually specified in advance and not treated as free parameters to be estimated (Harrell, 2015). Although knots as free parameters enhances the flexibility of the function, it is common practice to specify knots *a priori* because knots as free parameters result in instability of estimates and statistical inference problems (Harrell, 2015), and location of knots in an RCS GCM is not crucial, as the fit depends much more on the choice of k, the number of knots (Harrell, 2015; Stone, 1986).

In most cases, three to seven knots offer good flexibility for modelling complex growth trajectories and provide an adequate fit to the data (Harrell, 2015). Increasing the number of knots can result in loss of precision caused by overfitting (Harrell, 2015; Stone, 1986). A common strategy is to place knots at the fixed quantiles (percentiles) of the predictor's (age) marginal distribution, which ensures that enough data points are available in each interval, thus minimising the influence of outliers on knot placement (Harrell, 2015). The quantiles recommended by Harrell (2015) are as follows (quantiles in parentheses): three knots (0.10, 0.5, 0.90), four knots (0.05, 0.35, 0.65, 0.95), five knots (0.05, 0.275, 0.5, 0.725, 0.95), six knots (0.05, 0.23, 0.41, 0.59, 0.77, 0.95), and seven knots (0.025, 0.1833, 0.3417, 0.5, 0.6583, 0.8167, 0.975). Figure 4-3 shows RCS GCMs with different numbers of knots applied to a hypothetical data.

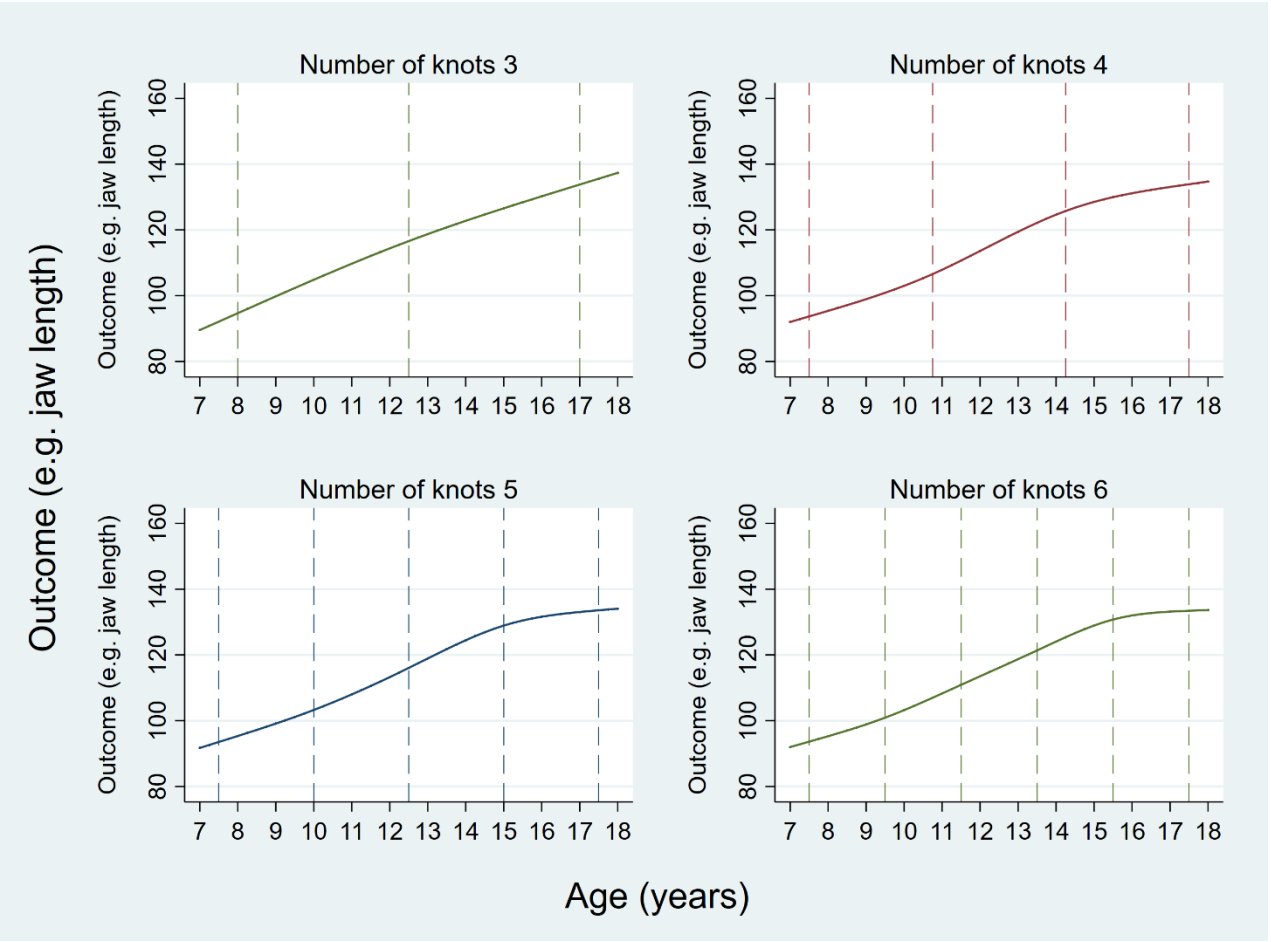


Figure 4-3. An illustration of population-average growth trajectories estimated by the restricted cubic spline (RCS) growth curve model. Vertical dashed lines show location of knots.

4.2.3. Nonlinear growth curve models

Like the linear GCM, the nonlinear GCM allows both fixed and random effects. However, in contrast to the linear GCM, the nonlinear GCM involves one or more nonlinear combinations of fixed and random effects (Lindstrom & Bates, 1990; Pinheiro & Bates, 1995a, 2000). Typically used for growth modelling, the assumptions about the individual-specific random effects and the within-individual residuals in a nonlinear GCM are identical to those of a linear GCM (Lindstrom & Bates, 1990; Pinheiro & Bates, 2000).

Superimposition by translation and rotation (SITAR)

The SITAR GCM (Cole et al., 2010) fits a mean curve to the data on the assumption that individual-specific growth curves differ from the mean curve in three ways (Figure 4-4): the size relative to the mean growth curve (vertical shift), the timing of the adolescent growth spurt relative to the mean APGV (horizontal shift), and the intensity of the growth velocity, i.e., the relative rate at which individuals grow during the adolescent growth spurt in comparison to the mean growth intensity (horizontal stretch) (Cole et al., 2010; Cole & Mori, 2017).

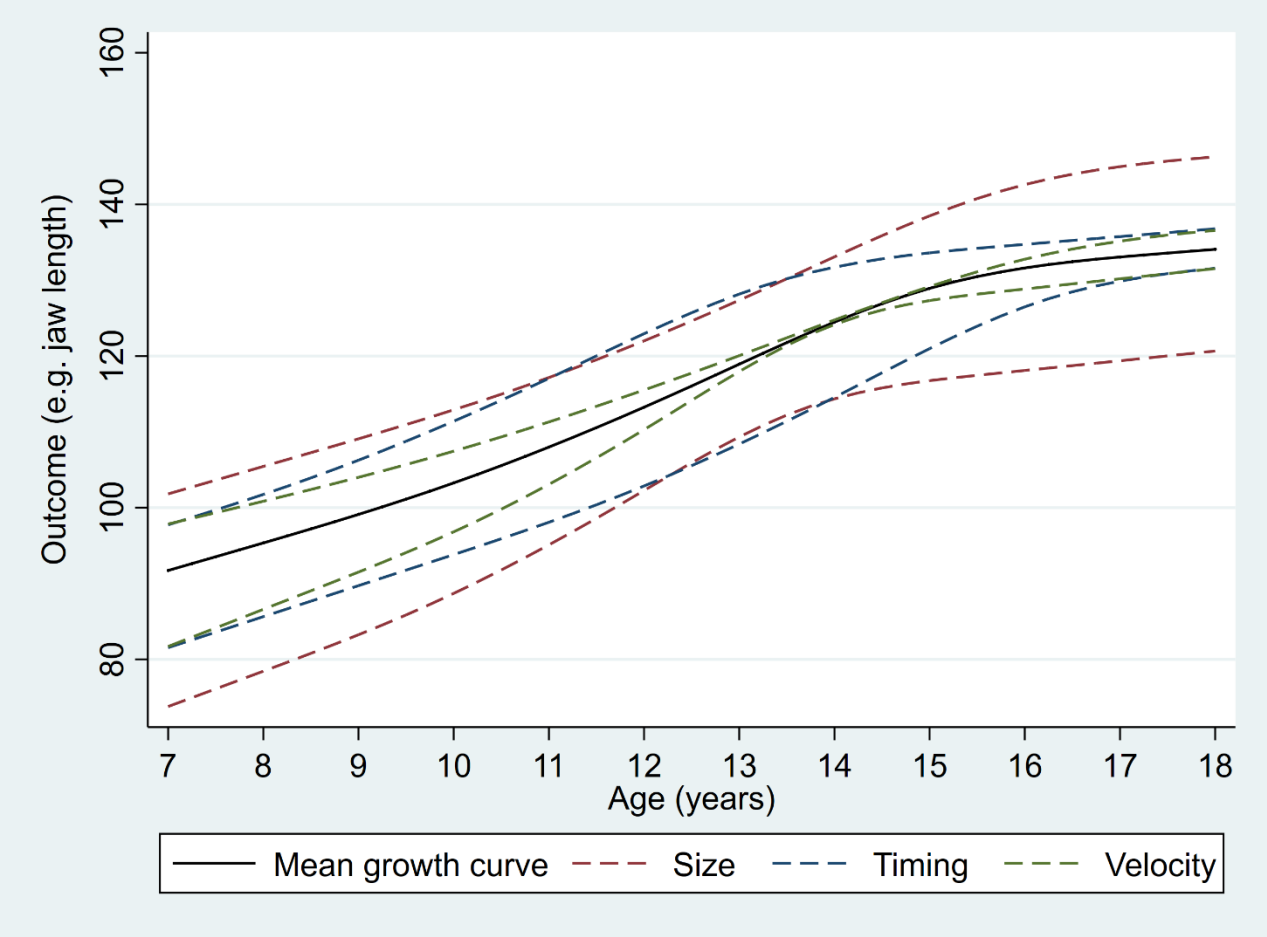


Figure 4-4. A geometrical illustration of parameters (size, timing and intensity) estimated by the superimposition by translation and rotation (SITAR) growth curve model. The solid black line is the mean growth curve. Paired maroon dashed lines indicate a vertical shift in the curve corresponding to an increase or decrease in size. Paired blue dashed lines indicate a horizontal age shift corresponding to early or late timing of the adolescent growth spurt. Paired green dashed lines show shrinking or stretching of the age scale due to increase or decrease in growth velocity. Redrawn from Cole et al. (2010). SITAR – a useful instrument for growth curve analysis. International Journal of Epidemiology, 39(6), 1558–1566. doi:10.1093/ije/dyq115.

The SITAR GCM has been extensively used for modelling height growth data (Cole et al., 2016; Cole, 2017b; Cole et al., 2010; Cole & Mori, 2017; Riddell et al., 2017). The model uses the B-spline basis to construct a design matrix of restricted cubic splines (natural splines) (Cole, 2018; Cole et al., 2010) as proposed by Beath (2007). The present study used a truncated power basis to construct the design matrix of the restricted cubic splines as proposed by (Harrell, 2015).Section 4.2.2 (see RCS GCM) described these two approaches which are commonly used to construct the design matrix of the restricted cubic splines. The restricted cubic splines created by either approach (B-spline basis or truncated power basis) are essentially the same as spline segment before the first knots and after the last knot are linear (see Chapter 6, Section 6.2.8 for further details). The SITAR GCM is expressed as (Cole et al., 2016):

$$y_{ij} = \alpha_0 + \alpha_j + S\left(\frac{age_{ij} - (\beta_0 + \beta_j)}{e^{-(\gamma_0 + \gamma_j)}}\right) + e_{ij} \qquad (4\text{-}10)$$

Here $S(.)$ is an RCS function as described earlier in Section 4.2.2 (Equation ((4-8)). The model (4-10) can be written as:

$$y_{ij} = \alpha_0 + \alpha_j + s_1\left(\frac{age_{ij} - (\beta_0 + \beta_j)}{e^{-(\gamma_0 + \gamma_j)}}\right) + \sum_{p=2}^{k_p-2} s_p . S\left(\frac{age_{ij} - (\beta_0 + \beta_j)}{e^{-(\gamma_0 + \gamma_j)}}\right) + e_{ij} \qquad (4\text{-}11)$$

The $S(.)$ function creates $k_p - 2$ spline terms where k_p is the number of knots (see below). Model (4-11) estimates $3 + k_p - 1$ (α_0, β_0, γ_0, $s_1, \dots, s_p$) fixed effects where α_0, β_0, γ_0 are population average size, timing and intensity parameters. The three individual-specific random effects for size (α_j), timing (β_j) and intensity (γ_j) growth parameters describe how an individual's growth differs from the mean growth curve collectively defined by $\alpha_0, \beta_0, \gamma_0, s_1, \dots, s_p$. The e_{ij} are residuals. The parameterisation of γ_j as exponential (e) ensures that both positive and negative values of velocity are permissible, with zero corresponding to average velocity (Cole et al., 2010).

Like linear GCMs, it is assumed that the random effects and residuals are mutually independent and follow normal distributions. The random effects (α_j , β_j and γ_j) follow a multivariate normal distribution with 0 mean vector and a 3×3 dimensional covariance matrix $\mathbf{\Omega_2}$.The residuals are assumed to be independent and normally distributed with 0 mean vector and a $n_j \times n_j$ dimensional identity covariance matrix $\mathbf{\Omega_1}$ with a diagonal constant variance parameter, σ_e^2.

$$\frac{\left(\left(\frac{age_{ij}-(\beta_0+\beta_j)}{e^{-(\gamma_0+\gamma_j)}}\right)-t_p\right)_+^3}{\left(t_{k_p}-t_1\right)^2}-\frac{\left(\left(\frac{age_{ij}-(\beta_0+\beta_j)}{e^{-(\gamma_0+\gamma_j)}}\right)-t_{k_p-1}\right)_+^3\left(t_{k_p}-t_p\right)}{\left(t_{k_p}-t_{k_p-1}\right)\left(t_{k_p}-t_1\right)^2}$$
$$+\frac{\left(\left(\frac{age_{ij}-(\beta_0+\beta_j)}{e^{-(\gamma_0+\gamma_j)}}\right)-t_{k_p}\right)_+^3\left(t_{k_p-1}-t_p\right)}{\left(t_{k_p}-t_{k_p-1}\right)\left(t_{k_p}-t_1\right)^2} \quad (4\text{-}12)$$

The parameters α_0, β_0, γ_0, α_j, β_j, γ_j are same as described above (4-11).

Preece–Baines (PB)

Preece and Baines (1978) proposed three models with five (Model 1), and six (Model 2 and Model 3) parameters. In their study, Preece and Baines (1978) reported that Model 1 and Model 3 have better convergence properties than Model 1.

As Model 1 is more parsimonious (five parameters) than Model 3 (six parameters) and has a particularly simple functional form (Preece & Baines, 1978), it is the most widely used model for modelling height data (Banik et al., 2017; Grimm et al., 2011; Hauspie et al., 2004; Sayers et al., 2013; Simpkin et al., 2017). The model is designed to fit adolescent growth starting from childhood (Hauspie et al., 2004).

The present study used PB Model 1 (hereafter referred to as the PB GCM) for modelling jaw growth data. Grimm et al. (2011) implemented the random effect form of the PB GCM (Preece & Baines, 1978), which can be written as:

$$y_{ij} = (s_{max} + s_{max\,j}) - \frac{2\left((s_{max} + s_{max\,j}) - (s_{\theta} + s_{\theta\,j})\right)}{e^{(s_0 + s_{0\,j})(age_{ij} - (\theta + \theta_j))} + e^{(s_1 + s_{1\,j})(age_{ij} - (\theta + \theta_j))}} + e_{ij} \quad (4\text{-}13)$$

where $s_{\max}$, s_{θ} , s_0 , s_1 and θ are the five fixed effect parameters describing the mean growth pattern for a group of individuals. $s_{\max}$ denotes the maximum adulthood size and s_{θ} is the size at age θ, where parameter θ denotes the adolescent growth spurt during puberty. s_0 and s_1 are growth-rate constants parameters related to prepubertal and pubertal growth velocities. $s_{max\,j}$, $s_{\theta\,j}$, $s_{0\,j}$, $s_{1\,j}$ and θ_j are corresponding individual-specific random effects. The e_{ij} are residuals. Similar to linear GCMs, it is assumed that the random effects and residuals are mutually independent and follow a normal distribution. The random effects follow a multivariate normal distribution with 0 mean vector and covariance matrix $\boldsymbol{\Omega_2}$with 15 distinct variance-covariance parameters. Residuals follow

a normal distribution with 0 mean vector and a $n_j \times n_j$ dimensional identity covariance matrix $\mathbf{\Omega_1}$ with a diagonal constant variance parameter, σ_e^2. Figure 4-5 shows population average growth trajectory estimated by PB GCM applied to a hypothetical data.

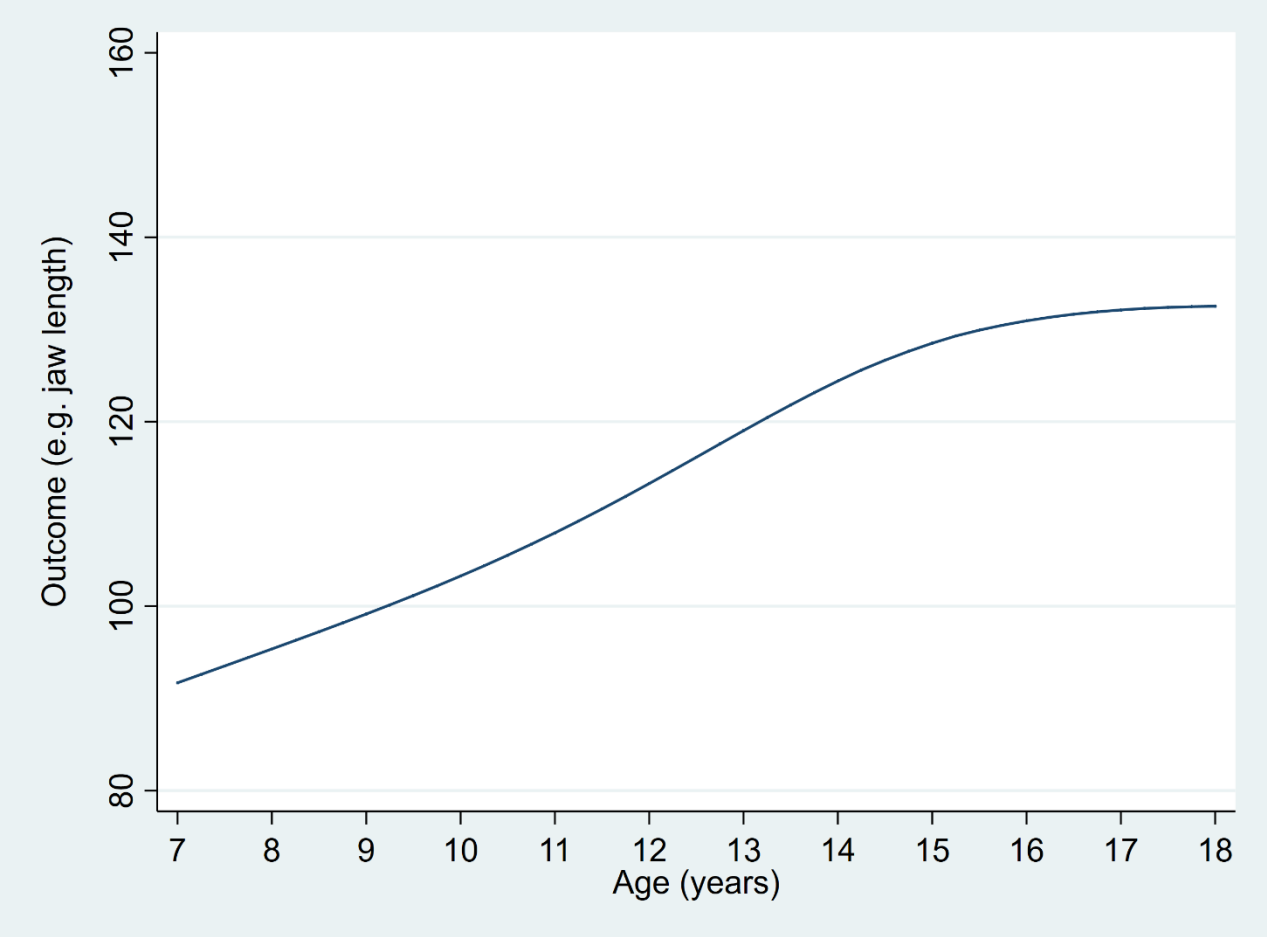


Figure 4-5. An illustration of population-average growth trajectory estimated by the Preece–Baines (PB) growth curve model.

4.2.4. Estimation methods

The full information maximum likelihood (FIML, which as simply referred as ML) method of estimation (Corbeil & Searle, 1976; Goldstein, 2011b; Hartley & Rao, 1967; Harville, 1977a, 1977b; Laird & Ware, 1982; Lindstrom & Bates, 1990; Pinheiro & Bates, 1995a) is most widely used in growth curve modelling as it has many favourable properties. When sample size is large, the estimates are consistent and usually robust against mild violations of the normality assumption (Hox, 2010), which is useful when making statistical inferences (Bollen, 1989; Raudenbush, 2001). As ML uses all available response data even if an individual only has one observed response measurement, it is a very useful estimation method for analysis of missing response data, as it does not lead to listwise deletion of missing data (Hox & Roberts, 2011; Mehta & Neale, 2005).

Depending on the parameters included in the likelihood function, the ML estimation method is classified into full ML and restricted maximum likelihood (REML) methods (Goldstein, 2011b; Hox, 2010). In full ML estimation, both the regression coefficients and the variance-covariance parameters are included in the likelihood function (Goldstein, 2011b; Hox, 2010; Searle et al., 2006). In the REML estimation, only the variance-covariance parameters are included in the likelihood function, and the regression coefficients are functions of the covariance parameter estimates (Goldstein, 2011b; Hox, 2010; Searle et al., 2006).

The difference between ML and REML estimation is usually small if the analysis is based on an adequately large sample size (Browne, 1998; Longford, 1993). However, when the sample size is small, the ML-based variance-covariance parameters are downwardly biased, because estimation does not take into account the degrees of freedom used in estimation of the fixed effects (Gumedze & Dunne, 2011). Studies have shown that REML performs much better than ML, and is therefore preferred over ML estimation, when the sample size is small (Bell et al., 2014; Browne & Draper, 2006; Gumedze & Dunne, 2011).

For nonlinear GCMs, the likelihood estimation requires numerical integration because the random effects enter the model nonlinearly (Harring & Liu, 2016; Lindstrom & Vorperian, 2005; Pinheiro & Bates, 2000). The linearizing approach, which is also termed the LME approximation approach, approximates the likelihood through linearizing the nonlinear function using the ML or REML. This approach, first proposed by Lindstrom and Bates (1990) and later extended by Pinheiro and Bates (1995a), is widely implemented in several mainstream statistical software packages (Pinheiro et al., 2018; Pinheiro & Bates, 2000; StataCorp, 2017a).

Bayesian estimation is an alternative to the frequentist approach of GCM estimation (Goldstein, 2011b; Stegmueller, 2013). Although Bayesian modelling has some advantages over the frequentist approach, especially when sample size is small, the frequentist approach is popular for fitting GCMs (de Valpine, 2012; Stegmueller, 2013). A simulation study has shown that, except when the level 2 sample size is too small (less than 15 units), both Bayesian and frequentist approaches perform well (Stegmueller, 2013). This thesis uses the frequentist approach (REML method).

4.3. Distance and derivatives

This section first briefly reviews the concepts of distance, velocity and acceleration curves, and their role in studying growth dynamics including APGV and PGV. It then describes estimation of distance and its derivatives (velocity and acceleration) for GCMs and their graphical interpretation, which is essential for finding APGV and PGV.

4.3.1. Distance (size) and derivatives (velocity and acceleration)

Age-specific growth changes can be described and visualised graphically by using the distance, velocity and acceleration curves as shown in Figure 4-6 (Hauspie & Roelants, 2012; Hauspie et al., 2004). The population average distance, velocity and acceleration curves shown in Figure 4-6 were obtained by fitting PB model to a hypothetical dataset. The *distance* is used to describe increase in size (such as height and jaw length), and is easy to visualise as it shows how far a child has progressed towards adulthood (Cameron & Bogin, 2012). The term *velocity* was coined by Tanner (1951) and expresses growth increments in a given period (Hauspie & Roelants, 2012). Distance curve is often used for reference charts as it shows a cumulative increase in size, but it does not provide information about growth dynamics (Hauspie et al., 2004). The term *acceleration* denotes the rate of change in velocity as a function of age. Even though knowing acceleration helps in locating the local minima and maxima (such as APGV) along the velocity curve (see Section 4.3.3), the acceleration curve itself is not easily interpretable in the context of growth (Hauspie et al., 2004).

The preferred tool to study growth is the velocity curve, which in a clear and intuitive way shows the areas of interest such as APGV (Hauspie et al., 2004). The velocity curve is central to a clinically relevant understanding of jaw growth (Dean et al., 2016, pp. 375-389)).

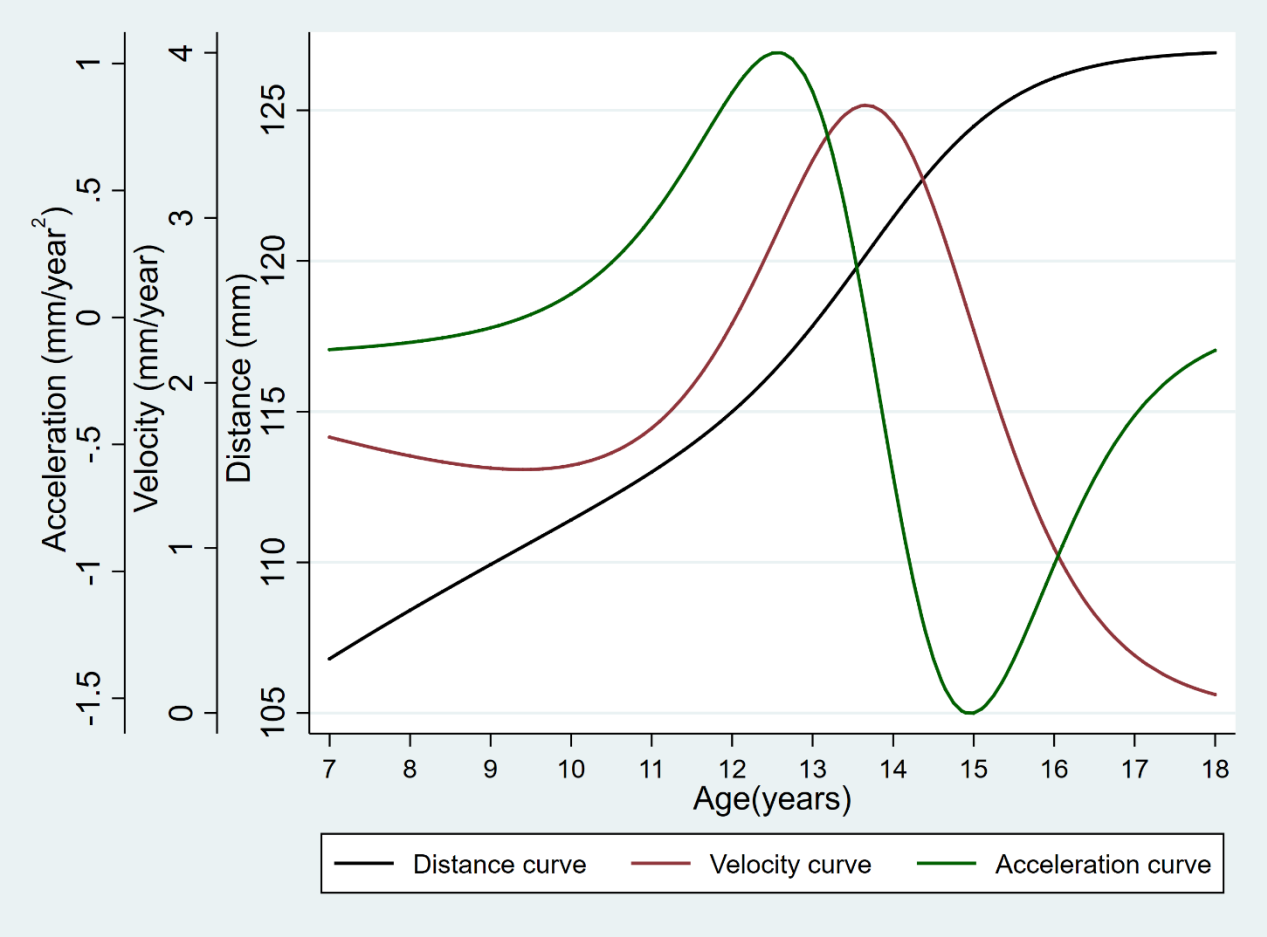


Figure 4-6. An illustrative example showing distance, velocity and acceleration curves.

Human physical maturational events can be described using landmarks on a velocity curve (Hauspie et al., 2004). The landmarks used to define stages of the adolescent growth spurt and growth parameters mapped on to the velocity curve are shown in Figure 4-7 are (Abbassi, 1998; Hauspie & Roelants, 2012; Hauspie et al., 2004; Taranger & Hagg, 1980): 1) age at take-off, which corresponds to age at minimal growth velocity and marks the onset of the adolescent growth spurt; 2) PGV, measuring the maximum growth velocity during the adolescent growth spurt; 3) APGV, denoting the timing of maximum growth velocity; 4) duration of the adolescent growth spurt, which lasts from the age at take-off until the end of the adolescent growth spurt, marked by age at the maximal deceleration in the growth during the adolescent growth spurt; and 5) the contribution of the adolescent growth spurt to the final adult size (termed adolescent gain), which is the difference between adult size and size at take-off. The present study focused on APGV and PGV, which are the most important factors used in clinical decisions for treating skeletal malocclusions (Buschang et al., 2017; Proffit, 2006; Proffit et al., 2014; Wheeler et al., 2006). Curves shown in Figure 4-7 were obtained by fitting the PB model to a hypothetical dataset.

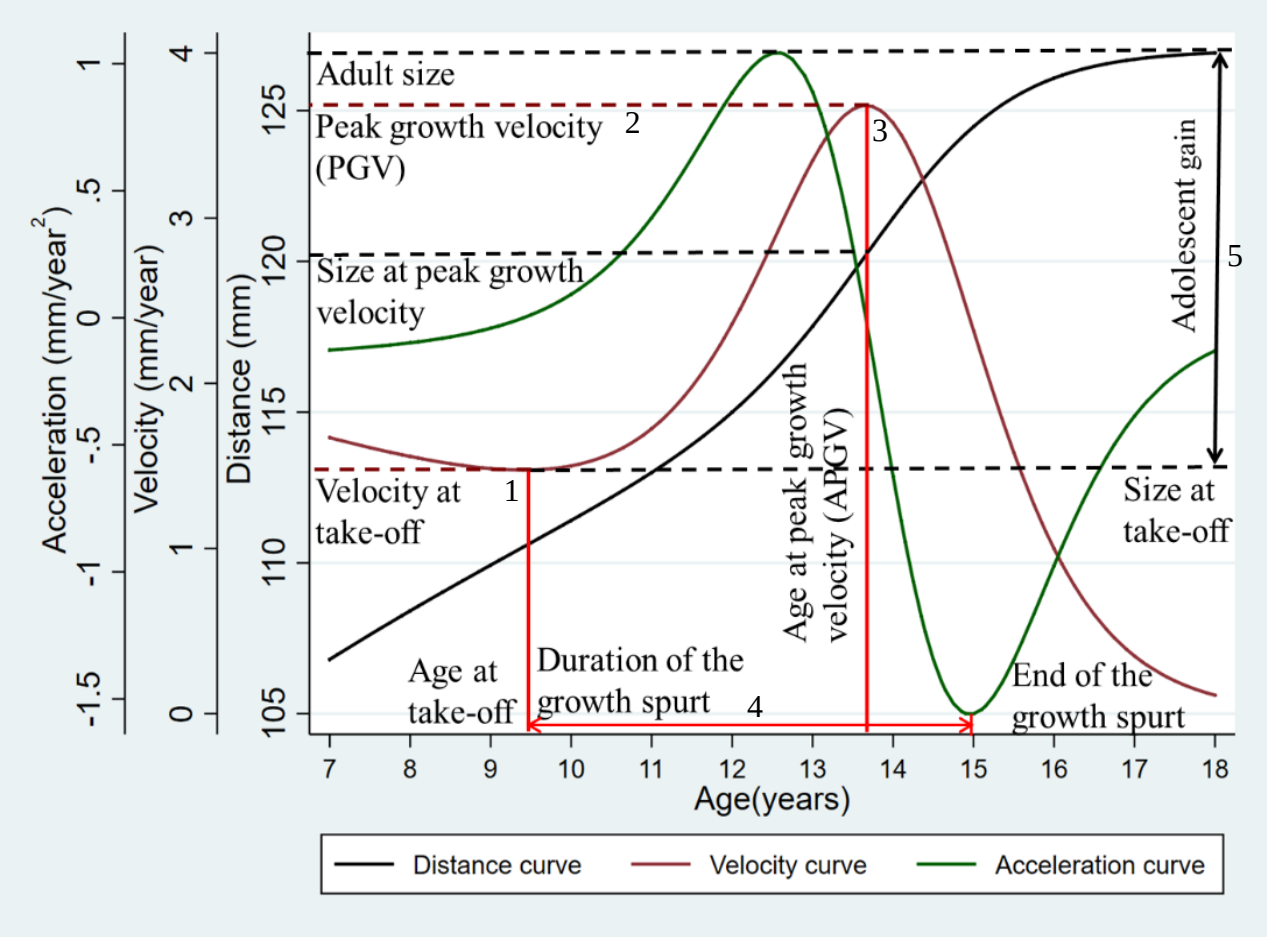


Figure 4-7. A graphical illustration of adolescent growth spurt parameters mapped on the velocity curve between seven and 18 years of age.

4.3.2. Estimation of distance and derivatives

The first derivative (velocity), which is defined as an instantaneous *rate of change in distance* as a function of age, is denoted by $f'(age)$ or $d_{distance}/d_{age}$. For a function $f(age)$, the second derivative of its first derivative ($f'(age)$) is written as $f''(age)$ or $d^2_{distance}/d_{age^2}$. The second derivative ($f''(age)$) is called the acceleration and shows the *rate of change in the velocity* as a function of age.

GCMs estimate population-average (fixed effects) and individual-specific (random effects) growth parameters. Therefore, population-average and individual-specific growth curves can be obtained. Subscript $_0$ denotes population-average distance ($f_0(age)$), its first derivative ($f_0'(age)$) and its second derivative $f_0''(age)$. Likewise, subscript $_j$ is used to denote individual-specific distance ($f_j(age)$), first derivative ($f_j'(age)$) and second derivative ($f_j''(age)$).

An illustrative example below shows distance and derivatives (first and second) estimation for a third degree i.e., a cubic CP GCM ($p = q = 3$) described earlier (Section 4.2.2, Equation (4-1)). To easily demonstrate how linear and nonlinear terms are included in the derivatives estimation, a CP GCM (4-1) is rewritten as:

$$y_{ij} = \beta_0 + \beta_1 age_{ij}^1 + \sum_{r=2}^{p} \beta_r age_{ij}^r + u_{0j} + u_{1j} age_{ij}^1 + \sum_{s=2}^{q} u_{sj} age_{ij}^s + e_{ij} \qquad (4\text{-}14)$$

As distance is a function itself, and the function linking the outcome to the predictors for a linear GCM is linear in its parameters (see Section 4.2.2), the regression coefficients are multiplied by respective functions (polynomial terms) and added up to estimate population-average distance (size) as:

$$f_{CP_0}(age) = \beta_0 + \beta_1 \{age_{ij}^1\} + \sum_{r=2}^{p} \beta_r \{age_{ij}^r\} \qquad (4\text{-}15)$$

To estimate individual-specific distance, the random effects are multiplied by respective polynomial terms and added to the population-average distance as:

$$f_{CP_j}(age) = \beta_0 + \beta_1\{age_{ij}^1\} + \sum_{r=2}^{p} \beta_r \{age_{ij}^r\} + u_{0j} + u_{1j}\{age_{ij}^1\} + \sum_{s=2}^{q} u_{sj}\{age_{ij}^s\} \quad \textit{(4-16)}$$

By using the basic differentiation rules (Strang & Herman, 2016a) which include the *constant* rule, the *power* rule, the *constant multiple* rule and the *sum* rule, the population-average (4-17) and individual-specific (4-18) first derivatives for conventional polynomial functions are calculated as:

$$f'_{CP_0}(age) = \beta_1 + \sum_{r=2}^{p} \beta_r \left\{r(age_{ij})^{(r-1)}\right\} \quad (4\text{-}17)$$

$$f'_{CP_j}(age) = \beta_1 + \sum_{r=2}^{p} \beta_r \left\{r(age_{ij})^{(r-1)}\right\} + u_{1j} + \sum_{s=2}^{q} u_{sj} \left\{s(age_{ij})^{(s-1)}\right\} \quad (4\text{-}18)$$

The second derivative is the derivative of the first derivatives. To calculate the second derivative, the same rules are applied to the first derivative. The population-average (4-19) and individual-specific (4-20) second derivatives of conventional polynomial functions are shown below:

$$f''_{CP_0}(age) = \sum_{r=2}^{p} \beta_r \left\{r(r-1)(age_{ij})^{(r-2)}\right\} \quad (4\text{-}19)$$

$$f''_{CP_j}(age) = \sum_{r=2}^{p} \beta_r \left\{r(r-1)(age_{ij})^{(r-2)}\right\} + \sum_{s=2}^{q} u_{sj} \left\{s(s-1)age_{ij}^{(s-2)}\right\} \quad (4\text{-}20)$$

Using the same differentiating approach and a few advanced rules, derivatives for FP, RCS and the SITAR GCMs are estimated. In addition to the basic differentiation rules applied above to the CP GCM, derivative estimation for fractional polynomial functions requires *logarithm* and *product* rules. RCS functions are essentially a combination of linear (before the first knot and after the last knot) and cubic (between first and the last knot) conventional polynomials. Therefore, the derivative estimation for the RCS GCM follows the basic differentiation rules that are applied to conventional polynomial functions. For the SITAR GCM, differentiation of RCS functions is the same as the RCS GCM. In addition, estimating derivatives for the SITAR GCM requires differentiation of an exponent (the intensity parameters; see Equation (4-10) in Section 4.2.3) using the *exponential* rule.

The PB GCM is a *solution* to a double logistic *differential equation* (Morgan, 1987; Preece & Baines, 1978; Verzelen et al., 2012). A differential equation relates an unknown function with one or more of its derivatives (Strang & Herman, 2016b). The derivative of a differential equation is calculated by first finding the function (also known as the *solution to the differential equation*) using integration, and then using the product rule to differentiate the function (Strang & Herman, 2016b). Preece and Baines (1978) provided a detailed description of different stages involved in finding the solution to the equation and estimating its first and second derivatives.

4.3.3. Finding change points (maxima and minima) using derivatives

Local maxima and minima occur where a function takes larger or smaller values compared with nearby points. For global maxima and minima, a function is larger or smaller than all other points along the entire interval of the function. Thus, a function can have multiple local minima and maxima but unique global maxima and minima. The local minima and maxima, which are of interest to the present study, are also known as critical or turning points (Hughes-Hallett et al., 2014). At a

critical point, the first derivative (f') is zero and the slope has different signs on either side of the point (Hughes-Hallett et al., 2014).

While the critical points are located on the first derivative (velocity curve), knowing inflection points using the second derivative (acceleration curve) helps in locating and testing whether a critical point is a local maximum or peak, or a local minimum or trough (Hughes-Hallett et al., 2014). At an inflection point, the second derivative is zero and the slope of the first derivative reverses sign either side of the inflection point (Hughes-Hallett et al., 2014).

4.4. Timing and intensity of the adolescent growth spurt

4.4.1. An overview

The APGV and the PGV can be obtained directly from the model-estimated parameters and from the derivatives. Although strictly speaking the APGV and the PGV are not directly estimated, I use the term "estimated" instead of "calculated" because the derivatives are derived using estimated parameters of the GCM model.

Recently, Cao et al. (2018) proposed two new methods for estimating population-average and individual-specific APGV and PGV. These methods are the quadratic function method and the numerical method. The authors reported a simulation study that compared the performances of these two methods with a model-based approach for the SITAR GCM (i.e., estimating APGV and PGV from model-estimated parameters). Results showed that both methods perform equally well or better than the model-based method (Cao et al., 2018).

In the numerical method, the velocity curve is approximated by numerically differentiating the distance curve: $\Delta\, height\, /\Delta\, age$, where $\Delta\, height$ and $\Delta\, age$ are the first-order derivatives of height and age, respectively. PGV is the largest ratio of $\Delta\, height$ to $\Delta\, age$. The age corresponding to PGV is the APGV.

The quadratic function method uses the velocity curve directly obtained by differentiating the growth functions. This method is computationally less intensive than the numerical method and is recommended by Cao et al. (2018). This thesis uses the quadratic function method. Details of determining PGV and APGV are provided in Section 4.4.2.

As frequent growth measurements are required to estimate an individual's PGV and APGV accurately, both quadratic function and numerical methods interpolate age measurements before

estimating distance and derivatives. In other words, after fitting the model to the original data, age is evenly interpolated by creating additional age measurements for an individual in their age range (Cao et al., 2018). Using the fixed and random effects parameters estimated from the original data, distance and derivatives are then estimated based on the new interpolated age measurements. Cao et al. (2018) recommend 12 and 365 age measurements per year for the quadratic function method and the numerical method, respectively. The recommended approach for the quadratic function method was followed for the present study.

4.4.2. Quadratic function method

The quadratic function method is based on the concept that the age corresponding to maximum velocity can be approximated by a quadratic polynomial function fitted to the velocity curve.

In the first step, Cao et al. (2018) numerically differentiated the velocity with respect to age to determine the acceleration. Instead of numerically differentiating the velocity curve, the present study used the second derivative estimated from the GCM (see the discussion of derivative estimation in Section 4.6).

In the second step, second derivative–based inflection points were used to mark the peak and trough areas on the velocity curve (Figure 4-8). Within the peak area, a crude estimate of the age corresponding to maximum velocity was obtained.

In the third and final step, a simple quadratic polynomial regression model was fitted on the velocity curve, using two neighbouring pairs on each side of the crude estimate of age (the peak region pr_0 shown as shaded grey area in Figure 4-8) as:

$$f'_{\mathrm{pr_0}} = \beta_{v0} + \beta_{v1} age_{pr_o} + \beta_{v2} age^2_{pr_o} + e \qquad (4\text{-}21)$$

where f'_{pr_0} is population-average velocity and age_{pr} is age at peak region denoted by the subscript *pr*. The e is the residual. The population-average age at peak growth velocity ($APGV_o$) is then computed as:

$$APGV_o = -\beta_{v1}/(2\beta_{v2}) \quad (4\text{-}22)$$

The population-average peak growth velocity (PGV_o) is calculated by substituting the value of $APGV_o$ in (4-21) as:

$$PGV_o = \beta_{v0} + \beta_{v1}APGV_o + \beta_{v2}APGV_o^2 \quad (4\text{-}23)$$

Individual-specific age at peak growth velocity ($APGV_j$) and peak growth velocity (PGV_j) were calculated similarly by following these three steps. Instead of using the population-average first and second derivatives, individual-specific derivatives were used to locate the peak region (pr_j). The quadratic conventional polynomial regression model was then fitted on the individual-specific velocity curve (f'_{pr_j}) and age (age_{pr_j}) corresponding to the peak region (pr_j) for estimating $APGV_j$, and PGV_j, as described in the above equations.

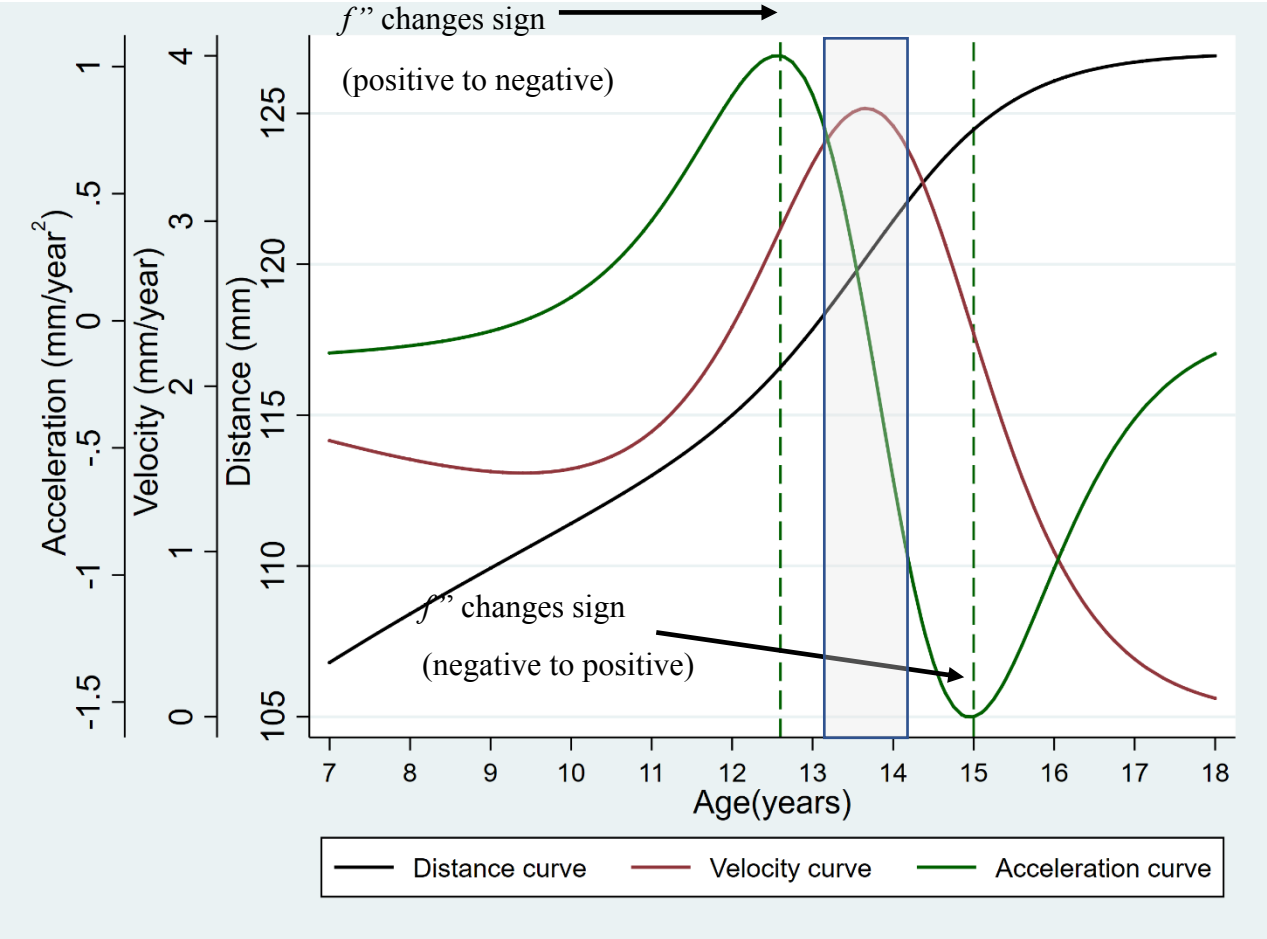


Figure 4-8. An illustrative example showing septation of peaks and troughs on the population-average velocity curve by using derivatives. The second derivative (f'')-based inflection points are denoted using vertical dashed lines. Once the peak region (pr_0; shaded grey region) is identified within the peak area (area between vertical dashed lines), the quadratic function method (regressing velocity on age corresponding to the pr region) is used to estimate APGV and PGV (see Section *4.4.2*).

4.5. Methodological and clinical research work

This thesis compared three linear (CP, FP and RCS) and two nonlinear (SITAR and PB) GCMs, focusing on evaluating their fit to the jaw growth data, their ability to model jaw growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing i.e., APGV and intensity i.e., PGV). All five GCMs were estimated using the REML method (see Section 4.2.4 for details). The thesis comprised of three research chapters. The first two chapters (Chapter 5 and Chapter 6) are methodological whereas the third chapter (Chapter 7) is clinical. Each research chapter addressed one of the three specific research questions posed in this thesis (see Chapter 1, Section 1.3).

4.5.1. Research Question 1 (Chapter 5)

Over the past few decades, a range of fit criteria have been developed to aid researchers in model selection. These criteria can be classified into two types: information criteria and prediction criteria (Miller, 2002). In the context of GCMs, the most popular information criteria are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) (Gurka & Edwards, 2007; Singer & Willett, 2003; Steele, 2013). Among prediction criteria, the R^2 statistic and the concordance correlation coefficient (CCC) are the most commonly used (Steele, 2013; Wang & Schaalje, 2009). Each of these criteria have relative strengths and weaknesses, and no single criterion has been established as the 'best' when selecting GCMs (Steele, 2013).

Several previous simulation studies have been conducted to evaluate the performance of information criteria and prediction criteria (Gurka, 2006; Jaeger et al., 2017; Vallejo et al., 2011; Vallejo et al., 2014; Wu & West, 2013; Wu et al., 2009; Xu, 2003). However, further research is needed for the following two reasons.

First, there is a lack of evidence about the comparative performance of information versus prediction criteria when selecting GCMs. Nearly all previous studies have focused only on

information criteria (Gurka, 2006; Vallejo et al., 2011; Vallejo et al., 2014), or only on prediction criteria (Jaeger et al., 2017; Wu & West, 2013; Wu et al., 2009; Xu, 2003). Only one study (Wang & Schaalje, 2009) has simultaneously evaluated both sets of criteria, but was carried out in a limited setting. The study only considered balanced data settings (10 and 20 individuals; three and five repeated measurements per individual) and compared only two models, which differed for a single fixed effect (Wang & Schaalje, 2009).

Second, although it is now well accepted that AIC and BIC are valid for comparison of REML estimated GCMs that differ for random as well as fixed effects (Gurka, 2006; Vallejo et al., 2011; Vallejo et al., 2014; Wang & Schaalje, 2009), researchers still disagree over what should be the recommended penalty adjustments for these criteria (Gurka, 2006; Vallejo et al., 2011).

The first research chapter (Chapter 5) addressed the above issues. A simulation study was conducted to evaluate and compare the performances of information criteria and prediction criteria under different simulation conditions by varying the number of individuals (level 2 sample size), the number of repeated measurements per individual (level 1 sample size), model complexity (fixed and random effects) and whether the data were balanced or unbalanced. The aim was to select a criterion that was then used to compare the fit of linear and nonlinear GCMs applied to the jaw growth data (Chapter 6).

4.5.2. Research Question 2 (Chapter 6)

The linear and nonlinear GCMs are methods of choice for analysing longitudinal growth data. Different linear and nonlinear GCMs have been used for estimating the adolescent growth spurt parameters for height (Cole et al., 2014; Gasser et al., 1991; Goldstein, 1986a; Goldstein et al., 2018; Grimm et al., 2011; Simpkin et al., 2017; James M. Tanner, 1981). In contrast, a systematic review of the literature (Chapter 2, Section 2.2.3) shows that all previous studies analysing jaw growth data have focused exclusively on a linear GCM involving conventional polynomial functions for modelling growth trajectories (Buschang et al., 2013; Buschang et al., 1999; Buschang et al.,

1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996).

In Chapter 6, I compared three linear (CP, FP and RCS) and two nonlinear (SITAR and PB) GCMs in terms of their fit to the jaw growth data and ability in modelling covariate adjusted growth trajectories and adolescent growth spurt parameters. No previous study has compared these five GCMs for modelling physical growth data and estimating adolescent growth spurt parameters. Simpkin et al. (2017) compared the FP, SITAR and PB GCMs for modelling height growth trajectories and estimating timing of the peak height velocity (PHV). However, unlike the FP and SITAR GCMs, which were applied as mixed effect models, the authors fitted a fixed effect PB model to analyse data separately for each child. Grimm et al. (2011) applied a mixed effects PB GCM (including all five random effects) to height data and compared its performance with the CP GCM. No study has compared RCS GCM head-to-head with any of the other four GCMs (CP, FP, SITAR or PB) for modelling growth trajectories and estimating adolescent growth spurt parameters.

Estimating covariate-adjusted growth trajectories and adolescent growth spurt parameters necessitates that covariates should be included in estimation of distance and derivatives (velocity and acceleration). Unlike distance, which is a function itself (see Section 4.2), inclusion of covariates into derivatives estimation involves differentiation of covariate coefficients. Below I briefly summarise the current status of derivatives estimation for covariate adjusted GCMs and outline my plan of extending the earlier work on derivatives estimation for GCMs.

Of the five GCMs included in this thesis, derivatives estimation for conventional polynomial functions is simplest as it involves basic differentiation rules (see Section 4.3.2). Recently, Simpkin et al. (2018) described in detail first (velocity) and second (acceleration) derivatives estimation for CP GCMs. However, the authors focused on derivative estimation for unadjusted (without covariate) CP GCMs (Simpkin et al., 2018).

Unlike conventional polynomials, derivative estimation for fractional polynomial functions is more complicated as it requires advanced differentiation rules for logarithmic fractional polynomial terms (see Section 4.3.2). Building on the earlier work by Long and Ryoo (2010), Simpkin et al. (2017) recently described methodology for estimating first (velocity) and second (acceleration) derivatives for unadjusted (without covariate) FP GCMs.

Derivative estimation of truncated cubic spline functions requires the basic differentiation rules applied to the conventional polynomials (see Section 4.3.2). Buis (2009) introduced the Stata package 'postrcspline' to estimate distance and first derivative (velocity) for simple RCS GCMs (simple linear regression). While the 'postrcspline' package provides a covariate-adjusted distance curve, it excludes covariates when estimating the first derivative. Recently, Simpkin et al. (2018) described methodology to differentiate fixed and random effects estimated using the unadjusted truncated quartic spline basis functions. The authors did not include any covariates in differentiating truncated quartic spline basis functions.

The 'sitar' package (Cole, 2018), by default, provides a covariate-adjusted distance curve when covariates are part of the fitted model. However, it is unclear how covariates are handled when estimating derivatives. I am not aware of any study reporting on covariate-adjusted or covariate-specific derivatives for the SITAR GCM. Recently, Cao et al. (2018) fitted the SITAR GCM using the 'sitar' package to demonstrate their new approach to estimating APGV and PGV (see Section 4.3). The authors developed a function using R software to calculate APGV and PGV by using the distance and the first derivative (velocity) estimated by the 'sitar' package. They (Cao et al., 2018) explicitly stated that the SITAR GCM should be fitted *without covariates*.

The seminal paper by Preece and Baines (1978) provided a detailed description of derivatives estimation for the PB GCM. Recently, Sayers et al. (2013) found an error in the second derivative (acceleration) equation and published an erratum to correct it. Sayer et al. (2013) also

wrote a Stata package to fit fixed effect PB GCM ('pbreg'), estimate first and second derivatives ('pbpredict'), and calculate timing and intensity of the adolescent growth spurt ('pbpoint'). However, like the original work by Preece and Baines (1978), Sayers et al. (2013) focused on derivatives estimation for the unadjusted PB GCM fitted separately for each individual.

4.5.3. Research Question 3 (Chapter 7)

Studying sex differences in growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing and intensity) for upper and lower jaw length and total face height between normal skeletal jaw relationship (Class I) and skeletal malocclusions (Class II and Class III) is essential to correct skeletal discrepancies for males and females (Proffit et al., 2014).

Chapter 1 (Section 1.1) provides details on correction of skeletal discrepancies using growth modification. Briefly, growth modification is best achieved when the timing of treatment is synchronised with the timing of the adolescent growth spurt (Buschang et al., 2017; De Clerck & Proffit, 2015; Mitani, 1977; Proffit et al., 2014). However, the literature on the timing of growth modification procedures remains controversial (De Clerck & Proffit, 2015; Dolce et al., 2007; Grippaudo et al., 2013; Kallunki et al., 2018; Koretsi et al., 2014; Nowak et al., 2019; Perinetti & Contardo, 2017; Perinetti et al., 2015; Proffit, 2006; Thiruvenkatachari et al., 2013; Wheeler et al., 2006).

Among the various reasons for the lack of consensus on treatment timing, the major factor is inability to reliably assess skeletal maturity (Perinetti & Contardo, 2017). Clinicians and researchers have used different approaches such as assessment of sexual maturation, chronological age, dental age, and radiographic methods to determine the timing of the adolescent growth spurt for jaw growth (see Chapter 1, Section 1.1 for details). However, each of these approaches have limited reliability in assessing the skeletal maturation (see Chapter 1, Section 1.1).

As reported earlier in Chapter 1 (Section 1.1), assessment of APGV using longitudinal information pertaining to jaw growth velocity is more accurate than assessment of sexual maturation, chronological age, dental age and radiographic methods (Buschang et al., 2017; Flores-Mir et al., 2004). The PGV is a somatic biological maturity indicator reflecting the maximum skeletal growth velocity during adolescence (Hauspie et al., 2004). The APGV is considered the gold standard to objectively assess the skeletal maturity of jaw bones (Flores-Mir et al., 2004).

The linear and nonlinear GCMs form the methodological backbone of studying human physical growth such as stature height. It is considered important to evaluate and compare the performance of linear and nonlinear GCMs and select the model which fits best to the data and perform well in estimating growth trajectories and adolescent growth spurt parameters. However, all previous studies applying GCMs to jaw growth data focused exclusively on a linear GCM, the CP GCM. The Chapter 6 addresses this issue by comparing alternative linear (FP and RCS) and nonlinear (SITAR and PB) with the CP GCMs. Results showed that the RCS GCM fits best to the jaw growth data and perform well in estimating growth trajectories and adolescent growth spurt parameters.

Chapter 7 fitted the RCS GCM to the upper jaw length, lower jaw length and total face height data for males and female to estimate and class differences in (Class I vs Class II, Class I vs Class III, and Class II vs Class III) jaw growth trajectories and adolescent growth spurt parameters for males and females.

4.6. Software

The MLwiN software was used for fitting linear GCMs (CP, FP and RCS). The restricted iterative generalised least squares (RIGLS) method, which is equivalent to the REML method of estimation (Goldstein, 1989b), was used for estimation.

I used the 'runmlwin' command (Leckie & Charlton, 2013b) to call MLwiN from within Stata. Stata's 'fp generate' and 'mkspline' commands were used for generating fractional polynomials (FP GCM) and creating RCS design matrix (for RCS GCM). The 'mkspline' implements equally spaced quantiles for creating the RCS design matrix, as recommended by Harrell (2015).

While MLwiN is very efficient in fitting linear GCMs, it does not allow for estimating nonlinear GCMs (Jones, 2004). I fitted the SITAR GCM using the well-documented 'sitar' R package (Cole, 2017a). The 'sitar' package calls the 'nlme' R package (Pinheiro et al., 2017), which fits the NLME models. To fit the SITAR GCM using the truncated RCS design matrix instead of the B-spline design matrix (see Section 4.2.3), the 'rms' R package (Harrell Jr, 2017) was used. Like Stata's 'mkspline' command, it implements equally spaced quantiles as recommended by Harrell (2015). For the PB GCM, Stata's 'menl' command is used, which was introduced recently in June 2017 (version 15.0 onwards; (StataCorp, 2017a). The REML method was used for estimating the SITAR and PB GCMs. Both 'nlme' and 'menl' implement the linearization approach (Lindstrom & Bates, 1990; Pinheiro & Bates, 1995a) for model estimation (see Section 4.2.4 for details).

Full details on software used for answering each of the three research questions are provided in Appendix A, Section A.2 (Research Question 1), Appendix B, Section B.2 (Research Question 2), and Appendix C, Section C.1 (Research Question 3).

4.7. Summary

This chapter has reviewed linear (CP, FP, RCS) and nonlinear (SITAR and PB) GCMs, which are popular in modelling human physical growth data. No previous study has compared them for modelling growth data including jaw growth and height. This thesis aims to compare linear and nonlinear GCMs in terms of their fit to the jaw growth data and ability to model jaw growth trajectories and estimate derivatives based APGV and PGV.

To meet these research objectives, I conduct a simulation study (Chapter 5), and compare covariate-adjusted linear and nonlinear GCMs (Chapter 6). The methodological work (Chapter 5 and Chapter 6) allows selection of a model that best fits to the jaw growth data. The best-fitting model is then used to answer the clinical research questions relating to class (Class I, Class II and Class III) differences in jaw growth trajectories and the adolescent growth spurt parameters for males and females (Chapter 7).

CHAPTER 5. GROWTH CURVE MODEL SELECTION – A SIMULATION STUDY

5.1. Background and aim

Model selection is an important part of any statistical analysis. The information criteria and prediction criteria are commonly used for growth curve model (GCMs) selection. The widely used information criteria for GCM selection are Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The R^2 statistic (i.e., measure of variance explained) and the concordance correlation coefficient (CCC) are popular prediction criteria to evaluate and compare the fit of GCMs. Literature shows that no comprehensive study has compared the performance of information criteria and prediction criteria (see Chapter 4, Section 4.5.1 for details).

The aim of this chapter was to evaluate and compare the performance of information criteria and prediction criteria when specifying the CP GCMs. A simulation study was designed and conducted to compare these criteria across a range of data settings that varied the number of individuals, the number of repeated measurements per individual, model complexity, and whether the data were balanced or unbalanced.

5.2. Methodology

5.2.1. Growth curve model

The CP GCM was used to evaluate and compare performance of information criteria and prediction criteria. The CP GCM has been described in Chapter 4, Section 4.2.2. The objective was to select the model that captured the main features of the data, and to avoid under-fitting and over-fitting (Burnham & Anderson, 2002, pp. 29-37). Under-fitted models that ignore important structures in the data can result in bias, leading to misleading conclusions, and over-fitted models that estimate more parameters than necessary may give imprecise estimates (Burnham & Anderson, 2002, pp. 29-37). To facilitate the model selection process, a variety of fit criteria were proposed by different authors.

5.2.2. Information criteria and prediction criteria

Information criteria

Adding predictors such as polynomial terms to a GCM generally improves the fit of the model (i.e., how well a model fits a set of observations) but increases the model complexity (Miller, 2002, pp. 155-160). Information criteria attempt to resolve this problem by introducing a penalty for model complexity (Burnham & Anderson, 2002; Miller, 2002). The present study included AIC (Akaike, 1974) and BIC (Schwarz, 1978) as these are the two most commonly used information criteria for GCM selection (Singer & Willett, 2003, p. 121; Steele, 2013). Two different types of AIC were calculated, which differed in the number of parameters included in the penalty term, along with four different types of BIC, which differed in the number of parameters as well as the sample size included in the penalty term. See Table 5-1 for further details. The likelihood ratio test (LRT) was not considered as it can only be used to compare nested models (Hox, 2010; Singer & Willett, 2003).

Table 5-1. Fit criteria included in the simulation study.

Criteria	Formulae
Information criteria	
AICk	$-2logL + 2k$
AICq	$-2logL + 2q$
BICkN	$-2logL + k(\log(N))$
BICqN	$-2logL + q(\log(N))$
BICkL2	$-2logL + k(\log(J))$
BICqL2	$-2logL + q(\log(J))$
Prediction criteria	
$CCC_{VC[c]}$	$1 - \left(\frac{N}{N-k}\right)\left(1 - \left(\frac{\sum_{j=1}^{J}(y_{ij} - y_{ij(c)})^2}{\sum_{j=1}^{J}(y_{ij} - \bar{y})^2 + \sum_{j=1}^{J}(y_{ij} - \bar{y}_{(c)})^2 + N(\bar{y} - \bar{y}_{(c)})^2}\right)\right)$
$R^2_{VC[c]}$	$1 - \left(\frac{N}{N-k}\right)\left(1 - \left(\frac{\sum_{j=1}^{J}(y_{ij} - y_{ij(c)})^2}{\sum_{j=1}^{J}(y_{ij} - \bar{y})^2}\right)\right)$
$R^2_{NS[c]}$	$\frac{var\left(\sum_{r=0}^{p}\beta_r age_{ij}^r\right) + \sigma_{u0}^2}{var\left(\sum_{r=0}^{p}\beta_r age_{ij}^r\right) + \sigma_{u0}^2 + \sigma_e^2}$
$R^2_{JNS[c]}$	$\frac{var\left(\sum_{r=0}^{p}\beta_r age_{ij}^r\right) + trace\left(Z_{ij}\boldsymbol{\Omega_2}Z_{ij}'\right)/N}{var\left(\sum_{r=0}^{p}\beta_r age_{ij}^r\right) + trace\left(Z_{ij}\boldsymbol{\Omega_2}Z_{ij}'\right)/N + \sigma_e^2}$

Here J is the total number of individuals (level 2 units) and N is the total number of observations. k is the total number of parameters, $p + q$, where p is the number of fixed effect parameters and q represents the total number of random effect parameters, comprising q(q+ 1) /2 variance-covariance parameters plus a single residual variance parameter (see Section 3.1). The logL is the restricted log-likelihood estimate of the fitted model. y_{ij} is the observed outcome at i-th occasion for individual j and $\bar{y}$ is the grand mean of y_{ij} $\left(\bar{y} = \sum_{j=1}^{J} y_{ij}/N\right)$. $y_{ij(c)}$ is the model-predicted individual-specific (conditional) mean, and $\bar{y}_{(c)}$ is the grand mean of $y_{ij(c)}$, $\left(\bar{y}_{(c)} = \sum_{j=1}^{J} y_{ij(c)}/N\right)$. $var\left(\sum_{r=0}^{p}\beta_r age_{ij}^r\right)$ is the fixed effect predictors' variance, σ_{u0}^2 is the random intercept variance, σ_e^2 is the residual variance, $\boldsymbol{\Omega_2}$ is the variance covariance matrix, and $trace\left(Z_{ij}\boldsymbol{\Omega_2}Z_{ij}'\right)/N$ is the mean of random effects variance where Z_{ij} is the random effect design matrix.

Prediction criteria

Prediction criteria, which aim to minimise the mean squared error of the predicted outcome (Miller, 2002, pp. 111-164; Wang & Schaalje, 2009), are commonly used for evaluating the fit of GCMs (Gurka & Edwards, 2007). The predicted outcome can be marginal (based on fixed effects only) or conditional on the random effects (based on both the fixed effects and the random effects), and so marginal and conditional versions of prediction criteria have been developed (Wang & Schaalje, 2009). Only the conditional versions of prediction criteria were considered, because they have been strongly recommended for assessing the combined fit of marginal and conditional mean structures of GCMs (Vonesh et al., 1996; Wu & West, 2013; Wu et al., 2009).

The conditional prediction criteria included in this study are (see Table 5-1): adjusted (for the total number of parameters included in the model) versions of the conditional concordance correlation $CCC_{VC(c)}$ and $R^2_{VC(c)}$ proposed by Vonesh et al. (Vonesh & Chinchilli, 1996; Vonesh et al., 1996), random-intercept-based $R^2_{NS(c)}$ developed by Nakagawa and Schielzeth (2013) and its random-slope-based version, the $R^2_{JNS(c)}$ introduced by Johnson (2014). With the exception of $R^2_{NS(c)}$, all other prediction criteria include random slopes, where specified, in their formulae (see Table 5-1). Versions of $CCC_{VC(c)}$ and $R^2_{VC(c)}$ that adjust only for the number of variance-covariance parameters were not considered because an earlier simulation study (Wang & Schaalje, 2009) has shown that inclusion of the total number of parameters (k) or the number variance-covariance parameters (q) makes little difference in the performnaces of REML-based $CCC_{VC(c)}$ and $R^2_{VC(c)}$.

These four prediction criteria were the focus because (i) performances of $CCC_{VC(c)}$ and $R^2_{VC(c)}$ have not been assessed via any large-scale simulation study (Gurka & Edwards, 2007); (ii) the performance of recently introduced $R^2_{NS(c)}$ and $R^2_{JNS(c)}$ have not been directly compared, or compared with $CCC_{VC(c)}$ or $R^2_{VC(c)}$; and (iii) inclusion of both $R^2_{NS(c)}$ and $R^2_{JNS(c)}$ provides an

opportunity to evaluate the effect of inclusion and exclusion of random slopes on the performance of these criteria.

Other prediction criteria commonly used to quantify the level 1 or total variance explained by the GCMs were not considered, such as (i) the level 1 R^2 statistic presented by Raudenbush and Bryk (2002), as it "must be calculated using random intercept models" (LaHuis et al., 2014); (ii) the level 1 R^2 statistic by Snijders and Bosker (1994), because the authors themselves did not recommend inclusion of random slopes while calculating R^2 as it "will not lead to important changes in the estimate values" (Snijders & Bosker, 1994); (iii) the multilevel variance partitioning (MVP)–based R^2 statistic introduced by LaHuis et al., as they, like Snijders and Bosker (1994), did not recommend inclusion of random slopes when calculating R^2 statistic because it often results in negative values (LaHuis et al., 2014).

A total of 10 fit criteria were thus included in this study (Table 5-1): six information criteria (AICk, AICq, BICkN, BICqN, BICkL2 and BICqL2) and four prediction criteria ($CCC_{VC(c)}$, $R^2_{VC(c)}$, $R^2_{JNS(c)}$ and $R^2_{NS(c)}$). For all six information criteria, a lower value indicates a better fit, while for all four prediction criteria, a higher value indicates a better fit.

5.2.3. Simulation study

Data generation

Data simulation used three different data generation GCMs: (i) a linear-linear model, which included an intercept and a linear term age_{ij} for time in both the fixed and random parts of the model; (ii) a quadratic-linear model, which included an intercept as well as linear and quadratic terms age_{ij} and age^2_{ij} for time in the fixed part of the model, and an intercept and a linear term age_{ij} for time in the random part of the model; and (iii) a quadratic-quadratic model, which included an intercept and linear and quadratic terms age_{ij} and age^2_{ij} for time in both the fixed and random

parts of the model. Population parameters (true values) were obtained by fitting each of three data generation models to the upper jaw length (COPAD) growth data for females.

Balanced and unbalanced data were simulated under four different sample size conditions: (i) 100 individuals with five repeated measurements per individual, (ii) 1,000 individuals with five repeated measurements per individual, (iii) 100 individuals with 20 repeated measurements per individual, and (iv) 1,000 individuals with 20 repeated measurements per individual. To simulate unbalanced data, balanced data was first simulated, then 20 percent of the total number of repeated measurements were randomly dropped. The average number of repeated measurements per individual in the unbalanced data were therefore 16 for the large level 1 sample size, and four for the small level 1 sample size (see Appendix A, Table A-2). For each of the three data generation models, 500 independent simulated datasets were generated using Stata version 14.2 (StataCorp, 2015b).

Data analysis

Each of the three sets of simulated datasets were analysed by the following six data analysis GCMs: (i) linear-linear, (ii) quadratic-linear, (iii) quadratic-quadratic, (iv) quadratic-intercept, (v) cubic-intercept, and (vi) cubic-linear. The first three models mirror the data generation models, and the other three vary as follows.

The quadratic-intercept model includes an intercept and linear and quadratic terms age_{ij} and age_{ij}^2 for time in the fixed part of the model, and only an intercept in the random part.

The cubic-intercept model includes an intercept and linear, quadratic and cubic terms age_{ij}, age_{ij}^2 and age_{ij}^3 for time in the fixed part of the model, and only an intercept in the random part.

The cubic-linear model included an intercept and linear, quadratic and cubic terms age_{ij}, age_{ij}^2 and age_{ij}^3 for time in the fixed part of the model, and an intercept and a linear term age_{ij} for time in the random part.

Thus, out of six data analysis models, one exactly matches the model used to generate each dataset, and the remaining five are either over- or under-specified with respect to the fixed and/or random effect(s). For example, under the linear-linear data generation model, the linear-linear data analysis model is the true model, whereas the other five models are over- or under-specified. The cubic-intercept model is over-specified in terms of the fixed part but is under-specified in terms of the random part. A complete comparison of the data generation and analysis models is provided in Table 5-2.

Table 5-2. Comparison of data generation and data analysis models.

Data generation models	Data analysis models					
	Linear-linear [k=6(2+4)] *	Quadratic-linear [k=7(3+4)] *	Quadratic-quadratic [k=10(3+7)] *	Quadratic-intercept [k=5(3+2)] *	Cubic-intercept [k=6(4+2)] *	Cubic-linear [k=8(4+4)] *
Linear-linear [k=6(2+4)] *	True	Over-specified fixed effects (+)	Over-specified fixed effects (+); over-specified random effects (+)	Over-specified fixed effects (+); under-specified random effects (-)	Over-specified fixed effects (++); under-specified random effects (-)	Over-specified fixed effects (++)
Quadratic-linear [k=7(3+4)] *	Under-specified fixed effects (-)	True	Over-specified random effects (+)	Under-specified random effects (- -)	Over-specified fixed effects (+); under-specified random effects (-)	Over-specified fixed effects (+)
Quadratic-quadratic [k=10(3+7)] *	Under-specified fixed effects (-); under-specified random effects (-)	Under-specified random effects (-)	True	Under-specified random effects (- -)	Over-specified fixed effects (+); under-specified random effects (- -)	Over-specified fixed effects (+); under-specified random effects (-)

The + sign shows over-specified fixed and/or random effects, and relative extent of over-specification is denoted by + and ++ signs. The - sign denotes under-specified fixed and/or random effects, and relative extent of under-specification is denoted by - and - - signs.
* Number of parameters: k is total number of parameters (fixed parameters [p] + variance-covariance parameters [q] + a residual variance parameter)

Evaluation of information and prediction criteria

After fitting an analysis model to a simulated dataset, the fit criteria were calculated based on the formulae provided in Table 5-1. The performance of each criterion in terms of its ability to select the true model and reject over- and under-specified models was assessed based on the proportion of the 500 simulated datasets in which the fit criterion selected the true model as the best-fitting model.

5.2.4. Software

To calculate different information criteria and the prediction criteria included in the simulation study, I wrote a new Stata program (‘gcmfit’) was written, as no suitable existing package or program was available (see Chapter 4, Section 4.5.3). Details on the ‘gcmfit’ are provided in Appendix A (Section A.2). To calculate the bias and Monte Carlo standard error for each estimated parameter, I used Stata command ‘simsum’ written by White (2010).

5.3. Results

5.3.1. Balanced data

Linear-linear data generation model

Table 5-3 reports the performance results for the six data analysis models fitted to the data simulated using the linear-linear data generation model. Out of the six data analysis models fitted to the data the linear-linear model was the true model. The information criteria based on the total number of parameters (AICk, BICkN and BICkL2) performed better than those calculated using only the number of variance-covariance parameters (AICq, BICqN and BICqL2).

When level 1 sample size was large (irrespective of the level 2 sample size), BICkN (99.6–100%) and BICkL2 (96.8–99.4%) performed better than AICk (74.2–76.0%) in selecting the true model. However, when both level 2 and level 1 sample sizes were small, AICk (58.8%) performed better than BICkN (35.0%) and BICkL2 (47.2%). Instead of selecting the true model, both BICkN and BICkL2 selected the quadratic-intercept model (BICkN 63.6%; BICkL2 50.2%), which is an over-specified model with respect to fixed effects and under-specified with respect to random effects.

Regardless of the level 2 and level 1 sample sizes, the prediction criteria performed poorly in selecting the true linear-linear model (success rate less than 30.6%). While random slope-based criteria ($R^2_{VC(c)}$, $CCC_{VC(c)}$ and $R^2_{JNS(c)}$) selected the quadratic-quadratic model (42.8–51.4%), a model with over-specified fixed and random effects, the random intercept-based criterion ($R^2_{NS(c)}$) selected the cubic-linear model (45.2–48.8%) with over-specified fixed effects, or the cubic-intercept model (47.2–52.2%) with over-specified fixed effects and under-specified random effects.

Table 5-3. Results for balanced data generated using the linear-linear model.

Level 2 [a]	Level 1 [b]	Data analysis models	Information criteria						Prediction criteria			
			AICk	AICq	BICkN	BICqN	BICkL2	BICqL2	$CCC_{VC[c]}$	$R^2_{VC[c]}$	$R^2_{JNS[c]}$	$R^2_{NS[c]}$
			(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
1000	20	Linear-linear	**74.2**	*0*	**100**	*0*	**99.4**	*0*	25.6	26.2	27.0	*0*
1000	20	Quadratic-linear	11.4	1.8	*0*	1.8	0.6	2.4	10.2	10.8	10.0	1.8
1000	20	Quadratic-quadratic	6	9	*0*	*0*	*0*	*0*	**50.4**	**49.8**	**51.4**	2.2
1000	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	2.6
1000	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**47.2**
1000	20	Cubic-linear	8.4	**89.2**	*0*	**98.2**	*0*	**97.6**	13.8	13.2	11.6	46.2
1000	5	Linear-linear	**76**	*0*	**100**	*0*	**99.4**	*0*	23.8	24.0	26.2	*0*
1000	5	Quadratic-linear	9.6	1.8	*0*	1.8	0.6	1.8	11.0	11.0	10.0	0.4
1000	5	Quadratic-quadratic	6.8	8.8	*0*	*0*	*0*	*0*	**47.8**	**47.4**	**48.4**	1
1000	5	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	0.2
1000	5	Cubic-intercept	*0*	*0*	*0*	1.6	*0*	0.6	*0*	*0*	*0*	**49.6**
1000	5	Cubic-linear	7.6	**89.4**	*0*	**96.6**	*0*	**97.6**	17.4	17.6	15.4	48.8
100	20	Linear-linear	**74.4**	*0*	**99.6**	*0*	**96.8**	0.2	29.6	30.6	27.8	*0*
100	20	Quadratic-linear	13	1.8	0.4	2.2	2.4	2.0	12.6	12.6	11.2	0.6
100	20	Quadratic-quadratic	5.8	7.8	*0*	*0*	*0*	*0*	**42.8**	**41.6**	**47.2**	0.8
100	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	1.2
100	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**52.2**
100	20	Cubic-linear	6.8	**90.4**	*0*	**97.8**	0.8	**97.8**	15.0	15.2	13.8	45.2
100	5	Linear-linear	**58.8**	*0.2*	35.0	*0*	47.2	*0*	24.2	25.4	26.2	*0*
100	5	Quadratic-linear	7.0	1.4	0.2	*0*	0.4	*0*	14.2	14.0	11.6	*0*
100	5	Quadratic-quadratic	*4.0*	6.0	*0*	*0*	*0*	*0*	**43.8**	**43.0**	**46.4**	0.2
100	5	Quadratic-intercept	20.6	2.4	**63.6**	5.2	**50.2**	5.2	2.0	2.0	1.6	0.2
100	5	Cubic-intercept	4.4	32.4	1.2	**83.0**	2.2	**72.8**	*0.8*	*0.8*	*0.6*	**51.0**
100	5	Cubic-linear	5.2	**57.6**	*0*	11.8	*0*	22.0	15.0	14.8	13.6	48.6

Note: Highest and lowest values are highlighted using bold and italics fonts, respectively.
[a] Level-2: Number of individuals; [b] Level-1: Average number of repeated measurements per individual.

Quadratic-linear data generation model

Table 5-4 shows the performance results for the six data analysis models fitted to the data simulated from the quadratic-linear data generation model. Out of a total six data analysis models fitted to the data, the quadratic-linear model was the true model. Similar to the linear-linear data generation model, the information criteria based on the total number of parameters (AICk, BICkN and BICkL2) performed substantially better than their counterparts based on the number of variance-covariance parameters (AICq, BICqN and BICqL2) in selecting the true model. Compared to AICk (73.6–74.6%), BICkN (99.4–99.8%) and BICkL2 (96.6–99.0%) performed better when level 1 sample size was large.

When level 1 sample size was small, the success rates of both BICkN and BICkL2 in selecting the true quadratic-linear model dropped dramatically (BICkN 0.0–6.6%; BICkL2 0.8–11.8%), regardless of the level 2 sample size. Instead of selecting the true model, both BICkN (93.2%) and BICkL2 (88%) selected the linear-linear model with under-specified fixed effects (BICkN 93.2%; BICkL2 88.0%) when the level 2 sample was large, and the quadratic-intercept model (BICkN 67.6%; BICkL2 55.8%) with under-specified random effects when the level 2 sample was small. When both level 1 and level 2 sample sizes were small, the true model selection rate of AICk (8.8%) also dropped, favouring the selection of the linear-linear model with under-specified fixed effects (50.8%).

As with the results reported earlier for the linear-linear data generation model, the prediction criteria performed poorly in selecting the true quadratic-linear model (success rate less than 35.2%), irrespective of the level 2 and level 1 sample sizes. The random slope-based prediction criteria ($R^2_{VC(c)}$, $CCC_{VC(c)}$, and $R^2_{JNS(c)}$) selected the quadratic-quadratic model (42.2–50.8%) with over-specified random effects. The random intercept-based criterion $R^2_{NS(c)}$ selected the cubic-linear model (46.2–53.6%) or the cubic-intercept model (44.4–53.4%).

Table 5-4. Results for balanced data generated using the quadratic-linear model.

Level 2 [a]	Level 1 [b]	Data analysis models	Information criteria						Prediction criteria			
			AICk	AICq	BICkN	BICqN	BICkL2	BICqL2	$CCC_{VC[c]}$	$R^2_{VC[c]}$	$R^2_{JNS[c]}$	$R^2_{NS[c]}$
			(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
1000	20	Linear-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
1000	20	Quadratic-linear	**74.6**	2.6	**99.8**	2.6	**99.0**	2.6	35.2	35.0	35.2	1.4
1000	20	Quadratic-quadratic	10.4	7.4	*0*	*0*	*0*	*0*	**46.2**	**46.0**	**47.8**	1.8
1000	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	2.0
1000	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**48.4**
1000	20	Cubic-linear	15.0	**90.0**	0.2	**97.4**	1	**97.4**	18.6	19.0	17.0	46.4
1000	5	Linear-linear	38.4	*0*	**93.2**	*0*	**88.0**	*0*	9.4	9.6	10.4	*0*
1000	5	Quadratic-linear	**40.6**	1.6	6.6	2.4	11.8	2.2	25	25.2	24.6	0.8
1000	5	Quadratic-quadratic	9.4	8.2	*0*	*0*	*0*	*0*	**50.2**	**50.0**	**51.6**	0.8
1000	5	Quadratic-intercept	*0*	*0*	0.2	*0*	0.2	*0*	*0*	*0*	*0*	0.4
1000	5	Cubic-intercept	*0*	*0*	*0*	1	*0*	0.2	*0*	*0*	*0*	44.4
1000	5	Cubic-linear	11.6	**90.2**	*0*	**96.6**	*0*	**97.6**	15.4	15.2	13.4	**53.6**
100	20	Linear-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
100	20	Quadratic-linear	**73.6**	2	**99.4**	2.8	**96.6**	2.2	34.2	35.2	30.2	1.0
100	20	Quadratic-quadratic	11.4	9	*0*	*0*	*0*	*0*	**44.0**	**42.8**	**50.8**	1.2
100	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	1.2
100	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	45.0
100	20	Cubic-linear	15.0	**89.0**	0.6	**97.2**	3.4	**97.8**	21.8	22.0	19.0	**51.6**
100	5	Linear-linear	**50.8**	*0*	32.2	*0*	43.0	*0*	23.0	23.8	24.8	*0*
100	5	Quadratic-linear	8.8	2.6	*0*	0.4	0.8	0.8	16.4	15.8	14.4	*0*
100	5	Quadratic-quadratic	8.0	8.8	*0*	*0*	*0*	0.2	**43.0**	**42.2**	**46.0**	0.2
100	5	Quadratic-intercept	24.2	2.4	**67.6**	5.6	**55.8**	4.8	2.0	2.2	1.8	0.2
100	5	Cubic-intercept	5.0	36.6	0.2	**82.0**	0.4	**73.0**	*1.0*	*1.2*	*0.8*	**53.4**
100	5	Cubic-linear	*3.2*	**49.6**	*0*	12.0	*0*	21.2	14.6	14.8	12.2	46.2

Note: Highest and lowest values are highlighted using bold and italics fonts, respectively.
[a] Level-2: Number of individuals; [b] Level-1: Average number of repeated measurements per individual.

Quadratic-quadratic data generation model

Table 5-5 presents the performance results for the six data analysis models fitted to the data simulated using the quadratic-quadratic data generation model. Out of a total six data analysis models fitted to the data, the quadratic-quadratic model was the true model. When level 1 sample size was large (irrespective of the level 2 sample size), all six information criteria calculated using either the total number of parameters (AICk, BICkN and BICkL2) or the number of variance-covariance parameters (AICq, BICqN and BICqL2) performed alike in selecting the true quadratic-model (100% success rate in selecting the true model).

However, when level 1 sample size was small, all four versions of the BIC (BICkN, BICqN, BICkL2 and BICqL2) performed poorly in selecting the true model (0.0–3.8%), regardless of the level 2 sample size. In contrast, when level 1 sample size was small, but the level 2 sample size was large, both versions of the AIC performed substantially better than any of the four versions of the BIC in selecting the true model (AICk 56.0%; AICq 54.6%). When both level 2 and level 1 sample sizes were small, the success rates of AICk (10.4%) and AICq (11.4%) in selecting the true model also dropped but were still higher than any of the four versions of the BIC.

All random slope-based prediction criteria ($R^2_{VC(c)}$, $CCC_{VC(c)}$, $R^2_{JNS(c)}$) selected the true quadratic-quadratic model (100%) when the level 1 sample size was large. When level 1 sample size was small (regardless of the level 2 sample size), the success rates of random slope-based prediction criteria in selecting the true model dropped considerably (45.0–65.8%). The random intercept-based $R^2_{NS(c)}$ almost always failed to select the true model under any of the sample size conditions (0.0–2.0%).

Table 5-5. Results for balanced data generated using the quadratic-quadratic model.

evel 2 [a]	Level 1 [b]	Data analysis models	Information criteria						Prediction criteria			
			AICk	AICq	BICkN	BICqN	BICkL2	BICqL2	$CCC_{VC[c]}$	$R^2_{VC[c]}$	$R^2_{JNS[c]}$	$R^2_{NS[c]}$
			(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
1000	20	Linear-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
1000	20	Quadratic-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	3
1000	20	Quadratic-quadratic	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	2
1000	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	3.8
1000	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	43.6
1000	20	Cubic-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**47.2**
1000	5	Linear-linear	18.4	*0*	**91.2**	*0*	**84.6**	*0*	8.2	8.6	7.6	*0*
1000	5	Quadratic-linear	20.2	0.6	8.4	1.4	13.6	1.8	18	18.2	17.6	1.4
1000	5	Quadratic-quadratic	**56.0**	**54.6**	0.4	1.2	1.8	3.8	**61.8**	**61.0**	**65.8**	0.2
1000	5	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	1.6
1000	5	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	46.6
1000	5	Cubic-linear	5.4	44.8	*0*	**97.4**	*0*	**94.4**	12.0	12.2	9.0	**50.2**
100	20	Linear-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
100	20	Quadratic-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	1.0
100	20	Quadratic-quadratic	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	1.0
100	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	1.6
100	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**49.4**
100	20	Cubic-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	47.0
100	5	Linear-linear	**64.4**	*0.6*	**73.0**	*0*	**78.4**	*0.2*	22.0	22.8	24.8	*0*
100	5	Quadratic-linear	11.8	3.2	1.0	2.6	2.2	3.2	13.4	13.8	12.6	0.2
100	5	Quadratic-quadratic	10.4	11.4	*0*	0.2	0.6	0.8	**45.6**	**45.0**	**47.8**	*0*
100	5	Quadratic-intercept	4.4	*0.6*	25.8	3.2	17.2	2.2	0.4	0.4	0.2	0.2
100	5	Cubic-intercept	*0.4*	8.4	*0*	**48.0**	*0.4*	34.0	*0*	*0*	*0*	47.4
100	5	Cubic-linear	8.6	**75.8**	0.2	46.0	1.2	**59.6**	18.6	18.0	14.6	**52.2**

Note: Highest and lowest values are highlighted using bold and italics fonts, respectively.
[a] Level-2: Number of individuals; [b] Level-1: Average number of repeated measurements per individual.

5.3.2. Unbalanced data

In contrast to the balanced data setting where all six data analysis models converged on every dataset, the model convergence rates were affected by the sample size and model complexity in the unbalanced data settings. When both level 2 and level 1 sample sizes were small, all data analysis models except the quadratic-intercept and cubic-intercept models failed to show 100% convergence rates. The convergence rate of the linear-linear, quadratic-linear, quadratic-quadratic and cubic-linear models varied from 83.2–98.8%. This was perhaps because randomly dropping 20% of the repeated measurements for unbalanced data resulted in insufficient level 1 sample size for the estimation of random effect parameters of GCMs involving random slopes. This may also explain an observed greater effect of unbalanced data on the convergence rate of the quadratic-quadratic model (83.2%), which had two random slopes (giving six random effect parameters), compared with the convergence rates of the linear-linear, quadratic-linear and cubic-linear models (97–98.8%), which had one random slope (giving three random effect parameters).

The performance results of the fit criteria under the unbalanced data settings are provided in Appendix A for linear-linear (Table A-3), quadratic-linear (Table A-4) and quadratic-quadratic (Table A-5) true data generation models. Even though unbalanced data affected the performance of information criteria and prediction criteria in terms of reducing their absolute values, their model selection patterns were not influenced under any of the simulation conditions studied.

As with balanced data, the AIC and BIC calculated using the total number of parameters (AICk, BICkN and BICkL2) performed equally well as, or better than, those calculated using the number of variance-covariance parameters (AICq, BICqN and BICqL2). Also as with balanced data, the AIC and BIC calculated using the total number of parameters (AICk, BICkN and BICkL2) performed better than the random slope-based prediction criteria ($R^2_{VC(c)}$, $CCC_{VC(c)}$, $R^2_{JNS(c)}$). The random intercept-based criterion ($R^2_{NS(c)}$) once again performed worst amongst all the fit criteria studied.

5.4. Discussion

This study evaluated and compared the model selection performances of commonly used information and prediction criteria in fitting GCMs to continuous repeated measures growth data. It focused on fitting models by REML as this, in contrast to ML, produces unbiased estimates of the variance-covariance components.

Previous studies evaluating the performance of REML-based AIC and BIC have included either the total number of parameters (Gurka, 2006; Vallejo et al., 2011) or the number of variance-covariance parameters (Vallejo et al., 2008; Vallejo et al., 2014) in their penalty terms. Disagreement also exists in the literature as to which sample size should be included into the penalty term of the BIC: the number of repeated measures across all individuals, or just the smaller number of individuals (Vallejo et al., 2011).

The simulation results of the present study show that the REML-based AIC and BIC calculated using the total number of model parameters perform at least as well as, and often better than, their counterparts calculated using only the number of variance-covariance parameters. In contrast, the simulation findings show that the choice between the two possible sample sizes in the BIC penalty has little influence on its performance. This is perhaps because as sample size enters the penalty term of BIC as a log, a 10-fold increase in the sample size results in a very mild increase (by 1 log unit) in the strength of penalty adjustment factor, which multiplies with the number of parameters (k for BICkN, and q for BICqN). Arguably, the total number of parameters should therefore be included in the penalty terms of both the AIC and BIC (as is implemented in Stata, but not SAS), but choice of sample size matters far less. These findings represent a significant advance for understanding the performance of REML-based AIC and BIC in GCM model selection.

The simulation results also show that the sample size and the complexity of the variance-covariance structure influences the performance of the REML-based AIC and BIC in similar ways,

regardless of whether the total number of parameters or the number of variance-covariance parameters is used in calculating these two information criteria. There are two explanations for this. First, and in support of earlier studies (Vallejo et al., 2011; Wang & Schaalje, 2009), the findings of the present study show that an increase in the level 1 or level 2 sample size enhances the performance of the REML-based AIC and BIC, but both criteria perform much better when the level 1 sample size is large. Second, the results support studies (Gurka, 2006; Vallejo et al., 2011; Vallejo et al., 2014) showing that the AIC performs better than the BIC when GCMs involve more complex variance-covariance structures.

In contrast to the AIC and BIC, the random slope-based prediction criteria evaluated in this study did not provide protection against over-fitting, as they always selected the most complex model, regardless of true model or sample size. These results support the findings of an earlier study (Wu & West, 2013), which reported that the conditional prediction criteria are less sensitive to the sample size and perform well in correctly identifying the best-fitting model when compared against simpler under-specified models. One potential explanation for why prediction criteria such as $R^2_{JNS(c)}$ and $R^2_{NS(c)}$ result in over-fitting is that they lack any adjustment for the model complexity. However, surprisingly, the adjusted $R^2_{VC(c)}$ and $CCC_{VC(c)}$ included in this study (which do adjust for model complexity based on the total number of model parameters) also failed to select the true models, again favouring more complex models. This could be due to the small difference in the number of parameters among the competing models considered and a relatively large sample size (total number of observations), which minimised the strength of the adjustment factor included in the $R^2_{VC(c)}$ and $CCC_{VC(c)}$. Lastly, the random intercept-based statistic considered ($R^2_{NS(c)}$) performed worse than its random slope-based counterpart $R^2_{JNS(c)}$. Johnson's (Johnson, 2014) recommendation to favour the latter in applications of GCMs to longitudinal data is therefore supported.

In summary, considering that the jaw growth data analysed in this thesis are comprised of small number of level-2 units (128 males and 139 females) and a small average number of level-1 units per level-2 unit (median eight repeated measurements per individual), the AIC is chosen as the criterion of choice for model selection. Results (see Section 5.3) show that under the conditions used, the AIC performs better than the BIC. The prediction criteria performed substantially worse than the information criteria (AIC and BIC) under all simulation conditions.

In addition to the above-mentioned reasons, another factor which favoured the selection of the AIC over the BIC is the fact that the data analysed in this thesis are biological data and full information on potential covariates is not available. The core strength of the BIC is that it selects the 'true' model and its effectiveness in selecting the true model increases as the sample size increases. However, the concept of a true model is only useful if the full set of explanatory variables is available for a given dataset (Burnham & Anderson, 2004). When this condition is not satisfied, which is unlikely to ever be true in biological applications, it is recommended to instead use the AIC (Aho et al., 2014; Burnham & Anderson, 2004). Since jaw growth is influenced by a multitude of factors (see Chapter 2, Section 2.1), it is virtually impossible to include all such variables while analysing jaw growth data. In particular, the jaw growth data analysed in this thesis provides information on very few individual characteristics or other potentially relevant variables (see Chapter 3, Section 3.4).

5.5. Limitations

The present study has two important limitations that could have affected the results. First, when evaluating the performances of fit criteria in terms of their ability to select the true model and reject over- and under-specified models, the magnitude of differences in the calculated criteria was not considered. For instance, when comparing the fit of two competing models using the AIC or BIC, a difference of 0–2 points is often considered "weak" evidence in favour of the model with the smaller value (Bell et al., 2013). However, such judgements were not considered when assessing the performance of the AIC and BIC in selecting the true model. The model with the smaller value was favoured irrespective of how small the difference was.

Second, only the CP GCM was considered in the simulations. There are many other approaches for modelling nonlinear growth trajectories, including fractional polynomial and splines (Tilling et al., 2014). Fit evaluation of nonlinear GCMs was also not considered. These issues are further discussed in Chapter 8 (Section 8.3).

Thirdly, although in this thesis the AIC was universally preferred over the BIC for comparing GCMs applied to the jaw growth data (see Section 5.4 for reasons), this might not always be the case in other sample size settings. This is because unlike the AIC which does not take into account the sample size, the BIC penalizes the likelihood based on both the sample size as well as the number of parameters. As the sample size increases, the ability of BIC to select the 'true' model also improves. However, as discussed above (Section 5.4), it is important to consider the data at hand when deciding whether BIC or AIC should be used for model selection even if the sample size is large.

5.6. Conclusion

My study is the first to show that entering the total number of parameters rather than only the number of variance-covariance parameters in the penalty terms is necessary when calculating REML-based AIC and BIC. For BIC, either the total number of repeated measurements or the number of individuals can be entered in the penalty term; the choice makes little difference.

When the number of repeated measurements per individual is high (e.g., 20), the BIC is recommended for model selection as it performs better than the AIC. However, when the number of repeated measurements per individual is low (e.g., five), the AIC is recommended for GCM selection, as it is more robust in small sample size settings and additionally performs better than the BIC when GCMs involve complex variance-covariance structures (multiple random effect), as is often the case in longitudinal growth studies. As the prediction criteria resulted in over-fitting, their use in selecting GCMs for longitudinal data is not recommended.

For the remainder of this thesis, I will use the AIC to evaluate and compare the model fit. This is because the jaw growth data is comprised of relatively a small sample size (median number of repeated measurements per individual is eight) and GCMs applied to data involve complex variance-covariance structures.

CHAPTER 6. COMPARISON OF LINEAR AND NONLINEAR GROWTH CURVE MODELS FOR ESTIMATING TIMING AND INTENSITY OF THE ADOLESCENT GROWTH SPURT

6.1. Background and aim

Estimation of growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing and intensity) requires fitting an appropriate growth curve model (GCM) to the longitudinal growth data. This chapter applies three linear and two nonlinear GCMs to the jaw growth data and compares their fit to the data and ability in modelling covariate adjusted growth trajectories and adolescent growth spurt parameters. The three linear GCMs are conventional polynomial (CP), fractional polynomial (FP) and the restricted cubic spline (RCS). The two nonlinear GCMs are the superimposition by translation and rotation (SITAR) and the Preece-Baines (PB).

No previous study has compared these five GCMs for modelling physical growth data and estimating adolescent growth spurt parameters (see Chapter 4, Section 4.5.2 for details). A direct comparison of all five GCMs would help in understanding their relative strengths and weaknesses in modelling jaw growth data. This would then inform the choice of the GCM to answer clinical research questions relating to estimating class differences in jaw growth trajectories (distance and derivatives) and adolescent growth spurt parameters (timing and intensity).

This chapter therefore aimed to compare covariate-adjusted CP, FP, RCS, SITAR and PB GCMs in terms of their ability to model jaw growth trajectories and to estimate APGV and PGV using derivatives adjusted for covariate effects.

6.2. Methodology

6.2.1. Data

This chapter analysed longitudinal jaw growth data available from the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection. Data details are provided in Chapter 3 (see Section 3.3). Briefly, the data comprised of repeated measurements of the upper and lower jaw length and total face height of 128 males (mean age 11.67 years, standard deviation 2.92) and 139 females (mean age 11.60 years, standard deviation 2.88) between seven and 18 years of age. Data were pooled from following eight longitudinal growth studies: (i) Bolton-Brush growth study, (ii) Burlington growth study, (iii) Denver growth study, (iv) Forsyth growth study, (v) Iowa growth study, (vi) Mathews growth study, (vii) Michigan growth study, and (viii) Oregon growth study.

6.2.2. Data analysis plan

Handling incomplete and unbalanced jaw growth data

Longitudinal growth data is typically characterised by missing responses and irregularly spaced measurement occasions (Curran et al., 2010; Gibbons et al., 2010; Ware, 1985). The jaw growth data analysed in this thesis is unbalanced but almost complete, as the missing responses for any of the three outcomes (upper and lower jaw length and total face height) are less than 4% for both males and females (see Chapter 3, Section 3.3 for details). No data are missing for covariates (class of skeletal jaw relationship).

The GCMs can easily handle the unbalanced data and require no changes to the model fitting (Fitzmaurice et al., 2008; Goldstein, 2011b; Laird & Ware, 1982). Since missing data is less than 4% for any of the three outcomes, imputation of missing jaw measurements was unnecessary. The

GCMs handle missing data by ML under the MAR assumption (Allison, 2012; Fitzmaurice et al., 2008; Goldstein, 2011b; Laird & Ware, 1982).

Adjusting for potential study effects (growth studies)

As different growth studies contribute to the AAOF Craniofacial Growth Legacy Collection database, the data analysed follows a three-level hierarchical structure where repeated measurements (level 1) on individuals (level 2) are nested in growth studies (level 3). An ideal approach for incorporating study variation into analysis is to fit three-level linear and nonlinear GCMs and allow growth parameters to vary randomly across studies via the inclusion of study as random effects.

The jaw growth data comprised of repeated growth measurements on individuals who participated in eight different growth studies (see Chapter 3, Section 3.3). As the number of studies is small (eight studies) and some provide data for only a few individuals, it was difficult to estimate between study variance with any degree of precision. In such circumstances, it is recommended to fit two-level models and include the highest-level units into the fixed part of the model by using dummy variables (McNeish & Wentzel, 2017; Snijders & Bosker, 2012). A recent simulation study has shown that this is the best approach to account for the clustering when level 3 sample size is small (McNeish & Wentzel, 2017). Covariates can be added at level 1 or level 2 as for a standard two-level model (McNeish & Wentzel, 2017).

The present study followed this approach of fitting a two-level model and including study-specific dummy variables into the fixed part of the linear and nonlinear GCMs. To allow growth trajectories to vary across studies, cross-level interactions were included for linear GCMs and each growth parameter adjusted for nonlinear GCMs.

Inclusion of time-invariant covariate (class of skeletal jaw relationship)

The skeletal jaw relationship is classified into Class I, Class II or Class III, reflecting a normal skeletal jaw relationship (Class I) or skeletal malocclusion (Class II and Class III). As upper and lower jaw length and total face height vary between Class I, Class II and Class III skeletal jaw relationships (Proffit et al., 2014), class was included as a covariate in model fitting and estimation of growth trajectories and adolescent growth spurt parameters. In the AAOF Craniofacial Growth Legacy Collection database, class is recorded as a nominal time-invariant variable. Class-specific dummy variables were used in the fixed part of the linear and nonlinear GCMs. To allow growth trajectories to vary across studies, cross-level interactions were included for linear GCMs, and each growth parameter adjusted for nonlinear GCMs. Since no data were available for individuals with Class II and Class III skeletal malocclusions in some studies (see Chapter 3, Table 3-4), no interactions were included between study and class variables. From a modelling perspective, inclusion of class as a covariate using dummy variables and adjusting for study effects are the same.

Model fitting by including all possible random effects

During adolescence, each individual follows a distinct growth pattern characterised by unique phases of growth acceleration and deceleration (Bose, 2007; Cameron & Bogin, 2012; Hauspie et al., 2004; Rauber, 1990). Empirical evidence shows a strong between-individual variability in the timing and intensity of the adolescent growth spurt for jaw growth in both sexes.

Fully accounting for the between-individual variability in growth is recommended when analysing skeletal growth data during adolescence (Grimm et al., 2011). Fitting a statistical model to the skeletal growth data by constraining between-individual variation in growth to zero (i.e., excluding random effects) fails to fully capture the between-individual differences in growth changes (Collins, 2006; Grimm et al., 2011). As evidence suggest that individuals grow at different rates and show variability in the timing and intensity of the adolescent growth spurt for jaw growth,

a priori decision was made to fit each model with all possible individual-specific random effects, which is often called a 'maximal' random-effect model (Barr et al., 2013; Eager & Roy, 2017).

Model assumptions

As repeated jaw growth measurements were taken at least six months apart for all individuals (median age gap 1 year, IQR 0.15 years; see Chapter 4 for details), independence of level 1 residuals (no autocorrelation) was assumed. For growth data such as height, it is not until measurements are approximately two months apart that serial correlation effects are found significant (Browne & Goldstein, 2010; Goldstein et al., 2018). It was assumed reasonable to expect a similar trend for jaw growth.

Linear and nonlinear GCMs were fitted under the assumption that level 2 random effects and level 1 residuals are normally distributed. Level 1 residuals were also assumed to be homoscedastic with constant level 1 variance. These assumptions are assessed in Section 6.2.4.

6.2.3. Covariate-adjusted linear and nonlinear growth curve models

Covariate (class) adjusted two-level linear and nonlinear GCMs were fitted where study level effects were controlled by using dummy variables. A detailed description of covariate-adjusted CP GCM is provided below as an illustrative example. Details on covariate-adjusted FP, RCS, SITAR and PB GCMs are provided in Appendix B (Section B.1.1).

Using the standard notation (see Chapter 4, Section 4.2.1) where y_{ij} denotes the growth measurement for individual j ($j = 1, 2, 3, \dots, J$) at occasion i ($i = 1, 2, 3, \dots, n_j$) and age_{ij} is the age in years of the individual at that occasion, a standard two-level unadjusted CP GCM is written as:

$$y_{ij} = \beta_0 + \sum_{r=1}^{p} \beta_r age_{ij}^r + u_{0j} + \sum_{s=1}^{q} u_{sj} age_{ij}^s + e_{ij} \quad (6\text{-}1)$$

To include study- and class-specific dummy variables in the fixed part of the model, the above model (6-1) is written as (note that the random effects and the residual parts of the model

remain the same):

$$\begin{aligned} y_{ij} = \beta_0 &+ \sum_{r=1}^{p} \beta_r age_{ij}^r \\ &+ \sum_{d_s=1}^{d_s k-1} \beta_{d_s 0} X_{d_s ij} + \sum_{d_s=1}^{d_s k-1} \sum_{r=1}^{p} \beta_{d_s r} X_{d_s ij} age_{ij}^r \\ &+ \sum_{d_c=1}^{d_c k-1} \beta_{d_c 0} X_{d_c ij} + \sum_{d_c=1}^{d_c k-1} \sum_{r=1}^{p} \beta_{d_c r} X_{d_c ij} age_{ij}^r \\ &+ u_{0j} + \sum_{s=1}^{q} u_{sj} age_{ij}^s \\ &+ e_{ij} \end{aligned} \tag{6-2}$$

where $d_s k$ and $d_c k$ are the number of levels for study ($d_s k = 8$) and class ($d_c k = 3$) effects, and $X_{d_s ij}$ and $X_{d_c ij}$ are design matrices for study and class effects, respectively. The Bolton growth study and Class I were used as reference categories. Seven dummy variables ($d_s k - 1$) were created for study effects ($d_s 1, .., d_s 7$) as:

$$\boldsymbol{X_{d_s ij}} = \begin{cases} X_{d_s 1ij} = 1\text{f study} = \text{Burlington, and 0 otherwise} \\ X_{d_s 2ij} = 1\text{f study} = \text{Denver, and 0 otherwise} \\ X_{d_s 3ij} = 1\text{f study} = \text{Forsyth, and 0 otherwise} \\ X_{d_s 4ij} = 1\text{f study} = \text{Iowa, and 0 otherwise} \\ X_{d_s 5ij} = 1\text{f study} = \text{Mathews, and 0 otherwise} \\ X_{d_s 6ij} = 1\text{f study} = \text{Michigan, and 0 otherwise} \\ X_{d_s 7ij} = 1\text{f study} = \text{Oregon, and 0 otherwise} \end{cases} \tag{6-3}$$

Similarly, two dummy variables ($d_c k - 1$) were created for class effects, where $d_c 1$ and $d_c 2$ denote Class II and Class III, respectively.

$$\boldsymbol{X_{d_c ij}} = \begin{cases} X_{d_c 1ij} = 1 \text{ if Class } = \text{Class II, and 0 otherwise} \\ X_{d_c 2ij} = 1 \text{ if Class } = \text{Class III, and 0 otherwise} \end{cases} \tag{6-4}$$

In the above model (6-1), β_0 and $\sum_{r=1}^{p} \beta_r age_{ij}^r$ now denote an intercept and *p*-th order fixed polynomial (positive integer) trend for Class I and the Bolton study estimated by the fixed effects regression coefficients $\beta_0, \ldots, \beta_p$. Because the model has only main effects (no interactions) between study- and class dummy variables, $\beta_{d_s 0}, \ldots, \beta_{d_s \text{p}}$ show the differences in the growth trajectories between Bolton growth study and each of the other seven studies coded using indicator variables.

Similarly, the $\beta_{d_c0}, \ldots, \beta_{d_c\mathrm{p}}$ regression coefficients capture differences in the growth trajectories between Class I and indicator variable coded Class II and Class III. The $u_{0j} + \sum_{s=1}^{q} u_{sj} age_{ij}^{s}$ denote random effects, a random intercept (u_{0j}) and random polynomial terms (positive integer) up to q-th order (where $q \leq p$). The residual e_{ij} captures the variation in the observed measurements around the individual-specific trajectories. (see Chapter 4 for further details).

6.2.4. Model fitting and evaluation of model fit

A systematic approach was followed for model fitting and fit evaluation, which included the following steps: i) defining *a priori* criteria for model fit evaluation, ii) model fitting, iii) model fit evaluation and comparison, and iv) assessment of model assumptions.

Criteria for model comparison

The simulation study described in Chapter 5 showed that the AIC performs better than the BIC when models involve complex variance-covariance structures and the sample size is not large. The results also showed that R^2 measures (variance explained) always select the most complex model. Therefore, the AIC was selected for model comparison in this chapter.

The study used the AIC calculated using the total number of parameters (fixed + random) for model comparison. Simulation findings (see Chapter 5) showed that under the REML method of estimation, the AIC based on the total number of parameters performed better than the AIC calculated using only the number of fixed effect parameters. Comparing the relative fit of competing models used a cut-off value of 10, which indicates that the more complex model has essentially no support (Burnham & Anderson, 2004).

Model fitting

The 'maximal' random-effect modelling approach was used for model fitting. Therefore, all possible random effects were included for each model. In case a model failed to converge, there was no attempt to re-fit the model with a subset of random effects. Section 6.5.1 discusses this in further detail.

Each of the five GCMs was fitted as a univariate model to upper jaw length, lower jaw length and total face height, separately for males and females. Except for the PB GCM, age was mean centred. For the PB GCM, I used the recommended original metric of age (Grimm et al., 2011; Hauspie et al., 2004).

For FP GCM, unlike the other four GCMs, age is not mean centred by simply subtracting the mean of age from the age itself, but rather each term is centred by following the same rules (such as repeated powers are multiplied by log of age and power zero is the log of the age) as applied to the construction of FP terms (StataCorp, 2017b).[8] For example, to centre age_{ij} on its mean c (e.g., 12 years) for a third-degree FP model with powers 1 1 3, the fractional polynomial terms are created as $[\, age_{ij}^{(1,1,3)} - c^{(1,1,3)} \,]$, which results in first term as $[\, age_{ij}^{1} - c^{1} \,]$, the second term as $[age_{ij}^{1}\log(age_{ij}) - c^{1}\log(c)]$, and the third term as $(age_{ij}^{3} - c^{3})$.

Both linear (CP, FP and RCS) and nonlinear (SITAR and PB) GCMs were estimated using the REML method. Chapter 4 (Section 4.5) provides details on the ML and REML methods of estimation, and reasons for choosing REML over ML.

[8] See https://www.stata.com/manuals/rfp.pdf (pp. 13–14)

Up to the fifth degree of CP GCMs were fitted, because most previous studies analysing jaw growth data have often used higher degree (beyond cubic) conventional polynomials (see Chapter 2). This resulted in a total of five CP models.

For FP GCM, all possible models were explored up to the third degree of polynomial using recommended eight powers. This resulted in a total of 164 models (see Chapter 4, Section 4.2.2). As recommend (StataCorp, 2017b)[9], a scale option was used in generating fractional polynomial terms. As reciprocal squared and cubic powers were used, values of the predictor variable should not be too large or too close to zero. The scale option automatically selects the optimal values of a and b for scaling the predictor $[(age_{ij} + a)/b]$ before generating the fractional polynomial terms.

For the RCS and SITAR GCMs, three to seven knots were placed at the fixed quantiles (percentiles) of the predictor's (age) marginal distribution, as described in Chapter 4 (see RCS GCM, Section 4.2.2). Thus, five RCS models and five SITAR models were fitted. Unlike the CP, FP, RCS and SITAR models, only one PB model was fitted to the data.

Model fit evaluation

Using the AIC, one best-fitting model was selected for each of the models fitted for different numbers of polynomials (CP and FP) and knots (RCS and SITAR). Thus, one CP GCM was selected from five, one RCS GCM from five, and one SITAR GCM from five. Similarly, one best model was selected from 164 FP GCMs. For PB GCM, only one model was fitted to each outcome.

For each best-fitting model, the residual standard deviation (RSD) estimates and the extent of agreement between the observed and model predicted outcomes were evaluated. A model was considered to fit well to the data if 95% limits of agreement between observed and predicted outcome lay within 10% of the mean in yearly growth periods.

[9] See https://www.stata.com/manuals/rfp.pdf

The FP GCM selection involves selecting the best power(s) for a given degree and then comparing different degrees of fractional polynomials. For example, first select one best fitting model for degree 1 out of 8 models and one best fitting model for degree 2 out of 36 models (see Chapter 4, Section 4.2.2 for the total number of models for each degree) and then compare best fitting degree 1 and degree 2 models. The approach originally suggested by Royston and Altman (1994) was slightly modified (see below).

Royston and Altman (1994) recommended using deviance (–2 x log-likelihood) for model comparison. To choose best-fitting power(s) for a given degree of model, a model with the lowest deviance was identified as the best-fitting model (e.g., one model out of eight for the first-degree FP GCM, one best model out of 36 for a second-degree FP GCM, and so on). To compare different degrees of models (e.g., first-degree and second-degree), (Royston & Altman, 1994) compared difference in deviance using chi-square distribution with two degrees of freedom (as two extra parameters, a power and a regression coefficient, are estimated).

Unlike Royston and Altman (1994), I used the AIC to compare different degrees of FP GCMs. The AIC penalises the log-likelihood for the model complexity based on the total number of parameters (fixed and random). Although model complexity does not make a difference when selecting power(s) for a given degree of a FP GCM, as the number of parameters remains the same, the AIC was used mainly to be consistent in using the same criteria when comparing different degrees of FP GCMs.

Model assumptions

Graphical methods were used to evaluate the assumption of homoscedasticity and normality of the residuals, as they are more robust than numerical tests (Kozak & Piepho, 2018). Standardised residuals (z-score transformation) were used as they are better for checking assumptions than raw residuals (Kozak & Piepho, 2018).

The normal quantile-quantile (QQ) plots, which compare the quantiles of observed distribution versus quantiles of the normal distribution, were used to examine the normality of the random effects and the level-1 residuals. The QQ plots show no violation of normality assumption when all points are aligned along the 45-degree reference line. Level 1 residuals were plotted against the predicted outcome to visualise the spread of the residuals and to inspect heteroscedasticity.

6.2.5. Distance and derivatives estimation

Estimating distance and derivative for covariate adjusted GCMs necessitates that covariates should be included in predictions. Chapter 4 provides a detailed review of derivative estimation for linear and nonlinear GCMs and their role in estimating APGV and PGV. Here, derivative estimation for CP, FP, RCS, SITAR and PB GCMs is extended by including covariates.

As an illustrative example, below I describe the procedure to estimate population-average and individual distance and derivatives (velocity and acceleration) for the CP GCM. Appendix B (Section B.1.2) provides details on distance and derivative estimation for FP, RCS, SITAR and PB GCMs. A third degree ($p = q = 3$) i.e., a cubic CP GCM is considered (see Equation (6-2)). For ease of understanding how linear and nonlinear terms are included to the derivative estimation, the linear term is separated out from the nonlinear (quadratic and cubic) terms and their interactions with covariates. Model (6-2) is now re-written as (6-5):

$$
\begin{aligned}
y_{ij} = {} & \beta_0 + \beta_1 age_{ij}^1 + \sum_{r=2}^{p} \beta_r age_{ij}^r \\
& + \sum_{d_s=1}^{d_s k-1} \beta_{d_s 0} X_{d_s ij} + \sum_{d_s=1}^{d_s k-1} \beta_{d_s 1} X_{d_s ij} age_{ij}^1 + \sum_{d_s=1}^{d_s k-1} \sum_{r=2}^{p} \beta_{d_s r} X_{d_s ij} age_{ij}^r \\
& + \sum_{d_c=1}^{d_c k-1} \beta_{d_c 0} X_{d_c ij} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c 1} X_{d_c ij} age_{ij}^1 + \sum_{d_c=1}^{d_c k-1} \sum_{r=2}^{p} \beta_{d_c r} X_{d_c ij} age_{ij}^r \\
& + u_{0j} + \sum_{s=1}^{q} u_{sj} age_{ij}^s \\
& + e_{ij}
\end{aligned}
\tag{6-5}
$$

Distance (e.g., jaw size)

The distance is a function itself. Therefore, the regression coefficients are multiplied by respective functions (polynomial terms) and then added (see Chapter 4, Section 4.3.2 for details). The population-average distance was estimated as:

$$\begin{aligned} & \beta_0 + \beta_1\{age_{ij}^1\} + \sum_{r=2}^{p} \beta_r \{age_{ij}^r\} \\ & + \beta_{d_s0}\{X_{d_sij}\} + \sum_{d_s=1}^{d_sk-1} \beta_{d_s1}\{X_{d_sij}\}\{age_{ij}^1\} + \sum_{d_s=1}^{d_sk-1}\sum_{r=2}^{p} \beta_{d_sr}\{X_{d_sij}\}\{age_{ij}^r\} \\ & + \beta_{d_c0}\{X_{d_cij}\} + \sum_{d_c=1}^{d_ck-1} \beta_{d_c1}\{X_{d_cij}\}\{age_{ij}^1\} + \sum_{d_c=1}^{d_ck-1}\sum_{r=2}^{p} \beta_{d_cr}\{X_{d_cij}\}\{age_{ij}^r\} \end{aligned} \tag{6-6}$$

To find individual-specific distances, $f_{CP_j}(age)$, the random effects are multiplied by respective polynomial terms and added to the population-average distance as:

$$\begin{aligned} & \beta_0 + \beta_1\{age_{ij}^1\} + \sum_{r=2}^{p} \beta_r \{age_{ij}^r\} \\ & + \beta_{d_s0}\{X_{d_sij}\} + \sum_{d_s=1}^{d_sk-1} \beta_{d_s1}\{X_{d_sij}\}\{age_{ij}^1\} + \sum_{d_s=1}^{d_sk-1}\sum_{r=2}^{p} \beta_{d_sr}\{X_{d_sij}\}\{age_{ij}^r\} \\ & + \beta_{d_c0}\{X_{d_cij}\} + \sum_{d_c=1}^{d_ck-1} \beta_{d_c1}\{X_{d_cij}\}\{age_{ij}^1\} + \sum_{d_c=1}^{d_ck-1}\sum_{r=2}^{p} \beta_{d_cr}\{X_{d_cij}\}\{age_{ij}^r\} \\ & + u_{0j} + u_{1j}\{age_{ij}^1\} + \sum_{s=2}^{q} u_{sj}\{age_{ij}^s\} \end{aligned} \tag{6-7}$$

Derivatives (growth velocity and acceleration)

Derivatives are estimated by using the differential rules as described earlier in Chapter 4 (Section 4.3.2).

First derivative (growth velocity)

The population-average and individual-specific first derivatives are estimates below.

Population-average first derivative ($f'_{CP_0}(age)$):

$$\begin{aligned}&\beta_1 + \sum_{r=2}^{p} \beta_r \left\{ r(age_{ij})^{(r-1)} \right\} \\ &+ \sum_{d_s=1}^{d_s k-1} \beta_{d_s 1}\{X_{d_s ij}\} + \sum_{d_s=1}^{d_s k-1} \sum_{r=2}^{p} \beta_{d_s r}\{X_{d_s ij}\} \left\{ r(age_{ij})^{(r-1)} \right\} \\ &+ \sum_{d_c=1}^{d_c k-1} \beta_{d_c 1}\{X_{d_c ij}\} + \sum_{d_c=1}^{d_c k-1} \sum_{r=2}^{p} \beta_{d_c r}\{X_{d_c ij}\} \left\{ r(age_{ij})^{(r-1)} \right\}\end{aligned} \tag{6-8}$$

Individual-specific first derivative ($f'_{CP_j}(age)$):

$$\begin{aligned}&\beta_1 + \sum_{r=2}^{p} \beta_r \left\{ r(age_{ij})^{(r-1)} \right\} \\ &+ \sum_{d_s=1}^{d_s k-1} \beta_{d_s 1}\{X_{d_s ij}\} + \sum_{d_s=1}^{d_s k-1} \sum_{r=2}^{p} \beta_{d_s r}\{X_{d_s ij}\} \left\{ r(age_{ij})^{(r-1)} \right\} \\ &+ \sum_{d_c=1}^{d_c k-1} \beta_{d_c 1}\{X_{d_c ij}\} + \sum_{d_c=1}^{d_c k-1} \sum_{r=2}^{p} \beta_{d_c r}\{X_{d_c ij}\} \left\{ r(age_{ij})^{(r-1)} \right\} \\ &+ u_{1j} + \sum_{s=2}^{q} u_{sj} \left\{ s(age_{ij})^{(s-1)} \right\}\end{aligned} \tag{6-9}$$

Second derivative (growth acceleration)

The second derivative is the derivative of the first derivative. Estimation of population-average $f'_{CP_0}(age)$ (6-10) and individual-specific $f''_{CP_j}(age)$ (6-11) second derivatives are shown below:

$$\begin{aligned}&\sum_{r=2}^{p} \beta_r \left\{ r(r-1)(age_{ij})^{(r-2)} \right\} \\ &+ \sum_{d_s=1}^{d_s k-1} \sum_{r=2}^{p} \beta_{d_s r}\{X_{d_s ij}\} \left\{ r(r-1)(age_{ij})^{(r-2)} \right\} \\ &+ \sum_{d_c=1}^{d_c k-1} \sum_{r=2}^{p} \beta_{d_c r}\{X_{d_c ij}\} \left\{ r(r-1)(age_{ij})^{(r-2)} \right\}\end{aligned} \tag{6-10}$$

$$\sum_{r=2}^{p} \beta_r \left\{ r(r-1)(age_{ij})^{(r-2)} \right\} + \sum_{d_s=1}^{d_s k-1} \sum_{r=2}^{p} \beta_{d_s r} \{X_{d_s ij}\} \left\{ r(r-1)(age_{ij})^{(r-2)} \right\} + \sum_{d_c=1}^{d_c k-1} \sum_{r=2}^{p} \beta_{d_c r} \{X_{d_c ij}\} \left\{ r(r-1)(age_{ij})^{(r-2)} \right\} + \sum_{s=2}^{q} u_{sj} \left\{ s(s-1)(age_{ij})^{(s-2)} \right\} \quad (6\text{-}11)$$

Covariate-adjusted distance and derivatives

Covariate-specific trajectories are obtained by leaving dummy variables (applicable to both linear and nonlinear GCMs) and cross-level interactions (only relevant for linear GCMs) between dummy variables and polynomial (CP and FP) and spline (RCS) terms at their original values. Unlike linear GCMs, there are no interactions between dummy variables and the level 1 predictor (age) for nonlinear GCMs. To obtain mean average trajectories, each dummy variable is mean centred by subtracting the arithmetic mean. This applies to both linear (CP, FP and RCS) and nonlinear (SITAR and PB) GCMs. For cross-level interactions in linear GCMs (CP, FP and RCS), each interaction term is centred by replacing it with a new term created by multiplying the polynomial (CP and FP) and spline (RCS) terms with mean of the dummy variable.

Thus, depending on how each covariate (study and class) is included in the post-estimation distance and derivatives estimation, the following can be obtained: (i) study- and class-specific distance and derivatives, (ii) class-adjusted study-specific distance and derivative (mean averaging over class), (iii) study-adjusted class-specific distance and derivative (mean averaging over study), and (iv) study- and class-adjusted distance and derivatives (mean averaging over both study and class effects). This chapter estimates distance and derivatives adjusted for both study and class effects.

6.2.6. Timing and the intensity of the adolescent growth spurt

The quadratic function method was used to calculate population-average and individual-specific APGV and PGV (Cao et al., 2018). Chapter 4 (Section 4.4.2) describes full details of the quadratic function method. The same quadratic function method was used to estimate population-average and individual-specific APGV and PGV. However, instead of using the population-average first and second derivatives, individual-specific derivatives were used when estimating individual-specific APGV and PGV. Following the notation described earlier in Chapter 4, sub subscript *0* was used to denote population-average $APGV_0$ and PGV_0, and sub subscript *j* to denote individual-specific $APGV_j$ and PGV_j.

Depending on how derivatives are estimated (covariate-specific or covariate-adjusted, see Section 6.2.5), covariate-specific or covariate-adjusted population-average and individual-specific APGV and PGV can be obtained as estimation of APGV and PGV is directly dependent on derivatives.

6.2.7. Bootstrapped confidence intervals

To account for the uncertainty in the model-estimated growth parameters, the bootstrap technique was used to estimate confidence intervals (CIs). Bootstraps construct empirical estimates of the distribution function of the data, which are then used to make inferences instead of the distribution that actually generated the data (Field & Welsh, 2007).

The bootstrapping is particularly useful for estimating CIs when model assumptions that underlie analytical methods are uncertain (Carpenter & Bithell, 2000; Davison & Hinkley, 1997). The bootstrap technique (Efron, 1979; Efron & Tibshirani, 1993) was initially developed for parametric models applied to independent and identically distributed data. For nested data, it is important to replicate the hierarchical dependence in the data structure (Rasbash et al., 2016).

The parametric bootstrap, based on the assumed hierarchical random effects model, is a popular approach for nested data. However, when the model assumptions are violated, the parametric inference is questionable (Carpenter et al., 2003). Many alternative approaches are available for bootstrapping the clustered data, such as the semiparametric and the nonparametric bootstrap (Chambers & Chandra, 2012; Field & Welsh, 2007; Thai et al., 2013).

The semiparametric (residual) bootstrapping, in which the parametric bootstrap is used to generate the marginal (fixed) part of the model and the dependence structure of the random effects (residuals), is generated nonparametrically (Carpenter et al., 2003; Rasbash et al., 2016). Although the semiparametric bootstrap is less sensitive to the model assumptions than the parametric bootstrap, it still relies on the model-based best linear unbiased predictions of the level 2 random effects (Chambers & Chandra, 2012).

Nonparametric bootstrap (by cluster with replacement) involves resampling with replacement the entire level 2 units (e.g., individuals) with joint vectors of predictors and corresponding responses from the original data before fitting the model (Thai et al., 2013).Unlike parametric and semi-parametric bootstrapping approaches, the nonparametric bootstrap makes no assumptions on the model (Thai et al., 2013). Therefore, nonparametric bootstrap (using 1,000 replications) was used to derive CIs for the APGV and PGV. For each sample, GCMs were fitted, distance and derivatives estimated, and then APGV and PGV computed. The 2.5^{th} and 97.5^{th} percentiles of bootstrapped APGV and PGV were used as the 95% CI for these parameters.

6.2.8. Sensitivity analysis

Sensitivity analysis were performed for SITAR (effect of different spline basis on the model fit, optimising the model) and FP (effect of age scale on the model fit) GCMs.

Effect of different spline basis on the fit of the SITAR model

As described in Chapter 4 (see Section 4.2.3), the 'sitar' R package (Cole, 2017a; Cole, 2019) was modified for fitting the SITAR GCM. This was done mainly to implement the same methodology to generate restricted cubic splines for the RCS and SITAR GCMs. Below, the key changes are summarised, then the need to conduct the sensitivity analysis is discussed.

The 'sitar' package calls the 'ns' function of the 'splines' R package (Bates & Venables, 2011) to construct the restricted cubic splines (natural cubic splines) using the B-spline basis. The piecewise-cubic splines are generated with the user-specified sequence of interior knots (or degree of freedom) and boundary knots. If boundary knots and interior knots are not specified by the user, then the boundary knots are set to the extremes of the data (minimum and maximum), and the interior knots are chosen based on the quantiles of predictor (age) depending on the degree of freedom provided by the user. The 'sitar' R package by default adds a fractional extension (0.04) to the age range but follows the same quantiles approach for placing the interior knots.

Instead of using the B-spline basis to generate restricted cubic splines for the SITAR GCM, the truncated power basis was used (Harrell, 2015). This was achieved by modifying the 'sitar' R package, which now calls the 'rms' R package (Harrell Jr, 2017) instead of the 'ns' function for generating restricted cubic splines. Although B-splines basis are numerically more stable than the truncated power basis, they are more complex (Harrell, 2015). The truncated power basis is commonly used in major software, such as in constructing design matrices for cubic spline functions, and seldom presents estimation problems (Harrell, 2015).

Like the 'ns' function of the 'splines' package, the 'rms' R package imposes a constraint that the function is linear before the first knot and after the last. The major difference between these two approaches is that unlike the 'ns' function, which places boundary knots at extremes of the data (or a slightly extended range as implemented in the 'sitar' package), 'rms' places all knots at equally

spaced quantiles as recommended by Harrell (2015). Stata's 'mkspline' command (StataCorp, 2017b), used in this study to generate the restricted cubic splines design matrix for the RCS GCM, implements the same equally spaced quantiles approach as in the 'rms' R package.

As the approach followed to generate the restricted cubic splines design matrix using the modified 'sitar' package differed from the original 'sitar' package, it is possible that the findings differ from the original 'sitar' R package. This might include differences in the fit of the model and the adolescent growth spurt parameters (APGV and PGV), which are derived from the derivatives. To ascertain that modifying the package did not adversely affect the model fit, a sensitivity analysis was conducted.

Optimising the SITAR model

As well as selecting the optimal number of knots for the best-fitting cubic spline curve, different models should be explored as part of the sensitivity analysis to examine how transforming the age and/or outcome scales affects the fit of SITAR GCM (Cole & Mori, 2017). This exercise of comparing the fit of models with and without transformation, and selecting the best age and/or outcome scales, is termed "*optimising the SITAR model*" (Cole & Mori, 2017).

Following the recommended procedure (Cole et al., 2010; Cole & Mori, 2017), the fit of the SITAR GCM was compared with and without log and square root transformations of age, for all three outcomes (upper and lower jaw length and total face height), for both male and female samples. The 'AICadj' function in the 'sitar' package was used to calculate the AIC, which adjusts the deviance estimate in case outcomes are transformed (Cole, 2017a).

Effect of age scale on the fit of FP model

While generating fractional polynomials, it important to ensure that values of the predictor variable (i.e., age) are not too large or too close to zero (StataCorp, 2017b). This is because higher

order cubic powers and reciprocal squared powers are explored while selecting the best fractional polynomial powers. To study the effect of the scale of age on the model fit, results of FP GCMs fitted to the data with original and re-scaled age were compared. In addition, I also evaluate how re-scaling of age affects the estimates of derivatives based adolescent growth spurt parameters (APGV and PGV).

Although it is tempting to follow the same approach (re-scaling age) for the CP GCM as cubic and even higher powers are used, I am not aware of any such practice for fitting the CP GCM. Therefore, the sensitivity analysis was restricted to the FP GCM. If re-scaling of age has a significant effect on the model fit of the FP GCM for positive non-fractional powers, then it might be useful in the future to explore scale effect on the fit of the CP GCMs.

6.2.9. Software

Chapter 4 reviews software available for fitting linear and nonlinear GCMs, derivative estimation, and estimating APGV and PGV. The present study primarily used Stata. Full details on the implementation of model fitting and the nonparametric bootstrapping are provided in Appendix B (Section B.2).

As no software package was available to estimate covariate-adjusted derivatives, a suite of post-estimation Stata programs was written to implement the study methodology of estimating covariate-adjusted derivatives and the APGV and PGV. As post-estimation commands require key information on model estimates such as degree of conventional polynomials (CP), degree and powers of fractional polynomials (FP), and number and location of knots (RCS and SITAR), a set of new commands was written that were either new or based on Stata's available commands (see Appendix B, Section B.2 for details).

6.3. Results

Data were analysed separately for males and females. Each GCM (CP, FP, RCS, SITAR and PB) was fitted to three outcomes i.e., upper jaw length, lower jaw length and total face height.

To select one best-fitting CP, FP, RCS, SITAR and PB GCMs for each of the three outcomes analysed separately for males and females, 114 models were fitted which include five CP (up to fifth degree), three FP (up to third degree), five RCS (three to seven knots), five SITAR (three to seven knots), and a PB GCM. The three FP GCMs for each outcome were selected from 164 models (8 for first degree; 36 for second degree; 120 for third degree), see Section 6.2.4.

The PB GCM failed to converge for any of the three outcomes for both sexes. This occurred even though models were layered in such a way that variance-covariance estimates from simpler models were used as initial values for subsequent more complex models (see Section 6.2.9). Fitting the PB GCM with a subset of random effects was not pursued, because this would lead to untenable model assumptions. This point is further discussed in Section 6.4.1.

Exclusion of the PB GCM reduced the number of models from 114 (19 x 3 x 2) to 108 (18 x 3 x 2). From 108 models, the best-fitting CP, FP, RCS and SITAR GCMs (a total 24 models) were selected. Results presented in this section are for 24 (4 x 3 x 2) best-fitting models.

The sensitivity analysis results (see Appendix B, Section B.3.1 for detailed results) show that the SITAR GCM using the truncated power basis used for constructing the cubic spline design matrix fitted equally well (upper jaw length) or better (lower jaw length and total face height) than the B-spline basis used in the original ‘sitar’ package. The log and square root transformation of predictor (age) and outcome did not improve the model fit for any of the three outcomes in either males or females. Therefore, results reported below are for the SITAR GCM fitted to data using the truncated power basis for generating cubic spline design matrix with original age and outcome scales.

The sensitivity analysis results (see Appendix B, Section B.3.1 for detailed results) reveal that irrespective of sex and outcomes, FP GCMs fitted to the data with re-scaled age fit better than the original scale of the age. The optimal values automatically selected by 'fpgen' for re-scaling age were zero and 10 for scale factors a and b, respectively. Thus, for all three outcomes for both males and females, age was re-scaled as [(age+0]/10] before generating the fractional polynomial terms. Thus, results reported below are for the FP GCM fitted to the data with re-scaled age.

For the sake of brevity, detailed results are reported for CP, FP, RCS and SITAR GCMs fitted to the upper jaw length for both males and females. Results for the other two outcomes (lower jaw length and total face height) are provided in Appendix B (Section B.3) but summarised along with the upper jaw length results where appropriate.

6.3.1. Model fitting

Model convergence

Table 6-1 and Table 6-2 summarise the convergence of CP, FP, RCS and SITAR GCMs fitted to the male and female data, respectively. For all successfully converged models (labelled using the check mark ✓), there was no error (such as singular variance-covariance matrix) reported by the 'runmlwin' (CP, FP, RCS) or 'nlme' (SITAR) used to fit GCMs. For models that failed to converge (labelled using the cross-mark ×), 'runmlwin' and 'nlme' reported convergence failure. As reported earlier, no attempt was made to simplify the variance-covariance structure when a model failed to converge.

Except for lower jaw length and total face height for males, the CP GCM did not converge beyond third degree. For these two outcomes, a fourth-degree CP model successfully converged but a fifth-degree CP GCM failed to converge. A third-degree FP GCM successfully converged for all three outcomes for both sexes. By design, higher degree (beyond third-degree) FP GCMs were not fitted for any of the three outcomes for either males or females.

Table 6-1. Summary of linear and nonlinear growth curve models applied to male data.

		Upper jaw length (COPAD)				Lower jaw length (COPOD)				Total face height (TFHNP)			
Degree (CP and FP GCMs)	Knots (RCS and SITAR GCMs)	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR
1	3	✓	✓	✓	×	✓	✓	✓	×	✓	✓	✓	×
2	4	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	×
3	5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	6	×	NA	✓	×	✓	NA	✓	✓	✓	NA	✓	✓
5	7	×	NA	×	×	×	NA	✓	✓	×	NA	✓	✓

NA: Not applicable (model was not fitted); ✓ model converged successfully; × model did not converge.

Models: CP conventional polynomial; FP fractional polynomial; RCS restricted cubic spline; SITAR superimposition by translation and rotation.

Degree: Degree (number of polynomials) of CP and FP models; Knots: Number of knots for RCS and SITAR models.

Outcomes: COPAD Condyle–Point A distance in millimetres; COPOD condyle–pogonion distance in millimetres; TFHNP total face height in millimetres.

Table 6-2. Summary of linear and nonlinear growth curve models applied to female data.

		Upper jaw length (COPAD)				Lower jaw length (COPOD)				Total face height (TFHNP)			
Degree	Knots	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR
1	3	✓	✓	✓	×	✓	✓	×	×	✓	✓	×	×
2	4	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	×
3	5	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓
4	6	×	NA	×	✓	×	NA	×	✓	×	NA	×	✓
5	7	×	NA	×	×	×	NA	×	✓	×	NA	×	✓

NA: Not applicable (model was not fitted); ✓ model converged successfully; × model did not converge.

Models: CP conventional polynomial; FP fractional polynomial; RCS restricted cubic spline; SITAR superimposition by translation and rotation.

Degree: Degree (number of polynomials) of CP and FP models; Knots: Number of knots for RCS and SITAR models.

Outcomes: COPAD Condyle–Point A distance in millimetres; COPOD condyle–pogonion distance in millimetres; TFHNP total face height in millimetres.

Fit to the data

The AIC and RSD estimates for each of the 18 GCMs (five CP, three FP, five RCS and five SITAR) fitted to upper and lower jaw length and total face height, separately for males and females, are provided in Appendix B (Section B.3.2).

Except for the SITAR GCM, an increase in the model complexity (i.e., an increase in the degree of CP and FP GCMs, and number of knots for the RCS GCM) resulted in a decrease of both AIC and RSD estimates.

For the SITAR GCM, while the AIC decreased, the RSD remained almost the same. This is perhaps because, unlike the SITAR GCM for which the number of random effects remain the same (three random effects) irrespective of the number of knots, an increase in the degree of polynomial models (CP and FP GCMs) and the number of knots for the RCS GCM results in a richer variance-covariance structure, capturing between-individual variability more efficiently. For the SITAR GCM, this unexplained variance pools as residual variance.

The AIC and RSD estimates for the best-fitting CP, FP, RCS and SITAR GCMs applied to upper and lower jaw length and total face height data for males and females are summarised in Table 6-3.

Out of five CP GCMs considered (up to degree five), a third-degree CP GCM fits best to all three outcomes for females. For males, the degree of best-fitting CP GCMs differed for different outcomes. While a third-degree model fitted best to upper jaw length, a fourth-degree model was identified as the best-fitting model for lower jaw length and total face height for males.

Table 6-3. Comparison of linear and nonlinear growth curve models applied to male and female data.

	Upper jaw length (COPAD)			Lower jaw length (COPOD)			Total face height (TFHNP)		
	Degree / Knots	AIC	RSD	Degree /Knots	AIC	RSD	Degree /Knots	AIC	RSD
Male									
CP	3	3844.78	1.06	4	4361.73	1.35	4	4361.83	1.36
FP	3	3838.47	1.05	3	4433.16	1.39	3	4425.89	1.42
RCS	5	3762.95	0.94	7	4320.65	1.03	7	4312.34	1.01
SITAR	4	3774.42	1.01	7	4336.22	1.23	7	4335.07	1.23
Female									
CP	3	3439.85	0.83	3	4335.36	1.12	3	4180.73	1.04
FP	3	3430.38	0.83	3	4330.46	1.10	3	4165.04	1.02
RCS	4	3388.79	0.77	5	4245.63	0.95	5	4100.77	0.89
SITAR	4	3399.04	0.79	5	4275.69	1.02	5	4114.11	0.93

AIC: Akaike information criterion; RSD: residual standard deviation in millimetres.

CP Conventional polynomial; FP Fractional polynomial; RCS Restricted cubic spline; SITAR Superimposition by translation and rotation.

Degree /Knots: Degree (number of polynomials) of CP and FP models; number of knots for RCS and SITAR models.
Fractional polynomial powers (for both males and females, age is scaled as [(age+0)/10]; see Section 6.2.4 for details)
Male: COPAD 1 3 3; COPOD 3 3 3; TFHNP 3 3 3
Female: COPAD 1 1 1; COPOD 3 3 3; TFHNP 3 3 3

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Among the FP GCMs compared, a third-degree FP GCM fitted best to all three outcomes for both males and females. For the upper jaw length, powers selected by the best-fitting models were 1 3 3 and 1 1 1 for males and females, respectively. Surprisingly, the best-fitting powers for the FP GCM fitted to the lower jaw length and total face height for both males and females were same, 3 3 3.This is because for the other three GCMs, models applied to lower jaw length and total face height data differed for degree of polynomial (CP) or number of knots (RCS and SITAR) for males and females. This is perhaps an indication that higher degree FP GCMs or different fractional polynomial powers should be considered when analysing lower jaw length and total face height. A fourth-degree FP GCM was not explored (see Section 6.2.3 for reasons), restricted to a recommended set of powers (-2 -1 -0.5 0 0.5 1 2 3).

For females, the best-fitting RCS and SITAR GCMs included the same number of knots for the upper jaw length (four knots), lower jaw length (five knots) and total face height (five knots). For males, the best-fitting RCS and SITAR GCMs included the same number of knots for lower jaw

length (seven knots) and total face height (seven knots). However, for the upper jaw length for males, the best-fitting RCS GCM involved five knots, whereas four knots fitted best for the SITAR GCM.

Model comparison (Table 6-3) shows that amongst all models applied to the jaw growth data (CP, FP, RCS and SITAR), the RCS GCM was the best-fitting model for all three outcomes. This is because AIC estimates were lowest for the RCS GCM (by at least 10 AIC units). Furthermore, for all three outcomes, the RSD estimates were lowest for the RCS GCM. For all three outcomes, the RSD estimated by all four GCMs (CP, FP, RCS and SITAR) was higher for males than females. The sex difference in the RSD were greater for the lower jaw length and total face height than the upper jaw length.

The differences between the model-predicted and the observed upper and lower jaw length and total face height are provided in Appendix B (Section B.3.2). For all three outcomes, each GCM appears to fit well to the data, with 95% limits of agreement between observed and predicted outcome being within 10% of the mean in each growth period.

Model assumptions

As the minimum age difference between two consecutive growth measurements is six months (median one year), independence of residuals was assumed, and autocorrelation of residuals was not tested (see Section 6.2.2 for details). QQ and residual versus predicted plots were used to graphically check the normality and homoscedasticity assumptions, respectively.

Normality

The QQ plots for CP, FP, RCS and SITAR GCMs fitted to upper and lower jaw length and total face height are shown in Appendix B (Section B.3.3). The QQ plots used to assess the normality assumption of individual-specific random effects are based on one datapoint per individual. For all

three outcomes, the QQ plots show no violation of normality assumption as all points are aligned along the reference line.

The QQ plots of level 1 residuals (e_{ij}) for CP, FP, RCS and SITAR GCMs fitted to upper and lower jaw length and total face height are shown in Appendix B (Section B.3.3). Plots are based on all datapoints for all individuals. The QQ plots show no violation of normality assumption for any of the three outcomes as all points are aligned along the reference line.

Homoscedasticity

The fitted versus residual plots evaluating the homoscedasticity assumption of level 1 residuals (e_{ij}) for CP, FP, RCS and SITAR GCMs fitted to upper and lower jaw length and total face height are shown in Appendix B (Section B.3.3). For all three outcomes, there is no evidence for heteroskedasticity, as plots show no sign of a systematic pattern in the spread of residuals.

6.3.2. Distance (size) and derivatives (velocity and acceleration)

In this section, I present results for distance and growth velocity for upper jaw length, lower jaw length and total face height for males and females. Results for growth acceleration are presented in Appendix B (Section B.3.4).

Distance curves for the upper jaw length are shown in Figure 6-1 and Figure 6-2 for males and females, respectively. Unlike the RCS and SITAR GCMs, which approximate an asymptote, both CP and FP GCMs show a biologically implausible decline in jaw sizes towards the end of the growth period.

Figure 6-3 and Figure 6-4 show velocity curves for the upper jaw length for males and females, respectively. In contrast to the RCS and SITAR GCMs, both CP and FP GCMs estimated biologically implausible negative growth velocities.

For both males and females, growth trajectories (distance and derivatives) estimated by RCS and SITAR GCMs looked alike but differed substantially from the CP and FP GCMs. Differences between cubic spline–based linear (RCS) and nonlinear (SITAR) GCMs and polynomial-based linear GCMs (CP and FP) were more pronounced for derivatives compared to the distance curves. For example, while the distance curves for all four GCMs varied slightly, difference among velocity curves was much greater, and largest for the acceleration curves.

The difference in the acceleration curves between CP and FP GCMs is interesting (see Appendix B, Section B.3.4). For both sexes, the same (third) degree of CP (powers 1 2 3) and FP (powers 1 3 3 for males; 1 1 1 for females) GCMs fitted to the upper jaw length. However, second derivatives (acceleration) of these two GCMs are very different. This is because of the difference in powers. Unlike a third-degree CP GCM for which the second derivative is always linear, i.e., a straight line (see Section 6.2.5), differentiation of a third-degree FP GCM is nonlinear except when powers are nonzero, nonfractional and nonnegative, i.e., identical to a third-degree CP GCM (powers 1 2 3). These issues are further discussed in Section 6.4.

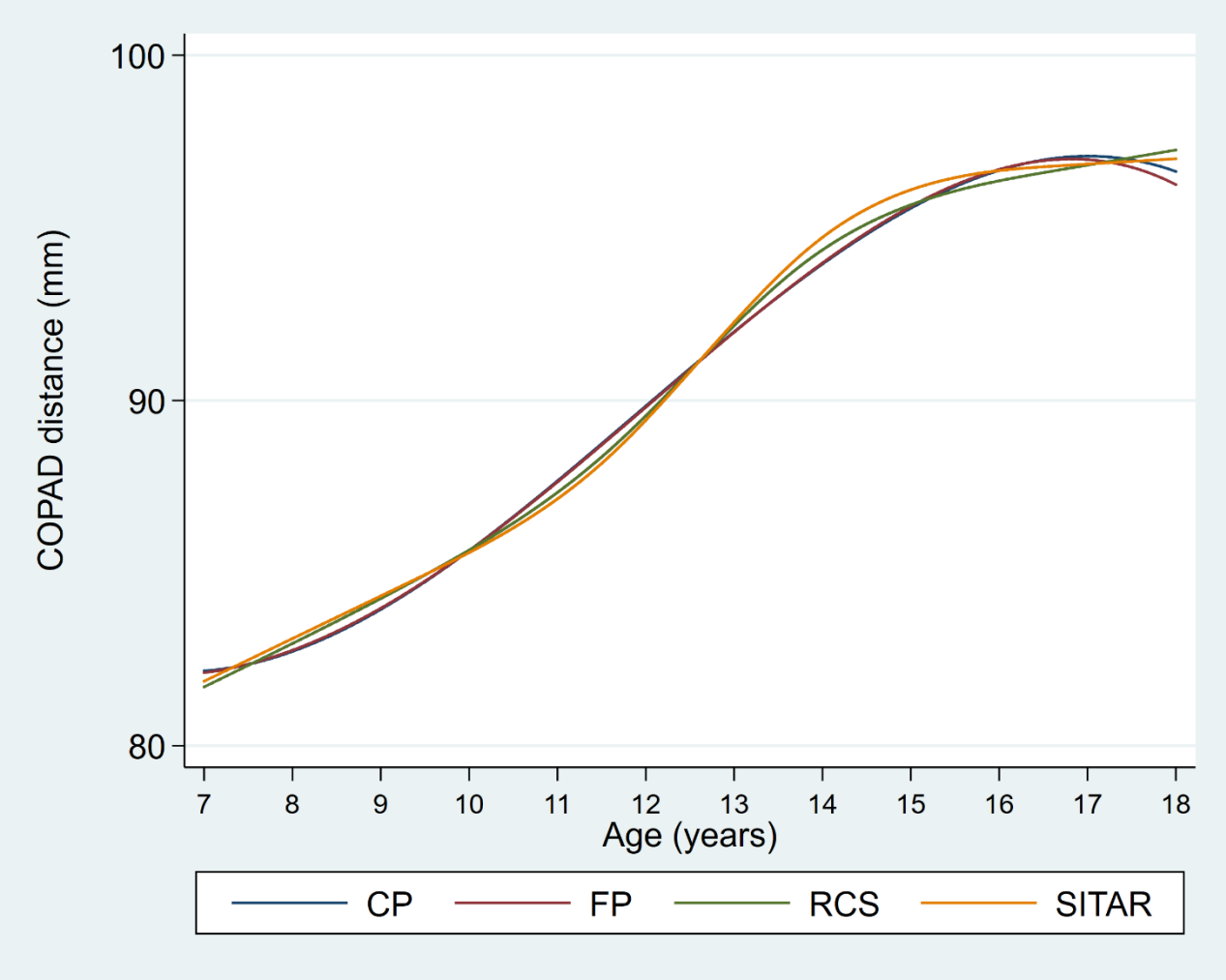


Figure 6-1. Population-average distance curves for the upper jaw length (COPAD) for males.

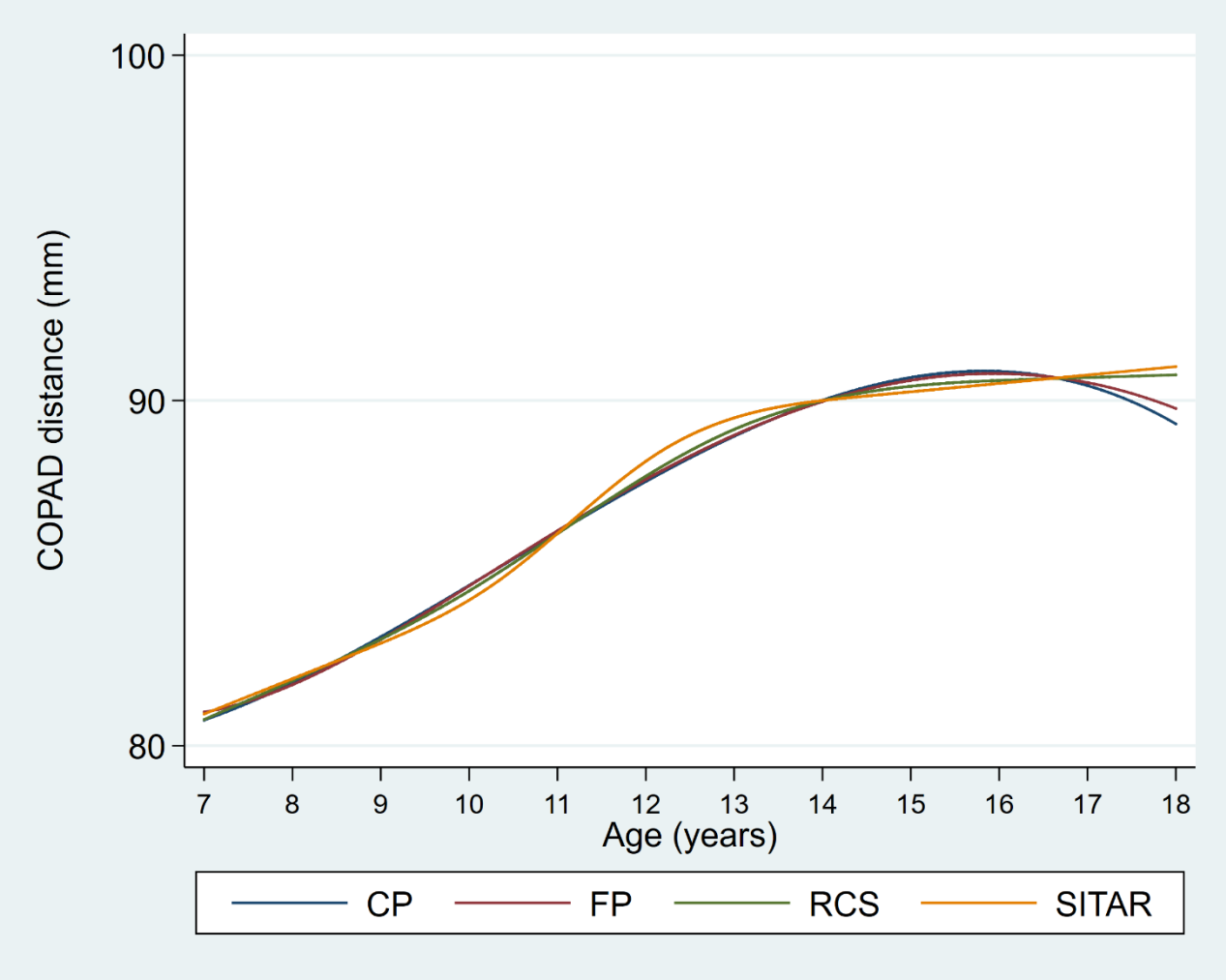


Figure 6-2. Population-average distance curves for the upper jaw length (COPAD) for females.

Figure 6-3. Population-average growth velocity curves for the upper jaw length (COPAD) for males.

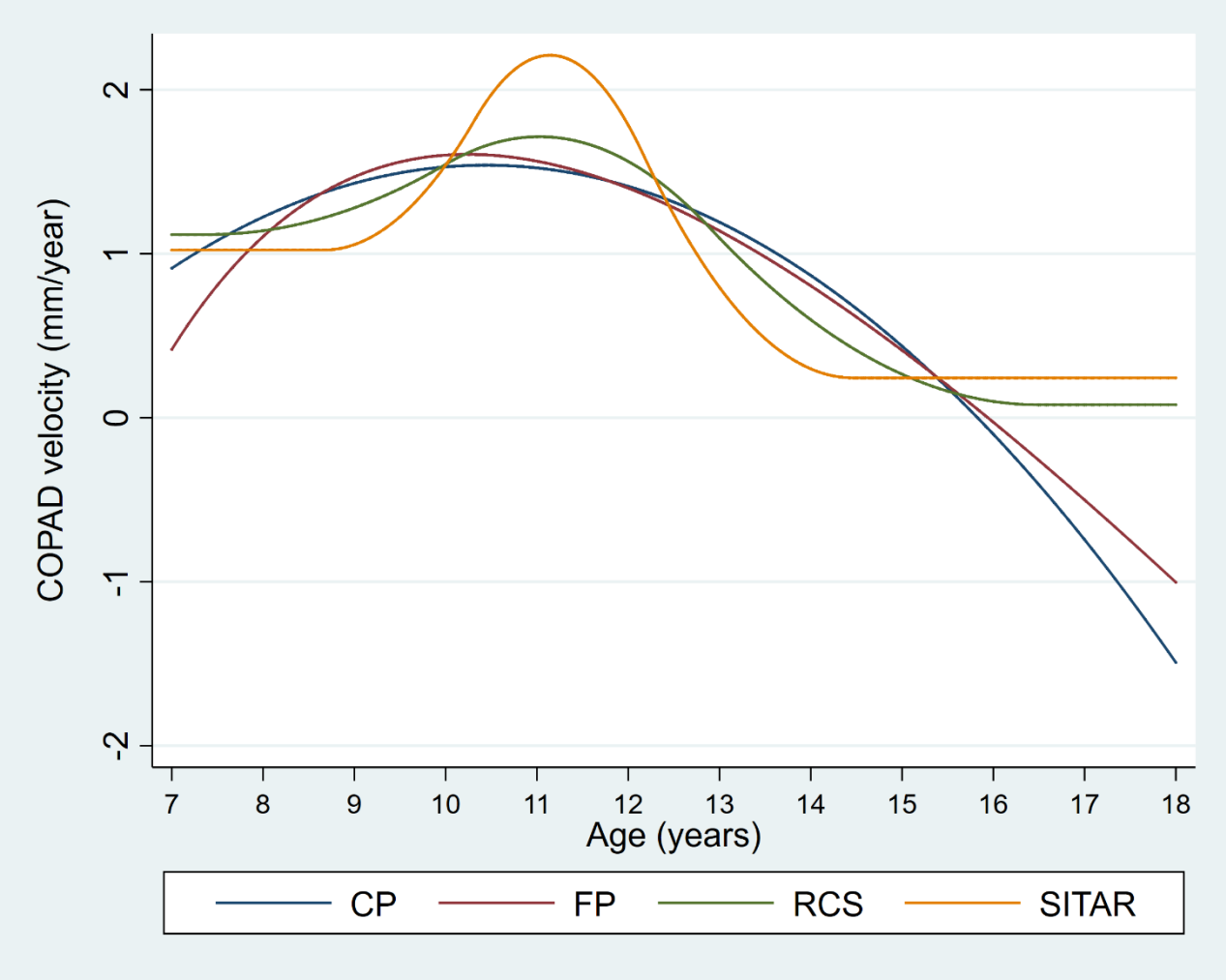


Figure 6-4. Population-average growth velocity curves for the upper jaw length (COPAD) for females.

Distance curves for lower jaw length are shown in Figure 6-5 and Figure 6-6 for males and females, respectively. Lower jaw length growth velocity curves for males are shown in Figure 6-7 whereas Figure 6-8 shows growth velocity curves for females. Figure 6-9 and Figure 6-10 show distance curves for total face height for males and females respectively. Growth velocity curves for total face height are presented in Figure 6-11 and Figure 6-12 for males and females respectively .

Like the upper jaw length, both RCS and SITAR GCMs approximated an asymptote for lower jaw length and total face height for both sexes. As for the upper jaw length, the growth trajectories and derivatives for the lower jaw length and total face height estimated by CP GCM showed a decline in jaw size and negative growth velocities for both males and females. While the FP GCM applied to female data performed like CP GCM in estimating negative growth velocities for lower jaw length and total face height, the FP GCM performed slightly better (lesser negative growth velocity) than the CP GCM for lower jaw length and total face height for males.

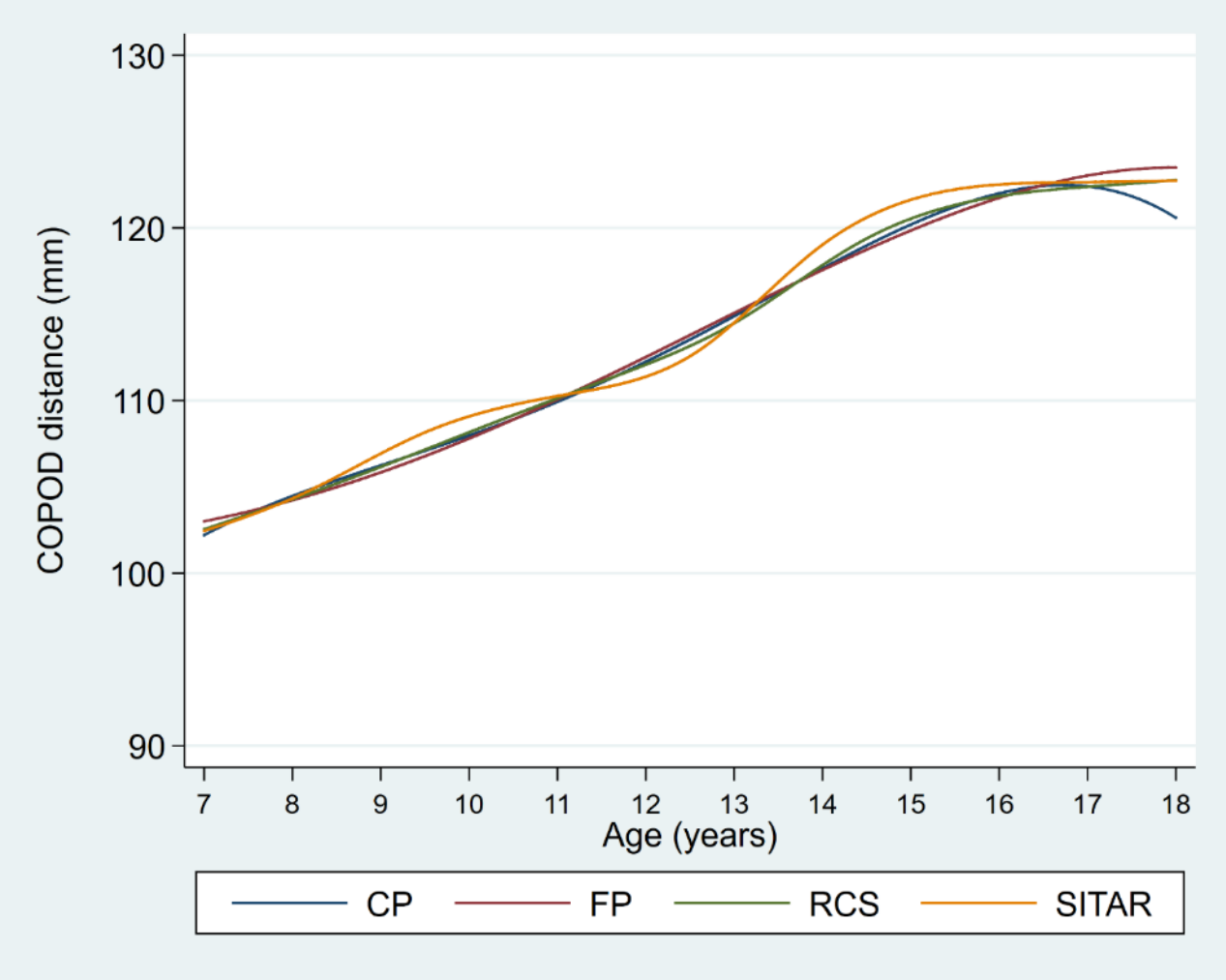


Figure 6-5. Population-average distance curves for the lower jaw length (COPOD) for males.

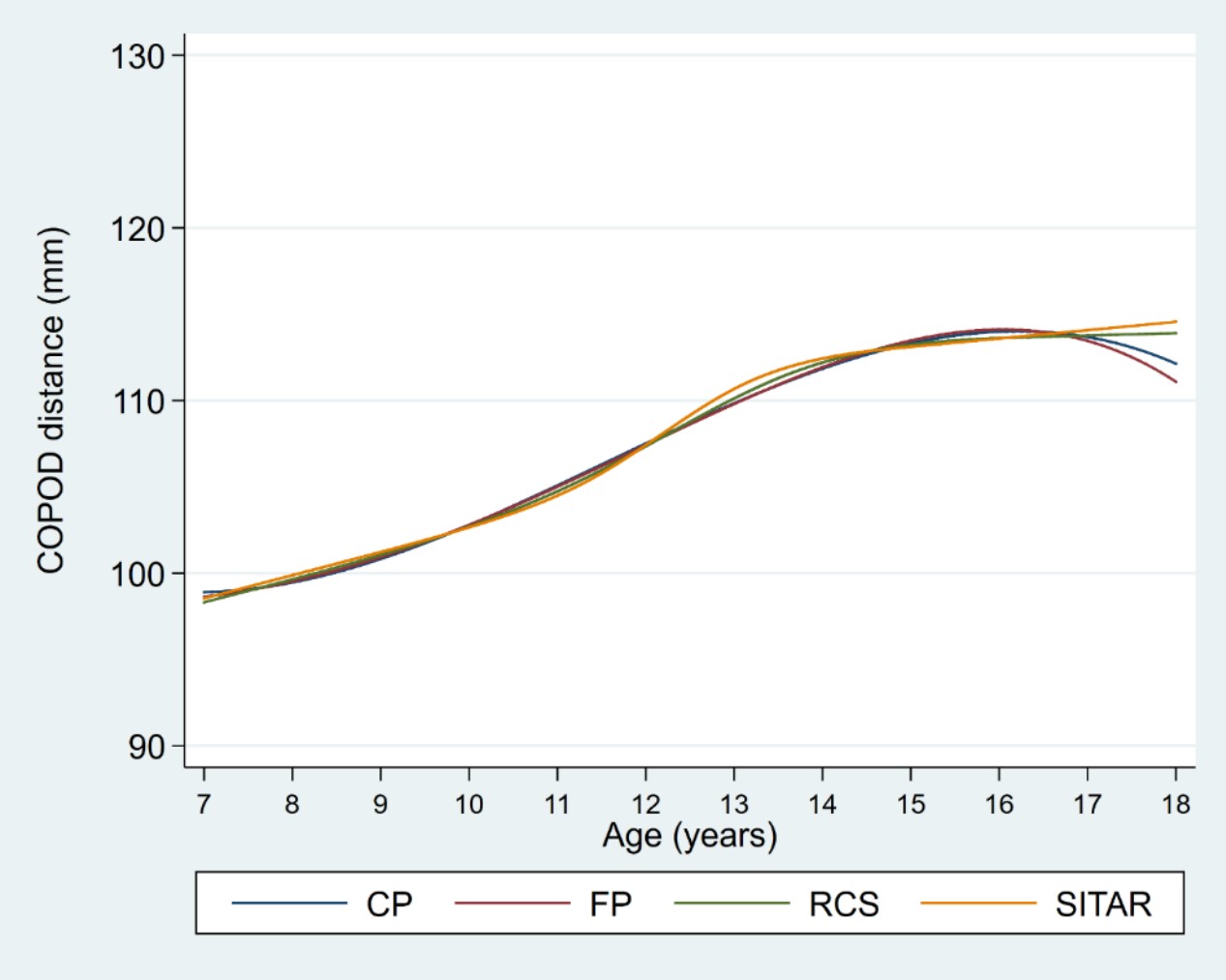


Figure 6-6. Population-average distance curves for the lower jaw length (COPOD) for females.

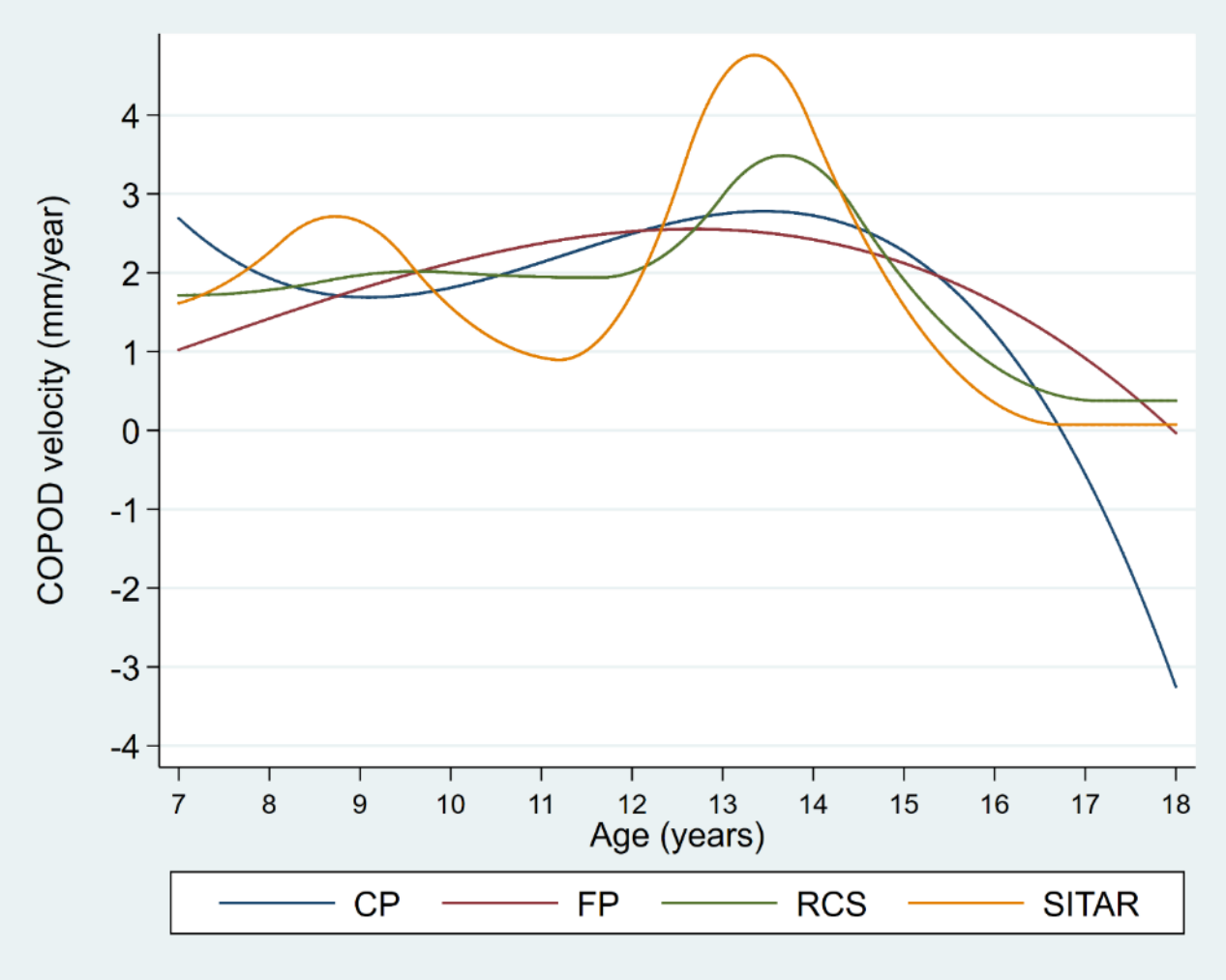


Figure 6-7. Population-average growth velocity curves for the lower jaw length (COPOD) for males.

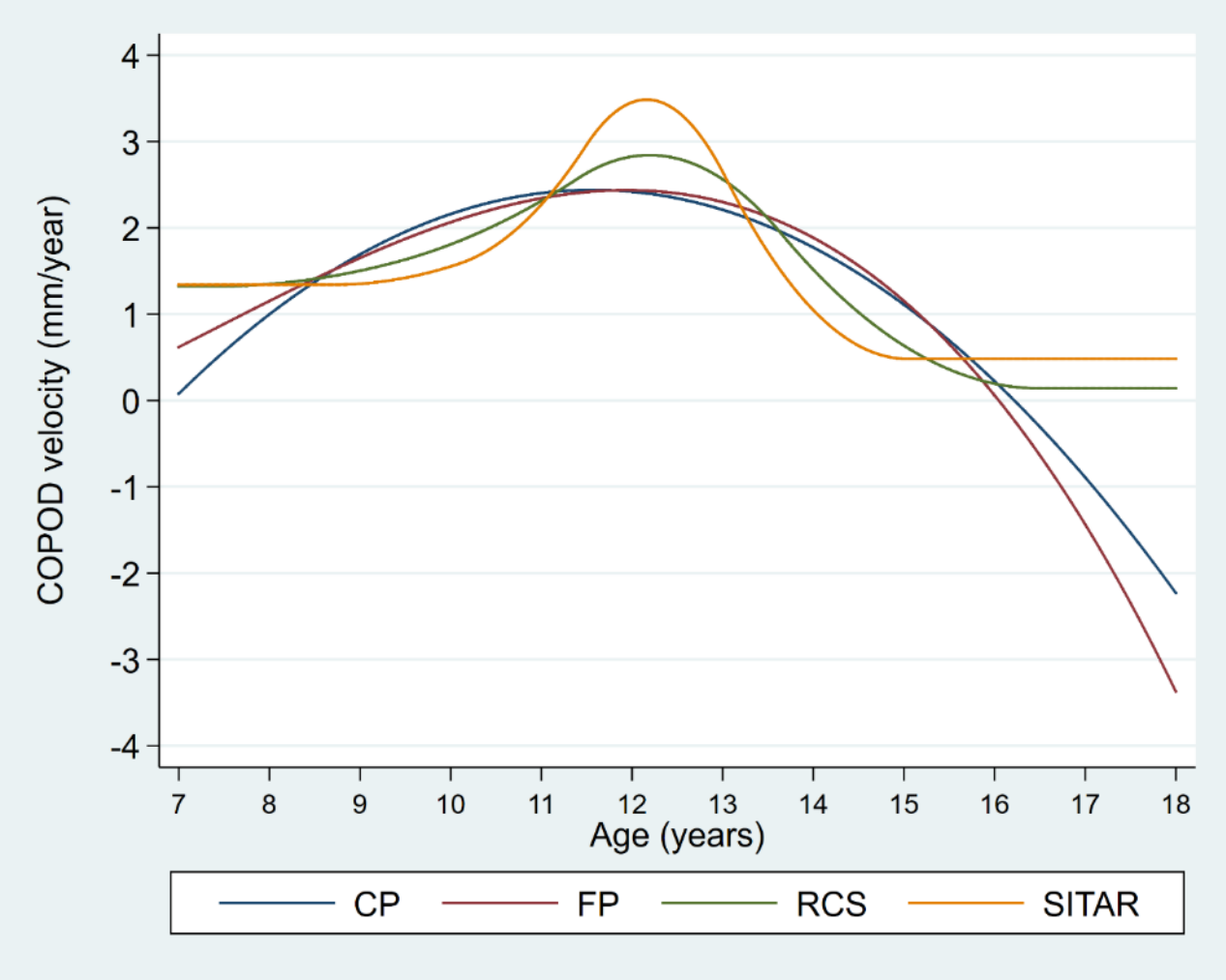


Figure 6-8. Population-average growth velocity curves for the lower jaw length (COPOD) for females.

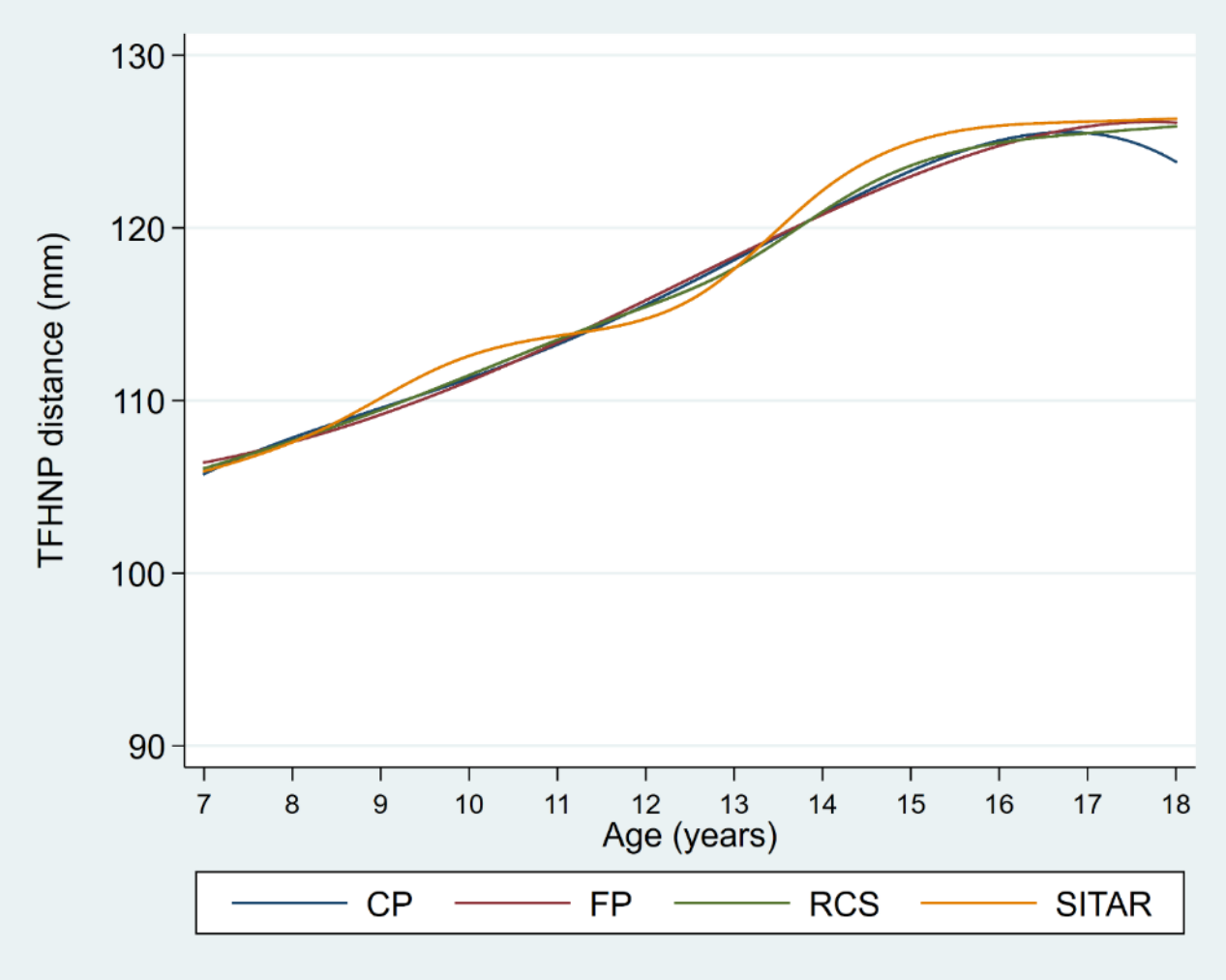


Figure 6-9. Population-average distance curves for the total face height (TFHNP) for males.

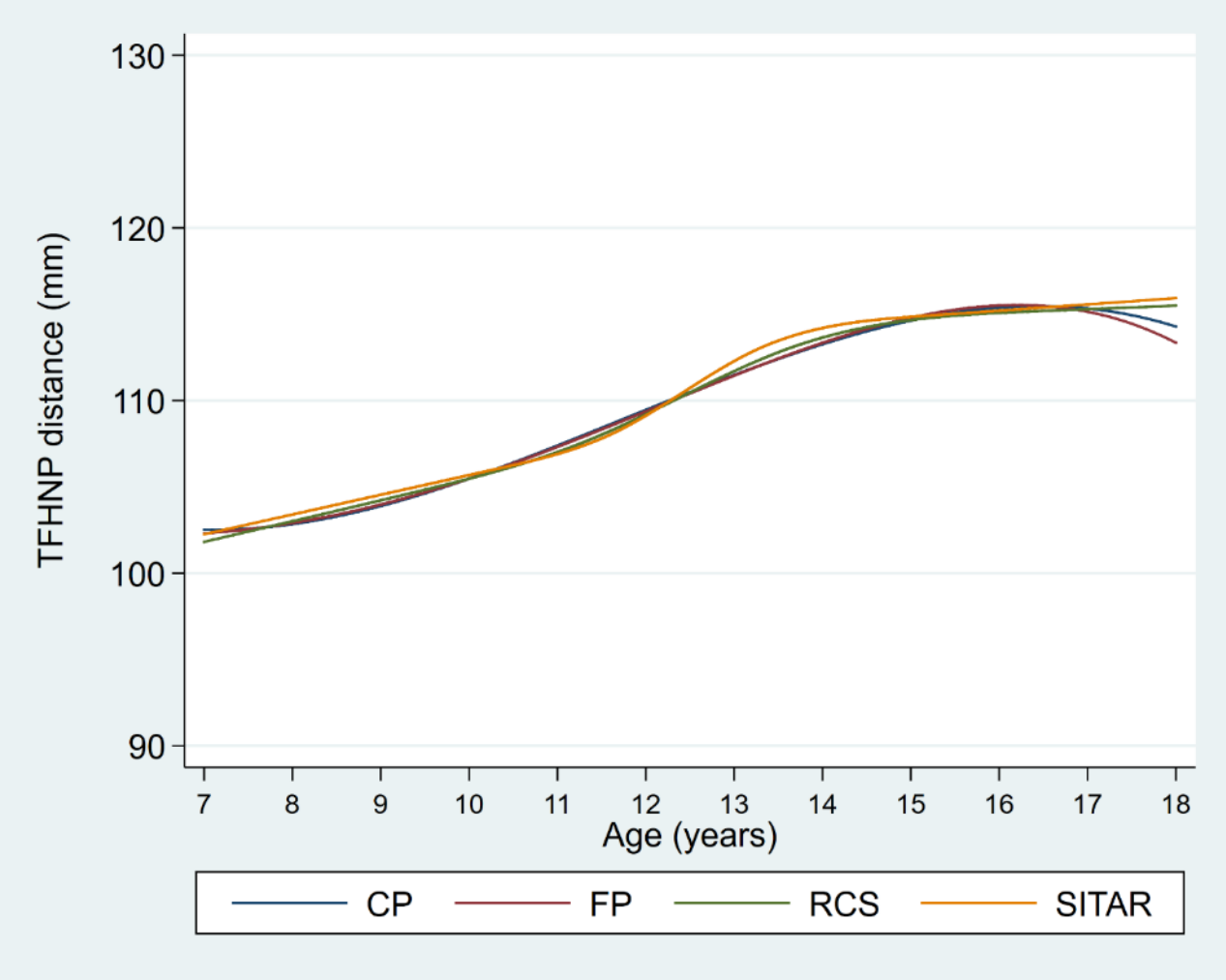


Figure 6-10. Population-average distance curves for the total face height (TFHNP) for females.

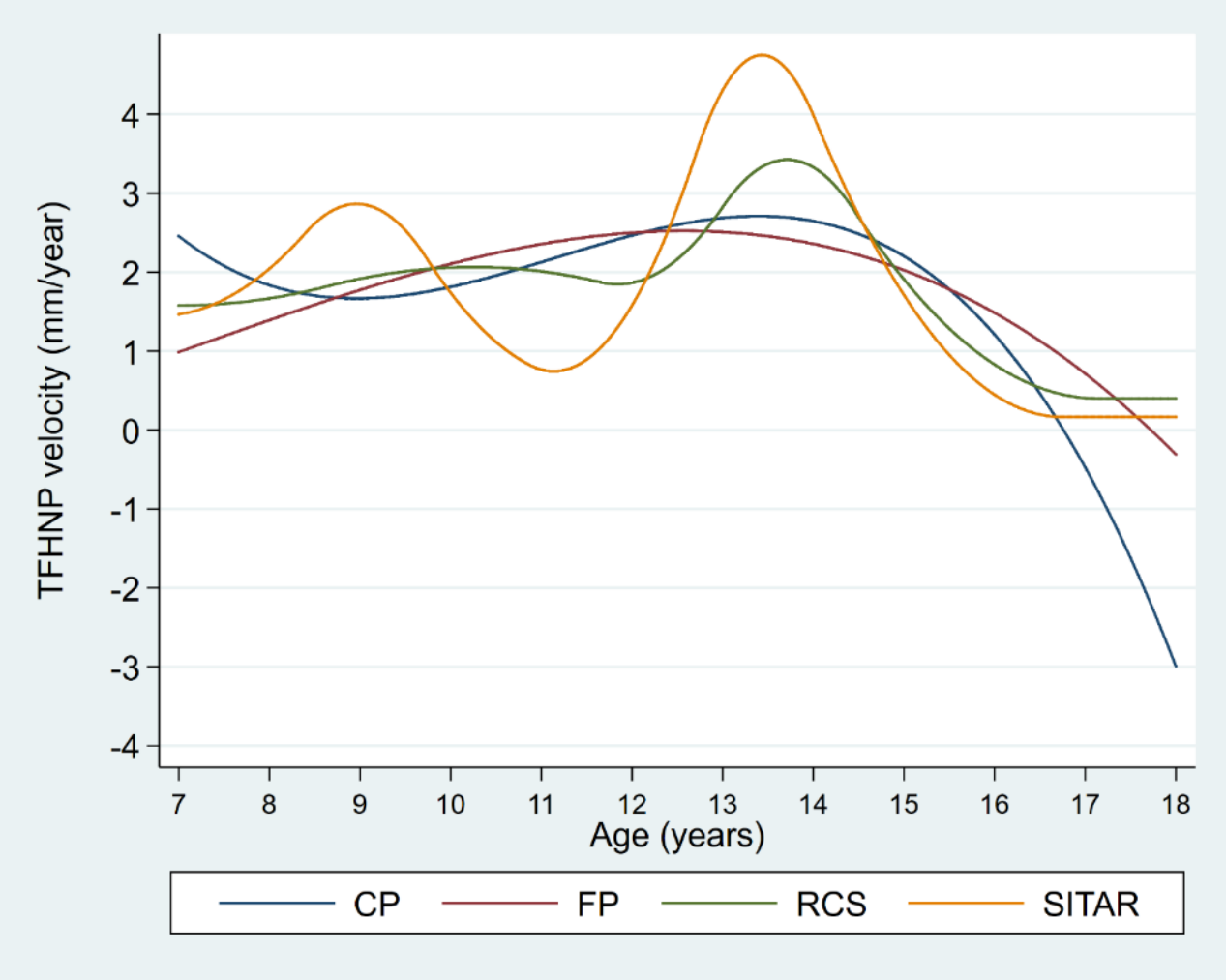


Figure 6-11. Population-average growth velocity curves for the total face height (TFHNP) for males.

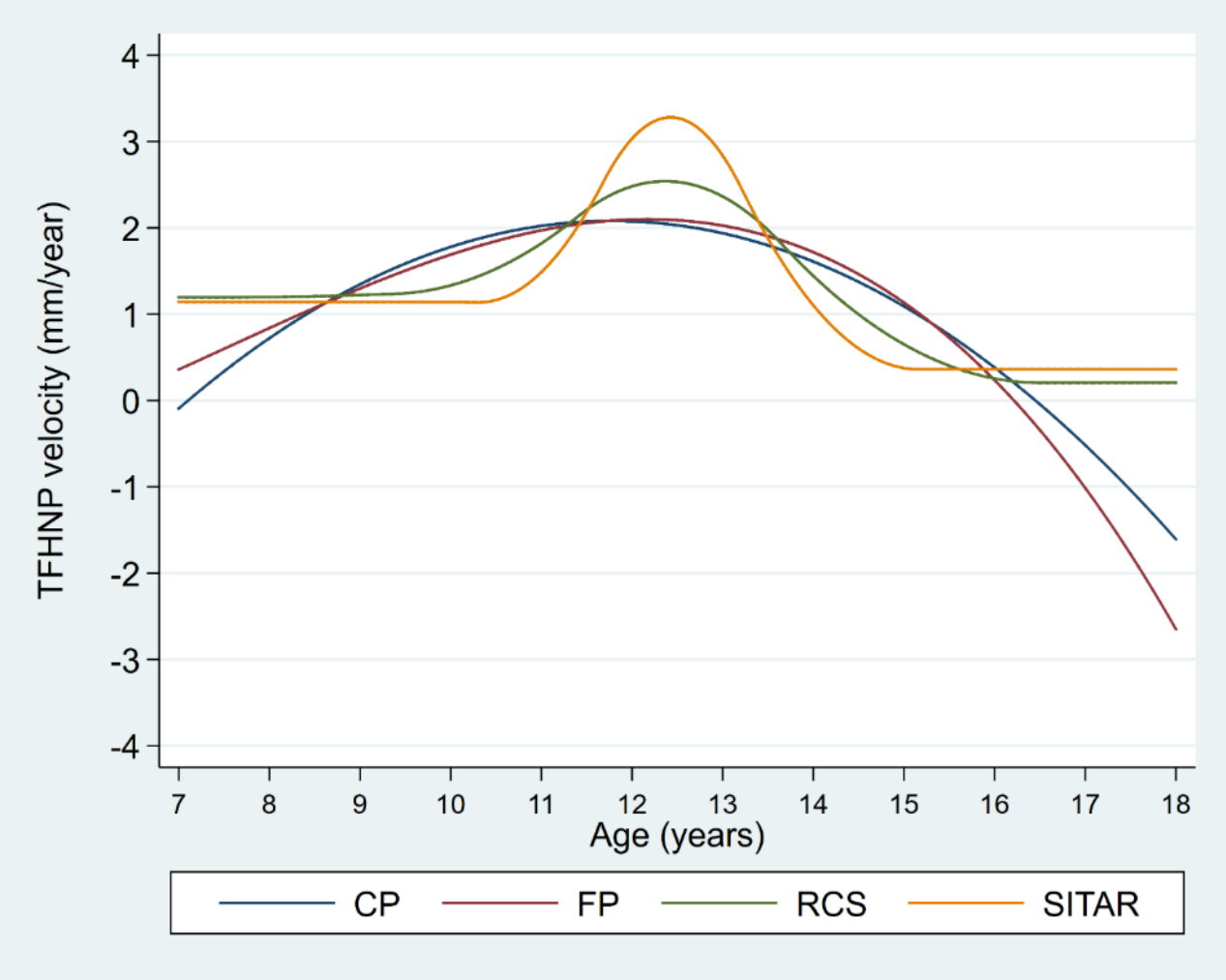


Figure 6-12. Population-average growth velocity curves for the total face height (TFHNP) for females.

6.3.3. Timing and intensity of the adolescent growth spurt

Population-average and individual-specific estimates of growth timing (APGV) and intensity (PGV) were calculated from derivatives using the quadratic function method (see Chapter 4 for details). Before summarising the APGV and PGV results, I first show population-average and individual-specific (for five randomly selected individuals) APGV and PGV estimates mapped on to the growth trajectories. In these illustrative figures, the APGV and PGV are denoted using vertical and horizontal dashed lines respectively.

For upper jaw length, population-average and individual-specific estimates of APGV and PGV for males are shown in Figure 6-13 and Figure 6-14. Population-average and individual-specific estimates of APGV and PGV for females are shown in Figure 6-15 and Figure 6-16.

Population-average and individual-specific estimates of APGV and PGV for lower jaw length are shown in Figure 6-17 and Figure 6-18 for males, and Figure 6-19 and Figure 6-20 for females. Results showing population-average and individual-specific estimates of APGV and PGV for total face height are displayed in Figure 6-21 and Figure 6-22 for males, and Figure 6-23 and Figure 6-24 for females.

For all three-growth measures (i.e., upper jaw length, lower jaw length and total face height), the timing of the adolescent growth spurt (APGV) is earlier for females than males. The intensity of the adolescent growth spurt (PGV) is higher for males than females. Sex differences in the timing and intensity of the adolescent growth spurt are greater for lower jaw length and total face height than for upper jaw length.

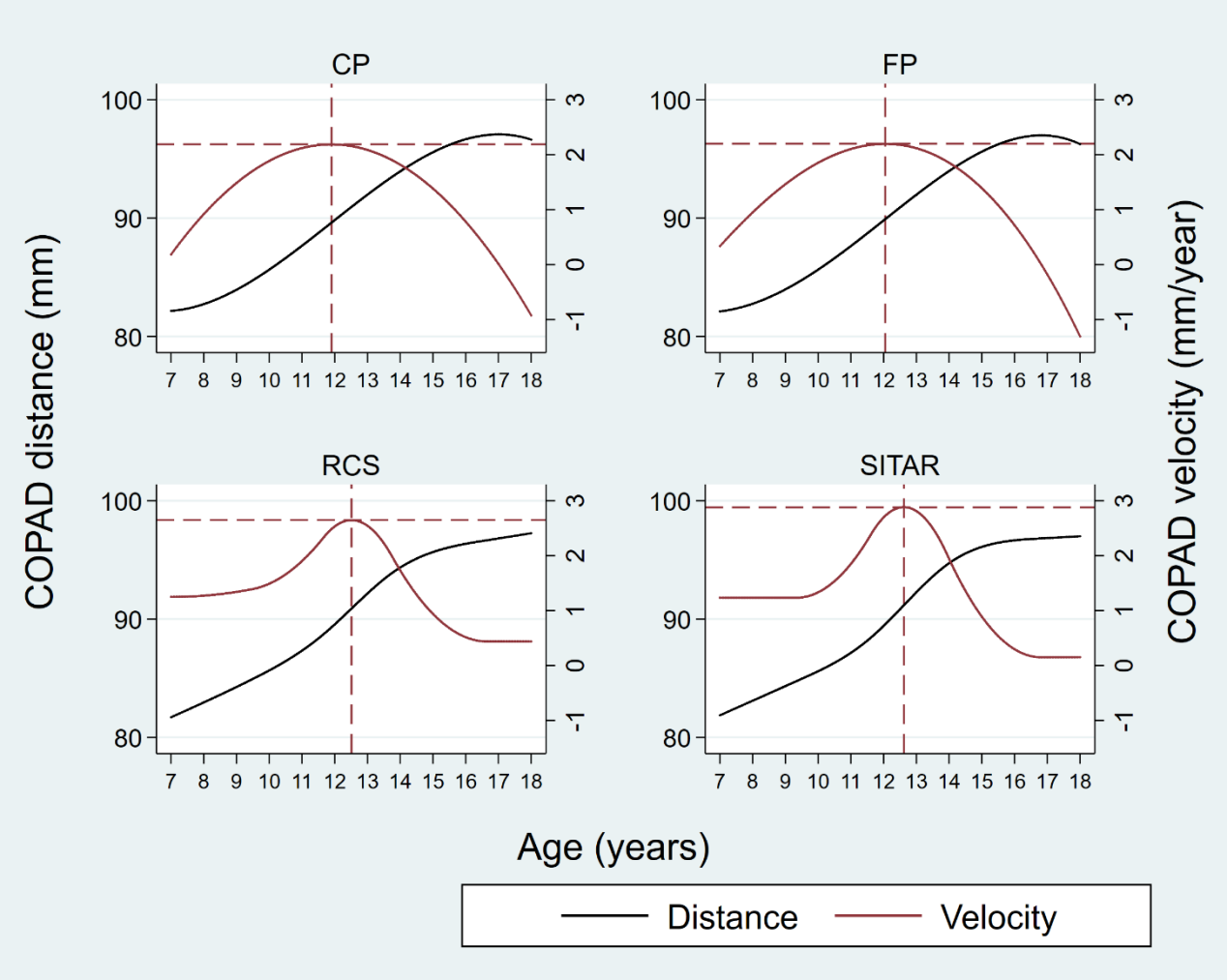


Figure 6-13. Population-average age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the upper jaw length (COPAD) for males. The APGV is denoted by vertical dash lines whereas the PGV is shown by the horizontal dash lines. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

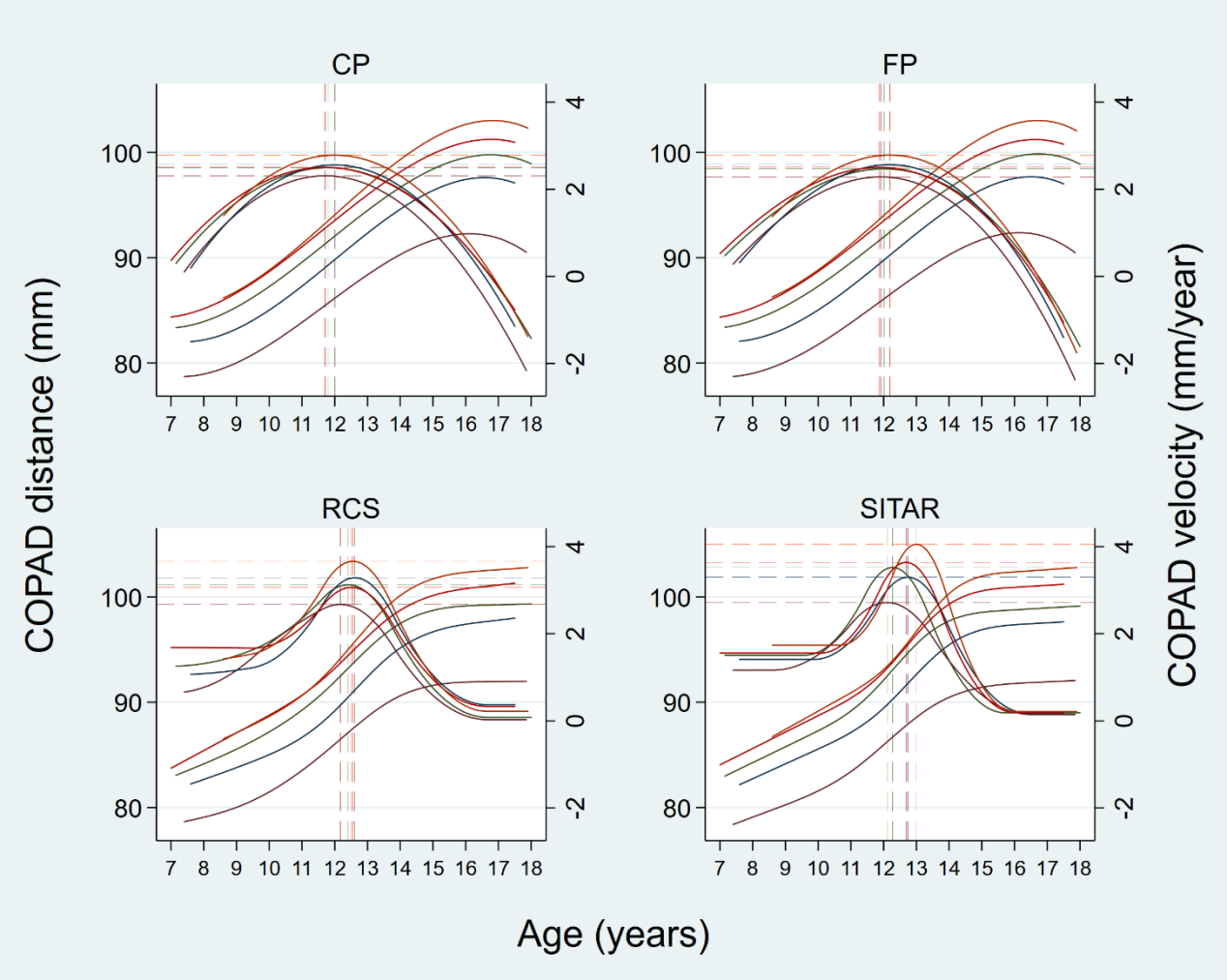


Figure 6-14. Individual-specific age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the upper jaw length (COPAD) for males. Results are shown for five randomly selected individuals. The colour-matched vertical and horizontal dash lines indicate the APGV and PGV for an individual. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

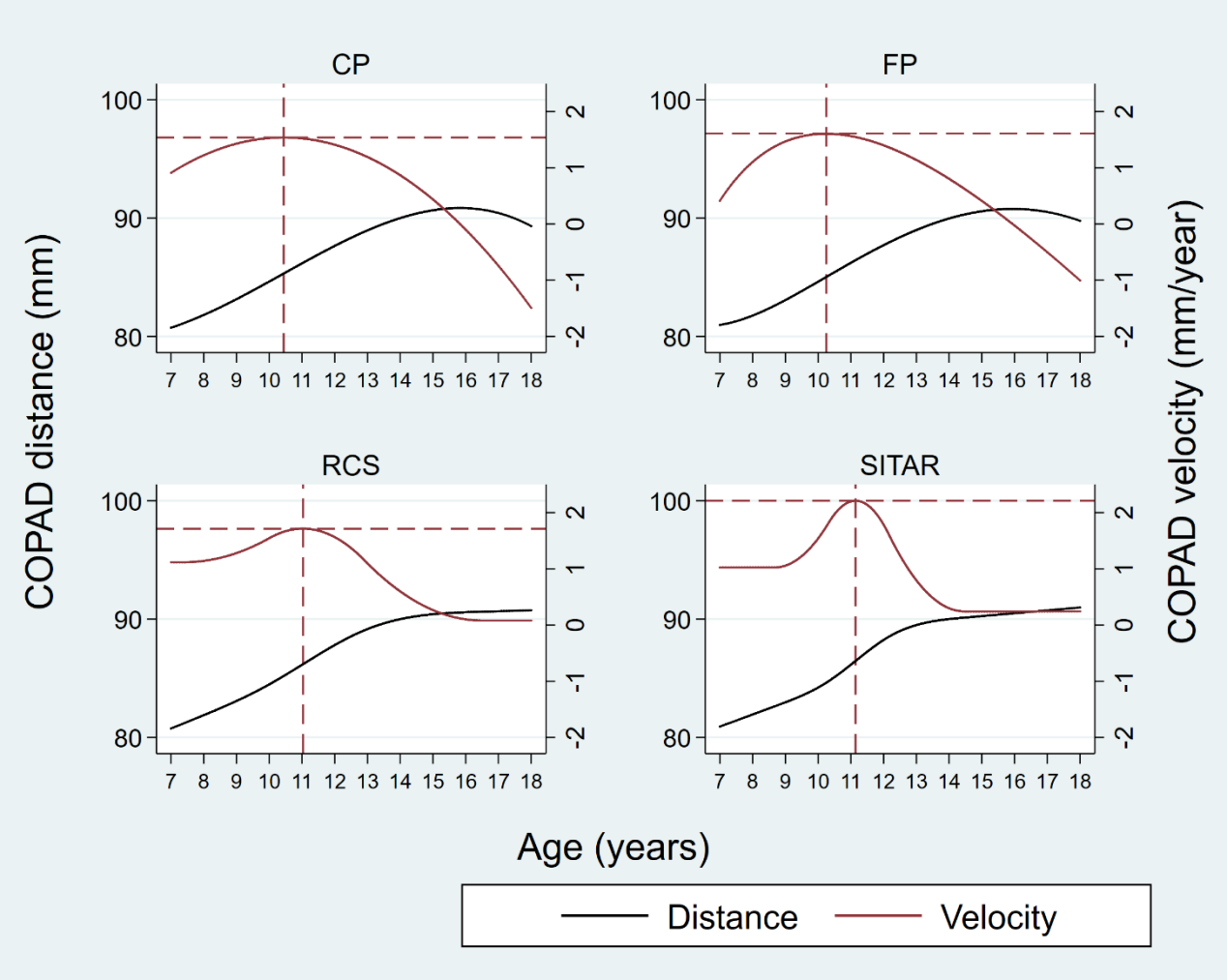


Figure 6-15. Population-average age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the upper jaw length (COPAD) for females. The APGV is denoted by vertical dash lines whereas the PGV is shown by the horizontal dash lines. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

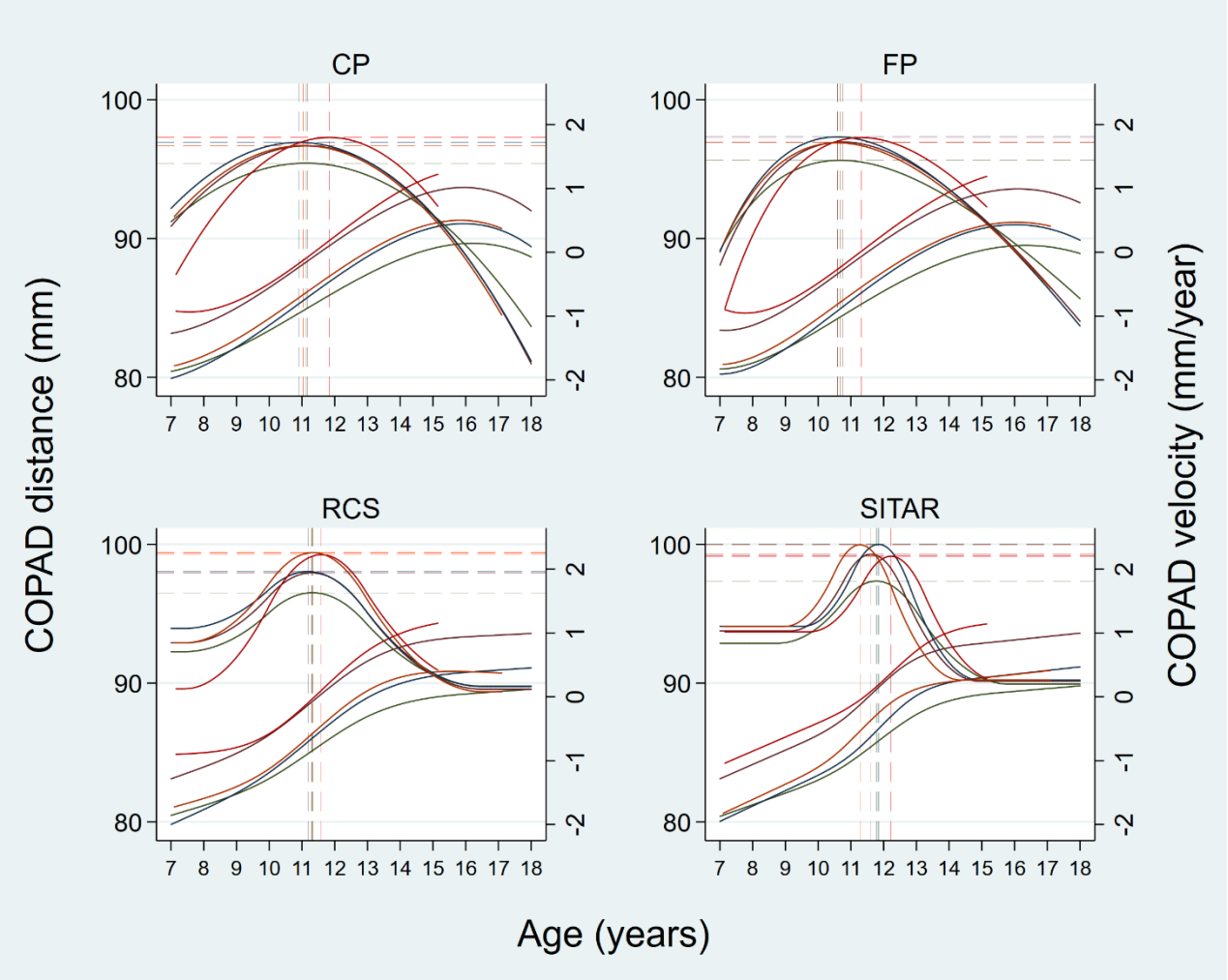


Figure 6-16. Individual-specific age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the upper jaw length (COPAD) for females. Results are shown for five randomly selected individuals. The colour-matched vertical and horizontal dash lines indicate the APGV and PGV for an individual. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

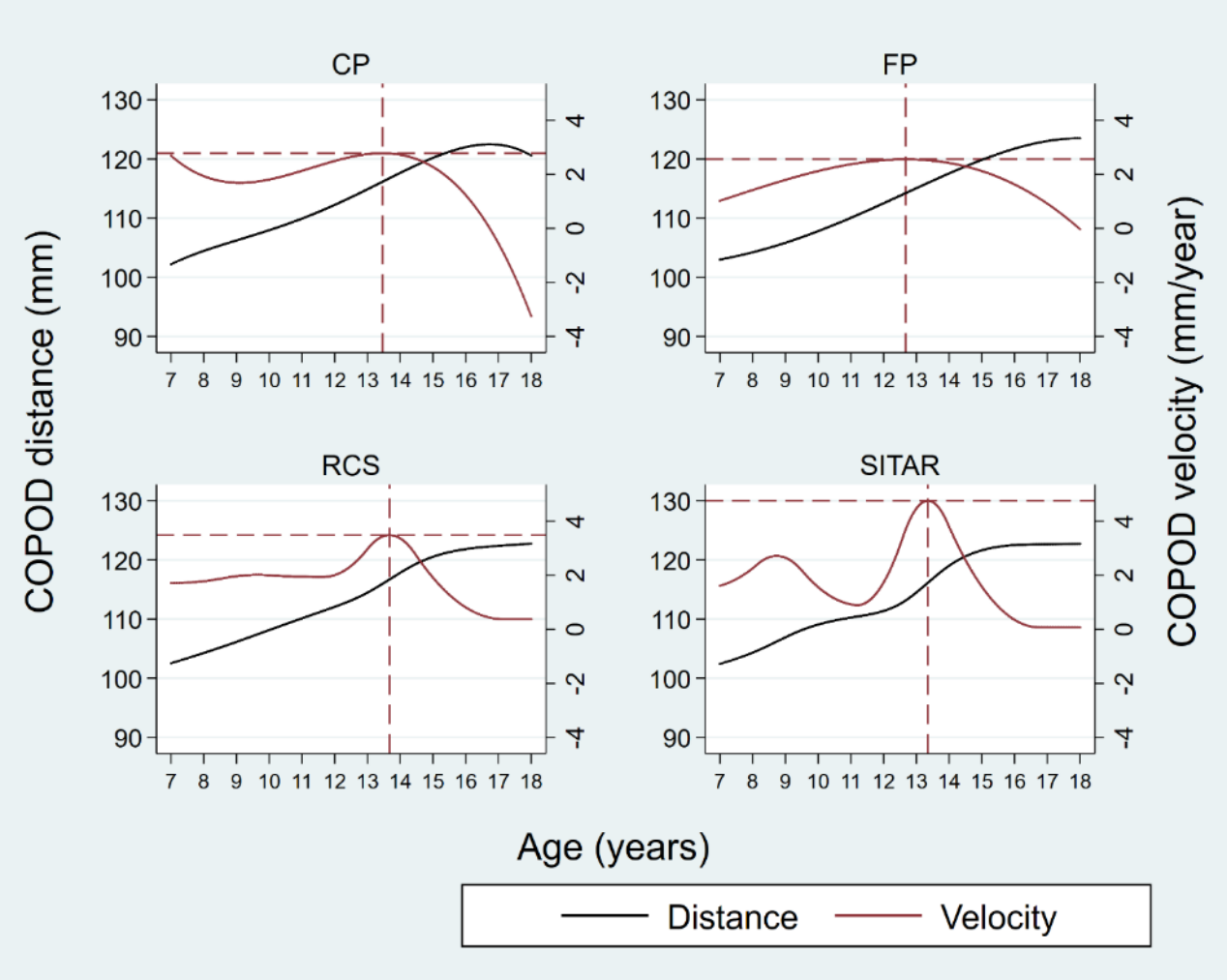


Figure 6-17. Population-average age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the lower jaw length (COPOD) for males. The APGV is denoted by vertical dash lines whereas the PGV is shown by the horizontal dash lines. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

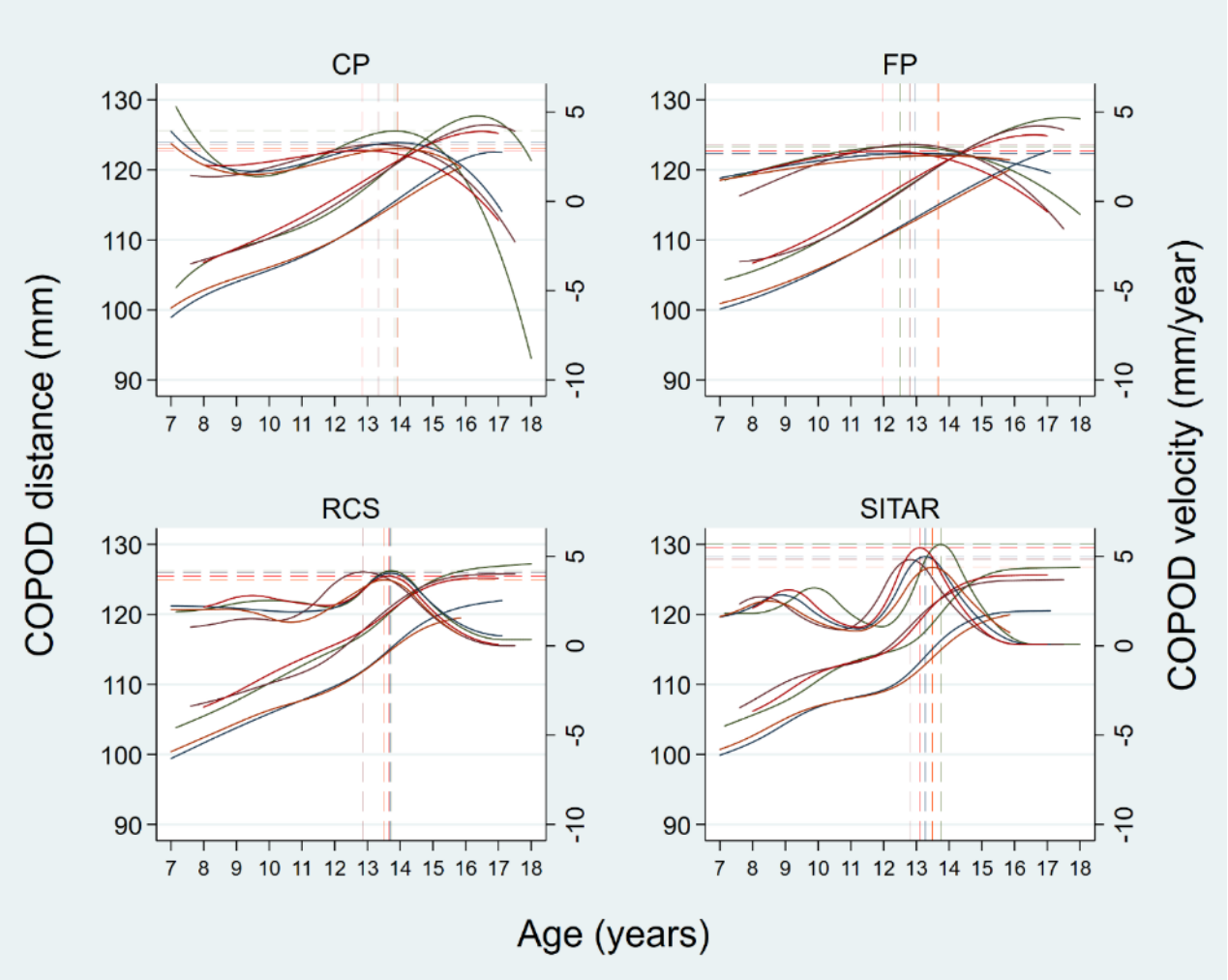


Figure 6-18. Individual-specific age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the lower jaw length (COPOD) for males. Results are shown for five randomly selected individuals. The colour-matched vertical and horizontal dash lines indicate the APGV and PGV for an individual. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

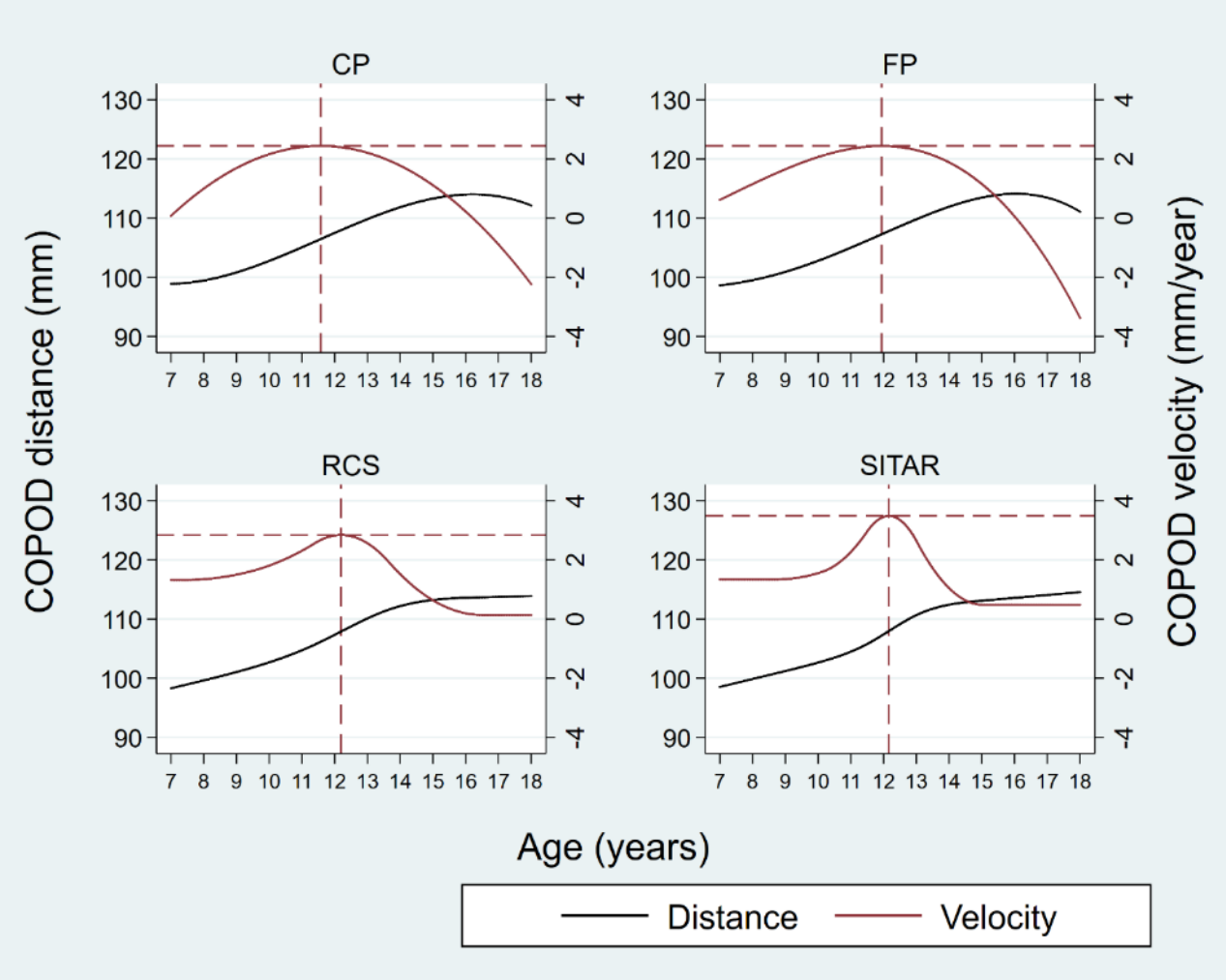


Figure 6-19. Population-average age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the lower jaw length (COPOD) for females. The APGV is denoted by vertical dash lines whereas the PGV is shown by the horizontal dash lines. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

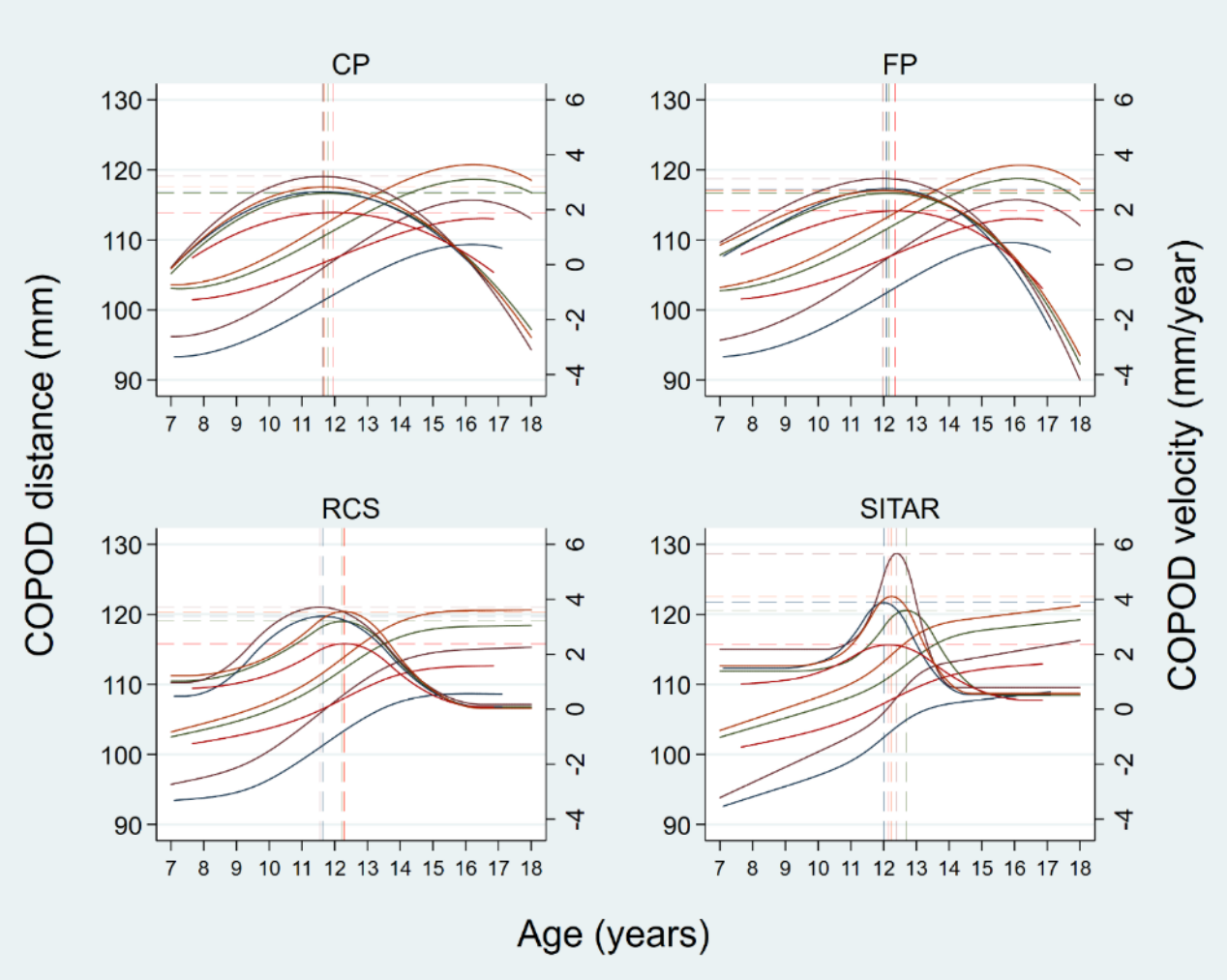


Figure 6-20. Individual-specific age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the lower jaw length (COPOD) for females. Results are shown for five randomly selected individuals. The colour-matched vertical and horizontal dash lines indicate the APGV and PGV for an individual. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

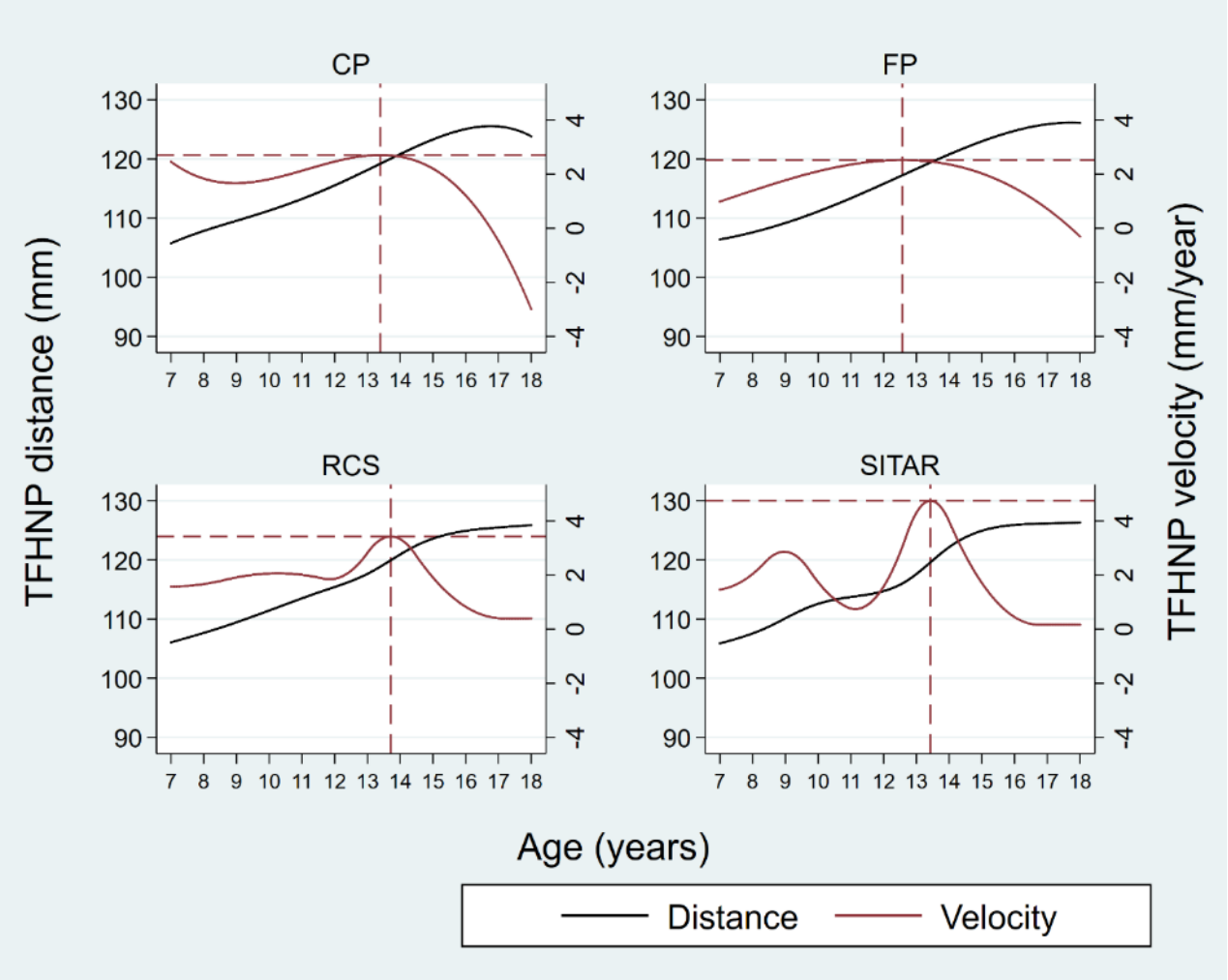


Figure 6-21. Population-average age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the total face height (TFHNP) for males. The APGV is denoted by vertical dash lines whereas the PGV is shown by the horizontal dash lines. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

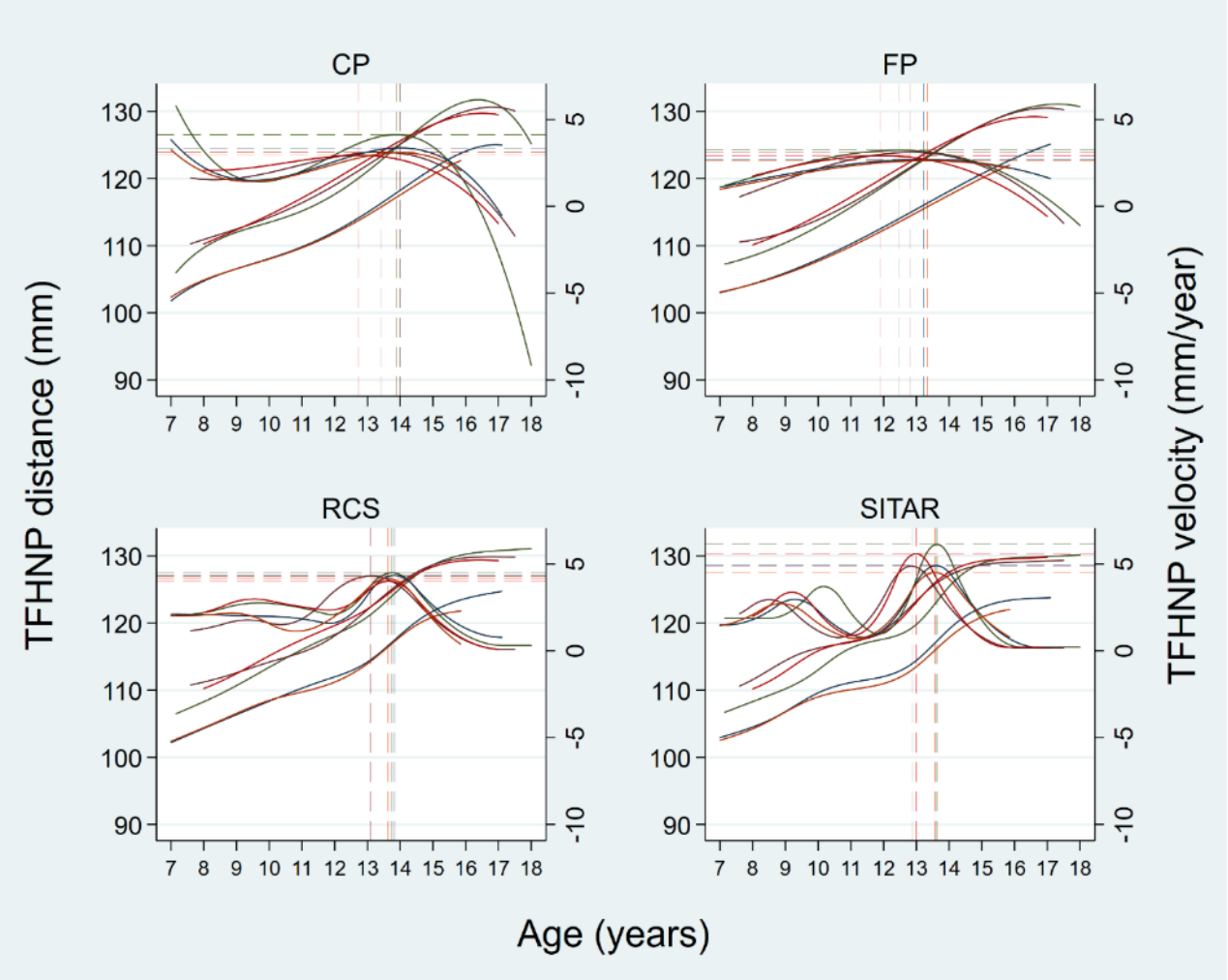


Figure 6-22. Individual-specific age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the total face height (TFHNP) for males. Results are shown for five randomly selected individuals. The colour-matched vertical and horizontal dash lines indicate the APGV and PGV for an individual. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

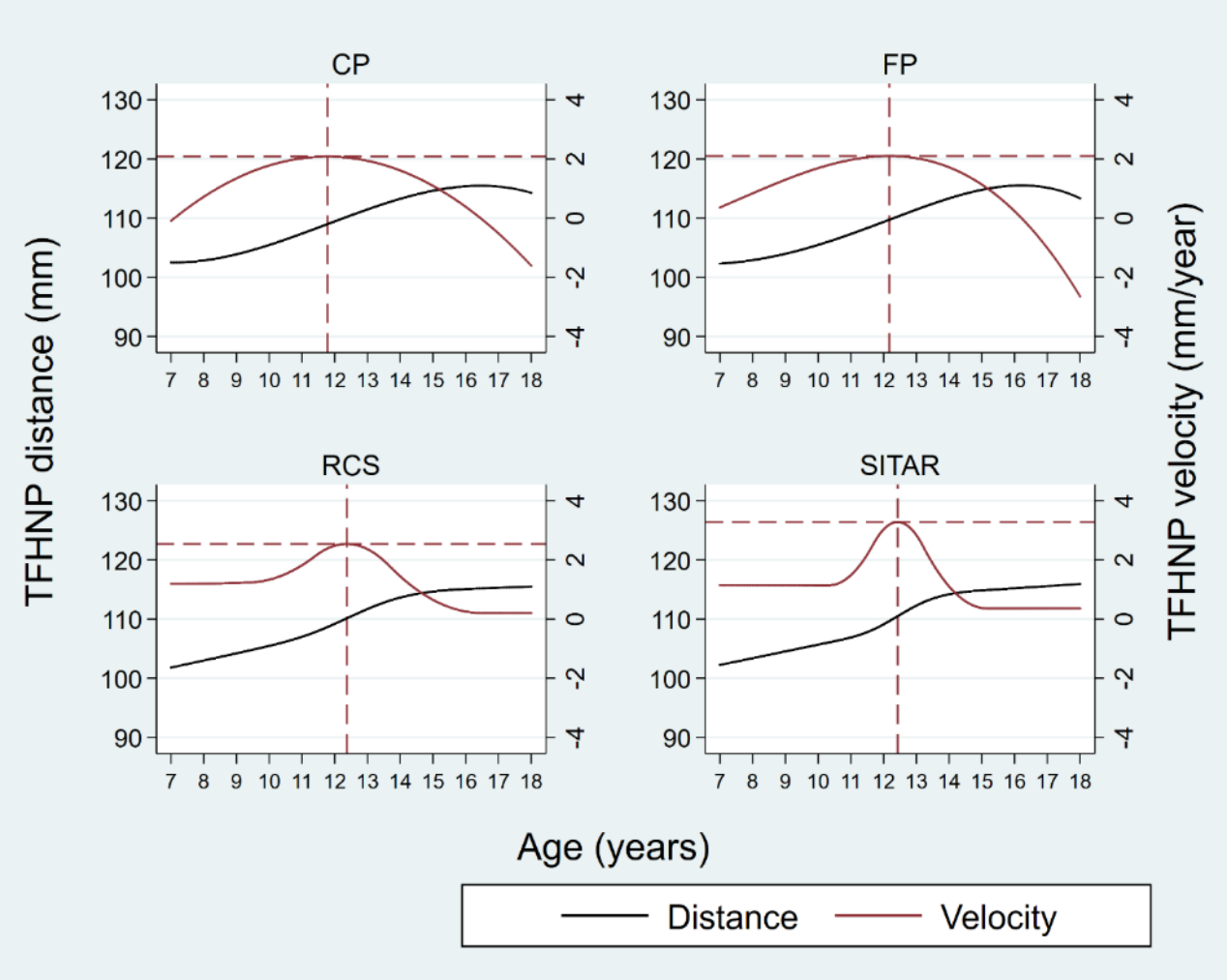


Figure 6-23. Population-average age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the total face height (TFHNP) for females. The APGV is denoted by vertical dash lines whereas the PGV is shown by the horizontal dash lines. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

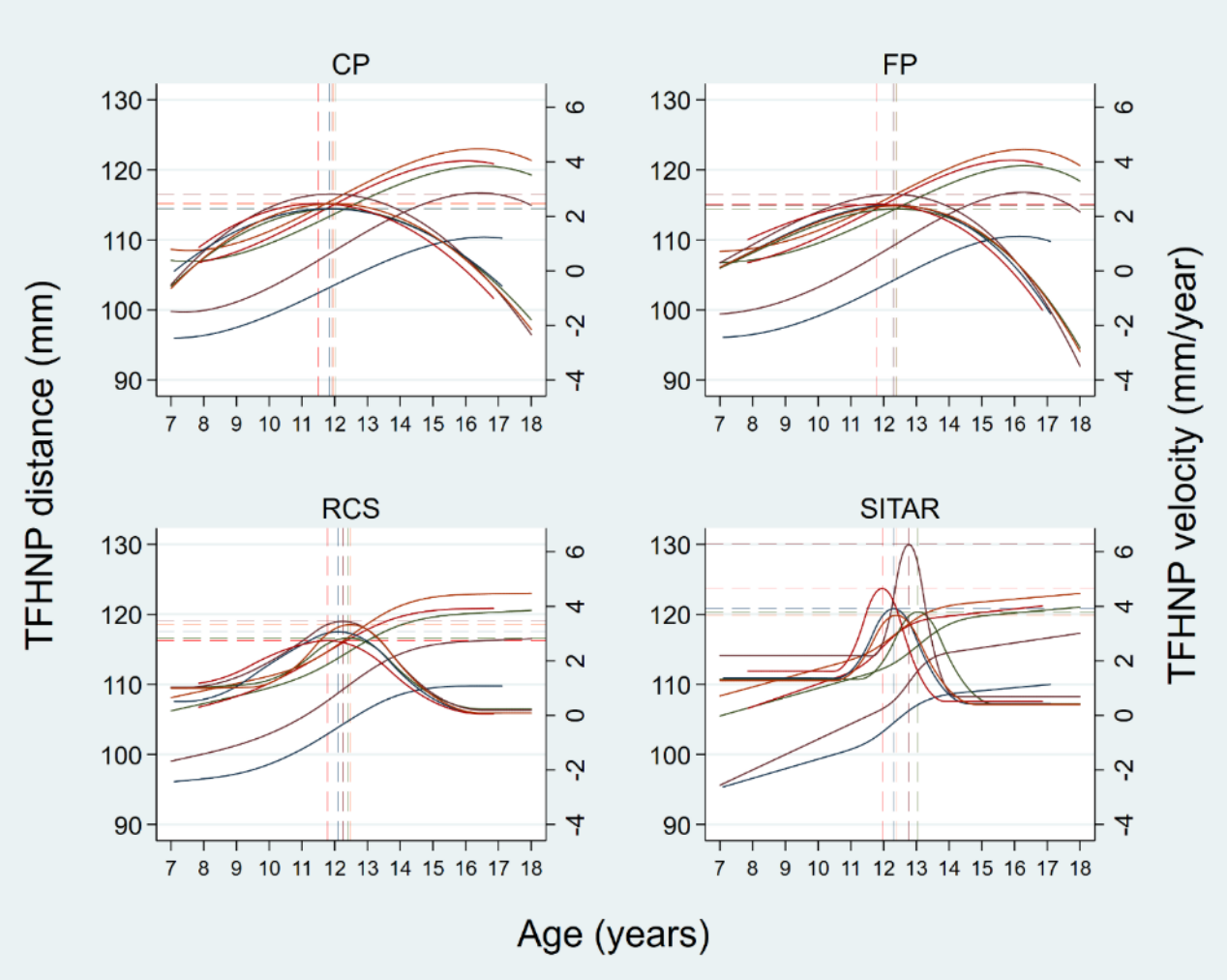


Figure 6-24. Individual-specific age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by linear and nonlinear growth curve models for the total face height (TFHNP) for females. Results are shown for five randomly selected individuals. The colour-matched vertical and horizontal dash lines indicate the APGV and PGV for an individual. The linear growth curve models are conventional polynomial (CP), fractional polynomial (FP) and restricted cubic spline (RCS). The nonlinear growth curve model is the superimposition by translation and rotation (SITAR).

Growth timing (APGV)

Population-average

Population-average estimates and bootstrapped 95% CIs for APGV are shown below in Table 6-4. Except for lower jaw length and total face height for males, the RCS and SITAR GCMs provided very similar estimates of APGV, with a difference of less than 1.5 months, i.e. 0.12 years. For lower jaw length and total face height for males, the difference between RCS and SITAR GCMs was 3.5–4 months (0.28–0.32 years). The APGV estimated by the SITAR GCM was lower than the RCS GCM.

Table 6-4. Summary of population-average age at peak growth velocity (APGV) estimated by linear and nonlinear growth curve models applied to male and female data.

Outcome	Age at peak growth velocity (APGV) in years (Bootstrapped 95% CIs)			
	CP	FP	RCS	SITAR
Male				
Upper jaw length (COPAD)	11.90 (11.51 to 12.11)	12.05 (11.27 to 12.71)	12.51 (12.32 to 12.74)	12.62 (12.29 to 13.04)
Lower jaw length (COPOD)	13.46 (12.74 to 13.70)	12.67 (12.20 to 13.40)	13.67 (13.42 to 13.86)	13.35 (12.71 to 13.78)
Total face height (TFHNP)	13.39 (12.91 to 13.97)	12.57 (12.11 to 13.19)	13.71 (13.51 to 14.01)	13.43 (12.78 to 13.91)
Female				
Upper jaw length (COPAD)	10.44 (9.48 to 10.83)	10.25 (9.12 to 10.85)	11.03 (10.85 to 11.23)	11.14 (10.69 to 11.15)
Lower jaw length (COPOD)	11.57 (11.29 to 11.74)	11.94 (11.43 to 12.39)	12.19 (11.97 to 12.38)	12.16 (11.79 to 12.53)
Total face height (TFHNP)	11.78 (11.59 to 12.58)	12.17 (11.97 to 12.91)	12.37 (12.21 to 12.67)	12.43 (11.99 to 12.81)

CP conventional polynomial; FP fractional polynomial; RCS restricted cubic spline; SITAR superimposition by translation and rotation.

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

The APGV estimated by the CP GCM for lower jaw length and total face height for males was closer to the RCS and SITAR GCMs (within 3.5 months or 0.32 years) than upper jaw length for males and all three outcomes (i.e., upper jaw length, lower jaw length and total face height) for females (approximately 6 months or 0.5 years). This could be because of different degree of CP GCM applied to different outcomes for males and females. Unlike a fourth-degree CP GCM which

fits best to lower jaw length and total face height for males, a third-degree CP GCM fits best to upper jaw length for males and all three outcomes for females.

For lower jaw length and total face height for males, the APGV estimated by the FP GCM was lower than those estimated by CP, RCS and SITAR GCMs. This difference in APGV was greatest between RCS and FP GCMs for lower jaw length (1.14 years), followed by total face height (almost 0.86 years). These findings show that the APGV estimated by a third-degree FP GCM, even with higher order repeated powers (powers 3 3 3), is substantially lower than a fourth-degree CP GCM and cubic spline-based linear (RCS) and nonlinear (SITAR) GCMs.

The range of bootstrapped 95% CI width across all outcomes for both sexes was smallest for the RCS GCM (0.38 to 0.50 years). Except for the RCS GCM, the width of CIs estimated by CP, FP and SITAR GCMs exceeded 0.50 years for all outcomes for both males and females.

Individual-specific

Table 6-5 below summarises the individual-specific APGV for males and females.

Table 6-5. Summary of individual-specific age at peak growth velocity (APGV) estimated by linear and nonlinear growth curve models applied to male and female data.

Outcome	Median age at peak growth velocity (APGV) in years (Interquartile range)			
	CP	FP	RCS	SITAR
Male				
Upper jaw length (COPAD)	11.97 (11.54 to 12.38)	12.11 (11.70 to 12.52)	12.52 (12.23 to 12.70)	12.56 (12.12 to 12.95)
Lower jaw length (COPOD)	13.30 (12.30 to 13.95)	12.60 (11.91 to 13.34)	13.42 (12.18 to 13.85)	13.14 (12.24 to 13.92)
Total face height (TFHNP)	13.24 (12.03 to 13.94)	12.58 (11.84 to 13.37)	13.35 (11.81 to 13.85)	12.99 (11.48 to 13.68)
Female				
Upper jaw length (COPAD)	10.51 (10.57 to 11.03)	10.28 (9.79 to 10.67)	11.06 (10.69 to 11.27)	11.18 (10.70 to 11.63)
Lower jaw length (COPOD)	11.56 (11.33 to 11.85)	11.91 (11.66 to 12.26)	12.18 (11.48 to 12.46)	12.21 (11.66 to 12.69)
Total face height (TFHNP)	11.80 (11.54 to 12.05)	12.19 (11.91 to 12.47)	12.37 (12.03 to 12.54)	12.41 (11.87 to 12.81)

CP conventional polynomial; FP fractional polynomial; RCS restricted cubic spline; SITAR superimposition by translation and rotation.

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

For males, the between-individual variability in APGV (the interquartile range: 75th percentile – 25th percentile) estimated by all four GCMs (CP, FP, RCS and SITAR) was lowest for the upper jaw length and highest for the total face height. The lowest and highest between-individual variability for females were for upper jaw length and lower jaw length, respectively. Comparison of GCMs showed no substantial difference in the between-individual variability for APGV estimated by CP, FP, RCS and SITAR GCMs. This is unlike within individual variability in growth (RSD estimate) which was lowest for the RCS GCM for all three outcomes for both sexes.

Table 6-6 and Table 6-7 show the Spearman correlation for individual-specific APGV for males and females. The correlation coefficients are shown in parentheses, and findings are summarised below.

Table 6-6. Spearman correlation (95% confidence interval) for age at peak growth velocity (APGV) estimated by linear and nonlinear growth curve models applied to male data.

	Upper jaw length (COPAD)				Lower jaw length (COPOD)				Total face height (TFHNP)			
	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR
CP	1				1				1			
FP	1.00 (0.99 to 1.00)	1			0.59 (0.45 to 0.70)	1			0.71 (0.64 to 0.80)	1		
RCS	0.67 (0.55 to 0.76)	0.67 (0.56 to 0.76)	1		0.68 (0.56 to 0.77)	0.44 (0.27 to 0.58)	1		0.75 (0.65 to 0.82)	0.66 (0.54 to 0.75)	1	
SITAR	0.76 (0.76 to 0.83)	0.77 (0.69 to 0.84)	0.88 (0.83 to 0.92)	1	0.60 (0.47 to 0.71)	0.45 (0.28 to 0.59)	0.58 (0.44 to 0.69)	1	0.68 (0.57 to 0.73)	0.57 (0.43 to 0.69)	0.58 (0.45 to 0.70)	1

CP conventional polynomial; FP fractional polynomial; RCS restricted cubic spline; SITAR superimposition by translation and rotation.

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Table 6-7. Spearman correlation (95% confidence interval) for age at peak growth velocity (APGV) estimated by linear and nonlinear growth curve models applied to female data.

	Upper jaw length (COPAD)				Lower jaw length (COPOD)				Total face height (TFHNP)			
	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR
CP	1				1				1			
FP	0.99 (0.98 to 0.99)	1			0.98 (0.97 to 0.99)	1			0.99 (0.99 to 0.99)	1		
RCS	0.88 (0.85 to 0.91)	0.85 (0.82 to 0.88)	1		0.78 (0.71 to 0.84)	0.80 (0.73 to 0.86)	1		0.71 (0.61 to 0.78)	0.73 (0.64 to 0.80)	1	
SITAR	0.83 (0.77 to 0.88)	0.84 (0.78 to 0.88)	0.90 (0.84 to 0.94)	1	0.73 (0.64 to 0.80)	0.74 (0.65 to 0.81)	0.88 (0.84 to 0.91)	1	0.63 (0.51 to 0.72)	0.64 (0.52 to 0.73)	0.79 (0.72 to 0.85)	1

CP conventional polynomial; FP fractional polynomial; RCS restricted cubic spline; SITAR superimposition by translation and rotation.
COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

For males, the individual-specific APGV estimates of upper jaw length by the CP and FP GCMs were perfectly correlated (1.00). Estimates of the SITAR GCM correlated more strongly with the CP and FP GCMs (0.76 to 0.77) than with the RCS GCM (0.67). For lower jaw length and total face height, the correlation between CP and RCS GCMs (lower jaw length 0.76; total face height 0.77) was stronger than between RCS and SITAR GCMs (0.58 for both lower jaw length and total face height).

For females, estimates of CP and FP GCMs for each of the three outcomes were almost perfectly correlated (0.98 to 0.99). The second-highest correlation for upper jaw length (0.90), lower jaw length (0.88) and total face height (0.79) was between the RCS and SITAR GCMs. Irrespective of the outcome, the APGV estimated by CP and FP GCMs correlated better with the RCS GCM than the SITAR model.

Growth intensity (PGV)

Population-average

The population-average estimates along with bootstrapped 95 per cent CIs for males and females are shown below in Table 6-8. For all three outcomes, estimates were consistently higher for the SITAR GCM than the all the other models for both sexes.

Table 6-8. Summary of population-average peak growth velocity (PGV) estimated by linear and nonlinear growth curve models applied to male and female data.

Outcome	Peak growth velocity (PGV) in mm/year (Bootstrapped 95% confidence intervals)			
	CP	FP	RCS	SITAR
Male				
Upper jaw length (COPAD)	2.19 (2.08 to 2.50)	2.20 (2.08 to 2.32)	2.65 (2.37 to 3.48)	2.88 (2.57 to 4.17)
Lower jaw length (COPOD)	2.78 (2.56 to 3.19)	2.56 (2.39 to 2.78)	3.49 (2.84 to 3.98)	4.76 (3.76 to 5.53)
Total face height (TFHNP)	2.71 (2.45 to 3.16)	2.52 (2.31 to 2.83)	3.43 (2.97 to 4.18)	4.75 (3.94 to 5.92)
Female				
Upper jaw length (COPAD)	1.54 (1.48 to 1.73)	1.61 (1.51 to 4.37)	1.71 (1.56 to 1.94)	2.21 (1.74 to 1.83)
Lower jaw length (COPOD)	2.44 (2.12 to 2.98)	2.44 (2.28 to 2.58)	2.84 (2.57 to 3.11)	3.49 (3.01 to 4.06)
Total face height (TFHNP)	2.08 (1.69 to 2.67)	2.10 (1.95 to 2.40)	2.54 (2.32 to 2.76)	3.28 (2.94 to 3.89)

CP conventional polynomial; FP fractional polynomial; RCS restricted cubic spline; SITAR superimposition by translation and rotation.

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Like APGV, the PGV estimated by the CP and FP GCMs for all three outcomes were lower than the RCS and SITAR GCMs for both sexes. Comparison of CP and FP GCMs shows that, except for lower jaw length and total face height for males (difference 0.19 to 0.22 mm/year), the difference was negligible (difference less than 0.07 mm/year). The width of bootstrapped CIs was in general narrower for CP and FP GCMs than RCS and SITAR GCMs. This is perhaps because unlike locally flexible restricted cubic spline functions (RCS and SITAR GCMs), global polynomial functions (CP and FP GCMs) do not easily adapt to a change in growth velocity.

Individual-specific

A summary of individual-specific PGV estimates is provided in Table 6-9. For males, the between-individual variability was lowest for upper jaw length and highest for TFHNP. For females, the highest and the lowest variability was for upper jaw length and lower jaw length, respectively. A similar pattern of between-individual variability in APGV for males and females was observed.

Table 6-9. Summary of individual-specific peak growth velocity (PGV) estimated by linear and nonlinear growth curve models applied to male and female data.

Outcome	Median peak growth velocity (PGV) in years (interquartile range)			
	CP	FP	RCS	SITAR
Male				
Upper jaw length (COPAD)	2.16 (1.93 to 2.54)	2.17 (1.94 to 2.53)	2.66 (2.29 to 3.16)	2.78 (2.48 to 3.38)
Lower jaw length (COPOD)	2.86 (2.42 to 3.38)	2.74 (2.31 to 3.21)	3.99 (3.10 to 4.87)	4.62 (3.66 to 5.38)
Total face height (TFHNP)	2.83 (2.27 to 3.31)	2.64 (2.14 to 3.13)	4.05 (3.05 to 4.98)	4.39 (3.35 to 5.50)
Female				
Upper jaw length (COPAD)	1.61 (1.51 to 1.72)	1.66 (1.55 to 1.75)	1.81 (1.66 to 1.96)	2.20 (2.03 to 2.39)
Lower jaw length (COPOD)	2.45 (2.13 to 2.76)	2.45 (2.12 to 2.76)	2.95 (2.49 to 3.42)	3.52 (3.01 to 3.99)
Total face height (TFHNP)	2.06 (1.78 to 2.37)	2.07 (1.83 to 2.39)	2.57 (2.09 to 3.01)	3.29 (2.79 to 3.75)

CP conventional polynomial; FP fractional polynomial; RCS restricted cubic spline; SITAR superimposition by translation and rotation.

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Like population-average PGV estimates: (i) the SITAR GCM provided higher estimates of individual-specific PGV than the RCS GCM for all three outcomes for both sexes, (ii) the difference between SITAR and the RCS GCMs was greatest for lower jaw length and total face height for males, (iii) irrespective of the outcome, the individual-specific estimates of CP and FP GCMs were smaller than those of the RCS and SITAR GCMs for both males and females, and (iv) the difference in the population-average and the individual-specific PGV estimated by CP and FP GCMs were greater than RCS and SITAR GCMs for lower jaw length and total face height for males.

Spearman correlation coefficients (along with 95% confidence intervals) between individual-specific PGV estimated by CP, FP, RCS and SITAR GCMs are provided in Table 6-10 and Table 6-11.

For males, the correlation was highest between CP and FP GCMs for all three outcomes. Estimates were perfectly correlated for upper jaw length (1.00), but the correlation was slightly weaker for total face height (0.91) and lower jaw length (0.87).

Table 6-10. Spearman correlation (95% confidence interval) for peak growth velocity (PGV) estimated by linear and nonlinear growth curve models applied to male data.

	Upper jaw length (COPAD)				Lower jaw length (COPOD)				Total face height (TFHNP)			
	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR
CP	1				1				1			
FP	1.00 (1.00 to 1.00)	1			0.87 (0.82 to 0.91)	1			0.91 (0.87 to 0.94)	1		
RCS	0.90 (0.85 to 0.93)	0.90 (0.86 to 0.93)	1		0.60 (0.46 to 0.71)	0.56 (0.41 to 0.67)	1		0.69 (0.58 to 0.78)	0.64 (0.49 to 0.74)	1	
SITAR	0.91 (0.88 to 0.94)	0.92 (0.88 to 0.94)	0.89 (0.85 to 0.93)	1	0.74 (0.64 to 0.82)	0.84 (0.77 to 0.89)	0.54 (0.40 to 0.67)	1	0.80 (0.73 to 0.86)	0.86 (0.80 to 0.90)	0.63 (0.50 to 0.73)	1

CP conventional polynomial; FP fractional polynomial; RCS restricted cubic spline; SITAR superimposition by translation and rotation.

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

For upper jaw length, correlation among all four GCMs (CP, FP, RCS and SITAR) was high (0.89 to 0.92). For lower jaw length and total face height, the weakest correlation was between the RCS and SITAR GCMs. Compared to the SITAR GCM, the correlation between RCS GCM and polynomial based linear GCMs (CP and FP) was weaker for lower jaw length (0.56 to 0.60) and total face height (0.64 to 0.69).

Table 6-11. Spearman correlation (95% confidence interval) for peak growth velocity (PGV) estimated by linear and nonlinear growth curve models applied to female data.

	Upper jaw length (COPAD)				Lower jaw length (COPOD)				Total face height (TFHNP)			
	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR	CP	FP	RCS	SITAR
CP	1				1				1			
FP	0.95 (0.93 to 0.96)	1			0.99 (0.99 to 0.99)	1			0.99 (0.99 to 1.00)	1		
RCS	0.80 (0.73 to 0.85)	0.84 (0.78 to 0.88)	1		0.86 (0.81 to 0.90)	0.89 (0.85 to 0.92)	1		0.89 (0.85 to 0.92)	0.91 (0.88 to 0.94)	1	
SITAR	0.73 (0.64 to 0.80)	0.80 (0.73 to 0.85)	0.63 (0.52 to 0.72)	1	0.95 (0.93 to 0.96)	0.93 (0.90 to 0.95)	0.78 (0.70 to 0.84)	1	0.93 (0.90 to 0.95)	0.92 (0.89 to 0.94)	0.83 (0.76 to 0.87)	1

CP conventional polynomial; FP fractional polynomial; RCS restricted cubic spline; SITAR superimposition by translation and rotation.

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Like male data, the correlation was weakest for the RCS and SITAR GCMs for lower jaw length (0.78) and total face height (0.83) for females. The lower jaw length estimates by CP and FP GCMs correlated better with the SITAR GCM (0.93 to 0.95) than with the RCS GCM (0.86 to 0.89). Similarly, for total face height, estimates of CP and FP GCMs correlate more strongly with the SITAR GCM (0.92 to 0.93) than with the RCS GCM (0.89 to 0.91).

6.4. Discussion

This chapter has compared five different GCMs in terms of their fit to the data and ability to describe jaw growth trajectories and estimate timing (APGV) and intensity of the growth spurt. Data analysed comprised the upper jaw length, lower jaw length and total face height measured repeatedly on 128 males and 139 females between seven and 18 years of age. The median number of repeated measurements per individual was 9 (IQR 7, 10) for males and 8 (IQR 7, 9) for females. Data were analysed separately for males and females.

The following section briefly reviews the GCM fitting, focusing on the concept of parsimony versus a 'maximal' random effects modelling approach, and how exclusion of random effects is difficult to justify when analysing biological data such as jaw growth. This explains the decision to drop the PB GCM, which failed to converge. Following is discussion of the model fit to the jaw growth data, and evaluation of the performances of the four remaining GCMs (CP, FP, RCS and SITAR) in modelling growth trajectories and estimating APGV and PGV.

6.4.1. Growth curve model fitting – an overview

Parsimony versus 'maximal' random effects modelling approach

Recently, there has been considerable debate on whether to include all possible random effects while fitting mixed effects, or to exclude some random effects on the grounds of parsimony. While Bates et al. (2015) and Matuschek et al. (2017) have advocated a parsimonious approach if model complexity is not supported by the data, Barr et al. (2013) strongly recommended a 'maximal' random-effect modelling approach by including all possible random effects. They suggested simplifying the random-effect structure only if the model "*fails to converge*" (Barr et al., 2013), and not merely because the model is complex.

An indication that information in the data is insufficient to support estimations of complex mixed effects models is a singular variance-covariance matrix (e.g., very small or zero variance-

covariance estimates). In such situations, Bates et al. (2015) and Matuschek et al. (2017) have advocated simplifying the variance-covariance structure by removing the random effects based on the log-likelihood ratio test (Matuschek et al., 2017) or the principal component analysis (Bates et al., 2015). In contrast, Barr et al. (2013) did not endorse removing random effects, even if the successfully converged maximal model is over-parameterised (i.e., some random effects parameters, particularly the correlations, are zero). However, findings of a recent simulation study (Eager & Roy, 2017) showed that when there are theoretically known non-zero variance-covariance parameters, excluding random effects to improve model convergence can lead to incorrect conclusions.

Model fitting to the biological data

Apart from model complexity and convergence issues, researchers have argued that the model assumptions must be biologically plausible when analysing biological data (Johnson & Omland, 2004; McDonald, 2014; Motulsky & Christopoulos, 2004). When a model fitted to the data does not carry biological meaning or is logically inconsistent with the empirical support, then these authors argue that it should be viewed as untenable (Johnson & Omland, 2004; Motulsky & Christopoulos, 2004). In the context of fitting models to growth data such as height and jaw growth, the random effects capture between-individual variability in growth. As individuals grow at different rates at different times of life (Bose, 2007; Cameron & Bogin, 2012; Rauber, 1990), it is erroneous not to fully account for the between-individual variability in all aspects of growth.

Since nonlinear GCMs, such as SITAR and PB, are structured on a strong underlying biological theory of growth, the issue of excluding random effects would seem more concerning than the linear GCMs. For instance, excluding timing and intensity random effects from the SITAR GCM implies that all individuals experience the adolescent growth spurt at the same age and grow at the same rate. Similarly, fitting the PB GCM without including the timing and growth rate

parameter as random effects implies that all children experience their adolescent growth spurt at the same age and follow the same rate of growth before and during this spurt.

6.4.2. Model fit

I followed the maximal random-effect modelling approach and did not consider simplifying the variance-covariance structure if a model failed to converge. Therefore, the PB GCM that failed to converge for any of the three outcomes for both sexes, was excluded.

I am aware of only one study that successfully fitted the PB GCM with all five random effects (Grimm et al., 2011). That study analysed height data comprising 30 repeated measurements per individual. It may be the case that the small number of jaw measurements per individual (median eight) in the data of the present study caused convergence problems for the PB GCM.

Grimm et al. (2011) compared the fit of the PB GCM that included all five random effects with a model that excluded random effects for timing and growth rate parameters. The results showed that excluding the random effects results in a worsening of the model fit (Grimm et al., 2011). This supports the decision for the present study not to pursue fitting the PB GCM with a subset of the random effects.

The AIC- and RSD-based comparison of models showed that the RCS GCM fitted best to all three outcomes for both males and females. Unlike the cubic spline functions (RCS GCM), which are locally flexible (i.e., can change shape in different sections of the growth trajectory), the conventional polynomials lack local flexibility and therefore cannot adapt to local changes in the growth trajectory (Simpkin et al., 2018). This could explain why the RCS GCM fitted better to the jaw data than the CP GCM.

The results for upper jaw length for both sexes support previous studies reporting that CP and FP GCMs of the same degree provide comparable fit to the data (Long & Ryoo, 2010; Royston &

Altman, 1994). However, the findings do not support an earlier study (Royston & Altman, 1994) that analysed foetal lower jaw length data and concluded that a second-degree FP GCM is sufficient in modelling jaw growth, and that cubic spline functions and fractional polynomials provide comparable fit to the data. In contrast to the findings reported by Royston and Altman (1994), the results of my study based on analysis of pubertal jaw growth data showed that second-degree fractional polynomials were inadequate in modelling jaw growth data as higher order fractional polynomials provided better fit to data, and that cubic spline functions fitted better to the data than fractional polynomials. The difference in findings reported by Royston and Altman (1994) and the present study could be explained as follows: a) unlike the early jaw growth data analysed by Royston and Altman (1994) when growth velocity falls monotonically and data can be well modelled by a second-degree fractional polynomial, higher degree fractional polynomials are required to capture growth changes during the pubertal growth spurt period (See comments by T. J. Cole; Royston & Altman, 1994, pp. 459-460), and b) extending the age range to include puberty adds complexities to the functional form (Ramsay & Silverman, 2005), and fractional polynomials cannot match the performance of cubic spline functions in approximating the complex functional forms (Binder et al., 2013).

I am not aware of any study that has compared the fit of SITAR GCM with the linear GCMs included in my study. Though one study compared SITAR and FP GCMs in estimating APGV (Simpkin et al., 2017), authors did not consider fit evaluation of these models. Therefore, the model fit results of the present study can be compared with the available literature.

My findings show that the RCS GCM fitted better to the jaw growth data than the SITAR GCM. This could be explained by the fact that being a structured nonlinear model, the SITAR GCM offers less flexibility than the RCS GCM. Interestingly, while the SITAR GCM fitted better than a third-degree CP GCM to the jaw growth data, its fit was worse than a fourth-degree CP GCM (see

lower jaw length and total face height results for males, Table 6-3). The difference in fit perhaps could be attributed to the large within-individual variability (RSD) for lower jaw length and total face height for males. A simulation study has shown that as the measurement error (RSD) increases, the performance of the SITAR GCM worsens (Simpkin et al., 2017).

Following the recommendations by (Cole et al., 2010), the fit of the SITAR GCM was evaluated and compared with and without transformation of age and outcomes. The results show that the SITAR GCM fitted best when modelling the outcome on its original scale (jaw measurements). This finding supports earlier studies reporting that models fitted to the original outcome scale (height) provide the best fit. Though one study found that the best transformation for age when fitting the SITAR GCM to the height data is the square root transformation (Cole & Mori, 2017), another study reported that models fitted to the log-transformed age or original age are better than those using the square root transformation of age (Cole et al., 2010). The results of my study show that the SITAR GCM fit better to the jaw data when original age was used and not the square root or log-transformed age.

6.4.3. Distance (size) and derivatives (velocity and acceleration)

When analysing data where theory suggests an asymptote (such as jaw growth data), it is advantageous to use a model with asymptotic-like effects (Long & Ryoo, 2010). The results show that for jaw growth data, cubic spline–based linear (RCS) and nonlinear (SITAR) GCMs performed better than polynomial-based linear GCMs (CP and FP).

Both CP and FP GCMs showed a biologically implausible decline in jaw size and estimated negative growth velocity towards the end of the growth period (adulthood). This behaviour of the CP GCM was observed for all three outcomes and for both sexes, indicating that both third- and fourth-degree CP GCMs failed to approximate an asymptote. Previous studies have also shown that

higher order conventional polynomials predict unstable trajectories towards the edges of the predictor (age) range (Long & Ryoo, 2010; Simpkin et al., 2018).

Compared to the CP GCM, growth trajectories estimated by the FP GCM differed across outcomes and for males and females. As an example, while a third-degree FP GCM showed CP GCM like growth trajectories for lower jaw length and total face height for females, a third-degree FP GCM fitted to lower jaw length and total face height for males showed less tendency for predicting decline in jaw size than the CP GCM. This occurred even though the same powers (powers 3 3 3) were chosen for lower jaw length and total face height for both males and females.

The difference in growth trajectories estimated by the CP and FP GCMs could perhaps be explained by the difference in differentiation of conventional polynomial powers and the fractional polynomial powers (see Chapter 4 Section 4.3.2 for full details). For example, unlike the conventional polynomials where the fixed and random effects for the first term (age^1) always enter the derivatives estimation as constants, the fixed and random effects for fractional polynomials enter as a slope, except when the power of the first term is 'one' (Long & Ryoo, 2010). Similar differences extend to higher degree models where derivatives of fractional polynomials, unlike conventional polynomials, are not necessarily symmetric (Long & Ryoo, 2010).

Due to the differences in powers and the way conventional polynomial and fractional polynomial functions are differentiated, the CP and FP GCMs can behave differently when estimating distance and derivatives. The fractional polynomial powers can better adapt to local changes in growth trajectories due to the difference in growth timing and the intensity. This could be because as lower jaw length and total face height for males show more intense growth and mature later, the FP GCM was more flexible than the CP GCM and adapted better to these conditions.

My finding on modelling growth trajectories disagrees with previous studies reporting that a major advantage of using FP GCM over CP GCM is that the former can approximate asymptotes

better than the latter (Long & Ryoo, 2010; Royston & Altman, 1994). The difference in findings could be explained by the fact that, unlike earlier studies that fitted FP model up to second-degree (Long & Ryoo, 2010; Royston & Altman, 1994), the present study's results are based on third-degree FP model. Like higher order conventional polynomials, which are known to be unstable (Simpkin et al., 2018), perhaps the stability of fractional polynomials also decreases as the degree of the FP model increases.

Unlike CP and FP GCMs, both RCS and SITAR GCMs estimate growth curves that are consistent with the notion of an asymptote. The results show that irrespective of the sex or the outcome analysed, growth curves estimated by RCS and SITAR GCMs always plateaued as individuals approach adulthood.

No previous study has compared RCS and SITAR GCMs for modelling growth data. Therefore, the findings cannot be directly compared with the literature. My results show that growth trajectories estimated by RCS and SITAR GCMs differ when the number of knots is high (see Section 6.3.2). For example, to successfully model the large within-individual variability in growth for the lower jaw length and total face data for males (see RSD estimates, Table 6-3), both RCS and the SITAR GCMs required seven knots (see number of knot, Table 6-3). For these two outcomes i.e., lower jaw length and total face data for males, the growth trajectories estimated by RCS and SITAR GCMs differed substantially (see Appendix B, Section B.3.4). Potential reasons that could explain the observed differences in growth trajectories estimated by RCS and SITAR GCMs are discussed below.

While both RCS and SITAR GCMs are based on cubic spline functions, they are conceptually different (see Chapter 4 Section 4.1 for full details). Unlike the RCS GCM, the SITAR GCM imposes a nonlinear mean curve as the velocity parameters enter the model within an exponential function. This nonlinearity in the mean curve allows the estimation of the timing parameter, which

shifts the age scale (Cole et al., 2010). While this interplay between the velocity and timing parameters provides a novel approach for understanding the biological relationship between growth velocity and timing, it rotates and stretches or shrinks growth curves in an attempt to match each individual's curve to a common underlying population-average curve. As the number of knots increases, the rotation and stretching effect also increases. This could explain why the SITAR GCM showed a greater rotation and stretching or shrinkage in growth trajectories of lower jaw length and total face height for males where the number of knots was seven.

6.4.4. Timing and intensity of the adolescent growth spurt

Growth timing (APGV)

Even though my findings support a recent simulation study reporting that conventional polynomials predict unrealistic growth trajectories towards the edges of the predictor range (Simpkin et al., 2018), quartic polynomials (not the cubic polynomials) were found to perform better than fractional polynomials in estimating APGV.

Though fractional polynomials have an advantage over conventional polynomials in terms of flexibility and derivatives estimation, here they do not perform better than conventional polynomials when the objective is to estimate APGV. Further, wider bootstrapped CIs for the FP GCM than the CP GCM suggest that fractional polynomials are less consistent in estimating APGV when compared with conventional polynomials. Recently, Simpkin et al. (2017) studied performance of fractional polynomials in estimating age at PHV and concluded that fractional polynomials perform poorly. My results support views expressed by Simpkin et al. (2017):

"*Despite the increased complexity and flexibility from using fractional polynomials opposed to conventional polynomials, models were not sufficiently flexible to capture an individual's underlying growth trajectory and accurately estimate their age at PHV.*"

I am not aware of any study comparing RCS and SITAR GCMs for estimating adolescent growth spurt parameters (APGV and PGV). The results of the present study show that compared to the RCS GCM, the SITAR GCM provide lower estimates of population-average and individual-specific APGV when within-individual variability is large (see lower jaw length and the total face height results for males, Section 6.3.3). Furthermore, as the within-individual variability increases, the SITAR GCM bootstrapped CIs for lower jaw length and total face height for males are much wider than the RCS GCM. For other outcomes, the CIs are also narrower for the RCS GCM than the SITAR GCM, but the difference is less evident.

A recent simulation study showed that as the measurement error (within-individual variability) increases, the performance of the SITAR GCM worsens (Simpkin et al., 2017). In such scenarios, the SITAR GCM tends to underestimate the APGV. This could perhaps explain the lower APGV estimated by the SITAR GCM for the lower jaw length and total face height for males.

As discussed in 6.4.3, the best-fitting RCS and the SITAR GCMs for these two outcomes (lower jaw length and total face height for males) for which APGV estimated by these two GCMs differed most, required high number of knots (seven knots). For both these outcomes i.e., lower jaw length and the total face height for males, the derivatives also differed substantially. Thus, as expected, the APGV which is estimated using derivatives, also differed substantially.

These findings suggest that perhaps a greater within-individual variability in growth requires fitting the SITAR GCM with a higher number of knots which in turn result in greater rotation and stretching or shrinkage in growth curves (see Section 6.4.3). This rotation and stretching or shrinkage in growth curves then affects the APGV estimates. However, as the simulation study (Simpkin et al., 2017) which investigated the effect of the measurement error (within-individual variability) on the performance of the SITAR GCM did not report the number of knots used to construct the cubic spline design matrix under varying conditions of measurement error, at present

no concrete evidence is available on this. Therefore, further research is required to study this hypothesis.

Growth intensity (PGV)

Unlike the APGV, a fourth-degree CP GCM provides lower estimates of PGV when compared to the cubic spline-based linear (RCS) and nonlinear (SITAR) GCMs (see the lower jaw length and the total face height results for males, Section 6.3.3). This is because, unlike a locally flexible cubic spline function, the global polynomial function fails to adapt to the sudden change in the growth velocity. Similar findings were reported in a simulation study that evaluated and compared performance of quartic conventional polynomial functions and quartic spline based functions (Simpkin et al., 2018).

Like APGV, PGV estimated by the FP GCM was lower than cubic spline-based linear (RCS) and nonlinear (SITAR) GCMs for all three outcomes and for both sexes. No previous study has evaluated and compared performances of FP GCM with any of the other three GCMs (CP, RCS or SITAR). The recent study (Simpkin et al., 2017) that compared FP and SITAR GCMs in estimating APGV did not focus on comparing PGV estimated by these two GCMs.

Comparison of PGV estimated by the RCS and the SITAR GCMs shows some surprising results. Unlike the APGV, the SITAR GCM provides higher estimates of population-average and individual-specific PGV especially when the within-individual variability is large (see lower jaw length and the total face height results for males, Section 6.3.3). Furthermore, the bootstrapped CIs for PGV are considerably wider for the SITAR GCM than the RCS GCM lower jaw length and the total face height results for males.

Since the previous study (Simpkin et al., 2017) that evaluated the performance of the SITAR GCM in modelling height data did not investigate growth velocity, it not possible to corroborate findings with empirical evidence. However, a possible explanation is that as the SITAR GCM is

based on the assumption that an early adolescent growth spurt is linked to a more intense growth spurt and vice versa, any factor (such as the measurement error) which lower the APGV will result in a higher PGV and vice versa.

6.4.5. A note on pre-adolescent growth spurt

In addition to the adolescent growth spurt, growth trajectories for lower jaw length and total face height show a pre-adolescent growth spurt for males at around nine years of age. Growth velocity curves and the growth acceleration curves for lower jaw length and total face height estimated by RCS and SITAR GCMs show a distinct sign of pre-adolescent growth spurt for males (see Appendix B, Section B.3.4).

Unlike RCS and SITAR GCMs, growth velocity and growth acceleration curves estimated by CP and FP GCMs show no evidence for a pre-adolescent growth spurt for upper jaw length (see Section 6.3.2), lower jaw length or total face height (see Appendix B, Section B.3.4) for males. This is perhaps because of the limited flexibility offered by a fourth-degree CP GCM and a third-degree FP GCM when compared with the greater flexibility (seven knots) of the RCS and SITAR GCMs fitted to lower jaw length and total face height data for males (see Table 6-3).

No evidence was found for a pre-adolescent growth spurt for upper jaw length for males or females (see Section 6.3.2), or for lower jaw length or total face height for females (see Appendix B, Section B.3.4).

The reason why upper jaw length for both sexes, and lower jaw length and total face height for females, did not show a pre-adolescent growth spurt is perhaps related to the age range (seven to 18 years). Assuming that the timing of the pre-adolescent growth spurt follows the same pattern as the timing of the adolescent growth spurt, the limited age range did not cover the pre-adolescent growth period for upper jaw length for both sexes, and the lower jaw length and total face height for females. This is because compared to lower jaw length and total face height for males, the adolescent

growth spurt for upper jaw length for both sexes, and lower jaw length and total face height for females, occurs one to two years earlier (see Table 6-4). A previous study (Buschang et al., 1988b) analysing lower jaw length data with an age range of six to 15 years found a pre-adolescent growth spurt for both sexes, with the pre-adolescent growth spurt for lower jaw length occurring one year earlier for females (7.7 years) than males (8.7 years). This is further discussed in Chapter 7, Section 7.4.2.

6.4.6. Comparison with previous studies using growth curve models

This section compares my findings with results reported in the earlier eight studies that used GCMs for modelling jaw growth data (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996). All eight studies used CP GCM for analysing jaw growth data (see Chapter 2, Section 2.2). Before discussing findings, key aspects of study design, data analysis, and the statistical methods applied in these studies are summarised.

Study characteristics and data

Chapter 2 (Section 2.2) provides full details on research focus, study design and data analysed, including sex-specific distribution of sample and age ranges.

Table 6-12 summarises the data analysed in each previous study and my study. No previous study included data for all three growth measures (upper and lower jaw length and total face height) and skeletal jaw relationships (Class I, Class II and Class III) for either males or females. The present study is the first to evaluate growth changes in all three measures for all three skeletal jaw relationships for both males and females.

Table 6-12. Characteristics of data analysed in previous studies which applied growth curve models to study jaw growth during adolescence, and in the present study.

Study (chronological order)	Sex [a]		Age (years)	Growth measure [b]			Class of skeletal jaw relationship		
	M	F	Range	Upper jaw	Lower jaw	Face height	Class I	Class II	Class III
(Buschang et al., 1988b)	✓	✓	10 to 15	×	✓	×	?	?	?
(Buschang et al., 1988a)	✓	✓	6 to 15	×	✓	×	✓	✓	×
(Buschang et al., 1989)	×	✓	6 to 15	×	✓	×	?	?	?
(van der Beek et al., 1996)	×	✓	6 to 15	×	×	✓	?	?	?
(Buschang et al., 1999)	✓	✓	6 to 16	×	✓	×	?	?	?
(Chvatal et al., 2005)	✓	✓	6 to 15	×	✓	×	?	?	?
(Buschang et al., 2013)	×	✓	10 to 15	✓	✓	✓	?	?	?
(Nahhas et al., 2014)	✓	✓	4 to 24	✓	✓	×	?	?	?
My study	✓	✓	7 to 18	✓	✓	✓	✓	✓	✓

? Information not provided; ✓Included ×Not included.

[a] M: Male; F: Female.

[b] Upper jaw: upper jaw length; Lower jaw: lower jaw length; Face height: total face height.

Data analysis

Table 6-13 provides a summary of the data analysis approach used in previous studies and the present study. All eight applied a CP GCM to the jaw growth data. The present study applied CP, FP, RCS and SITAR GCMs to jaw growth data. For RCS and SITAR GCMs, the number of restricted cubic spline terms (number of knots – 2) is shown rather than knots (see Table 6-3 for number of knots).

All previous studies fitted CP GCM with a subset of random effects. While one study did not report on the number of polynomial terms included in the random effects part of the model

(Nahhas et al., 2014), two studies fitted up to a fifth-degree CP GCMs for the fixed effects part of the model, but included only an intercept in the random effects part of the model (Buschang et al., 2013; Buschang et al., 1999). In contrast, I fit all four GCMs (CP, FP, RCS and SITAR) by including all possible individual-specific random effects.

Table 6-13. Comparison of growth curve models applied to the jaw growth between previous studies and the present study.

Study (chronological order)	GCM	Growth measure [a]											
		Upper jaw				Lower jaw				Face height			
		Male		Female		Male		Female		Male		Female	
		Fix	Ran	Fix	Ran	Fix	Ran	Fix	Ran	Fix	Ran	Fix	Ran
(Buschang et al., 1988b)	CP	NA	NA	NA	NA	4	3	3	2	NA	NA	NA	NA
(Buschang et al., 1988a)	CP	NA	NA	NA	NA	6	2	5	2	NA	NA	NA	NA
(Buschang et al., 1989)	CP	NA	NA	NA	NA	NA	NA	5	2	NA	NA	NA	NA
(van der Beek et al., 1996)	CP	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	4	2
(Buschang et al., 1999)	CP	NA	NA	NA	NA	4	0	5	0	NA	NA	NA	NA
(Chvatal et al., 2005)	CP	NA	NA	NA	NA	5	4	5	4	NA	NA	NA	NA
(Buschang et al., 2013)	CP	NA	NA	2	0	NA	NA	2	0	NA	NA	3	0
(Nahhas et al., 2014)	CP	5	NR	5	NR	5	NR	5	NR	NA	NA	NA	NA
My study	CP	3	3	3	3	4	4	3	3	4	4	3	3
	FP	3	3	3	3	3	3	3	3	3	3	3	3
	RCS	3	3	2	2	5	5	3	3	5	5	3	3
	SITAR	2	abc*	2	abc*	5	abc*	3	abc*	5	abc*	3	abc*

NA: not applicable (data not analysed); NR: not reported.

[a] Upper jaw: upper jaw length; Lower jaw: lower jaw length; Face height: total face height.

GCM: growth curve model; CP: conventional polynomial; FP: fractional polynomial; RCS: restricted cubic spline; SITAR: superimposition by translation and rotation.

Fix: number of polynomials terms excluding intercept (CP and FP) /number of spline terms excluding intercept (RCS and SITAR) for the fixed effects part of the GCMs.

The number of spline terms equals the number of knots – 2.

Ran: number of polynomials terms excluding intercept (CP and FP) /number of spline terms excluding intercept (RCS) for the fixed effects part of the GCMs.

* Random effects included for the SITAR model: a Size, b Timing, c Intensity.

The value '0' indicates that only random intercept was included in the random effects part for CP, FP or RCS GCMs.

Findings

Model fit

As in earlier studies (see Table 6-13), the results show that different degrees of CP GCM are required for modelling jaw growth trajectories for males and females. All previous studies (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996) fitted CP GCM with a subset of random effects to include higher order terms on the fixed effects part of the model (see Table 6-13). However, none of these eight studies provided any fit statistic to show how varying the number of terms in the fixed and random effects parts of the CP GCM changed fit to the data. Furthermore, the studies did not provide information on the fit of the best-fitting CP GCM applied to jaw growth data.

None of the previous studies assessed normality of random effects or level 1 residuals, or the homoscedasticity and independence of level 1 residuals (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996). Section 6.3.1 provides the results from the present study for normality and homoscedasticity assumption for each of the four GCMs (CP, FP, RCS and SITAR) applied to upper and lower jaw length and total face height measurements for males and females.

In line with my results for the CP GCM, previous studies reported that conventional polynomials resulted in biologically implausible growth trajectories (negative growth velocity(Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; van der Beek et al., 1996). My findings show that like conventional polynomials, growth trajectories estimated by fractional polynomials do not level off, thereby resulting in negative growth velocity. In contrast to CP and FP GCMs, restricted cubic spline-based linear (RCS) and nonlinear (SITAR) GCMs estimate biologically plausible growth trajectories as they successfully approximate an asymptote (see Section 6.4.3 for a detailed discussion).

Adolescent growth spurt

A summary of evidence reported in previous studies and my study is provided in Table 6-14. One study analysing upper and lower jaw length data for females (Buschang et al., 2013) reported findings that do not agree with results reported by others or the findings of my study for the adolescent growth spurt. For example, while Buschang et al. (2013) reported that females did not experience an adolescent growth spurt for upper and lower jaw length, another study (Nahhas et al., 2014) analysing upper and lower jaw length data for males and females found that both males and females experienced an adolescent growth spurt for the upper jaw length and the lower jaw length. Similarly, except Buschang et al. (2013), all other studies analysing lower jaw length for males and/or females found evidence of an adolescent growth spurt for lower jaw length in both sexes (Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Nahhas et al., 2014). My results show that both males and females experience an adolescent growth spurt for the upper jaw length and the lower jaw length (see Table 6-14).

Table 6-14. Summary of evidence reported in previous studies applying growth curve models to the jaw growth data, and in the present study.

Study (chronological order)		Growth measure [a]					
		Upper jaw		Lower jaw		Face height	
	GCM	Male	Female	Male	Female	Male	Female
(Buschang et al., 1988b)	CP	NA	NA	✓	✓	NA	NA
(Buschang et al., 1988a)	CP	NA	NA	✓	✓	NA	NA
(Buschang et al., 1989)	CP	NA	NA	NA	✓	NA	NA
(van der Beek et al., 1996)	CP	NA	NA	NA	NA	NA	✓
(Buschang et al., 1999)	CP	NA	NA	✓	✓	NA	NA
(Chvatal et al., 2005)	CP	NA	NA	✓	✓	NA	NA
(Buschang et al., 2013)	CP	NA	×	NA	×	NA	✓
(Nahhas et al., 2014)	CP	✓	✓	✓	✓	NA	NA
My study	CP	✓	✓	✓	✓	✓	✓
	FP	✓	✓	✓	✓	✓	✓
	RCS	✓	✓	✓	✓	✓	✓
	SITAR	✓	✓	✓	✓	✓	✓

NA: Not applicable (data not analysed); ✓Yes; ×No.

[a] Upper jaw: upper jaw length; Lower jaw: lower jaw length; Face height: total face height.

GCM: growth curve model; CP: conventional polynomial; FP: fractional polynomial; RCS: restricted cubic spline; SITAR: superimposition by translation and rotation.

Unlike upper and lower jaw length, Buschang et al. (2013) did report an adolescent growth spurt in total face height for females. This finding agrees with another study that analysed the total face height data for females and reported that females experienced an adolescent growth spurt for the total face height. My findings support the results reported in both these studies (Buschang et al., 2013; van der Beek et al., 1996). The results show that both males and females experience an adolescent growth spurt for total face height (see Table 6-14).

Unlike Buschang et al. (2013), who did not find evidence for the adolescent growth spurt in lower jaw length for females, two other studies (Buschang et al., 1989; Buschang et al., 1988b) analysing the lower jaw length data available from the same source (Human Growth Research Centre, University of Montreal, Canada) with an identical age range, six to 15 years (see Table 6-12) reported that females experienced an adolescent growth spurt in lower jaw length (see Table 6-14).

The one potential factor that could explain differences in findings reported by these three studies is the degree of CP GCM applied to the lower jaw length data for females (see Table 6-13). Buschang et al. (2013) applied a second-degree CP GCM to the lower jaw data for females but included only an intercept in the random effects part of the model. Age was centred at 12 years. The authors reported that "*a fifth-order polynomial was first fitted; the highest order term was sequentially eliminated if it was not statistically significant*" (Buschang et al., 2013). However, the authors did not report whether they tried the same approach for the random effects part of the model, or fitted models with varying degree of polynomials for the fixed effects part of the model, while including only an intercept in the random effects part of the model. The authors did not report whether they checked statistical significance for the regression parameters or the likelihood ratio test. The final best-fitting CP GCM was fitted with quadratic polynomials for the fixed effects part of the model and an intercept for the random effects part of the model (Buschang et al., 2013). Unlike Buschang et al. (2013), Buschang et al. (1988b) and Buschang et al. (1989) fitted a fifth-degree CP GCM to the lower jaw data for females, with an interpret, a linear term and a quadratic term included in the random effects part of the model (see Table 6-13).

Age was centred at 10 years (Buschang et al., 1989; Buschang et al., 1988b). Neither study (Buschang et al., 1989; Buschang et al., 1988b) mentioned how the model was built, i.e., whether any criteria were used to assess the fit of the model while including polynomials into the fixed or random effects parts of the model. The authors also did not mention why they restricted inclusion

up to a quadratic term in the random effects part of the model when the fixed effects part of the model involved up to a fifth-degree polynomial (Buschang et al., 1989; Buschang et al., 1988b).

Previous studies report a wide range of time and intensity of adolescent growth spurt parameters (see Chapter 2, Section 2.2.3 and subsection 'Studies using growth curve models'). For example, the APGV for lower jaw length varies from 13.41 to 14.30 years for males, and 10.84 to 12.70 years for females. The range of PGV for the lower jaw length reported in previous studies varies from 2.71 to 4.08 mm/year for males, and 2.20 to 2.51 mm/year for females. My results (see Section 6.3.3) for the lower jaw length are within the range of APGV and PGV reported in earlier studies.

The wide range of APGV and PGV reported across previous studies could be because all but one study (Buschang et al., 1988b) did not provide class-specific distribution of samples or whether and how adolescent growth spurt parameters were adjusted for class differences (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996). Therefore, it is unclear whether adolescent growth spurt parameters for the upper and lower jaw length and total face height reported in these seven studies are for Class I, Class II or Class III. The findings reported by Buschang et al. (1988b) are discussed in Chapter 7 (Section 7.4).

6.5. Limitations

One major limitation pertains to the data, and in particular the numbers of individuals and measurements. This influenced convergence of some of the models the study aimed to fit. Due to the limited sample size, particularly the number of growth measurements per individual, I could not successfully fit the maximal random effect PB GCM. The one previous study that successfully fit the PB GCM with all five random effects (Grimm et al., 2011) analysed longitudinal growth data comprising height measurements (30 repeated measurements per individual) made on 155 males and 167 females. Although the level 2 sample size for data in my study (128 males and 139 females) was comparable to the data analysed by Grimm et al. (2011), the level 1 sample size was considerably smaller (median eight measurements per individual in my study). Further, the unbalanced data in my study could have contributed to the convergence issues. One alternative could be to fit the PB GCM within the Bayesian framework, as it improves convergence (Eager & Roy, 2017).

A few previous studies using CP GCM for modelling jaw growth data have reported results for up to fifth- and sixth-degree of models (Buschang et al., 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014). However, unlike my study, these studies excluded random effects in pursuit of fitting higher degree CP GCM. As I followed a maximal random effects modelling approach, the CP GCM could not be successfully fitted beyond fourth degree for males and third degree for females (again, perhaps due to limited sample size).

Following the earlier recommendations, I fit up to the third-degree FP GCM. However, as the results show, even the third-degree FP GCM were inadequate in modelling the jaw growth data. It is therefore unclear how fitting higher order fractional polynomials could have influenced the results. The results indicate that perhaps exploring higher degree FP GCMs is better when analysing jaw

growth data. However, it is quite possible that a higher degree FP GCMs might result in a 'wigglier' growth trajectory, as observed for the fourth-degree CP model.

Although the jaw growth data analysed represents real-world data, perhaps it is not ideally suited for model comparison. This is because in addition to the limited sample size, the cephalometric measures are known to be less reliable as they are prone to measurement errors (Baughan et al., 1979b; Houston, 1983). Even though the analysed data from the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection have been meticulously checked for magnification errors and inter-observer variability, it is still possible that some measurement might be inaccurate due to the poor quality of radiographs (see Chapter 3 for details). Previous studies comparing different GCMs for estimating APGV have used height data. Due to its ease of measurement, human height is often considered a classic anthropometric quantitative trait (Jelenkovic et al., 2016). Therefore, my study's results might not be comparable to other studies. Further work is required to cross-validate the findings by running simulations and analysing height data.

For analysis growth data where theory suggests an asymptote, it is advantageous to use a model with asymptotic-like effects (Long & Ryoo, 2010). This study's results clearly show that RCS and SITAR GCMs were preferred over CP and FP GCMs. The SITAR GCM is a popular analytical tool for modelling height data. The RCS GCM is a valuable tool in handling data complexities and modelling nonlinear relationships between outcome and predictors for prognostic models (Collins et al., 2016) and demographic models (Dahlgren et al., 2011), but has attained less attention for modelling growth data. My study has shown that not only that the RCS GCM fits better to jaw growth data than the SITAR GCM, but also that it is efficient in estimating growth spurt parameters (narrower bootstrapped CIs). However, as discussed above, the jaw growth data analysed may not reflect the diversity of growth data often encountered in studying height and

weight. Therefore, a well-designed simulation is required to compare RCS and SITAR GCMs. I further discuss this topic in Chapter 8 (Section 8.3.1).

Lastly, unlike distance, derivatives (velocity and acceleration) are not observed directly but derived by differentiating the distance curve. Therefore, it is difficult to assess whether the best-fitting model for distance is also the best model for derivatives (Ramsay & Silverman, 2005; Simpkin et al., 2017). As APGV and PGV are estimated using derivatives, the same caveat applies to the adolescent growth spurt parameters. Thus, it is unclear whether the best-fitting RCS GCM is best for estimating derivatives and growth parameters. The same is true for the other three GCMs (CP, FP and SITAR) as well.

6.6. Conclusion

The RCS GCM fits best to the jaw growth data, and consistently produced narrower bootstrapped CIs for population average adolescent growth spurt parameters (APGV and PGV) than all other models, irrespective of outcome and sex.

Although both RCS and SITAR GCMs approximate asymptotes and estimate biologically plausible growth trajectories, the RCS GCM is less affected by the within-individual variability than the SITAR GCM. As within-individual variability increases, the SITAR GCM tends to provide lower estimates of population-average and individual-specific APGV, but higher estimates of population-average and individual-specific PGV.

Considering better fit to the data, estimating biologically plausible growth trajectories, robustness in handling noisy data, and narrow CIs for adolescent growth spurt parameters (APGV and PGV), the RCS GCM was identified as the model of choice for modelling jaw growth data and answering clinical research questions in the next study (Chapter 7).

CHAPTER 7. SEX DIFFERENCES IN THE TIMING AND INTENSITY OF THE ADOLESCENT GROWTH SPURT FOR NORMAL JAW GROWTH AND SKELETAL DISCREPANCIES

7.1. Background and aim

Skeletal jaw relationship is classified into Class I, Class II and Class III depending on how the upper and lower jaw are related to each other in the anteroposterior (horizontal) plane (Angle, 1899; Proffit et al., 2014). Class I represents normal skeletal jaw relationship whereas Class II and Class III denote abnormal jaw relationships and are collectively termed as skeletal malocclusions (Proffit et al., 2014). Class II skeletal malocclusion is characterised by longer upper jaw length, shorter lower jaw length and decreased total face height. In contrast, Class III skeletal malocclusion is characterised by shorter upper jaw length, longer lower jaw length, and increased total face height (see Chapter 1, Section 1.1 for further details).

Assessment of age at peak growth velocity (APGV) is essential to coordinate the timing of growth modification procedures used to correct skeletal discrepancies in jaw length and total face height (Perinetti & Contardo, 2017). The APGV denotes the age corresponding to the peak growth velocity (PGV) which is a biological indicator of skeletal maturity. Estimation of APGV and PGV involves fitting a growth curve model (GCM) and derivatives estimation (Hauspie et al., 2004). Please see Chapter 4 (Section 4.5.3) for further details.

Chapter 6 applied conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS), the superimposition by translation and rotation (SITAR)and the Preece-Baines (PB) models to the jaw growth data. The results showed that the RCS GCM fits best to male and female data. The aim of this chapter is to apply the RCS GCM to the jaw growth data for studying class differences in the growth trajectories and adolescent growth spurt parameters for upper and lower jaw length and total face height. No previous study has directly compared Class I, Class II and Class III for potential differences in growth trajectories for these outcomes or the timing and intensity of the adolescent growth spurt (see Chapter 2, Section 2.3 for details).

7.2. Methodology

7.2.1. Data

Full details on data are provided in Chapter 3 (Section 3.3) and summarized in Chapter 6 (Section 6.2.1). Briefly, the data comprised 128 males (mean age 11.66, SD 2.92, range 7–18 years) and 139 females (mean age 11.60, SD 2.88, range 7–18 years) who participated in eight growth studies. For each individual, three outcomes collectively describing the anteroposterior and vertical jaw growth changes were measured repeatedly. The three outcomes are the upper jaw length (COPAD), lower jaw length (COPOD) and total face height (TFHNP). Data are available from the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection. The class variable denoting the skeletal jaw relationship (Class I, Class II or Class III) is recorded as a nominal time-invariant variable.

7.2.2. Model fit

The model fitting strategy was the same as described earlier in Chapter 6 (see Section 6.3.2 for details). Briefly, data were analysed separately for males and females. A two-level RCS GCM was fitted to the data by including all possible individual-specific random effects (the maximal random effects approach). Study-specific and class-specific dummy variables and their respective interactions with age were included in the fixed part of the model to account for study- and class-level heterogeneity in the growth trajectories. Since in some studies no data were available for Class II or Class III skeletal jaw relationships (see Chapter 3, Table 3-4), no interactions were included between the study and class dummies for these studies. The normality and homoscedasticity assumptions for the RCS GCM fitted to the upper and lower jaw length and total face height have already been tested and shown to be tenable (Chapter 6).

7.2.3. Estimating class differences in growth parameters

Class-specific (Class I, Class II and Class III) growth trajectories (distance and derivatives) and adolescent growth spurt parameters (APGV and PGV) averaged over the growth studies were calculated. Chapter 6 provides full details on estimating study-adjusted class-specific growth trajectories (Section 6.3.3) and adolescent growth spurt parameters (Section 6.3.4). Briefly, study-specific dummy variables and their cross-level interactions were mean centred, whereas the class-specific dummy variables and cross-level interactions were left unaltered. Class-specific first and second derivatives were then used to estimate study-adjusted adolescent growth spurt parameters.

Class differences in each expected outcome (Class II versus Class I, Class III versus Class I and Class III versus Class II) in growth trajectories (distance and derivatives) were computed and tested at each year of age (i.e., 7 years, 8 years, ……., 18 years) via calculating linear combinations of coefficients at the 5 percent significance level. The 95% confidence intervals (CIs) were also computed.

To test class differences and construct confidence intervals for adolescent growth spurt parameters (APGV and PGV), I used the nonparametric bootstrapped approach (by cluster with replacement, a total 1,000 replications). The nonparametric bootstrap method does not make any assumption about the distribution of parameters (unlike the delta method) and is particularly useful for estimating standard error and constructing confidence intervals when sample size is limited. Chapter 6 (Section 6.2.7) provides full details on the nonparametric bootstrapping procedure. Briefly, at each replication, the RCS GCM was fitted, class specific APGV and PGV were estimated using derivatives, and class differences in the APGV and PGV were computed. The 2.5^{th} and 97.5^{th} percentiles of class differences in the bootstrapped APGV and PGV (over 1000 samples) were used as the 95% CIs for these parameters.

There is an ongoing debate on reporting statistically significant findings in scientific publications (Amrhein & Greenland, 2018; Amrhein et al., 2019; Baker, 2016; Benjamin et al., 2018; Ioannidis, 2018, 2019a, 2019b). While Ioannidis (2018, 2019a, 2019b) and Benjamin et al. (2018) favour reporting statistically significant findings, Baker (2016), Amrhein and Greenland (2018) and Amrhein et al. (2019) have made strong recommendations to completely abandon the practice of reporting statistical significance of research findings.

Authors voicing strong opinions against the use of p values (Amrhein & Greenland, 2018; Amrhein et al., 2019; Baker, 2016) have suggested that studies must report and discuss research findings in terms of their clinical significance rather than the statistical significance because "a P value cannot indicate the importance of a finding" (Baker, 2016). Drawing clinical inferences based on the point estimate and the range of CI around it is suggested (Amrhein et al., 2019). For a comparative study, it is unimportant and misleading to report separate CIs for each group (Altman, 2005). The correct approach is to present and discuss CI for the contrast (Altman, 2005).

In this study, the latter approach was chosen, and therefore findings were not interpreted solely in terms of their statistical significance. Rather, the CIs for class difference were used to support my discussions regarding the precision of the estimates and the overall level of evidence provided by the analyses.

7.2.4. Software

Chapter 6 (Section 6.2.9) provides details on software used to fit RCS GCM, estimate growth trajectories and the adolescent growth spurt parameters and to implement the nonparametric bootstrapping procedure. New to this chapter, to estimate class differences in yearly growth trajectories (distances and derivatives), a new Stata program ('covdiffs') was written. The 'covdiffs' program calls the Stata 'lincom' command to test a linear combination of parameter estimates. Full details on the covdiffs' program are provided in Appendix C (Section C.1).

7.3. Results

Data were analysed separately for males and females. For males, the best-fitting RCS GCM for the lower jaw length and total face height included seven knots. For the upper jaw length, a model with five knots provided the best-fit. For females, the RCS GCM with five knots provided best fit to the lower jaw length and total face height data. The best-fitting RCS GCM for the upper jaw length included four knots. Appendix B (Table B-9) provides the fit statistics (AIC) of the RCS GCM applied to the male and female data with different number of knots.

For the sake of brevity, I present graphical results only for distance and growth velocity. Detailed graphical results for growth acceleration are provided in Appendix C (Section C.2.3).

7.3.1. Distance (size)

Distance curves for upper and lower jaw length and total face height are shown in Figure 7-1, Figure 7-2 and Figure 7-3, respectively. The results show that for normal skeletal relationship (Class I) and skeletal malocclusions (Class II and Class III), the cumulative increase in the upper and lower jaw lengths and total face height was greater for males than females. Comparison of distance curves between Class I, Class II and Class III skeletal jaw relationships shows that compared to Class I, in Class II skeletal malocclusion for both males and females the upper jaw was larger while lower jaw length and total face height were smaller, whereas in Class III skeletal malocclusion for both males and females the upper jaw was smaller while lower jaw length and total face height were larger.

Class differences (along with 95% CIs) in the upper jaw length for males and females are shown in Figure 7-4 and Figure 7-5. The results for class differences in lower jaw length for males and females are shown in Figure 7-6 and Figure 7-7. Class differences in the total face height for males and females are shown in Figure 7-8 and Figure 7-9. See Appendix C (Section C.2.1) for

corresponding numerical results for class differences in the upper and lower jaw length and total face height for males and females.

The results show that differences in upper and lower jaw length and total face height between normal Class I skeletal relationship and skeletal malocclusions (Class II and Class III) increased with age for both sexes. However, class differences did not increase uniformly between seven and 18 years of age but emerged rapidly during the circumpubertal (age around puberty) growth period for both males and females.

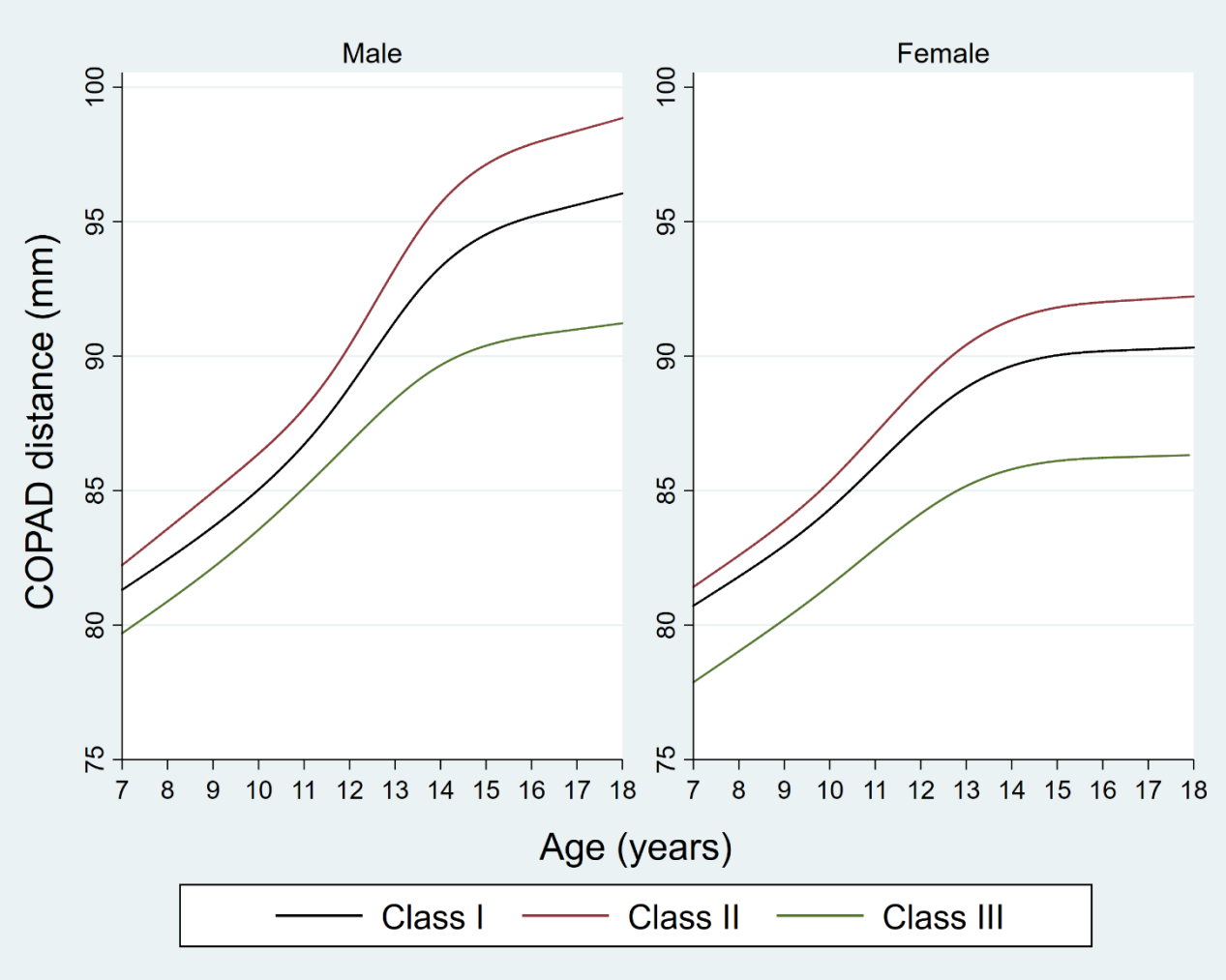


Figure 7-1. Class-specific distance curves for the upper jaw length (COPAD) for males and females.

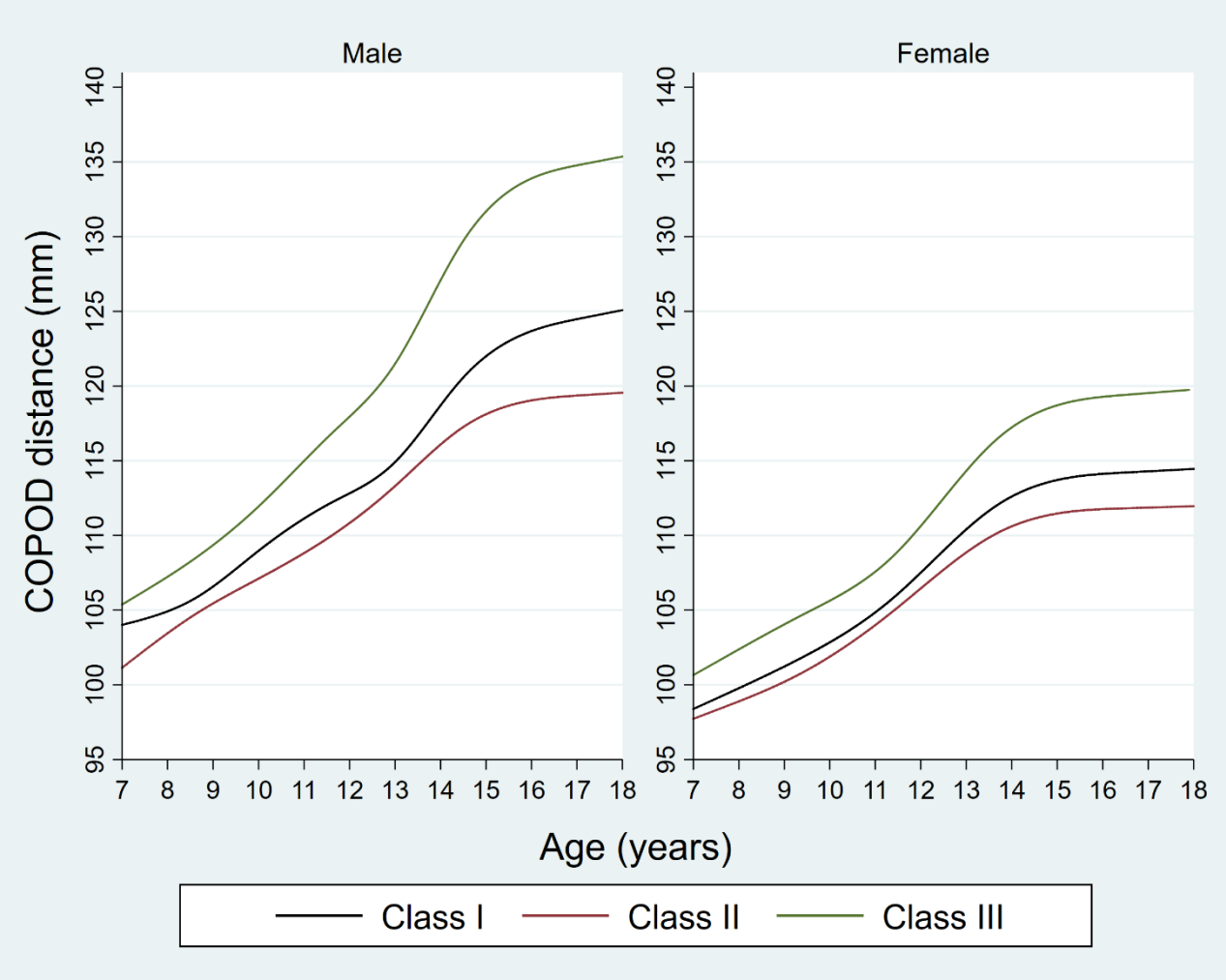


Figure 7-2. Class-specific distance curves for the lower jaw length (COPOD) for males and females.

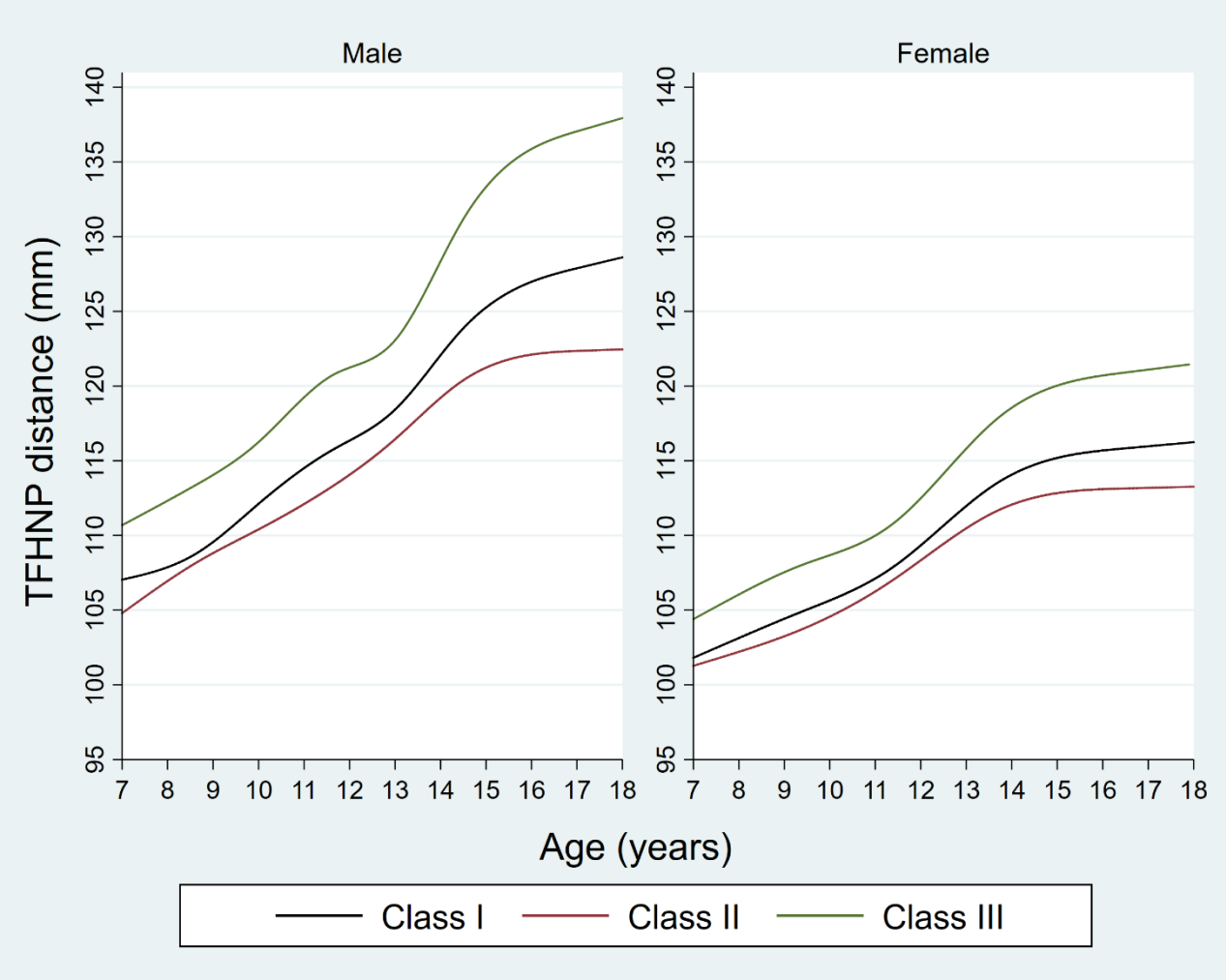


Figure 7-3. Class-specific distance curves for the total face height (TFHNP) for males and females.

Findings show that for both sexes, class differences (skeletal discrepancies) were greater for lower jaw length Figure 7-6 and Figure 7-7) and total face height Figure 7-8 and Figure 7-9) than for upper jaw length Figure 7-4 and Figure 7-5). As shown in Section 7.3.2, the circumpubertal growth velocity for lower jaw length and total face height was greater than for upper jaw length for both sexes. Furthermore, as males showed a higher growth rate around the circumpubertal growth period than females, class differences were greater for males than females. Sex differences were larger for lower jaw length and total face height than the upper jaw length because sex differences in peak growth velocity are greater for lower jaw length and total face height than for upper jaw length.

These findings suggest that sex-specific class differences in jaw sizes and face height are proportional to class differences in growth velocity between the upper and lower jaw length and total face height. For example (see Appendix C Section C.2.1), the difference between Class I and Class II (Class II – Class I) at 18 years of age is greater for lower jaw length (males -5.53 mm, 95% CI -8.77 to -2.28; females -2.49 mm, 95% CI -4.87 to -0.11) and total face height (males -6.16 mm, 95% CI -9.65 to -2.67; females -2.98 mm, 95% CI -5.33 to -0.63) than upper jaw length (males 2.80 mm, 95% CI 0.92 to 4.68; females 1.90 mm, 95% CI 0.61 to 3.20).

The difference between Class I and Class III (Class III – Class I) at 18 years of age is similarly greater for lower jaw length (males 10.28 mm, 95% CI 4.13 to 16.43; females 5.32 mm, 95% CI 1.52 to 9.12) and total face height (males 9.32 mm, 95% CI 2.48 to 16.17; females 5.24 mm, 95% CI 1.46 to 9.02) than upper jaw length (males -4.83 mm, 95% CI -8.32 to -1.34; females -4.00 mm, 95% CI (-6.06 to -1.93).

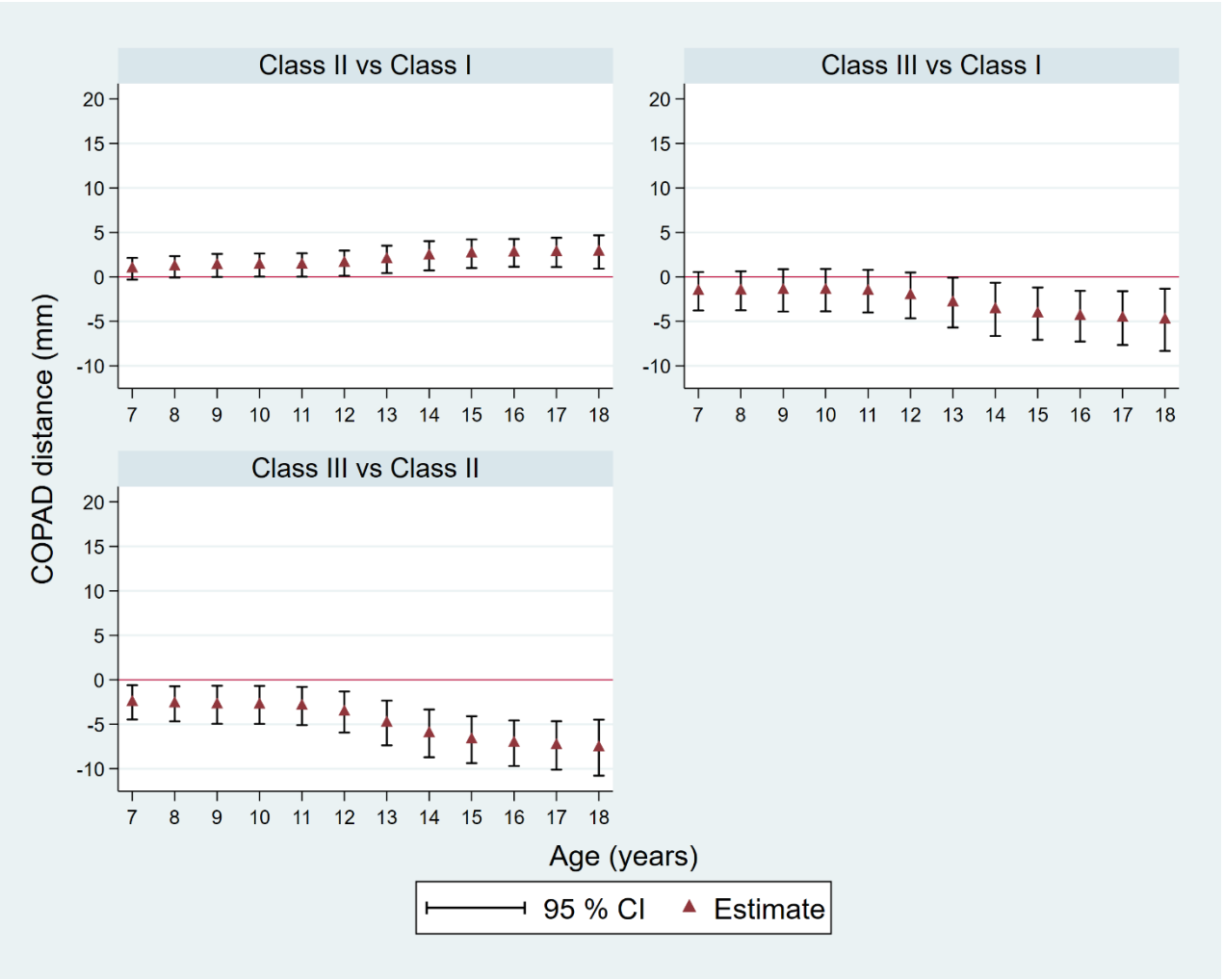


Figure 7-4. Class differences in the upper jaw length (COPAD) distance between seven and 18 years of age for males.

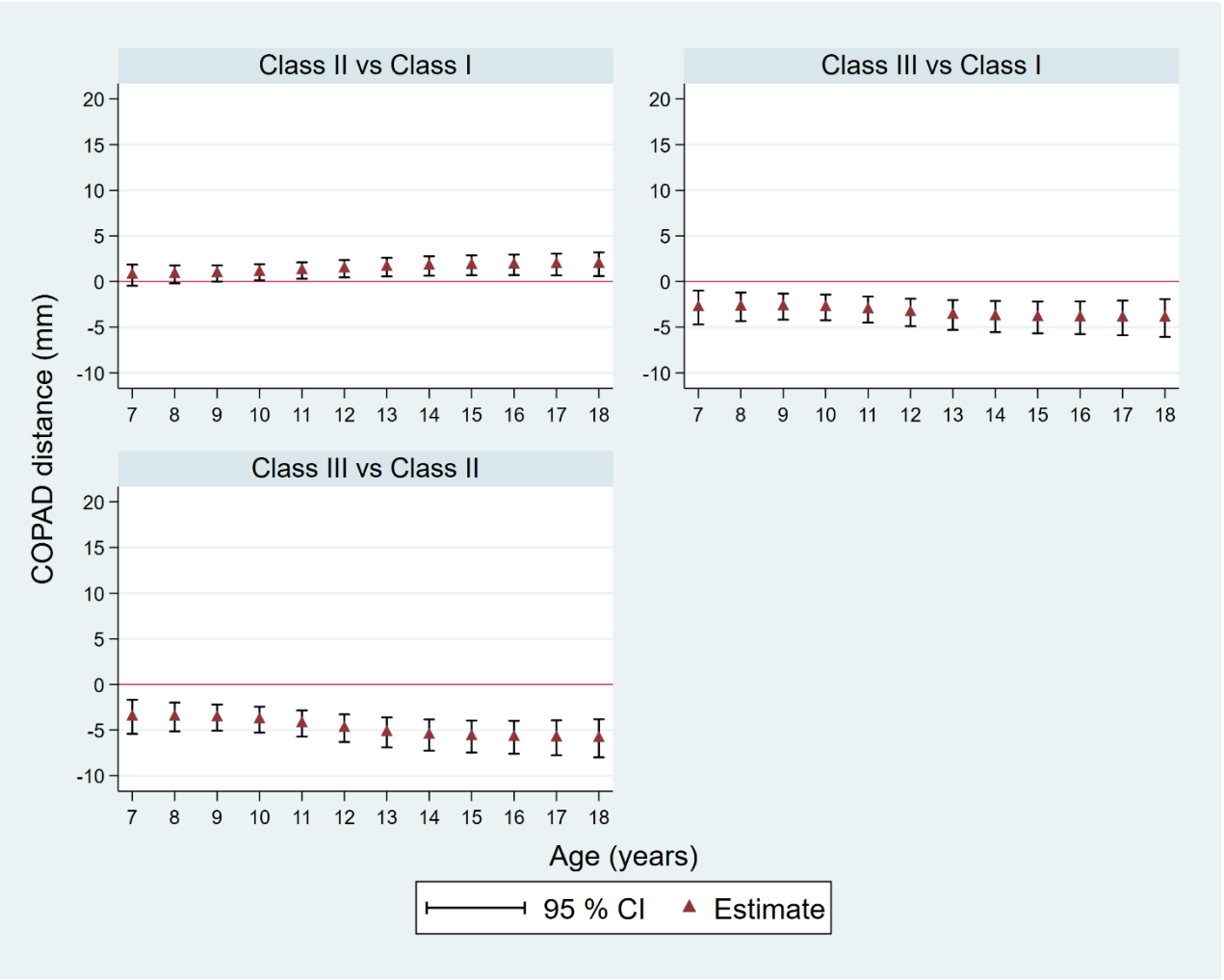


Figure 7-5. Class differences in the upper jaw length (COPAD) distance between seven and 18 years of age for females.

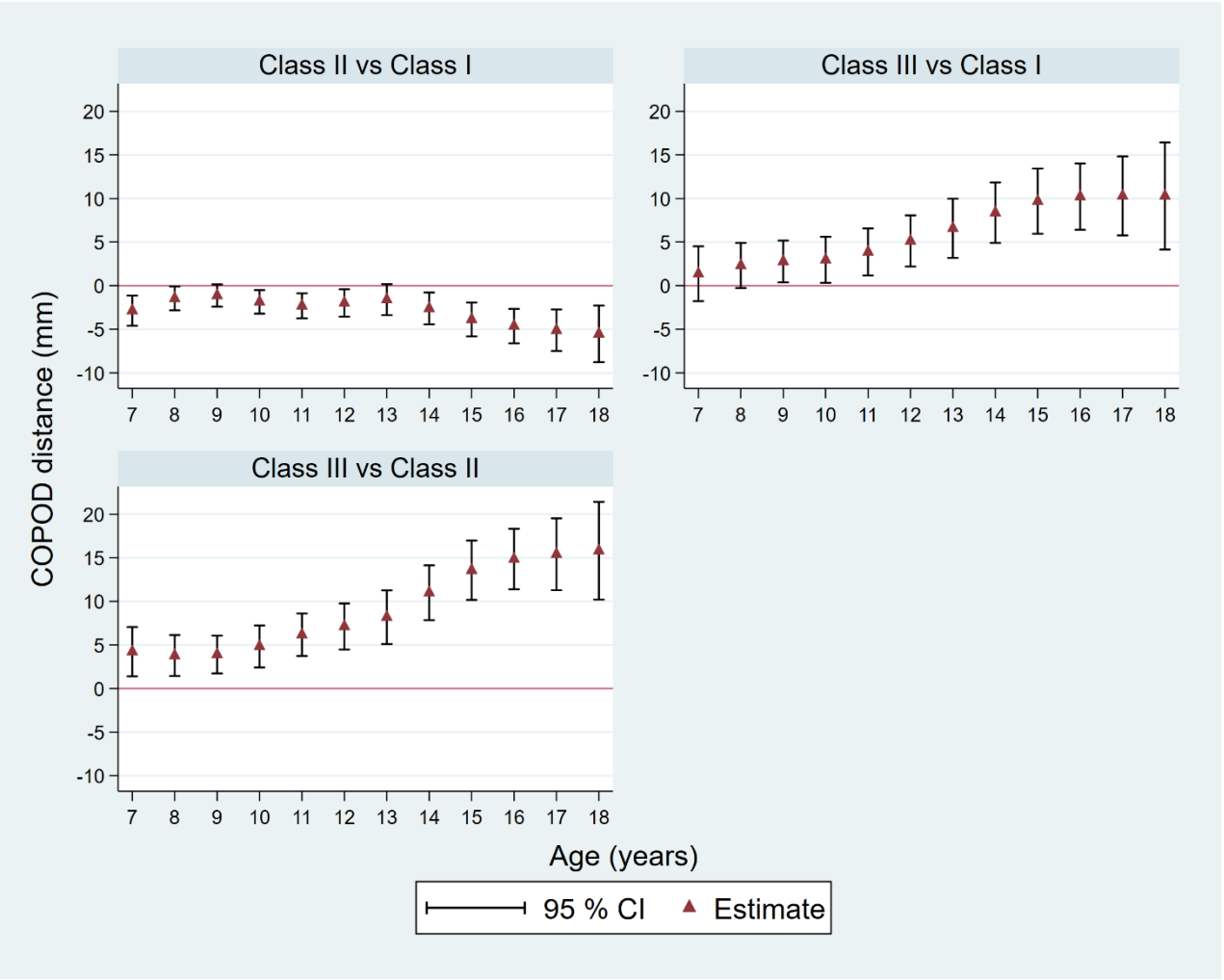


Figure 7-6. Class differences in the lower jaw length (COPOD) distance between seven and 18 years of age for males.

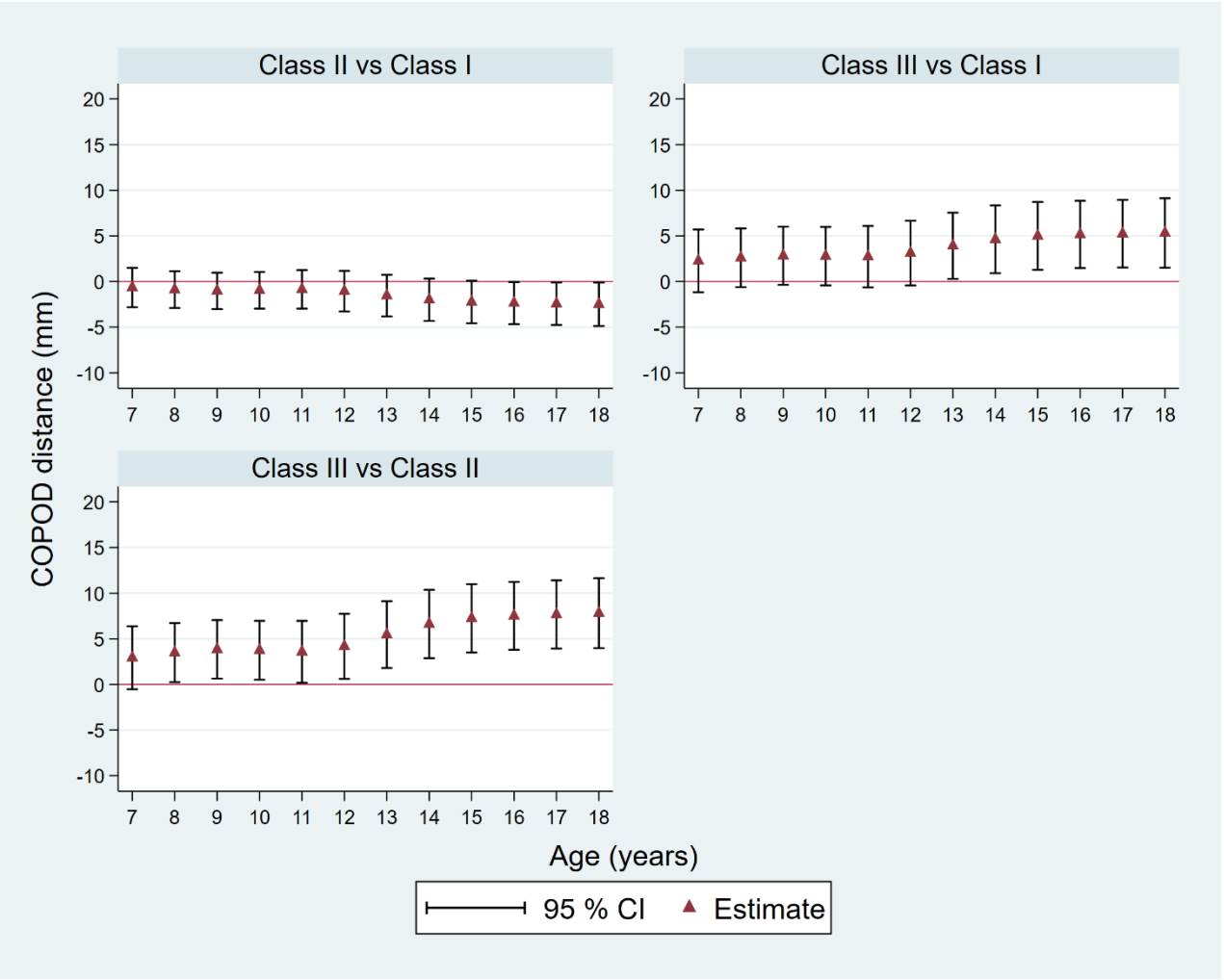


Figure 7-7. Class differences in the lower jaw length (COPOD) distance between seven and 18 years of age for females.

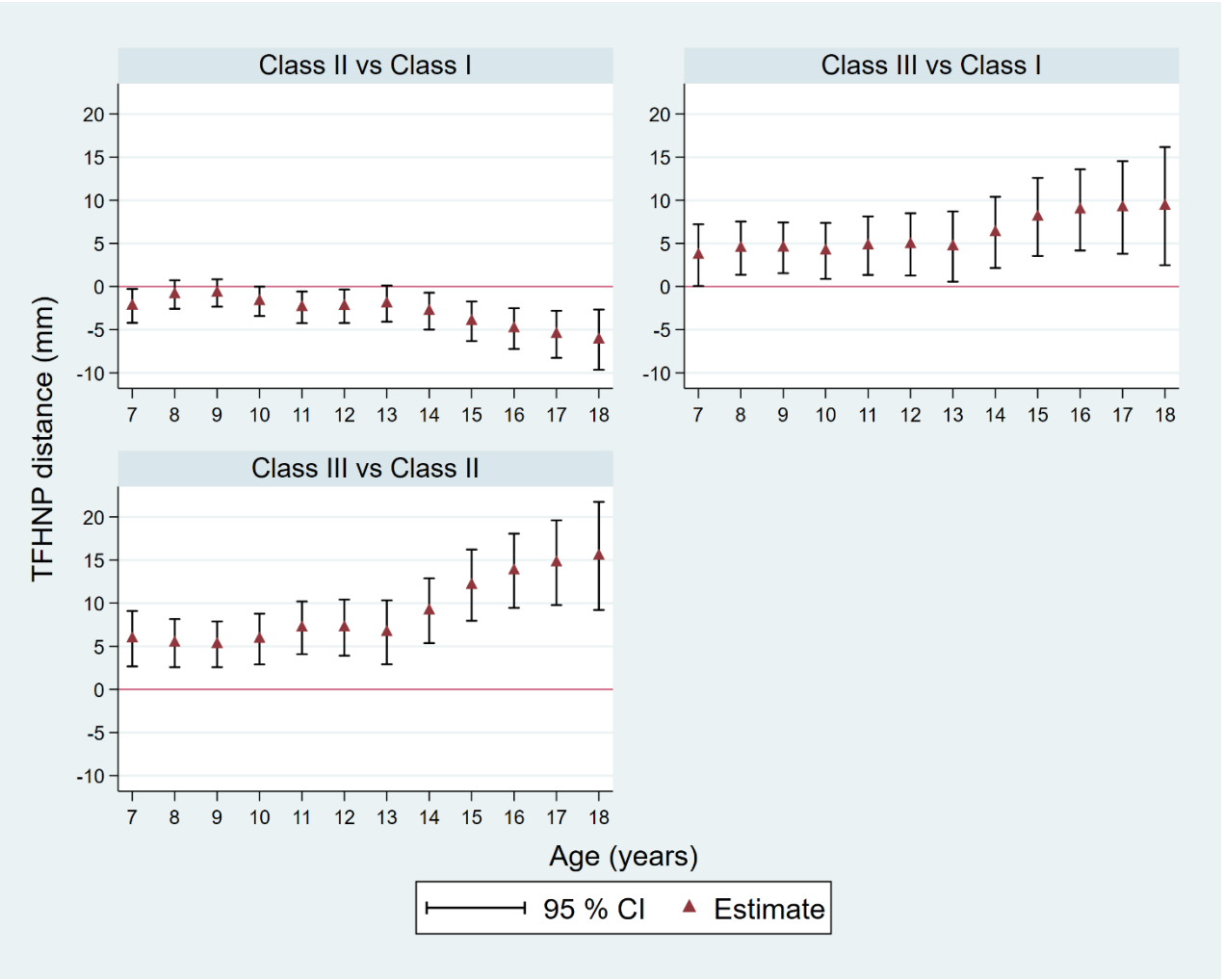


Figure 7-8. Class differences in the total face height (TFHNP) distance between seven and 18 years of age for males.

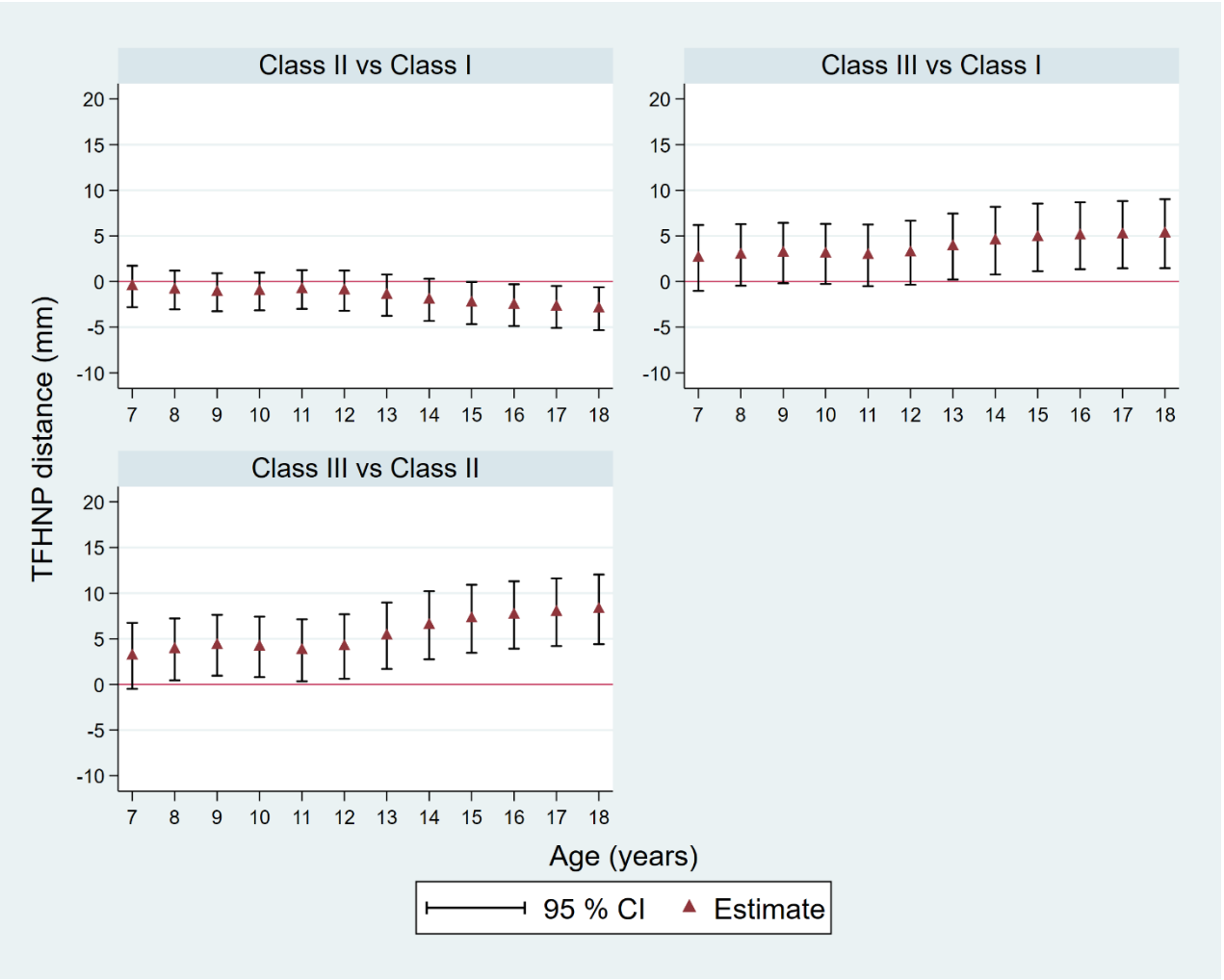


Figure 7-9. Class differences in the total face height (TFHNP) distance between seven and 18 years of age for females.

7.3.2. Growth velocity

Velocity curves (first derivative) for upper and lower jaw length and total face height are shown in Figure 7-10, Figure 7-11 and, Figure 7-12 respectively. For all three growth measures, growth velocity curves show a distinct peak, indicating an adolescent growth spurt for both males and females. Growth velocity curves for lower jaw length and total face height almost mirror each other for both sexes. Growth velocity curves for males show signs of a pre-adolescent growth spurt for lower jaw length and total face height, which are described following the results for the adolescent growth spurt.

For both males and females, the circumpubertal growth velocity for upper jaw length (around 12 to 14 years for males, and 10 to 12 years for females) was lower than for lower jaw length (around 13 to 15 years for males, and 11 to 13 years for females) and total face height (around 13 to 15 years for males, and 11 to 13 years for females). Circumpubertal growth velocity for females was lower than for males. The sex differences in the circumpubertal growth velocity were greater for lower jaw length (Figure 7-11) and total face height (Figure 7-12) than upper jaw length (Figure 7-10).

Class differences (along with 95% CIs) in the growth velocity for upper jaw length are shown as Figure 7-13 and Figure 7-14 for males and females, respectively. Results for class differences in lower jaw length growth velocity for males and females are shown in Figure 7-15 and Figure 7-16, while Figure 7-17 and Figure 7-18 show the class differences in total face height growth velocity for males and females. Appendix C (Section C.2.2) provides corresponding numerical results for class differences in growth velocity of upper and lower jaw length and total face height for males and females.

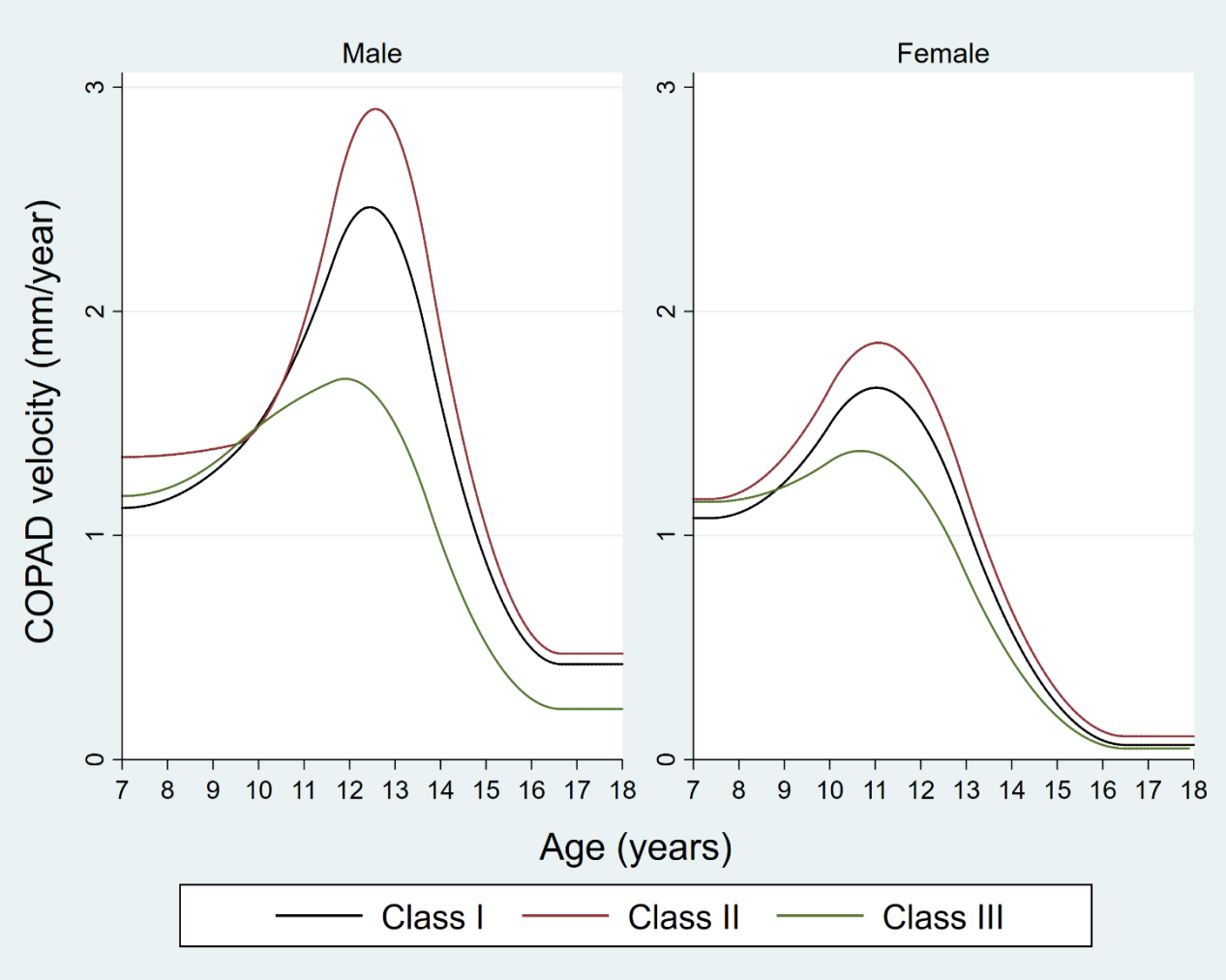


Figure 7-10. Class-specific growth velocity curves for the upper jaw length (COPAD) for males and females.

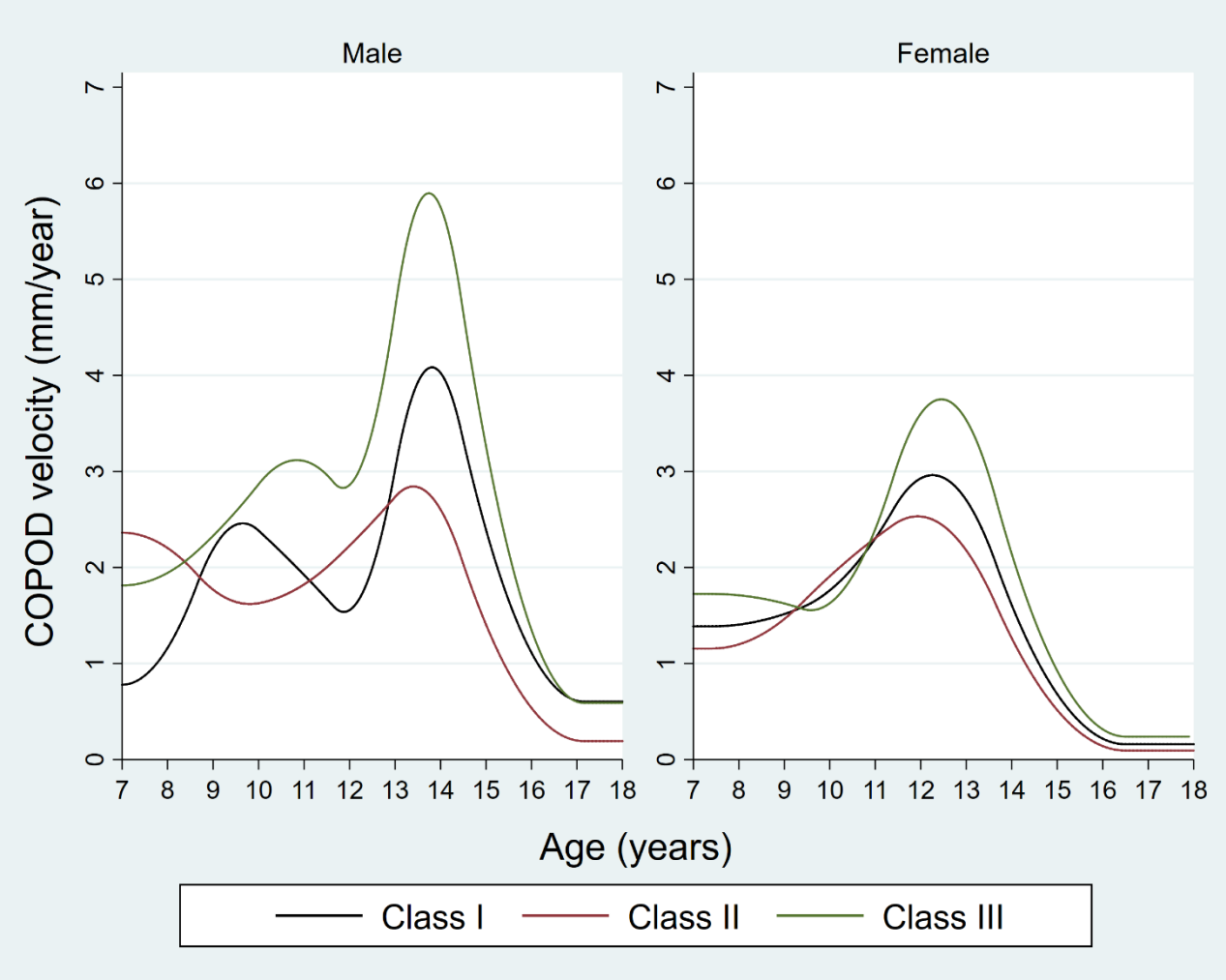


Figure 7-11. Class-specific growth velocity curves for the lower jaw length (COPOD) for males and females.

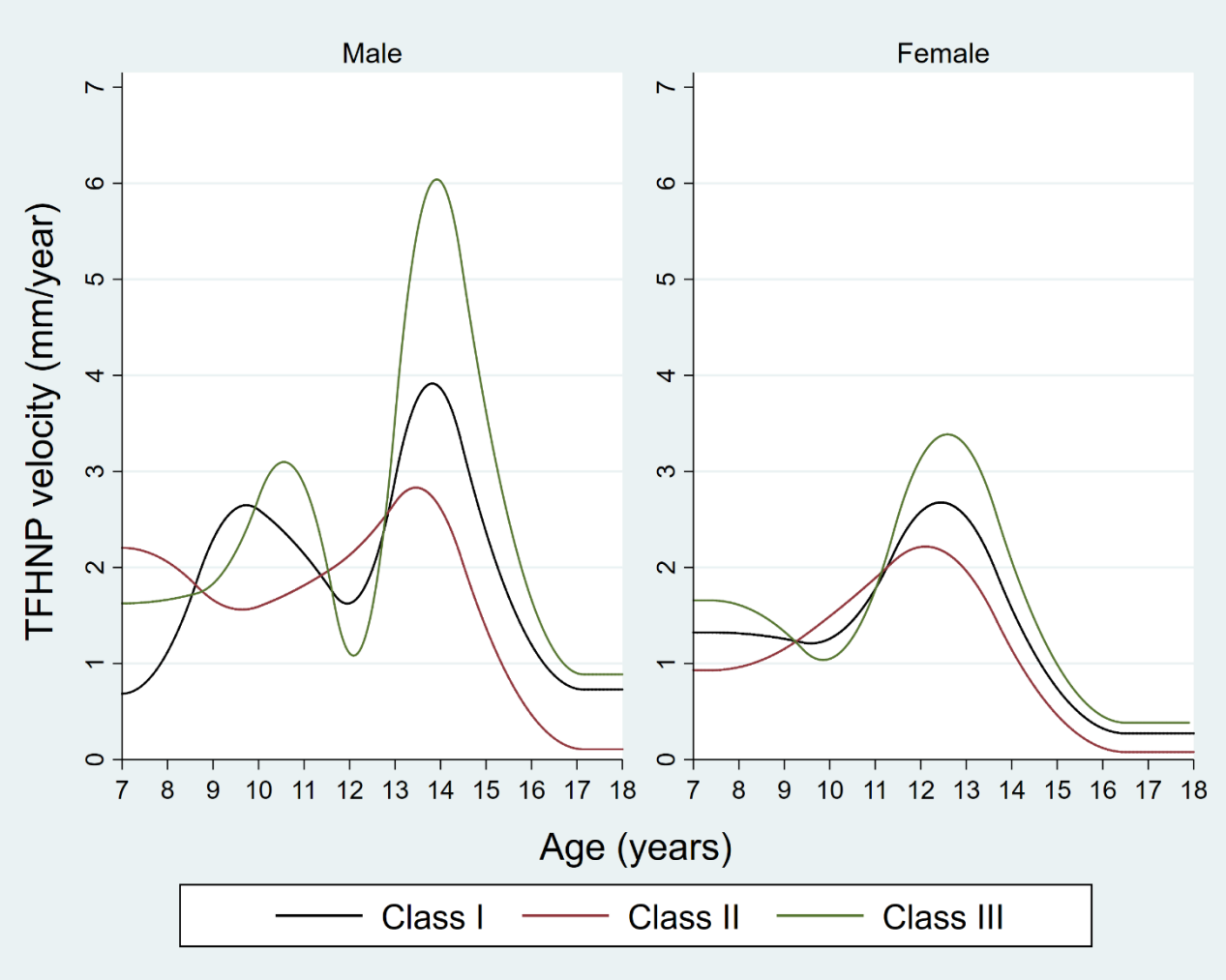


Figure 7-12. Class-specific growth velocity curves for the total face height (TFHNP) for males and females.

The findings show that for both sexes, class differences in circumpubertal growth velocity were smaller for upper jaw length (Figure 7-13 and Figure 7-14) than lower jaw length (Figure 7-15 and Figure 7-16) and total face height (Figure 7-17 and Figure 7-18). For both males and females, class differences emerged earlier for upper jaw length than lower jaw length and total face height. The timing of emergence and the magnitude of class differences in the circumpubertal growth velocity for lower jaw length and total face height are closely related for both males and females.

The results show sexual dimorphism in the timing of emergence and the magnitude of class differences in growth velocity for all three growth measures. Class differences for all three growth measures emerged earlier for females than males, but the magnitude of difference in growth velocity around the circumpubertal growth period was greater for males. Sex differences in the timing of emergence and the magnitude of differences was greater for lower jaw length and total face height than upper jaw length.

As an example of class differences in the circumpubertal growth velocity between growth measures and how they differed for males and females, below are reported class differences at approximately the middle of the circumpubertal growth period for upper jaw length (13 years for males; 11 years for females), lower jaw length (14 years for males; 12 years for females) and total face height (14 years for males; 12 years for females). Appendix C (Section C.2.2) includes full results on class differences in growth velocity between seven and 18 years of age for males and females.

The difference in growth velocity between Class I and Class II (Class II – Class I) was greater for lower jaw length (males -1.42 mm/year, 95% CI -2.40 to -0.45; females -0.52 mm/year, 95% CI -0.92 to -0.12) and total face height (males -1.25 mm/year, 95% CI -2.30 to -0.20; females -0.56 mm/year, 95% CI -0.92 to -0.20) than for upper jaw length (males 0.46 mm/year, 95% CI 0.08 to 0.84; females 0.20 mm/year, 95% CI 0.02 to 0.38).

Similarly, the difference in growth velocity between Class I and Class III (Class III – Class I) was greater for lower jaw length (males 1.73 mm/year, 95% CI -0.33 to 3.79; females 0.83 mm/year, 95% CI 0.17 to 1.50) and total face height (males 2.16 mm/year, 95% CI 0.23 to 4.54; females 0.74 mm/year, 95% CI 0.12 to 1.35) than for upper jaw length (males -0.85 mm/year, 95% CI -1.64 to -0.07; females -0.29 mm/year, 95% CI (-0.59 to -0.01).

Males showed a pre-adolescent growth spurt in lower jaw length (Figure 7-11) and total face height (Figure 7-12) for Class I at around 10 years of age and Class III at around 11 years of age. At both 10 and 11 years of age, the pre-adolescent PGV for Class III was greater than Class I (see Appendix C, Table C-9 and Table C-11). The difference in the pre-adolescent PGV between Class I and Class III at 10 and 11 years of age was greater for lower jaw length (10 years: 0.48 mm/year, 95% CI -0.55 to 1.51; 11 years: 1.19 mm/year, 95% CI -0.41 to 2.79) than for total face height (10 years: 0.12 mm/year, 95% CI -0.98 to 1.22; 11 years: 0.72 mm/year, 95% CI -.95 to 2.39).

Unlike Class I and Class III, the growth velocity curve for males with Class II skeletal jaw relationship showed no sign of a similar pre-adolescent growth spurt in lower jaw length and total face height. Additionally, the growth velocity curve for upper jaw length showed no sign of a pre-adolescent growth spurt for either Class I, Class II or Class III.

For females, no evidence was found for a pre-adolescent growth spurt for any of the three growth measures for either Class I, Class II or Class III.

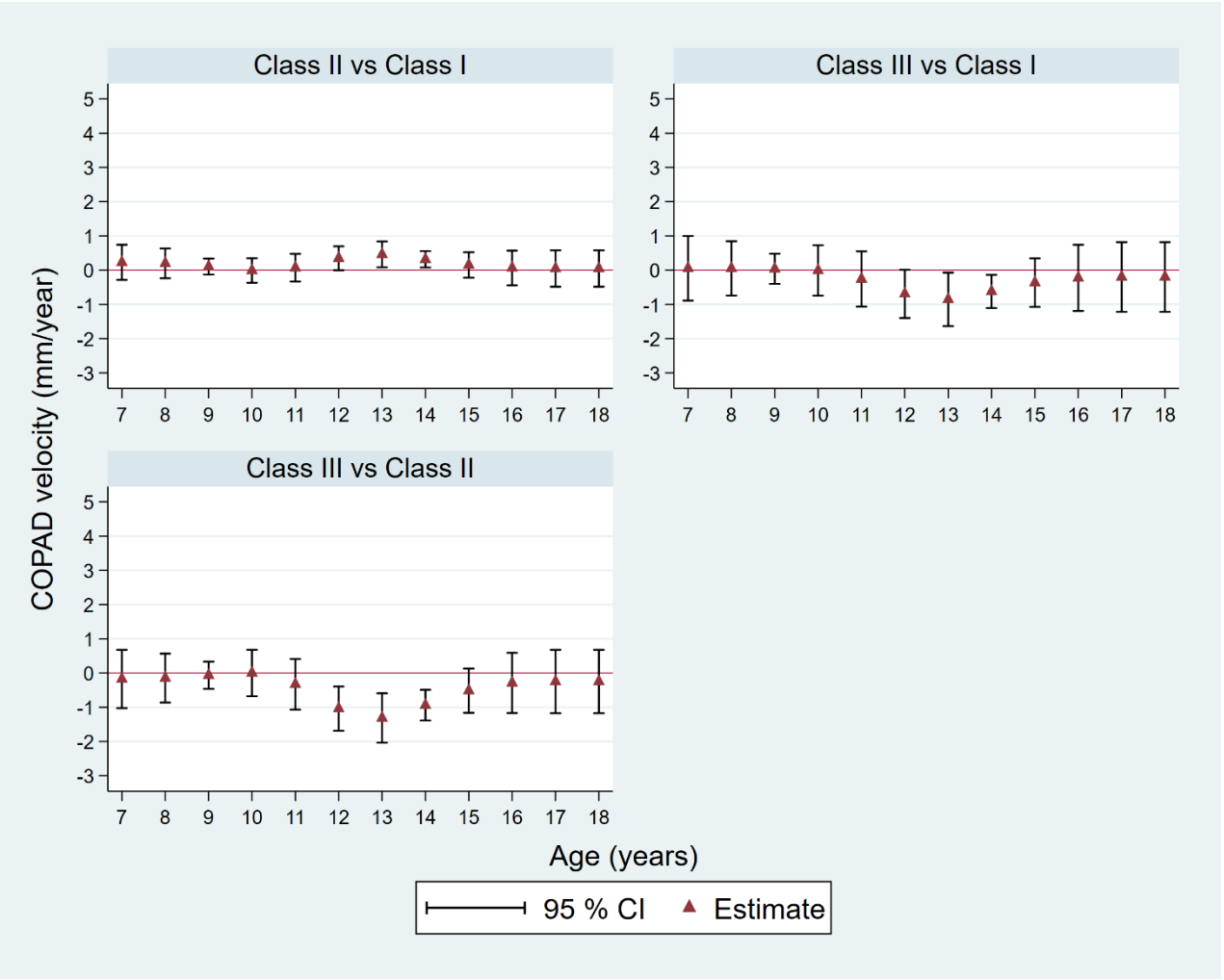


Figure 7-13. Class differences in the upper jaw length growth (COPAD) velocity between seven and 18 years of age for males.

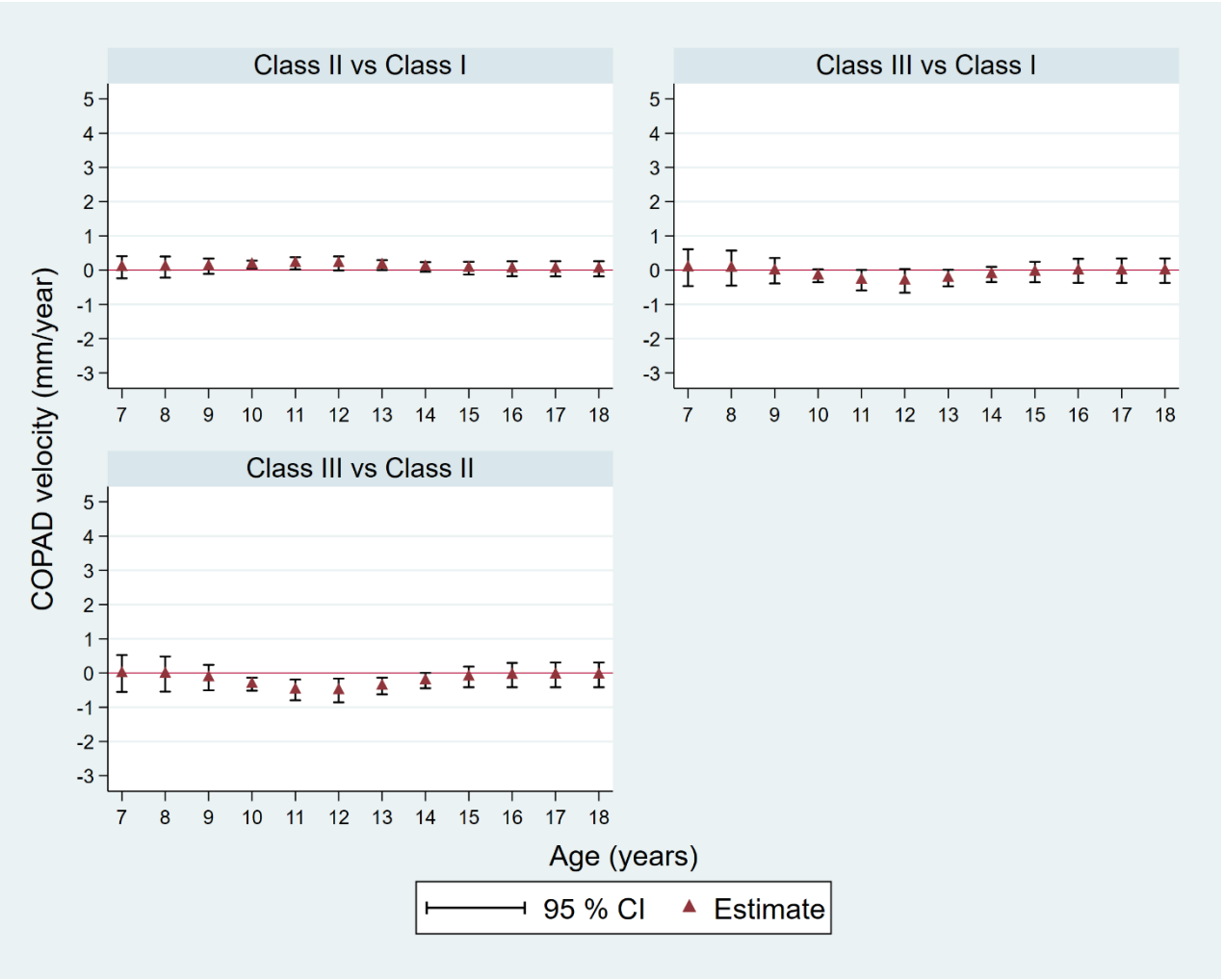


Figure 7-14. Class differences in the upper jaw length (COPAD) growth velocity between seven and 18 years of age for females.

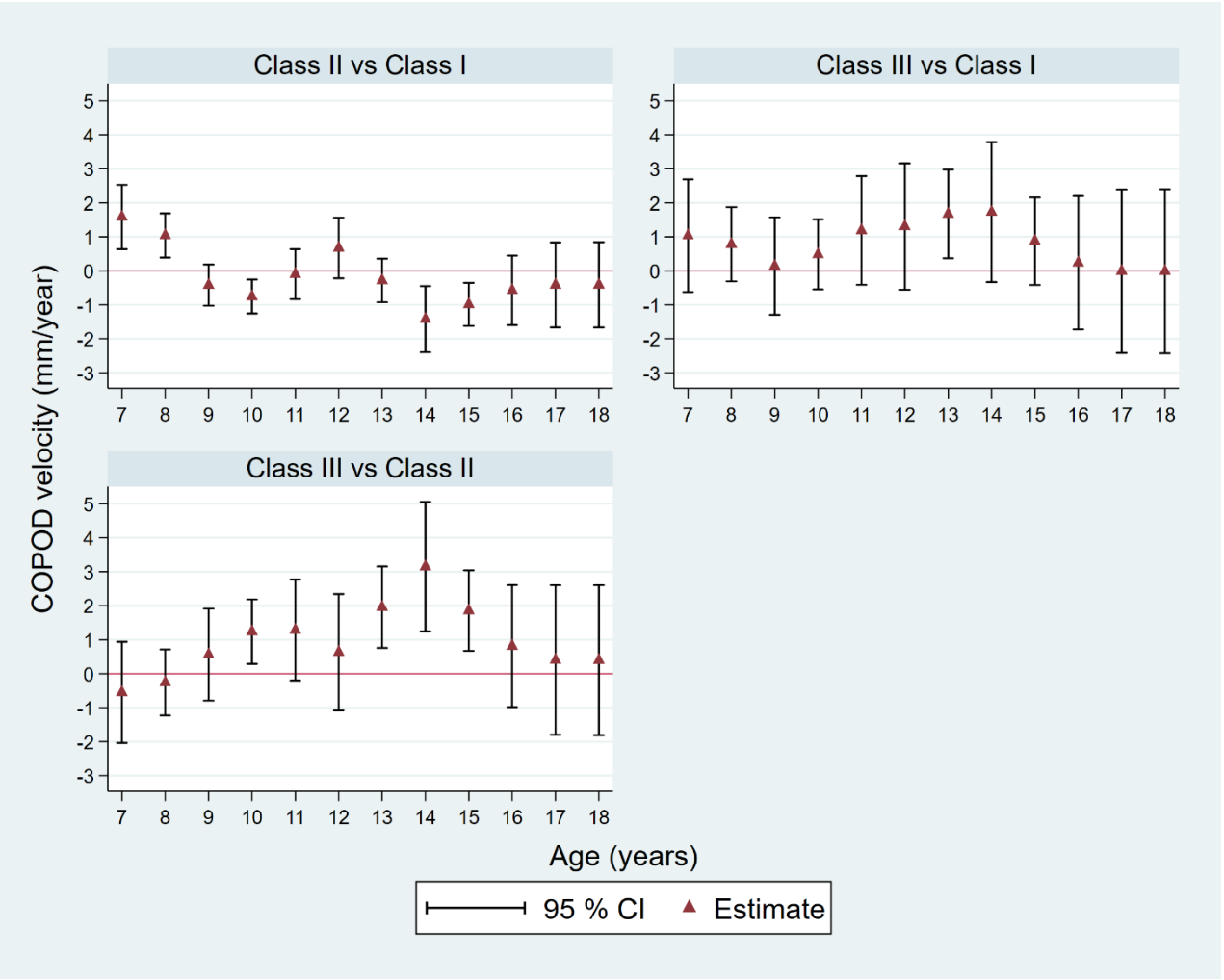


Figure 7-15. Class differences in the lower jaw length (COPOD) growth velocity between seven and 18 years of age for males.

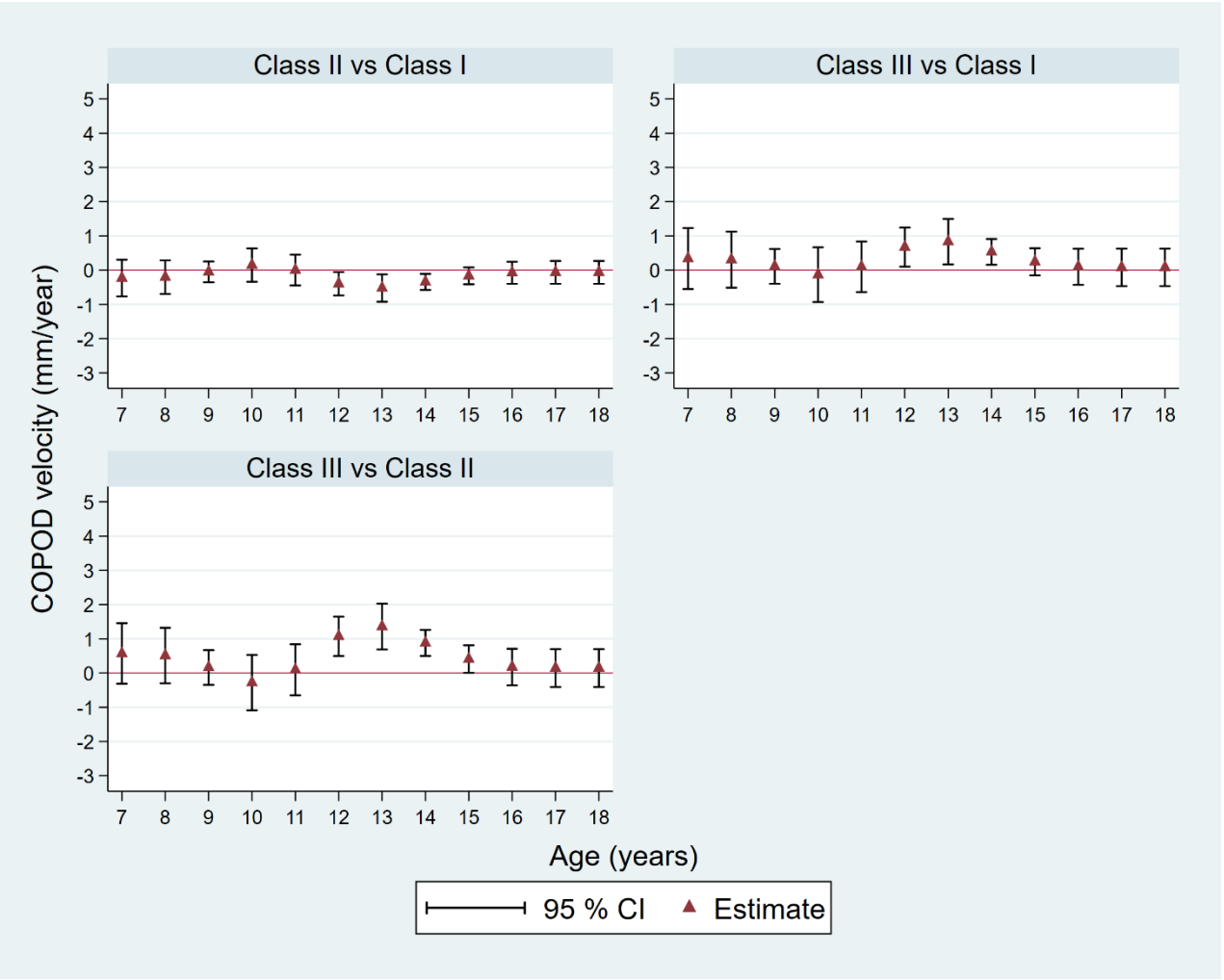


Figure 7-16. Class differences in the lower jaw length (COPOD) growth velocity between seven and 18 years of age for females.

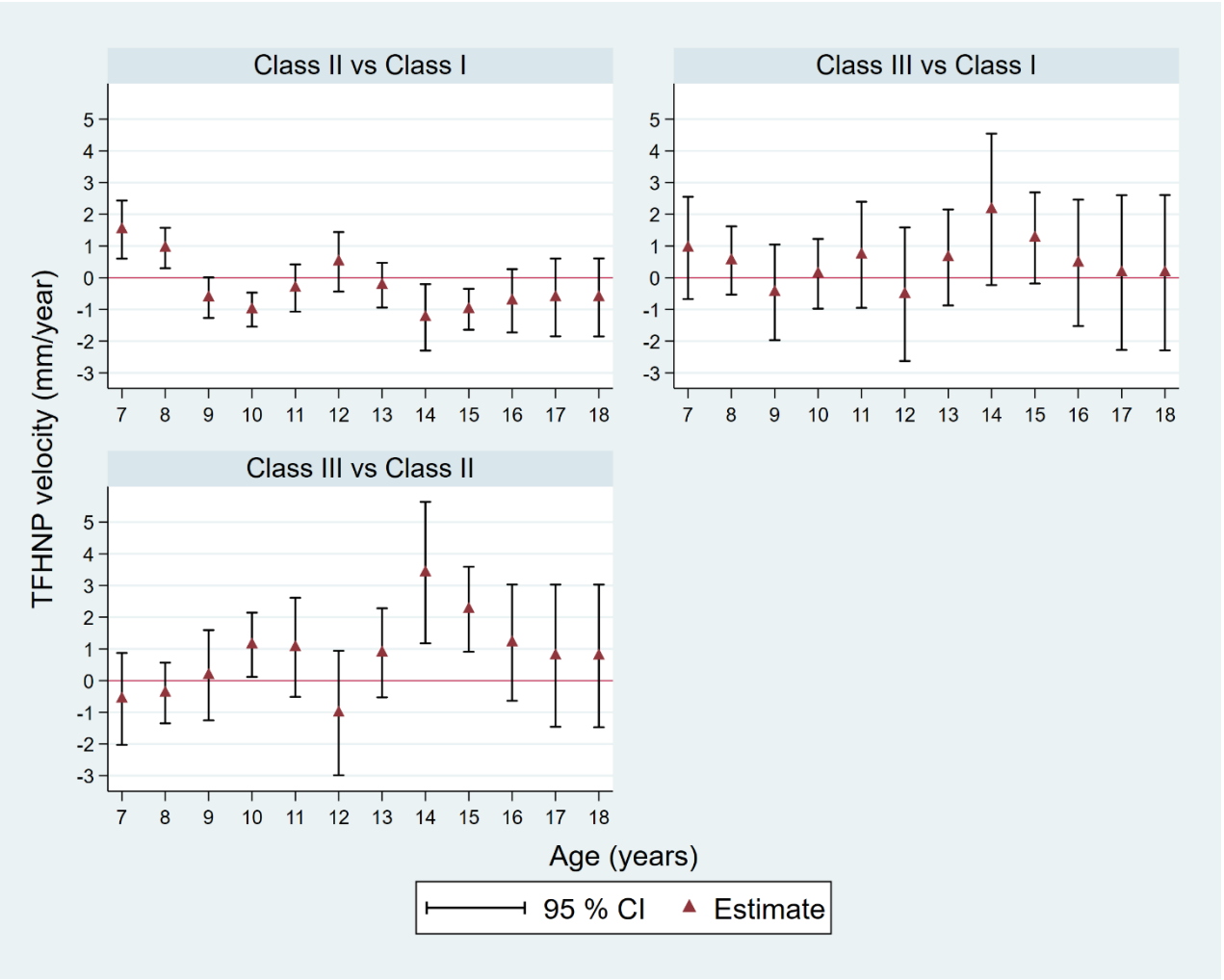


Figure 7-17. Class differences in the total face height (TFHNP) growth velocity between seven and 18 years of age for males.

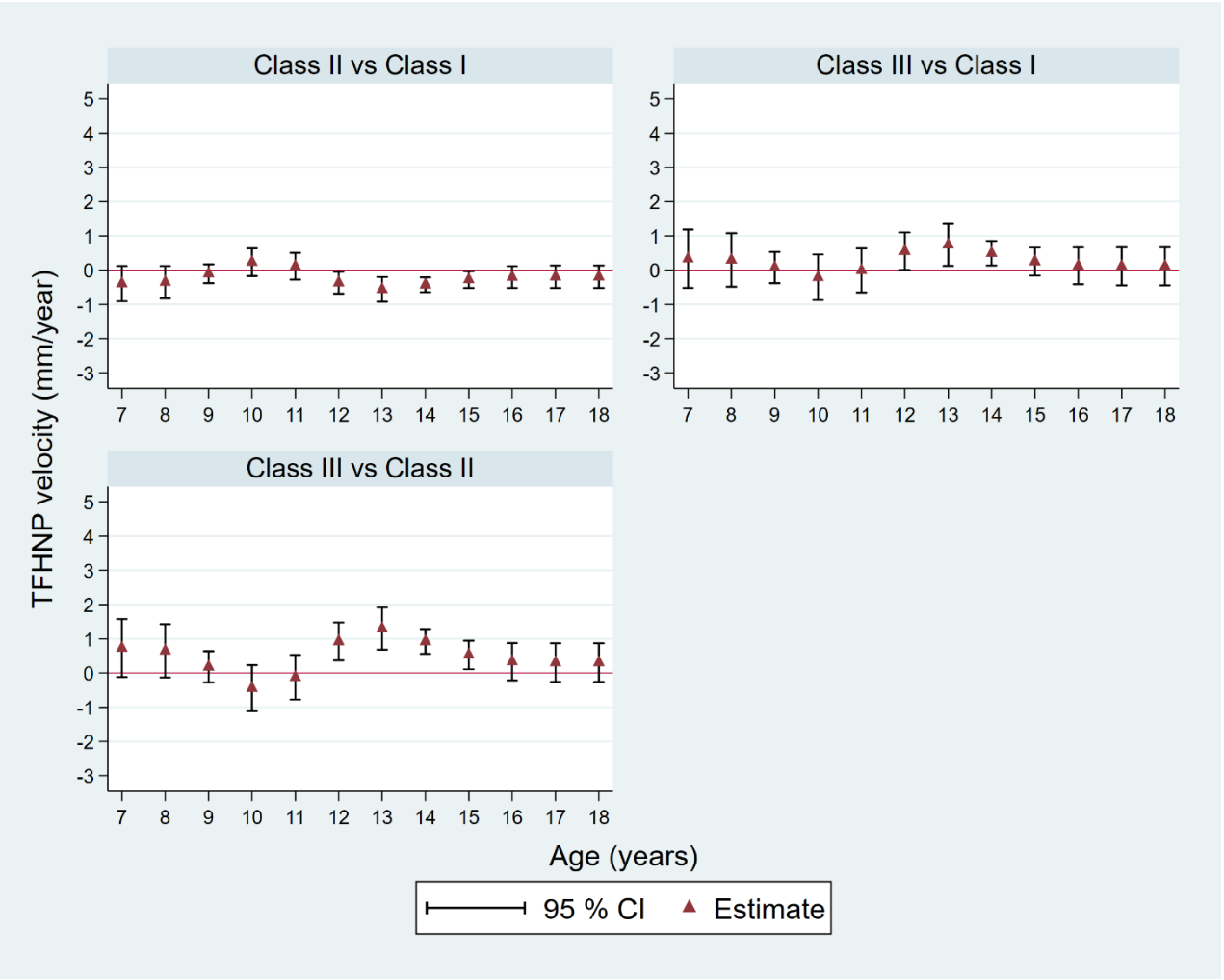


Figure 7-18. Class differences in the total face height (TFHNP) growth velocity between seven and 18 years of age for females.

7.3.3. Growth acceleration

Acceleration curves (second derivative) for upper and lower jaw length and total face height are shown in Appendix C (Section C.2.3). Since the second derivative of a cubic polynomial is linear (see Chapter 4, Section 4.3.2 for details), the growth acceleration curve for the RCS GCM is linear between consecutive knots (after the first knot and before the last knot). Furthermore, since the restricted cubic spline function is linear before the first knot and after the last knot, the second derivative is zero here (see Chapter 4, Section 4.3.2 for details).

As growth acceleration denotes the rate of change in growth velocity, the more intense the growth spurt, the greater the acceleration and deceleration in the growth rate. Therefore, growth acceleration curves for lower jaw length and total face height are steeper than upper jaw length (**Error! Reference source not found.**) for both sexes (see Appendix C, Section C.2.3). Similarly, g rowth acceleration curves are steeper for males than females, particularly for lower jaw length and total face height. Growth acceleration curves for the lower jaw length and total face height look alike.

For both males and females, the circumpubertal growth acceleration for upper jaw length (around 11 to 13 years for males, and nine to 11 years for females) was lower than that for lower jaw length (around 12 to 14 years for males, and 10 to 12 years for females) and total face height (around 12 to 14 years for males, and 10 to 12 years for females). Circumpubertal growth acceleration for females was slower than males, and differences in growth acceleration between males and females were greater for lower jaw length and total face height than for upper jaw length.

Appendix C (Section C.2.3) provides graphical and corresponding numerical results for class differences (along with 95% CIs) in the growth acceleration of upper and lower jaw length and total face height for males and females.

For both sexes, class differences in circumpubertal growth acceleration were smaller for upper jaw length than lower jaw length and total face height. For both males and females, class differences emerged earlier for upper jaw length than lower jaw length and total face height. The timing of emergence and the magnitude of class differences in circumpubertal growth acceleration for lower jaw length and total face height are closely related for both males and females.

The results show sexual dimorphism in the timing of emergence and the magnitude of class differences in the growth acceleration for all three growth measures. Class differences emerged earlier for females than males, but the magnitude of difference in growth acceleration around the circumpubertal growth period was greater for males. Sex differences in the timing of emergence and the magnitude of differences was greater for lower jaw length and total face height than upper jaw length.

As an example of class differences in the circumpubertal growth acceleration between growth measures and how they differed for males and females, following are reported class differences at approximately the middle of the circumpubertal growth period for upper jaw length (12 years for males; 10 years for females), lower jaw length (13 years for males; 11 years for females), and total face height (13 years for males; 11 years for females). Appendix C (Section C.2.3) includes full results on class differences in growth acceleration between seven and 18 years of age for males and females.

The difference in growth velocity between Class I and Class II (Class III – Class I) was greater for lower jaw length (males -2.08 mm/year^2, 95% CI -3.98 to -0.18; females -0.34 mm/year^2, 95% CI -0.72 to 0.04) and total face height (males -1.77 mm/year^2, 95% CI -3.77 to 0.23; females -0.38 mm/year^2, 95% CI -0.70 to -0.06) than for upper jaw length (males 0.25 mm/year^2, 95% CI -0.12 to 0.61; females 0.06 mm/year^2, 95% CI -0.16 to 0.28).

Similarly, the difference in growth velocity between Class I and Class III (Class III – Class I) was greater for the lower jaw length (males 0.62 mm/year^2, 95% CI -3.35 to 4.58; females 0.50 mm/year^2, 95% CI -0.13 to 1.12) and total face height (males 2.97 mm/year^2, 95% CI -1.47 to 7.42; females 0.47 mm/year^2, 95% CI -0.07 to 1.01) than upper jaw length (males -0.36 mm/year^2, 95% CI -1.10 to 0.37; females -0.18 mm/year^2, 95% CI -0.55 to 0.19).

Lastly, growth acceleration curves for males confirmed the presence of a pre-adolescent growth spurt at around nine to 10 years of age for lower jaw length and total face height for Class I and Class III (see Section 7.3.2 for corresponding growth velocity curves). Like the difference in age corresponding to circumpubertal periods of growth velocity and growth acceleration, the peak growth acceleration (advanced) for the pre-adolescent growth spurt was advanced by approximately one year when compared with PGV for the pre-adolescent growth spurt (see Section 7.3.2 for circumpubertal growth velocity periods).

The peak growth acceleration for pre-adolescent growth spurts for lower jaw length and total face height occurred earlier for Class I (at around nine years of age) and Class III (around 10 years of age) when compared with PGV for corresponding pre-adolescent growth spurts for Class I (at around 10 years of age) and Class III (around 11 years of age). The pre-adolescent peak growth acceleration for Class III was greater than Class I, reached at nine and 10 years of age, respectively (see Appendix C).

Unlike Class I and Class III, the growth acceleration curve for males with Class II skeletal jaw relationship showed no sign of a similar pre-adolescent growth spurt for lower jaw length and total face height. Additionally, the growth acceleration curve for upper jaw length showed no sign of a pre-adolescent growth spurt for either Class I, Class II or Class III.

For females, no evidence was found for a pre-adolescent growth spurt for any of the three growth measures for either Class I, Class II or Class III.

7.3.4. Timing and the intensity of the adolescent growth spurt

Table 7-1 and Table 7-2 summarise class differences in the population-average timing (APGV) and intensity (PGV) of the adolescent growth spurt for the upper and lower jaw length and total face height. Irrespective of the skeletal jaw relationship (Class I, Class II or Class III), the adolescent growth spurt for all three growth measures occurs approximately 1.25 years earlier for females than males. Additionally, the intensity of the adolescent growth spurt for all three measures is greater for males than females. Below I report class differences in the timing and intensity of the adolescent growth spurt for upper and lower jaw length and total face height (Class II versus Class I, and Class III versus Class I).

Table 7-1. Class differences in population-average age at peak growth velocity (APGV, years) and peak growth velocity (PGV, mm/year) for males.

Outcome	Estimate			Estimated difference (95% confidence intervals)		
Growth parameter	Class I	Class II	Class III	Class II vs Class I (Class II – Class I)	Class III vs Class I (Class III – Class I)	Class III vs Class II (Class III – Class II)
Upper jaw length (COPAD)						
APGV	12.44	12.57	11.92	0.13 (-0.80 to 1.47)	-0.52 (-1.59 to -0.09)	-0.65 (-1.66 to -0.17)
PGV	2.46	2.90	1.70	0.44 (0.10 to 1.19)	-0.76 (-1.51 to -0.11)	-1.20 (-2.06 to -0.62)
Lower jaw length (COPAD)						
APGV	13.81	13.40	13.85	-0.41 (-1.21 to -0.08)	0.04 (-1.56 to 0.45)	0.45 (0.17 to 1.18)
PGV	4.08	2.84	5.90	-1.24 (-1.96 to -0.39)	1.82 (0.19 to 3.11)	3.06 (1.18 to 4.20)
Total face height (TFHNP)						
APGV	13.82	13.46	13.92	-0.36 (-1.10 to -0.04)	0.10 (-1.69 to 1.44)	0.46 (0.11 to 1.29)
PGV	3.91	2.83	6.04	-1.08 (-2.06 to -0.46)	2.13 (0.38 to 3.42)	3.21 (1.40 to 4.97)

APGV age at peak growth velocity (years); PGV peak growth velocity (millimetres/year)

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Table 7-2. Class differences in population-average age at peak growth velocity (APGV, years) and peak growth velocity (PGV, mm/year) for females.

Outcome	Estimate			Estimated difference (95% confidence intervals)		
Growth parameter	Class I	Class II	Class III	Class II vs Class I (Class II – Class I)	Class III vs Class I (Class III – Class I)	Class III vs Class II (Class III – Class II)
Upper jaw length (COPAD)						
APGV	11.03	11.07	10.68	0.04 (-0.54 to 0.72)	-0.35 (-1.64 to -0.05)	-0.39 (-1.66 to -0.09)
PGV	1.66	1.86	1.38	0.20 (0.02 to 0.59)	-0.28 (-0.65 to -0.09)	-0.48 (-1.08 to -0.11)
Lower jaw length (COPAD)						
APGV	12.26	11.93	12.45	-0.33 (-1.26 to -0.02)	0.19 (-0.35 to 0.71)	0.52 (0.13 to 1.44)
PGV	2.96	2.53	3.75	-0.43 (-0.78 to -0.09)	0.79 (0.13 to 1.87)	1.22 (0.47 to 2.23)
Total face height (TFHNP)						
APGV	12.45	12.10	12.59	-0.35 (-1.00 to -0.08)	0.14 (-0.12 to 0.39)	0.49 (0.17 to 1.41)
PGV	2.68	2.22	3.39	-0.46 (-0.77 to -0.07)	0.71 (0.03 to 1.71)	1.17 (0.41 to 2.18)

APGV age at peak growth velocity (years); PGV peak growth velocity (millimetres/year)

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Compared to Class I, the timing of the adolescent growth spurt for lower jaw length was earlier for both males (-0.41 years, 95% CI -1.21 to -0.08) and females (-0.33 years, 95% CI -1.26 to -0.02). The intensity of the adolescent growth spurt for lower jaw length was weaker for Class II than Class I (males: -1.24 mm/year, 95% CI -1.96 to -0.39; females: -0.43 mm/year, 95% CI -0.78 to -0.09).

In contrast to lower jaw length, the adolescent growth spurt for upper jaw length was more intense for Class II than Class I (males: 0.44 mm/year, 95% CI 0.10 to 1.19; females: 0.20 mm/year, 95% CI 0.02 to 0.59). For females, the adolescent growth spurt for upper jaw length occurred almost at the same time as for Class I and Class II (0.04 years, 95% CI -0.54 to 0.72). For males, the adolescent growth spurt for upper jaw length occurred slightly later for Class II than Class I (0.13 years, 95% CI -0.80 to 1.47).

Like lower jaw length, the adolescent growth spurt for total face height occurred earlier in Class II than Class I (males: -0.36 years, 95% CI -1.10 to -0.04; females: -0.35 years, 95% CI -1.00 to -0.08), and the intensity of the spurt was weaker for Class II than for Class I (males: -1.08 mm/year, 95% CI -2.06 to -0.46; females: -0.46 mm/year, 95% CI -0.77 to -0.07).

For Class III, the adolescent growth spurt for upper jaw length occurred earlier than Class I (male: -0.52 year, 95% CI -1.59 to -0.09; females: -0.35 year, 95% CI -1.64 to -0.05), and the intensity of the growth spurt was weaker for Class III than Class I (males: -0.76 mm/year, 95% CI -1.51 to -0.11; females: -0.28 mm/year, 95% CI -0.65 to -0.09). Unlike upper jaw length, the adolescent spurt for lower jaw length was more intense for Class III than Class I (males: 1.82 mm/year, 95% CI 0.19 to 3.11; females: 0.79 mm/year, 95% CI 0.13 to 1.87).

For males, the adolescent growth spurt for upper jaw length occurred almost at the same time for Class I and Class II (.04 years, 95% CI -1.56 to 0.45). For females, the adolescent growth spurt for the upper jaw length occurred slightly later for Class II than Class I (0.19 years, 95% CI -0.35 to 0.71).

Similar to lower jaw length, the intensity of the adolescent growth spurt for total face height was greater for Class III than Class I (males: 2.13 mm/year, 95% CI 0.38 to 3.42; females: 0.71 mm/year, 95% CI 0.03 to 1.71) and the timing of the spurt was slightly after Class I (males: 0.10 years, 95% CI -1.69 to 1.44; females: 0.14 years, 95% CI -0.12 to 0.39).

Individual-specific estimates for the timing and intensity of the adolescent growth spurt for all growth measures are summarised in Table 7-3 for males, and Table 7-4 for females. Between-individual variability in both timing and intensity was larger for lower jaw length and total face height than upper jaw length, with a wider interquartile range. Compared to females, males showed a greater between-individual variability in the timing and intensity of the adolescent growth spurt, especially for lower jaw length and total face height, with a wider interquartile range.

Table 7-3. Summary of individual-specific age at peak growth velocity (APGV, years) and peak growth velocity (PGV, mm/year) for males. The summary statistics shown are median and interquartile range (IQR).

Outcome Growth parameter	Class I	Class II	Class III
Upper jaw length (COPAD)			
APGV	12.45	12.60	12.04
	(12.01 to 12.67)	(12.36 to 12.72)	(10.94 to 12.50)
PGV	2.43	2.92	1.78
	(2.12 to 2.91)	(2.51 to 3.47)	(1.33 to 1.90)
Lower jaw length (COPOD)			
APGV	13.67	12.69	13.73
	(12.25 to 13.95)	(12.14 to 13.66)	(13.00 to 13.80)
PGV	4.75	3.61	5.79
	(3.27 to 5.75)	(3.08 to 4.35)	(4.56 to 6.56)
Total face height (TFHNP)			
APGV	13.65	12.75	13.91
	(12.04 to 13.93)	(12.14 to 13.61)	(10.78 to 13.94)
PGV	4.51	3.68	5.66
	(3.41 to 5.48)	(3.03 to 4.32)	(3.54 to 7.34)

APGV age at peak growth velocity (years); PGV peak growth velocity (millimetres/year).

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Table 7-4. Summary of individual-specific age at peak growth velocity (APGV) and peak growth velocity (PGV) for females. The summary statistics shown are median and interquartile range (IQR).

Outcome Growth parameter	Class I	Class II	Class III
Upper jaw length (COPAD)			
APGV	11.06	11.11	10.68
	(10.64 to 11.27)	(10.68 to 11.28)	(10.34 to 11.06)
PGV	1.77	1.97	1.41
	(1.63 to 1.90)	(1.78 to 2.12)	(1.34 to 1.49)
Lower jaw length (COPOD)			
APGV	12.29	11.83	12.47
	(11.71 to 12.49)	(10.91 to 12.39)	(12.29 to 12.60)
PGV	3.08	2.66	3.53
	(2.63 to 3.48)	(2.37 to 3.19)	(3.40 to 4.15)
Total face height (TFHNP)			
APGV	12.46	12.2	12.62
	(12.10 to 12.59)	(11.32 to 12.41)	(12.45 to 12.64)
PGV	2.68	2.34	3.24
	(2.18 to 3.17)	(2.03 to 2.81)	(2.94 to 3.59)

APGV age at peak growth velocity (years); PGV peak growth velocity (millimetres/year).

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

7.3.5. A note on confidence intervals

CIs not including the value of 'zero effect' indicate statistically significant differences at the 5% level. As it is now recommended to report clinically significant findings rather than statistically significant results (see Section 7.2), this study did not focus on interpreting class differences in terms of their statistical significance. Rather, focus was on describing a general pattern of class differences in the distance, growth velocity, growth acceleration and adolescent growth spurt parameters (APGV and PGV).

As the literature suggests (see Section 7.2), CIs for class comparisons (contrasts) are reported along with point estimates. The CIs inform the precision of the estimates and help in evaluating the level of certainty or uncertainty of the evidence. The results show wide CIs for class differences in the distance, velocity, acceleration and adolescent growth spurt parameters. This suggests that the level of uncertainty of class differences is high. This is particularly true for comparisons involving Class III, because the sample size is very small when compared with the Class I and Class II sample sizes (see Chapter 6, Table 6-1). The CIs are wider for males than females because of following two reasons: i) the sample size is slightly smaller for males than females (139 females and 128 males), and ii) variability in growth parameters is greater for males than females.

As an example, the range of probable effects around the difference in upper jaw length between Class I and Class III at 18 years of age for males (-4.83 mm) is -8.32 to -1.34 mm (see Appendix C, Table C-1). This indicates that, though the sample average size of upper jaw is smaller in Class III than Class I (negative effect size), the difference could be very small (-1.34 mm) or substantial (8.32 mm).

7.4. Discussion

Normal downward and forward growth of the upper and lower jaw is essential to achieve a harmonious anteroposterior skeletal jaw relationship (Class I) and balanced total face height. Abnormal growth in either or both jaws can result in anteroposterior skeletal malocclusions (Class II and Class III), which are often accompanied by vertical skeletal discrepancies (Buschang et al., 2017; Proffit, 2006; Proffit et al., 2014).

Though many previous studies (a total 37 studies; see systematic review Chapter 2, Section 2.2 for details) have used longitudinal growth data to evaluate jaw growth during adolescence, no study has compared growth for upper and lower jaw length and total face height for all three classes of skeletal jaw relationships (Class I, Class II and Class III).

Out of a total 37 studies, the majority of the studies (29 studies) applied simple statistical methods (such as ordinary least square regression) to longitudinal data (see Chapter 2 Section 2.2 for details). Out of these 29 studies, 10 studies clearly reported the class of skeletal jaw relationship (Class I, Class II or Class III). While seven studies included individuals with normal skeletal jaw relationship i.e., Class I (Bishara, 1981; Bishara et al., 1981; Jamison et al., 1982; Mitani, 1977; Nanda, 1971; Ochoa & Nanda, 2004; Thilander et al., 2005), three studies evaluated jaw growth changes for Class I normal skeletal jaw and Class II skeletal malocclusion (Baccetti et al., 2011; Baughan et al., 1979b; Buschang et al., 1983). Out of these 10 studies, one study (Baughan et al., 1979b) included data for females only whereas remaining nine studies analysed data for both sexes.

Out of the total 37 studies, eight studies applied CP GCMs to jaw growth data (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996). Findings of all eight studies have been discussed in Chapter 6 (Section 6.4.6). Except for one study (Buschang et al., 1988b), authors did not report class-specific distribution of the sample or growth parameters (e.g., Class I, Class II or

Class III) . Buschang et al. (1988b) analysed and compared growth changes between Class I and Class II jaw relationships for males and females.

As findings of my study and previous studies (Baccetti et al., 2011; Baughan et al., 1979b; Bishara, 1981; Bishara et al., 1981; Buschang et al., 1983; Buschang et al., 1988b; Jamison et al., 1982; Mitani, 1977; Nanda, 1971; Ochoa & Nanda, 2004; Thilander et al., 2005) show that growth trajectories and adolescent growth spurt parameters differ for normal skeletal jaw relationship and skeletal malocclusions, findings of studies which did not report class specific information are not discussed. This is because it is unclear whether results reported in these studies are for Class I, Class II or Class III skeletal jaw relationship.

The next section reports and discusses the findings from studies that clearly reported results for Class I and/or Class II skeletal jaw relationships Where appropriate, I will also discuss and relate my findings with studies that used radiographic methods to assess the timing of the adolescent jaw growth spurt. Jaw growth changes are described first for normal skeletal jaw relationship (Class I) and then class differences (Class I versus Class II, and Class I versus Class III) are discussed in terms of growth trajectories and adolescent growth spurt parameters for all growth measures for males and females. The pre-adolescent growth spurt for males and females is discussed, along with the clinical implications of the findings.

7.4.1. Jaw growth during adolescence

Normal jaw growth (Class I)

The results of the present study show that for both sexes, the upper jaw experiences the adolescent growth spurt earlier than the lower jaw and total face height, but the intensity of the growth spurt is weaker than for the lower jaw. A previous study reported a similar maturational trend for the upper and lower jaw and total face height (Buschang et al., 1983). The authors

concluded that differences in the maturational patterns for the upper and lower jaw and total face height could be explained by the cephalocaudal gradient of growth.

Chapter 2 reviews the concept of the cephalocaudal gradient of growth in detail. Briefly, according to this concept, the skeletal structures closer to the head (e.g., the upper jaw) mature earlier than those placed more distally (e.g., the lower jaw). Further, the intensity of the adolescent growth spurt is greater for distal structures (e.g., the lower jaw) than the proximal structures (e.g., the upper jaw). As total face height is largely determined by the growth in lower jaw length, the same maturational pattern is observed for total face height (Bishara, 1981; Buschang et al., 1983).

Due to differences in the adolescent growth spurt parameters, the increase in size for lower jaw length and total face height is greater than for the upper jaw length. For example, mean increase in upper jaw length (males 14.50 mm; females 09.62 mm) between age seven and 18 years (see Appendix C, Section C.2.1) is less than the mean increase in the lower jaw length (males 21.04 mm; females 16.10 mm; see Appendix C, Section C.2.1). A similar trend for differences in the mean increase of all growth measurements has been reported in earlier studies (Bishara, 1981; Bishara et al., 1981).

My results show that for all three growth measures, the adolescent growth spurt is less intense for females than males and occurs earlier for females than males. Similar findings have been reported in earlier studies for upper and lower jaw lengths (Bishara, 1981; Bishara et al., 1981; Jamison et al., 1982; Mitani, 1977; Nanda, 1971; Ochoa & Nanda, 2004; Thilander et al., 2005) and total face height (Baccetti et al., 2011; Buschang et al., 1983; Nanda, 1971; Thilander et al., 2005). A recent study (Montasser, 2019) using radiographic methods to assess timing of the adolescent growth spurt for upper and lower jaw lengths reported similar findings that the adolescent growth spurt occurs earlier for females than males.

As males show a more intense growth spurt, differences in the upper and lower jaw lengths and total face height increase with age. This study's findings show that compared to males, females have a shorter upper and lower jaw length and total face height, and these differences are greatest at 18 years of age. Previous studies have reported similar findings of sexual dimorphism in the upper and lower jaw lengths and total face height (Bishara et al., 1981; Jamison et al., 1982; McNamara, 1984; Nanda, 1955; Ochoa & Nanda, 2004).

A few studies have explored the relationship between the timing of the adolescent growth spurt for jaw growth and the PHV (peak height velocity) for stature. It has been reported that the upper jaw length attains peak growth velocity earlier than the PHV, whereas the adolescent growth spurt for the lower jaw length and total face height occurs after the PHV (Baughan et al., 1979b; Buschang et al., 1983).

As stature height measurements are not available from the data analysed in this study (see Chapter 3), it is not possible to directly compare results of this study with the previous studies. However, the findings indirectly support (see below) that the adolescent growth spurt for upper jaw length precedes the PHV while the timing of the peak growth velocity for lower jaw length and total face height follows the PHV.

The mean age for PHV is 13.5 years for males and 11.5 for females (Abbassi, 1998; Kelly et al., 2014). The results of the present study (see Table 7-1) show that the adolescent growth spurt for the upper jaw length occurred before 13.5 years for males and 11.5 years for females; the timing of the peak growth velocity for lower jaw length and total face height is after 13.5 years for males and 11.5 years for females.

Skeletal malocclusions (Class II and Class III)

My findings show that Class II growth pattern (smaller lower jaw length, larger upper jaw length, and shorter total face height) is established at an early age (before puberty) for both sexes.

However, variations in the timing and intensity of the adolescent growth spurt for all growth measures accentuate the skeletal discrepancies seen in Class II skeletal malocclusions. These findings support previous studies that have evaluated jaw growth changes for Class II skeletal malocclusions and reported that discrepancies in upper and lower jaw length and total face height are present at an early age, then increased by the adolescent growth spurt (Baccetti et al., 2011; Baughan et al., 1979b; Buschang et al., 1983; Buschang et al., 1988b).

Class differences (Class II versus Class I) in the timing and intensity of the adolescent growth spurt (see Section 7.3.4) showed that clinical features of Class II skeletal malocclusion for both sexes are attributable to an early but less intense adolescent growth spurt for lower jaw length and total face height, and an excessive intensity of the adolescent growth spurt for upper jaw length. A previous study (Buschang et al., 1988b) applying CP GCM to jaw growth data reported similar findings that for both males and females, the adolescent growth spurt for lower jaw length was less intense for Class II skeletal malocclusion than Class I skeletal jaw relationship (see Chapter 6, Section 6.4.6 for details). The study (Buschang et al., 1988b) did not analyse data for upper jaw length or total face height. Other studies using simple statistical methods (such as ordinary least square regression) to analyse the upper and lower jaw length data malocclusion support findings reported in my study (Baccetti et al., 2011; Baughan et al., 1979b; Buschang et al., 1983). Recent studies using radiographic methods to ascertain the timing of the adolescent jaw growth spurt concluded that for both males and females, the adolescent growth spurt for lower jaw length occurred earlier for Class II skeletal malocclusion than for Class I normal skeletal jaw relationship (Jeelani et al., 2016; Salazar-Lazo et al., 2014).

Sexual dimorphism was found for class differences (Class II versus Class I) in the timing and intensity of the adolescent growth spurt. Class differences emerge earlier for females but the differences in intensity of the adolescent growth spurt for all three growth measures are greater for

males than females. Due to a greater difference in the intensity of the adolescent growth spurt, the skeletal discrepancies in upper and lower jaw length and total face height for Class II skeletal malocclusion are more severe for in males than females. My findings support Buschang et al. (1988b) who analysed the lower jaw length data using CP GCM and reported similar sex differences in the timing and intensity of the adolescent growth spurt. Compared to females, class differences in the lower jaw length emerge earlier for females and differences in the peak growth velocity between Class II skeletal malocclusion and Class I skeletal jaw relationship are greater for males than females. Studies using simple statistical methods (such as ordinary least square regression) for data analysis also concluded that class difference in the upper and lower jaw length emerge earlier for females than males and differences in the peak growth velocity are greater for males than females (Baccetti et al., 2011; Baughan et al., 1979b; Buschang et al., 1983).

Studies using radiographic methods to determine the timing of the adolescent jaw growth spurt also found sexual dimorphism in the timing between Class II skeletal malocclusion and Class I normal skeletal jaw relationship (Jeelani et al., 2016; Salazar-Lazo et al., 2014). The class differences (Class II versus Class I) emerge earlier for females than males.

Class III skeletal malocclusion, unlike Class II skeletal malocclusion, is characterised by shorter upper jaw length, longer lower jaw length, and increased total face height (see results, Section 7.3). My findings show that Class III growth patterns are established at an early age (before puberty) for both males and females, but variations in the timing and intensity of the adolescent growth spurt for all three growth measures accentuate skeletal discrepancies seen in Class II and Class III skeletal malocclusions. No previous study has quantified differences in growth trajectories between Class I and Class III skeletal malocclusion.

The results for class differences (Class III versus Class I) in the timing and intensity of the adolescent growth spurt (see Section 7.3.4) show that clinical features of Class III skeletal

malocclusion for males and females mainly develop due to an early but less intense adolescent growth spurt for upper jaw length and an excessive intensity of the adolescent growth spurt for lower jaw length and the total face height. Earlier studies using radiographs to evaluate the timing of the adolescent growth spurt for upper and lower jaw length for Class III skeletal malocclusion reported similar findings of early adolescent growth spurt for upper jaw length (Baccetti et al., 2007; Baccetti et al., 2005; Kuc-Michalska & Baccetti, 2010).

Like Class II skeletal malocclusion, evidence shows sexual dimorphism for class differences (Class III versus Class I) in the timing and intensity of the adolescent growth spurt. The timing and intensity of the adolescent growth spurt for upper and lower jaw length and total face height are greater for males than females. Due to a larger difference in the timing and intensity of the adolescent growth spurt, the skeletal discrepancies in growth measurements for skeletal malocclusion Class III are greater for males than females. No previous study has quantified sex differences in the timing and intensity of the adolescent growth spurt for upper and lower jaw length and total face height between Class I and Class III. Out of three studies (Baccetti et al., 2007; Baccetti et al., 2005; Kuc-Michalska & Baccetti, 2010) that used radiographs to evaluate the timing of the adolescent growth spurt for upper and lower jaw length for Class III skeletal malocclusion, one study did not report sex differences in the timing between Class I and Class III (Kuc-Michalska & Baccetti, 2010), and two did not include data for Class I (Baccetti et al., 2007; Baccetti et al., 2005).

In summary, normal timing and intensity of the adolescent growth spurt for upper and lower jaw length and total face height is essential to achieve a harmonious anteroposterior skeletal jaw relationship (Class I) and balanced total face height. Unlike previous studies, which were limited to comparing jaw growth changes between Class I and Class II skeletal malocclusion, or between Class I and Class III skeletal malocclusion for one or two growth measures, this study compared and

quantified class differences in the timing and intensity of the adolescent growth spurt for the upper and lower jaw length and total face height for both males and females.

7.4.2. Pre-adolescent growth spurt

The findings (see results, Section 7.3.2) show that males experienced a pre-adolescent growth spurt at around 10 to 11 years of age for lower jaw length and total face height for Class I and Class III. However, growth velocity curves for lower jaw length and total face height showed no sign of a similar pre-adolescent growth for Class II skeletal jaw relationship. Furthermore, growth velocity curves for upper jaw length showed no sign of a similar pre-adolescent growth for either Class I, Class II or Class III skeletal jaw relationship.

Unlike males, growth velocity curves for females (see results, Section 7.3.2) showed no sign of a pre-adolescent growth spurt for any of the three growth measures for either Class I, Class II or Class III skeletal jaw relationship.

A previous study (Buschang et al., 1988b) applying CP GCM to lower jaw length data (age range six to 15 years) found evidence for a pre-adolescent growth spurt for both sexes. The pre-adolescent growth spurt for lower jaw length occurred one year earlier for females (at around eight years of age) than males (at around nine years of age). This earlier study (Buschang et al., 1988b) did not analyse data for upper jaw length or total face height.

Many other studies examining lower jaw growth for males and females with Class I skeletal jaw relationships also found clear evidence for a pre-adolescent growth spurt for both sexes (Harris, 1962; Nanda, 1955; Woodside, 1968). All these studies typically enrolled children at an early age: 4 to 12 years (Harris, 1962); 3 to 20 years (Woodside, 1968); and 4 to 20 years (Nanda, 1955). Compared to lower jaw length, longitudinal studies examining growth of upper jaw length growth are limited (Laowansiri et al., 2013).

The age range analysed in this thesis is 7 to 18 years. Though my results also show that males experience a pre-adolescent growth spurt for lower jaw length, there is no evidence that females also experience a similar growth spurt for lower jaw length. As discussed below, these differences with regard to the pre-adolescent growth spurt could perhaps be due to differences in the timing of the pre-adolescent growth spurt and a limited range of data that did not cover the pre-adolescent growth period for all classes of skeletal jaw relationship for both males and females.

The observed difference in the childhood dynamics for males and females between upper and lower jaw length and total face height for males and females could be explained by the difference in the timing of the adolescent growth spurt and its relationship with the pre-adolescent growth spurt. Results show that (Section 7.3) the adolescent growth spurt for upper jaw length occurs earlier than lower jaw length and total face height for Class I, Class II, Class III. Assuming a similar pattern for timing of the pre-adolescent growth spurt, it is possible that a pre-adolescent growth spurt for upper jaw length occurred earlier than for lower jaw length and total face height and could not be captured because of the limited age range (seven to 18 years). The adolescent growth spurt for lower jaw length and total face height occurred earlier for Class II than Class I and Class III (Section7.3). Considering the same pattern for the pre-adolescent growth spurt timing, it is possible that the pre-adolescent growth spurt for lower jaw length and total face height occurred earlier for Class II than Class I and Class III, and therefore could not be identified because of the limited age range.

Regardless of the jaw relationship (Class I, Class II or Class III), the adolescent growth spurt for all growth measurements occurred earlier for females than males (Section7.3). It is possible that a similar sexual dimorphism exists for the timing of the pre-adolescent growth spurt. Thus, the limited age range included in this study was insufficient to find evidence for the pre-adolescent growth spurt for females.

7.4.3. Clinical implications

Skeletal malocclusion affects both sexes across ethnicities (Joshi et al., 2014; Proffit et al., 2014). The anteroposterior and vertical skeletal discrepancies adversely affect oral health quality, facial aesthetics and psychosocial wellbeing (Baskaradoss et al., 2013; Bollhalder et al., 2013; Dimberg et al., 2015; Ghafournia & Hajenourozali Tehrani, 2012; Joshi et al., 2014; Koroluk, 2016; Martins-Junior et al., 2012; Masood et al., 2013; Proffit et al., 2014). Therefore, treatment of skeletal discrepancies is a priority for clinical practices worldwide (Bernabe et al., 2008; Joshi et al., 2014).

The treatment of choice to correct skeletal malocclusions is modification of the jaw growth. The desired growth modification to correct Class II skeletal malocclusion is to stimulate the lower jaw growth but to restrain the upper jaw growth (De Clerck & Proffit, 2015; Nowak et al., 2019). As Class III skeletal malocclusion is the reverse of Class II skeletal malocclusion, correction of Class III malocclusion involves stimulation of the upper jaw growth and inhibition of the lower jaw growth (De Clerck & Proffit, 2015; Nowak et al., 2019). Growth modification procedures are most effective when the treatment timing is carefully coordinated with the timing of the adolescent growth spurt (Buschang et al., 2017; De Clerck & Proffit, 2015; Nowak et al., 2019; Proffit, 2006; Proffit et al., 2014; Turpin, 2000). To achieve best possible outcome, it is recommended that jaw growth modification should be attempted around the age when jaw growth velocity is at peak i.e., around the APGV (see Chapter 1, Section 1.1 for details).

The results show that the timing and intensity of the adolescent growth spurt for upper and lower jaw length and total face height differ for Class II and Class III skeletal malocclusions. Further, strong evidence was found for sexual dimorphism in the development of skeletal malocclusions and magnitude of the skeletal discrepancies. Therefore, the timing of growth modification procedures should be carefully chosen to maximise the success of treatment for males and females. As

differences in the intensity of the adolescent growth spurt are greater for males than females, the orthopaedic appliances used for growth modification should be appropriately calibrated.

For Class II skeletal malocclusion, the results suggest that growth modification procedures aimed to inhibit the forward growth of the upper jaw should, on average, be started at around 12.6 years for males and 11 years for females. Growth modification targeted at stimulating the forward growth of the lower jaw should be introduced at around 13.4 years for males and 11.9 years for females. To correct Class III skeletal malocclusion, stimulation of the upper jaw growth to increase its length should be initiated at earlier age (around 11.9 years for males and 10.7 years for females) than inhibition of the lower jaw growth (around 13.9 years for males and 12.5 years for females). However, as recommended (Björk, 1963; Bjork & Helm, 1967; Buschang et al., 2017; Proffit et al., 2014; Turpin, 2000), individual-specific variations in the timing and intensity of the adolescent growth spurt should be considered when planning treatment to correct skeletal discrepancies. This is particularly true for males as they show a greater between-individual variability in the timing and intensity of the adolescent growth spurt. As an example, the median APGV for the lower jaw length for males with Class II skeletal malocclusion is 12.69 years with IQR range as 12.14 to 13.66 years (see Table 7-3). This suggest that for male patients with Class II skeletal malocclusion requiring growth modification procedure to stimulate the lower jaw growth, 25% will attain APGV at an age less than 12.14 years, 25% between 12.14 and 12.69 years, 25 % between 12.69 and 13.66 years, and remaining 25% after the age of 13.66 years. Therefore, a clinical should be aware that the right age to start growth medication for male children may not necessarily at 12.69 years but could be at an age interval, as explained above. A typical feature of Class III skeletal malocclusion is increased total face height. As excessive vertical growth in the lower jaw is primarily responsible for increased face height, and the adolescent growth spurt for total face height occurs almost at the same time as for lower jaw length, it is strongly recommended that appliances used to inhibit the forward growth

of the lower jaw should be designed to restrict its vertical growth. Also, it is desirable to control the vertical growth of the upper jaw while attempting it to bring it forward.

In contrast to Class III skeletal malocclusion, the total face height is short in Class II skeletal malocclusion. Although results show that a short face height is mainly because of deficient growth in the lower jaw, caution should be exercised when attempting to increase face height by stimulating its vertical growth. This because a backward rotation of the lower jaw can worsen the Class II skeletal malocclusion (Proffit et al., 2014).

While some authors suggest an early intervention at around 10 years to correct Class III skeletal malocclusion (Baccetti & Tollaro, 1998; Battagel & Orton, 1995; Campbell, 1983; Kim et al., 1999), others have found no evidence to support early (pre-adolescent) treatment for Class III correction, and recommend a single phase of treatment during adolescence (Atalay & Tortop, 2010; Kapust et al., 1998; Zere et al., 2018). Though this current study did not aim to explore the potential impact of the pre-adolescent growth spurt on the development of skeletal malocclusion and its role in treatment planning, the results indicate that perhaps it is better to start Class III correction at an early age for males (at around 10 years). This because males experience a pre-adolescent growth spurt at around 10 years for lower jaw length and total face height (sees Figure 7-2 and Figure 7-3).

7.5. Limitations

Limitations pertaining to the data and the statistical method applied have been already discussed in Chapter 6. This section focuses on some key limitations that could have potentially affected the clinical findings.

The data analysed in this study comprised repeated jaw measurements on males and females who participated in eight different growth studies. Because two growth studies did not provide data for Class II and five growth studies did not include individuals with Class III (see Chapter 3, Section 3.3), the statistical model (RCS GCM) fitted to the data did not include interaction between study and class variables. Fitting a model without incorporating study–class interactions effects assumes that the class effects are same across all eight studies. As there is evidence for a secular trend in jaw growth (Antoun et al., 2015), it is possible that failure to include interactions between study and class effect could have potentially influenced the class estimates.

In addition to the timing (APGV) and intensity (PGV) of the adolescent growth spurt, its duration also makes important contributions to final adulthood jaw size and face height (García-Drago et al., 2014; Kuc-Michalska & Baccetti, 2010; Proffit et al., 2014). Duration of the adolescent growth spurt lasts from the age at take-off until the end of the adolescent growth spurt, marked by the age corresponding to maximal deceleration in post-adolescent growth (see Chapter 4, Section 4.3.1 for further details). To achieve smooth estimates of second derivative, a minimum of a quartic truncated spline basis is recommended (Simpkin et al., 2018). Since the restricted cubic spline function (i.e., the RCS GCM) does not provide a smooth second derivative (acceleration curve), the duration of the adolescent growth spurt was not estimated. An alternative is to use a quartic spline function which allows for estimation of a smooth second derivative. This is discussed further in Chapter 8 (Section 8.3.2).

Lastly, Class II skeletal malocclusion is further subclassified into Class II division 1 and Class II division 2 (Proffit et al., 2014). Although most of the differences between Class II division 1 and Class II division 2 are dental (malposition of teeth), these two subtypes are characterised by different vertical growth patterns. While Class II division 2 is typically associated with short face height, the face height is increased for Class II division 1 due to a backward rotation of the lower jaw (Proffit et al., 2014). However, data analysed in this study did not provide any information on the subclassification of Class II skeletal malocclusion. Therefore, it is impossible to comment on whether the total face height results presented in this chapter related to Class II division 1 or Class II division 2 skeletal malocclusion.

7.6. Conclusion

Clear evidence was found for an adolescent growth spurt for upper and lower jaw length and total face height, with females attaining peak adolescent growth velocity approximately 1.5 years earlier than males. The intensity of the adolescent growth spurt is greater for males than females. For both males and females, the adolescent growth spurt for upper jaw length occurs earlier than for lower jaw length and total face height.

The growth pattern for Class I, Class II and Class III is established at an early age (around seven years), but class differences in upper and lower jaw length and total face height emerge rapidly during adolescence. Strong evidence was found for sex differences in the development of skeletal malocclusions (Class II and Class III). Class differences emerge earlier for females than males, but the magnitude of skeletal discrepancies is greater for males. Clinical implications of research findings are discussed.

Further research involving a larger sample size is recommended to strengthen the evidence. It is also recommended that future studies focusing on the pre-adolescent growth spurt should include a wider age range covering the childhood growth period by starting at a younger age (five years for females and six years for males).

CHAPTER 8. DISCUSSION

The field of human growth research is broad. Different research questions have been posed to better understand physical growth changes in normally growing children and to learn about growth patterns and factors affecting them (Johnson, 2015). In dentistry, particularly the fields of orthodontics and dentofacial orthopaedics, studying dynamics of jaw growth during adolescence is a central issue. Some of the most pressing issues relate to the timing and intensity of the adolescent growth spurt for normal skeletal jaw relationship (Class I) and skeletal malocclusions (Class II and Class III) for males and females. This thesis focused on addressing these issues.

Chapter 1 (Section 1.1) provides details on the clinical characteristics of Class I, Class II and Class III skeletal jaw relationships. Briefly, Class I represents a harmonious relationship between upper and lower jaw lengths and a well-balanced total face height. In contrast, both Class II and Class III are associated with skeletal jaw discrepancies in upper and lower jaw length and total face height. Class II skeletal malocclusion is typically associated with a large upper jaw length, short lower jaw length, and short total face height. In contrast, Class III skeletal malocclusion is characterised by a short upper jaw length, large lower jaw length, and increased total face height.

As outlined in Chapter 1 (Section 1.3), research work done in this thesis included conducting a simulation study to select an *a priori* criterion to evaluate the fit of growth curve models (Chapter 5), comparing linear mixed effects (LME) and nonlinear mixed effects (NLME) growth curve models (GCMs) applied to the jaw growth data (Chapter 6), and applying the best-fitting GCM to the jaw growth data for answering clinical research questions (Chapter 7). The LME and NLME GCMs are referred to as linear and nonlinear GCMs, respectively.

8.1. Research findings

The specific research findings have already been discussed in detail in previous chapters (see Table 8-1 for a summary). This section discusses the broader methodological (Chapter 5 and Chapter 6) and clinical (Chapter 7) contributions made by this thesis.

Table 8-1. Summary of research work and key findings.

Chapter	Objective	Research work	Key findings	Conclusion
5	To identify a criterion for GCM selection	A simulation study conducted to compare performances of information (AIC and BIC) and prediction (R2 and CCC) criteria	- Information criteria perform consistently better than prediction criteria. - AIC performs better than BIC when sample size is small and GCM involves complex variance-covariance structure.	AIC selected as the criterion of choice to compare fit of GCMs
6	To compare linear and nonlinear GCMs and select a model that fits best to the jaw growth data	Fitted covariate-adjusted linear (CP, FP and RCS) and nonlinear (SITAR and PB) GCMs to the jaw growth data	- The PB model failed to converge for both sexes. - RCS model fits best to the male and female data. - Unlike RCS and SITAR models, CP and FP models estimate biologically implausible growth trajectories. - Compared to the SITAR model, the RCS model estimates lower RSD.	The RCS identified as the best fitting GCM to the jaw growth data
7	To answer clinical research questions by applying the best fitting GCM	Applied RCS model to estimate class (Class I, Class II and Class III) differences in jaw growth for males and females.	- Class differences in growth trajectories and adolescent growth spurt parameters for the upper jaw length, the lower jaw length and the total anterior face height for males and females.	Jaw growth differs between males and females, and varies between Class I, Class II and Class III

GCM: growth curve model; CP: conventional polynomial; FP: fractional polynomial; RCS: restricted cubic spline; SITAR: superimposition by translation and rotation; PB: Preece–Baines. AIC: Akaike information criterion; BIC: Bayesian information criterion; RSD: residual standard deviation. Class I: normal skeletal jaw relationship, Class II and Class III: skeletal malocclusions.

8.1.1. Methodological

The research work involved comparing fit of the linear and nonlinear GCMs estimated using the restricted maximum likelihood (REML) method. Therefore, the first objective was to select an *a priori* model selection criterion. In Chapter 5, a simulation study compared popular information criteria and prediction criteria. The information criteria included in the simulation study were the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The prediction criteria were the conditional concordance correlation (CCC) by Vonesh et al. (1996), the R^2 statistic proposed by Vonesh et al. (1996), the R^2 statistic introduced by Nakagawa and Schielzeth (2013) and its modification by Johnson (2014).

Research findings are discussed in detail in Chapter 5 (Section 5.4). Briefly, simulation results showed that under different conditions of sample size (level 2 and level 1), type of data (balanced and unbalanced) and varying model complexity (fixed and random effects), the information criteria (AIC and BIC) performed better than prediction criteria (CCC and R^2 statistics) regardless of the sample size, the model complexity and whether data were balanced or unbalanced. The AIC performed better than the BIC when level 1 sample size was small, and the model involves a complex variance-covariance structure. As these two conditions resemble more closely the jaw growth data analysed in the next study (Chapter 6), the AIC was selected as the criterion to evaluate and compare the fit of linear and nonlinear GCMs applied to the data.

Chapter 6 evaluated and compared the fit of linear (CP, FP and RCS) and nonlinear (SITAR and PB) GCMs applied to upper jaw length, lower jaw length and total face height data for males and females. Each GCM was fitted by including class of skeletal jaw relationship as a covariate (Class I, Class II and Class III). In addition, study variable, which denote eight growth studies (see Chapter 3), was also included in the model fitting to adjust growth trajectories and adolescent growth spurt parameters for potential study effects. All five GCMs were fitted by including all possible

individual-specific random effects. The REML method was used for model estimation. Chapter 6 (Section 6.2) provides full methodological details. Findings have been already discussed in Chapter 6 (Section 6.4); key findings are revisited here.

The PB GCM failed to converge for any of the three growth measures for both males and females. This was perhaps because of the small sample size, particularly the level 1 sample size. (see Chapter 6, Section 6.4 for details). The CP, FP, RCS and SITAR GCMs applied to male and female data successfully converged for all three growth measures. For all four GCMs (CP, FP, RCS and SITAR), assessment of model assumptions showed no violation of key assumptions of normality of random effects and level 1 residuals, or the homoscedasticity of level 1 residuals. Independence of level 1 was assumed and autocorrelation of residuals was not tested.

The results suggests that, like human height measurements, which approximately follow a normal distribution (Jelenkovic et al., 2016), jaw measurements also follow an approximate normal distribution. A recent study analysing height data reported that while it is more realistic to model complex level 1 variance (heteroskedasticity), it creates only small changes to the growth parameters because level 1 variation is relatively small (Goldstein et al., 2018). The results of the present study showed no detectable violation of homoscedasticity assumptions for the jaw growth data. Therefore, the findings suggest that assumptions of normality and homoscedasticity are valid for fitting linear and nonlinear GCMs to jaw growth data for males and females.

Literature on growth curve modelling of jaw growth data offered no guidelines on testing the autocorrelation of residuals (serial correlation of level 1 residuals). This is because none of the previous studies applying GCMs to the jaw growth data tested for autocorrelation of residuals (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996). However, literature on modelling height data suggests that serial correlation are found when height measurements are less

than approximately two months apart (Browne & Goldstein, 2010). For the jaw data analysed, the minimum time gap between two consecutive growth measurements for all three growth measures was six months. Therefore, I did not pursue testing this assumption.

Assessment of model fit to data (using AIC) showed that the RCS GCM fits best to all three growth measures, i.e., the upper jaw length, lower jaw length and total anterior face height, for both males and females. In addition, the RCS GCM estimates the lowest residual standard deviation (RSD) for all growth measures for both sexes. Since no previous study compared fit of CP, FP, RCS and SITAR GCMs to the longitudinal growth data (height or jaw growth data), the findings could not be related to the literature. A recent study evaluating the performance of FP, SITAR and PB GCMs in modelling height data did not report on the model fit (Simpkin et al., 2017).

In addition to evaluating the fit of each GCM, they were compared in terms of their ability to model covariate-adjusted growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing and intensity). Comparison of growth trajectories shows that CP and FP GCMs estimate biologically implausible negative growth velocity toward the end of growth period (adulthood) for the upper jaw length, lower jaw length and total face height for males and females. Previous studies analysing the jaw growth data have also shown that conventional polynomials estimate biologically implausible growth trajectories with a negative velocity (implausibly implying that size is decreasing) (Nahhas et al., 2014). Unlike CP and FP GCMs, both RCS and SITAR GCMs estimate biologically plausible growth trajectories as they approximate an asymptote.

As discussed in Chapter 6 (Section 6.4), a model that approximates an asymptote is preferred when growth theory suggests an asymptote (Grimm et al., 2011). The jaw growth follows a sigmoid pattern of growth characterised by an upper asymptote (Proffit et al., 2014). Therefore, both RCS and SITAR GCMs which approximate an asymptote are preferred over CP and FP GCMs. Though

both RCS and the SITAR GCMs estimate biologically plausible growth trajectories, the findings show that RCS and the SITAR GCMs performed differently under different conditions of measurement error (within-individual variability in growth).

As the within-individual variability (RSD) increased, differences in adolescent growth spurt parameters (timing and intensity) estimated by RCS and the SITAR GCMs also increased. The difference in the timing and intensity of the adolescent growth spurt estimated by these two GCMs was highest for lower jaw length and total anterior face height measurement for males, and lowest for upper jaw length for females (see Chapter 6, Section 6.3.3). The difference in RSD estimated by RCS and SITAR GCMs was highest for lower jaw length and total anterior face height measurement for males, and lowest for upper jaw length for females.

Since true parameters are unknown for real-life jaw growth data, it is impossible to comment on which models underestimate or overestimate adolescent growth spurt parameters. However, a recent simulation study (where true parameters are known) has shown that as the measurement error i.e., within-individual growth variability, RSD increases, the performance of the SITAR GCM worsens. Under these conditions, the SITAR model tends to underestimate the timing of the adolescent growth spurt (Simpkin et al., 2017). This indirect evidence suggests that perhaps the SITAR GCM provided lower estimates for the timing of the adolescent growth spurt for the jaw data analysed in the present study.

As RSD influences the performance of the SITAR GCM (Simpkin et al., 2017), caution needs to be exercised when generalising these findings to modelling height data. This is because unlike human height, which is a "*classic anthropometric quantitative trait for its ease of measurement*" (Jelenkovic et al., 2016), assessment of jaw growth using radiographic cephalometric techniques is prone to measurement errors (Baughan et al., 1979b; Houston, 1983). Even though data available from the American Association of Orthodontists Foundation (AAOF) Craniofacial

Growth Legacy Collection have been meticulously checked for magnification errors and inter-observer variability, it is still possible that some measurement might be inaccurate. Errors and inaccuracy in measurements are common in medical radiology (Brady, 2017).

Based on the findings, it was concluded that among the three linear GCMs (CP, FP and RCS) included in this thesis, only the RCS performed well in modelling jaw growth data. Of the two nonlinear GCMs (SITAR and PB), the PB GCM failed to converge. The SITAR GCM performed well in modelling jaw growth trajectories. However, results suggest that compared to the restricted cubic spline based linear RCS GCM, the nonlinear SITAR GCM is less robust to modelling a 'noisy' data characterised by a larger RSD. As discussed earlier in Chapter 6 (Section 6.4.3), this is perhaps because as the variance covariance structure of RCS GCM get richer with an addition of spline terms in the fixed- and the random effects structures, it captures between-individual variability more efficiently. Unlike RCS, the random effects structure of the SITAR GCM remains same (three random effects) irrespective of the fixed effects structure. As a result, the between-individual variability is not modelled as efficiently as the RCS GCM. This unexplained variance pools as residual variance (RSD). However, as discussed above, the results for the comparative performance of the RCS and SITAR GCMs in estimating the adolescent growth spurt parameters are inconclusive. This issue is discussed further in Section 8.3.1.

Researchers should be aware of the trade-off between the statistically best fitting model and the biological plausibility of the model estimated growth parameters. It is important that when analysing biological data such as height and jaw growth, one should not focus only on the statistical fit of the model but also consider the biological plausibility of model estimated growth parameters.

Compared to linear GCMs (CP, FP and RCS), nonlinear GCMs (SITAR and PB) that are based on an underlying biological theory of growth have a more interpretable model formulation and growth parameters (Cole et al., 2010; Grimm et al., 2011; Pinheiro & Bates, 2000). For instance,

a negative correlation between the timing and intensity of adolescent growth spurt estimated by the SITAR model maps well on to the theory of growth indicating that early maturing individuals have a more intense growth spurt and vice-versa (Cole et al., 2010).

Though I evaluated and discussed (see Chapter 6, Section 6.4 for full details) the biological plausibility of growth trajectories estimated by linear and nonlinear GCMs, I did not consider it as the sole criterion when selecting model. Two reasons underly this decision. First, I decided, a priori, to use an objective criterion (such as AIC) to evaluate and compare the fit of linear and nonlinear GCMs applied to the jaw growth data. Second, the objective of my research was to evaluate and compare the performance of linear and nonlinear GCMs in terms of their ability in estimating APGV and PGV statistics. As discussed in Chapter 4 (Section 4.4), the approach used to estimate APGV and PGV statistics for linear and nonlinear GCMs (quadratic function method) was identical.

In summary, considering the better fit to upper and lower jaw length and total anterior face height measurement for males and females, the RCS GCM was identified as the model of choice for answering the clinical research questions in the next study (Chapter 7).

8.1.2. Clinical

Chapter 7 applied the best-fitting RCS GCM (identified in Chapter 6) to male and female data for estimating class (Class I, Class II and Class III) differences in growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing and intensity) for upper and lower jaw length and total anterior face height. The growth trajectories and adolescent growth spurt parameters were adjusted for potential growth study effects. Findings have been already discussed in Chapter 7 (Section 7.4); key findings are summarised below.

The results describe sex-specific growth changes (from seven to 18 years of age) in upper and lower jaw length and total anterior face height for normal skeletal jaw relationship (Class I) and skeletal malocclusions (Class II and Class III). Results show sexual dimorphism in the growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing and intensity) for normal skeletal jaw relationship (Class I) and skeletal malocclusions (Class II and Class III). For all three growth measures, the adolescent growth spurt was less intense for females than males and occurred earlier in females than males for Class I, Class II and Class III jaw relationships.

Chapter 6 (Section 6.4.6) discussed and related the findings with previous studies applying GCMs to jaw growth. Here I briefly discuss a few key issues noticed while evaluating the methodological details reported in earlier studies using GCMs for data analysis. A total of eight studies analysed jaw growth data using GCMs (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996). All eight studies used a CP GCM for modelling jaw growth trajectories and estimating adolescent growth spurt parameters.

Like the present study (RCS GCM), all previous studies analysed (CP GCM) unbalanced longitudinal jaw growth data. As one of the key strengths of GCMs is their ability to handle

unbalanced data, this does not pose analytical challenges (Goldstein, 2011b; Laird & Ware, 1982; Liu, 2015). However, it is good practice to report summary statistics (such as median and interquartile range) for the number of repeated measurements per individual (level 1 sample size) in level 2 sample size. Only two studies (Buschang et al., 1989; Nahhas et al., 2014) reported this relevant information. The remaining six studies did not report any information for the number of repeated measurements per individual (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a; Buschang et al., 1988b; Chvatal et al., 2005; van der Beek et al., 1996).

For modelling jaw growth data, empirical evidence supports sex and class (Class I, Class II and Class III) differences in growth trajectories and adolescent growth spurt parameters (Buschang et al., 2017; Proffit et al., 2014). Therefore, two important time-invariant covariates to be included in GCMs are the sex of the individual and the class of skeletal jaw relationship. Since all studies analysed data separately for males and females (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996), the one potential covariate that needed to be included in the modelling framework was class of skeletal jaw relationship.

Out of a total of eight studies that applied a CP GCM to jaw growth data (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996), only one study (Buschang et al., 1988b) clearly mentioned that class was included as a covariate into the GCM. The authors (Buschang et al., 1988b) analysed and compared lower jaw length growth between Class I normal skeletal occlusion and Class II skeletal malocclusion. The remaining seven studies provided no such information (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996).

The study comparing lower jaw length growth between Class I and Class II reported important class differences in growth trajectories and the adolescent growth spurt parameters for both males and females (Buschang et al., 1988b). The results of the present study (see Chapter 7, Section 7.3) show that growth trajectories and the adolescent growth spurt parameters for upper and lower jaw length and total face height differed between Class I, Class II and Class IIII skeletal jaw relationships for both sexes. This highlights the importance of including class as a covariate in GCMs. As estimating class differences has important clinical implications (see Chapter 7, Section 7.4.3), it is far more informative to study their respective growth trajectories and adolescent growth spurt parameters than a pooled average.

A core strength of GCMs is their ability to model between-individual variability in growth (Curran & Bauer, 2011; Curran et al., 2010; Goldstein, 2011b; Johnson, 2015; McNeish & Matta, 2018). Once the between-individual variability is appropriately modelled by including random effects (Verbeke & Molenberghs, 2000), researchers may also consider modifying the residual variance structure to account for heterogeneity and autocorrelation of residuals (Goldstein et al., 1994; Goldstein et al., 2018). It is recommended that residual (within-individual) covariance structure should be relaxed "when there is no evidence for the presence of additional random effects" (Verbeke & Molenberghs, 2000, p. 26).

The importance of modelling between-individual variability in growth has emerged forcefully in modelling longitudinal data (Collins, 2006; Grimm et al., 2011). This because individuals differ with respect to each other in biological parameters of skeletal growth (e.g., height), such as growth trajectories and the rate of growth (Grimm et al., 2011). During adolescence, which is characterised by large between-individual variability in growth trajectories as well as adolescent growth spurt parameters, constraining the between-individual variation in growth trajectories by removing random effects (i.e., setting them to zero) is a biologically untenable assumption. When a

model fitted to the data is logically inconsistent with the empirical support, then model should be viewed as untenable (Johnson & Omland, 2004; Motulsky & Christopoulos, 2004). Chapter 6 (Section 6.4.1) discusses this issue in detail.

Fitting a higher degree CP GCM to capture mean growth trajectory by removing random effects "*does not allow for or capture between-person differences in the patterns of change"* (Grimm et al., 2011). For example, fitting a CP GCM with only random intercept assumes that rate of growth for each individual is exactly the same. A recent study (Buschang et al., 2013) fitted up to a fifth-degree CP GCM to jaw growth data by including only an intercept into the random effects part of the model. Another recent study did not report on the random effect structure of the CP GCM fitted to the jaw growth data but mentioned that the "best-fitting model may not necessarily include individual-level variation for every polynomial coefficient, nor correlation between every random effect" (Nahhas et al., 2014). This suggests that the authors (Nahhas et al., 2014) followed the same approach of fitting the CP GCM with a subset of random effects as was reported in the other six studies (Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; van der Beek et al., 1996).

All previous studies applying CP GCM to jaw growth data thus fitted the model with a subset of random effects to include higher order polynomials into the fixed effects part of the model. Although one school of thought, the parsimonious approach (Bates et al., 2015; Matuschek et al., 2017), recommends fitting GCMs with a subset of random effects when a model with a complex variance-covariance structure fails to converge due to a small sample size effect (see Chapter 6 Section 6.4.1), the implementation of the parsimonious approach requires a systematic step-by-step procedure to decide which variance-covariance parameters need to be excluded (Bates et al., 2015; Matuschek et al., 2017). For example, Matuschek et al. (2017) recommend using the AIC or the log-likelihood ratio test to guide which variance-covariance needs to be excluded. However, authors

fitting CP GCM with a subset of random effects to the jaw growth data have not reported any details on how they implemented their approach (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996).

Unlike previous studies applying GCMs (CP) to jaw growth data that fitted model with a subset of random effects, the present study fitted linear (CP, FP and TCS) and nonlinear (SITAR and PB) GCMs by including all individual-specific random effects (see Chapter 6 for details). As discussed by Grimm et al. (2011), excluding random effects for higher-order terms of linear GCM fails to model between-person differences in growth trajectories. Since nonlinear GCMs (such as SITAR and PB) are structured on a strong underlying biological theory of growth, the issue of excluding random effects would seem more concerning than the linear GCMs. For instance, excluding timing and intensity random effects from the SITAR GCM implies that all individuals experience adolescent growth spurt of same intensity and at the exact same age for everyone.

It is true that often a small sample size complicates the successful estimation of GCMs with a complex random effect structure. However, instead of fitting a GCM with untenable model assumptions, it is better to collect more data that support successfully estimation of a GCM with complex random effect structure. The Bayesian framework is an alternative approach for fitting a GCM with complex random effect structure when sample size is small. In this thesis, an *a priori* decision was made not to fit a linear or nonlinear GCM with a subset of random effects and restricted to frequentist approach for model estimation. Therefore, fitting the PB GCM that failed to converge was not pursued (see Section 8.1.1). This point is further discussed in Section 8.3.1.

Lastly, it is important to test and report on key model assumptions such as the normality of random effects and level 1 residuals, and the independence and homoscedasticity of level 1 residuals (Goldstein, 2011b; Goldstein et al., 2018; Snijders & Bosker, 2012). However, as discussed earlier

(Section 8.1.1), none of the previous studies applying a CP GCM to jaw growth data reported assessment of model assumptions (Buschang et al., 2013; Buschang et al., 1999; Buschang et al., 1988a, 1989; Buschang et al., 1988b; Chvatal et al., 2005; Nahhas et al., 2014; van der Beek et al., 1996).

8.2. Contributions to the literature

This thesis makes some novel contributions to the methodological literature on modelling physical growth data (Chapter 5 and Chapter 6) and the clinical literature on jaw growth during adolescence (Chapter 7).

As discussed below, new Stata programs have been written to implement methodology presented in each research chapter (Chapter 5, Chapter 6, and Chapter 7). Availability of the analytical tools to implement the methodological advances made in this thesis would allow researchers to apply these methods to their own work. Readily available programs will be particularly helpful to clinically oriented researchers less familiar with the underpinnings of statistical methods.

8.2.1. Methodological

The findings fill the existing gaps in the literature on selection of GCMs using the REML method. The simulation study presented in Chapter 5 demonstrates that regardless of the sample size, the type of data or the model complexity, the information criteria (AIC and BIC) performed consistently better than the prediction criteria (CCC and R^2 statistics). These findings will help researchers in decision-making when selecting an appropriate criterion for GCM selection. Clearly, the prediction criteria are not recommended for GCM selection.

I have shown that AIC and BIC calculated using the total number of model parameters perform at least as well and often better than their counterparts calculated using only the number of variance-covariance parameters. For BIC, the choice between the two possible sample sizes (level 2 sample size, or the total number of observations) has little influence on its performance. A comparison of AIC and BIC under different conditions of sample size, type of data and model complexity provided new insights into their strengths and weaknesses in GCM selection. These

findings will prove useful to researchers when selecting an appropriate information criterion (AIC or BIC) for GCM section, considering the sample size and the model complexity.

Chapter 6 makes several original contributions to the methodological literature on analysing human physical growth data. This study is in line with a recent emerging trend to evaluate and compare performance of different GCMs rather than focusing on a single model for examining human growth processes (Crozier et al., 2019; Simpkin et al., 2017).

This study is the first in the literature to compare five different GCMs (CP, FP, RCS, SITAR and PB) for modelling growth trajectories and estimating adolescent growth spurt parameters. Furthermore, it has developed the methodology to include covariates when estimating population-average and individual-specific growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing and the intensity) for linear (CP, FP, RCS) and nonlinear (SITAR and PB) GCMs (see Chapter 6, Section 6.2).

Comparison of different GCMs, some of which have been introduced recently (SITAR), provided new understandings of their relative strengths and weaknesses. Unfortunately, the PB GCM failed to converge (see Chapter 6, Section 6.2.9. The results for other four GCMs showed that polynomial based linear GCMs (CP and FP) performed poorly in modelling jaw growth trajectories when compared with restricted cubic spline-based linear (RCS) and nonlinear (SITAR) GCMs. The RCS GCM proved particularly useful for modelling noisy jaw growth data. More specifically, this study demonstrated that statistical methods (CP GCM) used in earlier studies are inadequate in describing the complexities of jaw growth during adolescence.

To calculate different information criteria and the prediction criteria included in the simulation study (Chapter 5), a new Stata program (‘gcmfit’) was written. While Stata (StataCorp, 2015a) users can specify a different sample size (such as the level 2 sample size or the total number of observations) to calculate BIC, there is no option to change the number of parameters (total

number of parameters or the number of variance-covariance parameters) included in the penalty terms of AIC and BIC. Though other software such as the Statistical Analysis System (SAS) software (SAS Institute Inc., 2015) provides greater flexibility in calculating different versions of AIC and BIC by changing the number of parameters included in their penalty terms, Stata software was selected for this study.

Like AIC and BIC, no Stata package was available to calculate prediction criteria that include the concordance correlation coefficient (CCC) by Vonesh et al. (1996) and three R^2 statistics proposed by Vonesh et al. (1996), Nakagawa and Schielzeth (2013), and Johnson (2014). Vonesh (2012, see Appendix D of the book) provided code for the goodness-of-fit (GOF) SAS macro to calculate CCC by Vonesh et al. (1996) and the R^2 statistic by Vonesh et al. (1996). Nakagawa and Schielzeth (2013) and Johnson (2014) provided R scripts to calculate their R^2 statistics using the R software (R Core Team, 2015).

Chapter 6 estimates growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing and the intensity) for covariate-adjusted linear (CP, FP and RCS) and nonlinear (SITAR and PB) GCMs. However, no program in either Stata (StataCorp, 2015a) , SAS (SAS Institute Inc., 2015) or R (R Core Team, 2015) software was available to estimate covariate-specific or covariate-adjusted derivatives (velocity and acceleration). Similarly, no program was available to compute covariate-specific or covariate-adjusted adolescent growth spurt parameters (timing and intensity). To fill these gaps, a new set of Stata programs was written. These programs allow inclusion of covariates when estimating distance, first derivative (velocity), second derivative (acceleration), and the timing and intensity of the adolescent growth spurt. Appendix B (Section B.2) provides full details on the working of each program. The programs provide an option to estimate covariate-specific or covariate-adjusted growth trajectories and adolescent growth spurt parameters.

8.2.2. Clinical

The third study (Chapter 7) makes important advances in understanding dynamics of jaw growth during adolescence. This is the first study to investigate and compare class differences in the growth trajectories (distance, velocity and acceleration) and the adolescent growth spurt parameters (timing and the intensity) for upper and lower jaw length and total anterior face height for both males and females. The findings enhance clinical understanding of the development of anteroposterior (upper and lower jaw length) and vertical (total anterior face height) skeletal jaw discrepancies for Class II and Class III skeletal malocclusions (see Chapter 7, Section 7.4).

Unlike previous studies applying CP GCMs to jaw growth data and reporting conflicting findings for the adolescent jaw growth spurt (see Chapter 6, Section 6.2.9), this study's results clearly show that both males and females experienced an adolescent growth spurt for upper and lower jaw length and total anterior face height. The study provides sex-specific estimates for the timing of the adolescent growth spurt for all growth measurements for Class II and Class III skeletal malocclusions. This will help clinicians in planning treatment to correct anteroposterior (upper and lower jaw length) and vertical (total anterior face height) skeletal jaw discrepancies for males and females (see Chapter 7, Section 7.4.3 for details).

8.3. Limitations and future work

Though this thesis makes some novel contributions to the methodological and clinical literature on modelling jaw growth data (see Section 8.2), there are inevitably various limitations to the research. Due to time constraints or the nature of the jaw growth data analysed in the thesis, some methodological and clinical research questions could not be answered.

Limitations pertaining to simulation study for GCM selection (Chapter 5, Section 5.5), comparison of linear and nonlinear GCMs (Chapter 6, Section 6.5), evaluation of jaw growth during adolescence (Chapter 7, Section 7.5) have been discussed earlier in detail. This section focuses on highlighting key points that would help in guiding future methodological and clinical research work.

8.3.1. Methodological

Table 8-2 summarise the key methodological limitations (Chapter 5 and Chapter 6) and outline potential future research work. Due to time constraints, the simulation study Chapter 5) was restricted to the CP GCM. While it seems reasonable to assume that these results will also apply when using alternative approaches for modelling growth trajectories (such as FP and RCS), it is recommended to confirm this assumption by conducting further simulations.

Though researchers may consider including information criteria and prediction criteria in future simulation studies, it is recommended to focus only on the information criteria. This because my results show that for both balanced and unbalanced data, information criteria (AIC and BIC) perform consistently better than prediction criteria (CCC and R^2 statistics) under different conditions of sample size and model complexity. Previous simulation studies also concluded that CCC and R^2 statistics do not perform well in selecting the GCMs (Wu & West, 2013; Yuan et al., 2019). Thus, there is an emerging consensus that CCC and R^2 statistics are inadequate for GCM selection.

Table 8-2. Methodological limitations and future research work.

No.	Limitation/unanswered research question	Reason/possible explanation	Future work
1	Simulation study conducted to evaluate the performance of fit criteria in GCM selection limited to CP GCM (Chapter 5)	Time constraint	To conduct simulation study to evaluate and compare the performance of information criteria for: i.) linear GCMs using restricted cubic splines for modelling functional form of growth trajectories (RCS) ii.) nonlinear GCMs (PB and SITAR)
2	As PB GCM failed to converge, it could not be compared with CP, FP, RCS and SITAR GCMs (Chapter 6)	Small sample size	To fit PB GCM to jaw growth data by: i) using a bigger sample size, especially the level 1 sample size, or ii) considering Bayesian approach
3	Except for the fit to the jaw growth data, evidence on the comparative performance of RCS and SITAR GCMs is inconclusive (Chapter 6)	Large measurement error in the jaw growth data	To compare RCS and the SITAR GCMs by: i) conducting a simulation study ii) applying these two GCMs to height data

GCM: growth curve model; CP: conventional polynomial; FP: fractional polynomial; RCS: restricted cubic spline; SITAR: superimposition by translation and rotation; PB: Preece-Baines.

Focusing on information criteria would increase the value of research and reduce waste by avoiding repeating research findings. In future simulation studies, it would be interesting to compare AIC and BIC with other information criteria available for GCM selection, which include the AICC (Burnham, 2002; Hurvich & Tsai, 1989), CAIC (Bozdogan, 1987), ABIC (Rissanen, 1978; Sclove, 1987), and HQIC (Hannan & Quinn, 1979). Unlike Stata, the Statistical Analysis System (SAS) software (SAS Institute Inc., 2015) has already developed the capability to calculate these six different information criteria. I am not aware of an R software (R Core Team, 2015) package that

computes all six information criteria. The ‘gcmfit’ program which I have written for Stata software computes all of these information criteria.

Information criteria are commonly used to assess the fit of nonlinear GCMs (Pinheiro & Bates, 1995b, 2000; Vonesh & Chinchilli, 1996). However, no previous study that compared different information criteria for the selection of nonlinear GCMs. A detailed study investigating the effect of varying simulation conditions (such as the sample size and type of data) on the performances of different information criteria would make an important contribution to the literature on nonlinear GCM selection. It would be interesting to investigate how different penalty adjustments for REML-estimated AIC and BIC perform for a nonlinear model such as the SITAR GCM. The results of my study showed that AIC outperforms BIC when a linear GCM (CP) involves complex variance-covariance structure. However, unlike linear GCM, the variance-covariance structure of the SITAR GCM remains same (assuming all three random effects are included in competing models), while the fixed effects structure may vary depending on the number of knots used to construct the restricted cubic spline design matrix. For a structural nonlinear model such as the PB GCM, both the variance-covariance structure (assuming all five random effects are included) and the fixed effect structure remain same. Thus, the potential factors that may impact the performance of AIC and BIC for PB GCM selection are the sample size and type of data.

Chapter 6 applies different linear and nonlinear GCMs to real-life longitudinal jaw growth data. Although this research makes novel contributions to the methodological literature on modelling physical growth data, two important limitations were identified. First, the performance of the PB GCM could not be evaluated as it failed to converge. Second, though the AIC statistic showed that the RCS GCM fitted better to the jaw growth data than the SITAR GCM, further work is required before generalising the findings to other settings such as modelling height data.

The PB GCM applied to all three growth measures (upper and lower jaw length and total anterior face height) failed to converge for both males and females. This is likely because of the small sample size, which did not support the estimation of the complex variance-covariance structure of the PB GCM fitted by including all five random effects. Fitting the model with a subset of random effects was not pursued (see Chapter 6, Section 6.4for reasons). Therefore, the performance of the PB GCM could not be evaluated and compared with the other four GCMs (CP, FP, RCS and SITAR) included in the thesis.

Unlike linear GCMs, the estimation of nonlinear GCMs is more complex. This is because nonlinear dependency of the conditional mean on the random effects requires multidimensional integration (Harring & Liu, 2016; Serroyen et al., 2009). A previous simulation study has shown that fitting nonlinear mixed effects models to data with a small sample size (50 to 100 individuals) often results in numerical problems and convergence failure (Harring & Liu, 2016).

Bayesian methods are better equipped to handle small sample situations than frequentist methods when fitting GCMs (Baldwin & Fellingham, 2013; Browne & Draper, 2006; Gelman, 2006; McNeish & Stapleton, 2016). Unlike frequentist methods, Bayesian methods do not rely on large samples because data are considered fixed while the parameters are random (McNeish, 2016). With sampling-based Bayesian methods such as Markov chain Monte Carlo (MCMC), this means that the quality of inference is controlled not by sample size approaching infinity, but rather by the number of samples taken approaching infinity (Kruschke, 2014; Lee & Song, 2004).

Future studies should thus either fit the PB GCM to a bigger data set (particularly the level 1 sample size) or consider fitting the model within the Bayesian framework. It is important to be careful with the former (especially for the level 2 covariance matrix) because estimates for Bayesian models applied to small sample sizes are particularly sensitive to the specification of the prior distribution (McNeish, 2016).

Though the results clearly show that the RCS GCM is better than the SITAR GCM for modelling jaw growth data, further work is required before generalising the findings to other research areas such as modelling height data. The SITAR GCM has proved its usefulness in modelling height growth data (Cole et al., 2010; Cole & Mori, 2017). As no previous study has compared these two GCMs (RCS and SITAR) head-to-head for modelling height data, it is unclear how the SITAR GCM will perform in comparison to the RCS GCM under those conditions.

An ideal approach to compare RCS and SITAR GCMs would be to first compare them using a well-designed simulation study and then apply them to height data. A simulation study would help in understanding how sample size (level 2 and level 1), type of data (balanced and unbalanced), measurement error (within individual variability in growth), and so on affect the performances of the RCS and SITAR GCMs. An important factor to be investigated is measurement error (within-individual variability in growth). This is because the results (see Chapter 6, Section 6.2.9) show that as the within-individual variability in growth increases (e.g., the lower jaw length data for males), the difference in the adolescent growth spurt parameters estimated by the RCS and SITAR GCMs increases.

8.3.2. Clinical

The key clinical limitations (Chapter 7) and outline of potential future research work are summarised below in Table 8-3.

Table 8-3. Clinical limitations and future research work.

No.	Limitation/unanswered research question	Reason/possible explanation	Future work [a]
1	Evidence on class difference in growth trajectories and adolescent growth spurt parameters is less certain	Small sample size	To analyse upper jaw length, lower jaw length and total face height data with a bigger sample size (males and females) with an adequate representation of Class I, Class II and Class III skeletal jaw relationships
2	Inconclusive evidence on the pre-adolescent (childhood) jaw growth spurt for all three classes of skeletal jaw relationship for both sexes	Inadequate age range to cover the childhood growth for Class I, Class II and Class III skeletal jaw relationship for both sexes	To analyse upper jaw length, the lower jaw length and the total face height data with adequate age range, at least 6 to 18 years for males, and 5 to 18 years for females
3	Estimation of adolescent growth spurt parameters restricted to timing and the intensity of the growth spurt	The RCS GCM does not provide a smooth second derivative (acceleration curve), which is required to estimate other adolescent growth spurt parameters such as the duration of the adolescent growth spurt	To apply quartic spline model upper jaw length, the lower jaw length and the total face height data for males and females, which would require extending derivatives estimation for the quartic spline to include covariate(s)
4	Did not explore multivariate GCM to simultaneously model growth trajectories for the upper jaw length, the lower jaw length and the total face height	Small sample size	To fit a multivariate GCM to jointly estimate growth trajectories for the upper jaw length, the lower jaw length and the total face height for males and females

[a] The GCM of choice as well as the modelling framework (frequentist or Bayesian) will depend on the sample size and the specific methodological work.
GCM: growth curve model; RCS: restricted cubic spline.

Because of the small sample size, the evidence for class differences in jaw growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing and intensity is less certain, with wide CIs. This is particularly true for comparisons involving Class III skeletal jaw relationship because compared to Class I and Class II skeletal jaw relationships, the sample size for Class III skeletal jaw relationship was very small (see Chapter 3, Section 3.3). Therefore, to strengthen the evidence, future studies should analyse a bigger data set involving larger numbers of males and females with adequate representation of each class (Class I, Class II and Class III) in the sample. An alternative would be to conduct a meta-analysis of longitudinal studies to combine effect sizes reported at multiple time points. Methodological advances have made it possible to conduct a meta-analysis of longitudinal effect sizes by accounting for dependence between effect sizes, both within and between studies (Ishak et al., 2007; Musekiwa et al., 2016; Trikalinos & Olkin, 2012).

This study focused on modelling jaw growth during adolescence, analysing jaw growth data between seven and 18 years of age (see Chapter 1, Section 1.2 for details). Though evidence was found for the pre-adolescent growth spurt for lower jaw length and total anterior face height for males, meaningful conclusions could not be drawn for class differences in the pre-adolescent growth spurt for males. This because unlike Class I and Class III, growth velocity curves for the lower jaw length and the total anterior face height did not show the pre-adolescent growth spurt for Class II (see Chapter 7, Section 7.3). Furthermore, male growth velocity curves for the upper jaw length did not show any sign of pre-adolescent growth spurt for either Class I, Class II or Class III. Unlike males, results for females showed no evidence for the pre-adolescent growth spurt for any of the three growth measures for either Class I, Class II or Class III (see Chapter 7, Section 7.3).

As discussed in Chapter 7 (Section 7.4.2), the variability in the evidence on the pre-adolescent growth spurt between males and females, and for Class I, Class II and Class III for males,

is perhaps due to an inadequate age range (seven to 18 years) included in the analysis. Since females experienced adolescent growth spurt earlier than males for all three growth measures regardless of the class of skeletal jaw relationship, an inadequate age range for the childhood growth for females could have obscured the detection of the pre-adolescent growth spurt. As the adolescent growth spurt for the lower jaw length and the total anterior face height occurred earlier for Class II than Class I and Class III, a limited age range of data for males also could have resulted in failure to find the pre-adolescent growth spurt for lower jaw length and total anterior face height for Class II. Likewise, the limited data for males did not allow detection of the adolescent growth spurt for upper jaw length, which occurs earlier than lower jaw length and total anterior face height for all three classes of skeletal jaw relationship (Class I, Class II and Class III).

Due to the potential clinical implications of knowing the timing and the intensity of the pre-adolescent growth spurt and its relationship with adolescent growth spurt, it is advised that future studies should focus on analysing jaw growth data with a wider age range covering the childhood and the adolescence growth periods (males 6–18 years; females 5–18 years). The childhood jaw growth data is available from the American Association of Orthodontists Foundation (AAOF) Foundation Craniofacial Growth Legacy Collections. The data used in this thesis (7 to 18 years) was also obtained from the AAOF Foundation Craniofacial Growth Legacy Collections (see Chapter 3, section 3.1). Since the sample size available from the AAOF Foundation Craniofacial Growth Legacy Collections data is limited from age four to seven years, it would be better to consider a Bayesian approach, which has advantages over the frequentist approach when sample size is small (see Section 8.3.1).

Establishing a relationship between timings of the pre-adolescent and the adolescent growth spurts has important potential clinical implications. This would help in predicting the timing of the adolescent growth spurt before it occurs. For example, a positive correlation between timings of the

pre-adolescent and the adolescent growth spurts would imply that a child experiencing an early pre-adolescent growth spurt would also experience an early adolescent growth spurt. Similarly, the intensity of the pre-adolescent growth spurt can inform the intensity of the adolescent growth spurt. I am not aware of any work done in this direction in the fields of jaw growth or height.

As reviewed in Chapter 4 (see Section 4.3), the adolescent growth spurt parameters comprise: (i) age at take-off (onset of the spurt), (ii) age at peak growth velocity (timing), (iii) peak growth velocity (intensity), (iv) end of the spurt, (v) duration of the spurt (time gap between age at take-off and end of the spurt), and (vi) the adolescent gain, which denotes the increase in size from age at take-off to the adulthood size. This thesis focused on timing and intensity of adolescent growth spurt. The duration of the adolescent growth spurt also makes important contributions to final adulthood jaw sizes and face height (García-Drago et al., 2014; Kuc-Michalska & Baccetti, 2010; Proffit et al., 2014).

The duration of the adolescent growth spurt is defined as the time from the age at take-off until the end of the adolescent growth spurt marked by the age corresponding to the maximum deceleration in the post-adolescent growth rate. However, this could not be estimated because the RCS GCM does not provide a smooth second derivative (acceleration curve), required to find the age corresponding to the maximum deceleration in the post-adolescent growth rate. To achieve smooth estimates of second derivative, it is recommended to use a minimum quartic spline basis function (Simpkin et al., 2018).

Simpkin et al. (2018) have provided derivatives estimation for quartic spline basis function based unadjusted (without covariates) mixed effect model. To estimate class-specific duration of the adolescent growth spurt, future studies need to develop methodology on including covariates when differentiating quartic truncated spline basis function. The truncated power basis is one of the two approaches to construct the spline design matrix. The other approach is the B-spline basis (see

Chapter 4, Section 4.2.2). This can be easily done as the methodological framework is similar to inclusion of covariates in the derivatives estimation for the cubic spline basis function (see Chapter 6).

Lastly, the results show that the upper and lower jaw length and total face height measures did not share a common growth trajectory, yet a coordinated growth amongst these three structures is required to achieve a normal skeletal jaw relationship (Class I). Therefore, to better understand the intricacies of these three growth measures, a need exists to jointly model them by fitting a multivariate GCM. Multivariate GCMs allow for the simultaneous estimation of growth processes in two or more constructs (e.g., upper and lower jaw length and total face height(Goldstein, 2011b). This was not pursued because of the small sample size, which could not support the model complexity. As the number of outcomes (growth measures) increases, the model complexity also increases. A multivariate GCM for three outcomes is considerably more complex than the model with two outcomes (Baldwin et al., 2014).

As multivariate GCMs involve more complex variance-covariance structure, it is important to take into account sample size considerations. As discussed in Section 8.3.1, one attractive approach is to use a Bayesian framework to implement a multivariate GCM. Bayesian methods are becoming popular in fitting GCMs and some authors recommend using the Bayesian statistical framework for growth curve modelling as it provides flexibility in fitting GCMs with various levels of complexity (Oravecz & Muth, 2017). For multivariate GCMs, Bayesian methods are particularly useful for interval estimation for correlations among the random effects (Baldwin et al., 2014).

CHAPTER 9. CONCLUSION

9.1. A summary

This thesis began with an overview of jaw growth during adolescence and the importance of studying jaw growth dynamics in correcting skeletal jaw discrepancies. A systematic review found that all previous studies applying growth curve models (GCMs) to the jaw growth data have focused on conventional polynomials for modelling growth trajectories. Therefore, this research work was undertaken to compare the performance of conventional polynomial (CP) GCM with two alternatives, linear and two nonlinear GCMs, both of which are popular in modelling height data. The two linear GCMs are fractional polynomial (FP) and restricted cubic spline (RCS). The two nonlinear GCMs are superimposition by translation and rotation (SITAR) and the Preece-Baines (PB). The aim was to select a best-fitting linear or nonlinear GCM and apply it to male and female data to answer clinical research questions relating to class differences (Class I, Class II and Class III) in growth trajectories (distance, velocity and acceleration) and adolescent growth spurt parameters (timing and intensity) for the upper jaw length, lower jaw length and total face height. To achieve this aim, three separate studies were conducted.

The first study (Chapter 5) used simulations to compare performance of information criteria (Akaike information criterion, AIC; Bayesian information criterion, BIC) and prediction criteria (measure of variance explained, R2; concordance correlation coefficient, CCC) under different conditions of sample size, balanced or unbalanced, and model complexity. The CP GCM was selected for the simulation study. The results showed that both AIC and BIC performed better than prediction criteria. Of the information criteria, the AIC performed better when sample size was small, and the model had a complex variance-covariance structure. The real-life jaw growth data analysed in this thesis comprised a small sample size, and GCMs applied to the data involved complex

variance-covariance structure. Therefore, the AIC was identified as the criterion of choice to evaluate and compare the fit of GCMs applied to the jaw growth data (Chapter 6).

The second study (Chapter 6) used the AIC to compare fit of linear (CP, FP and RCS) and nonlinear (SITAR and PB) GCMs applied to the jaw growth data, comprising 128 males and 139 females who participated in eight growth studies. To compare linear and nonlinear models under the same conditions as in the next clinical study (Chapter 7), each model was fitted to the upper jaw length, lower jaw length and total anterior face height measurements by including class as a covariate in the fixed part of the model. Each linear and nonlinear model was fitted as a two-level model with growth study-specific effects adjusted by following a fixed effect approach. Data were analysed separately for males and females. Linear and nonlinear models were estimated using the REML method. In addition to their fit to the jaw growth data assessed, GCMs were compared in terms of their abilities to estimate covariate-adjusted growth trajectories and adolescent growth spurt parameters. The results showed that the RCS GCM fitted best to the jaw growth data for both sexes. Unlike the RCS and SITAR GCMs, both CP and FP GCMs estimated biologically implausible growth trajectories for both sexes. Compared to the SITAR, the RCS estimated lower residual standard deviation suggesting that it performs better in modelling within-individual variability in growth. Therefore, the RCS GCM was selected as the model of choice for the next clinical study (Chapter 7).

The third study (Chapter 7) applied the RCS GCM to estimate growth study-adjusted class (covariate) differences in the growth trajectories and the adolescent growth spurt parameters for upper and lower jaw length and total anterior face height for males and females. Data were analysed separately for males and females, and models were fitted using the REML method. Results showed that both males and females experienced an adolescent growth spurt for the upper jaw length, lower jaw length and total face height. Clear evidence was found for sexual dimorphism in the growth

trajectories and adolescent growth spurt parameters. For all three outcomes, the adolescent growth spurt occurred earlier in females than males, but the intensity of the growth spurt was greater for males. Class differences in the timing and intensity of the adolescent growth spurt for all three measurements were greater for males than females. Clinical implications of the findings were discussed.

9.2. Concluding remarks

This thesis identified and filled existing gaps in the methodological and clinical literature on modelling longitudinal jaw growth data. While pursuing the research goal, novel contributions were made to the methodological literature on GCM selection and comparison of covariate-adjusted linear and nonlinear GCMs for modelling longitudinal growth data. To demonstrate the methodology on how to include covariates when estimating growth trajectories and adolescent growth spurt parameters for linear and nonlinear GCMs, jaw growth data was used as an example. Several new programs for Stata were written to implement the research methodology.

From a clinical viewpoint, the research findings improve understanding of the development of normal skeletal jaw relationship (Class I) and how the adolescent growth spurt plays an important role in the development of skeletal jaw discrepancies in the upper jaw length, lower jaw length and total face height for Class II and Class III skeletal malocclusions for males and females. The findings may prove useful to clinicians when planning treatment to correct skeletal jaw discrepancies for males and females. The methodological and clinical implications of research findings are discussed, along with potential topics for further methodological and clinical research work.

References

Abbassi, V. (1998). Growth and normal puberty. *Pediatrics, 102*(2), 507-511.

Adams, C. P. (1972). Changes in occlusion and craniofacial pattern during growth. *Transactions of the European Orthodontic Society*, 85-96.

Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: the worldviews of AIC and BIC. *Ecology, 95*(3), 631-636.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on, 19*(6), 716-723. doi:10.1109/TAC.1974.1100705

Allison, P. D. (2012). *Handling missing data by maximum likelihood.* Paper presented at the SAS global forum, Haverford, PA, USA.

Altman, D. G. (2005). Why we need confidence intervals. *World Journal of Surgery, 29*(5), 554-556. doi:10.1007/s00268-005-7911-0

American Psychological Association. (2002). *Developing Adolescents: A Reference for Professionals*. Washington, DC.

Amrhein, V., & Greenland, S. (2018). Remove, rather than redefine, statistical significance. *Nature Human Behaviour, 2*(1), 4-4. doi:10.1038/s41562-017-0224-0

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567*(7748), 305-307. doi:10.1038/d41586-019-00857-9

Angle, E. H. (1899). Classification of malocclusion. *Dental cosmos, 41*, 248-264.

Antoun, J. S., Cameron, C., Sew Hoy, W., Herbison, P., & Farella, M. (2015). Evidence of secular trends in a collection of historical craniofacial growth studies. *The European Journal of Orthodontics, 37*(1), 60-66.

Atalay, Z., & Tortop, T. (2010). Dentofacial effects of a modified tandem traction bow appliance. *European Journal of Orthodontics, 32*(6), 655-661. doi:10.1093/ejo/cjp153

Baccetti, T., Franchi, L., & McNamara, J. A. (2007). Growth in the Untreated Class III Subject. *Seminars in Orthodontics, 13*(3), 130-142. doi:https://doi.org/10.1053/j.sodo.2007.05.006

Baccetti, T., Franchi, L., & McNamara, J. A., Jr. (2011). Longitudinal growth changes in subjects with deepbite. *American Journal of Orthodontics and Dentofacial Orthopedics, 140*(2), 202-209. doi:10.1016/j.ajodo.2011.04.015

Baccetti, T., Reyes, B. C., & McNamara, J. A., Jr. (2005). Gender differences in Class III malocclusion. *Angle Orthodontist, 75*(4), 510-520. doi:10.1043/0003-3219(2005)75[510:GDICIM]2.0.CO;2

Baccetti, T., & Tollaro, I. (1998). A retrospective comparison of functional appliance treatment of Class III malocclusions in the deciduous and mixed dentitions. *European Journal of Orthodontics, 20*(3), 309-317.

Baker, M. (2016). Statisticians issue warning over misuse of P values. *Nature, 531*(7593), 151. doi:10.1038/nature.2016.19503

Baldwin, S. A., & Fellingham, G. W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods, 18*(2), 151-164. doi:10.1037/a0030642

Baldwin, S. A., Imel, Z. E., Braithwaite, S. R., & Atkins, D. C. (2014). Analyzing Multiple Outcomes in Clinical Research Using Multivariate Multilevel Models. *Journal of Consulting and Clinical Psychology*. doi:10.1037/a0035628

Ball, G., Woodside, D., Tompson, B., Hunter, W. S., & Posluns, J. (2011). Relationship between cervical vertebral maturation and mandibular growth. *American Journal of Orthodontics and Dentofacial Orthopedics, 139*(5), e455-e461. doi:http://dx.doi.org/10.1016/j.ajodo.2010.01.035

Bambha, J. K. (1961). Longitudinal cephalometric roentgenographic study of face and cranium in relation to body height. *Journal of the American Dental Association, 63*, 776-799.

Bambha, J. K., & Van Natta, P. (1963). Longitudinal study of facial growth in relation to skeletal maturation during adolescence. *American Journal of Orthodontics, 49*(7), 481-493. doi:http://dx.doi.org/10.1016/0002-9416(63)90203-3

Banik, S. D., Salehabadi, S. M., & Dickinson, F. (2017). Preece-Baines Model 1 to Estimate Height and Knee Height Growth in Boys and Girls From Merida, Mexico. *Food and Nutrition Bulletin, 38*(2), 182-195. doi:10.1177/0379572117700270

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi:10.1016/j.jml.2012.11.001

Barry, P. D. (2015). *Geometry with Trigonometry: Second Edition*.

Baskaradoss, J. K., Geevarghese, A., Roger, C., & Thaliath, A. (2013). Prevalence of malocclusion and its relationship with caries among school children aged 11-1 5 years in southern India. *Korean Journal of Orthodontics, 43*(1), 35-41. doi:10.4041/kjod.2013.43.1.35

Bates, D., Kliegl, R., & Vasishth, S. (2015). *Parsimonious mixed models*. Cornell University. Retrieved from https://arxiv.org/abs/1506.04967v2

Bates, D. M., & Venables, W. N. (2011). 'splines': Regression Spline Functions and Classes. version 3.6.0.

Battagel, J. M., & Orton, H. S. (1995). A comparative study of the effects of customized facemask therapy or headgear to the lower arch on the developing Class III face. *European Journal of Orthodontics, 17*(6), 467-482.

Baughan, B., Demirjian, A., Levesque, G., & Lapalme-Chaput, L. (1979a). The pattern of facial growth before and during puberty, as shown by French-Canadian girls. *Annals of Human Biology, 6*(1), 59-76.

Baughan, B., Demirjian, A., Levesque, G. Y., & Lapalme-Chaput, L. (1979b). The pattern of facial growth before and during puberty, as shown by French-Canadian girls. *Annals of Human Biology*, *6*(1), 59-76.

Baume, R. M., Buschang, P. H., & Weinstein, S. (1983). Stature, head height, and growth of the vertical face. *American Journal of Orthodontics*, *83*(6), 477-484.

Baumrind, S., & Curry, S. (2015). American Association of Orthodontists Foundation Craniofacial Growth Legacy Collection: Overview of a powerful tool for orthodontic research and teaching. *American Journal of Orthodontics and Dentofacial Orthopedics*, *148*(2), 217-225. doi:10.1016/j.ajodo.2015.06.002

Baumrind, S., & Miller, D. M. (1980). Computer-Aided Head Film Analysis - the University-of-California-San-Francisco Method. *American Journal of Orthodontics and Dentofacial Orthopedics*, *78*(1), 41-65. doi:Doi 10.1016/0002-9416(80)90039-1

Beath, K. J. (2007). Infant growth modelling using a shape invariant model with random effects. *Statistics in Medicine*, *26*(12), 2547-2564. doi:10.1002/sim.2718

Behrents, R. G., & Broadbent, B. H. (1984). *A chronological account of the Bolton-Brush growth studies: In search of truth for the greater good of man*. In. Retrieved from https://dental.case.edu/media/school-of-dental-medicine/departments--programs/bolton-brush/Chronological_Account_BoltonBrush.pdf

Bell, B. A., Ene, M., & Schoeneberger, J. (2013). *A multilevel model primer using SAS PROC MIXED.* Paper presented at the SAS Global Forum 2013: Statistics and Data Analysis.

Bell, B. A., Morgan, G. B., Schoeneberger, J. A., Kromrey, J. D., & Ferron, J. M. (2014). How Low Can You Go? An Investigation of the Influence of Sample Size and Model Complexity on Point and Interval Estimates in Two-Level Linear Models. *Methodology-European Journal of Research Methods for the Behavioral and Social Sciences*, *10*(1), 1-11. doi:10.1027/1614-2241/a000062

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., *et al.* (2018). Redefine statistical significance. *Nature Human Behaviour, 2*(1), 6-10. doi:10.1038/s41562-017-0189-z

Berman, A., Snyder, S. J., & Frandsen, G. (2016). Concepts of Growth and Development. In A. Berman, S. J. Snyder, & G. Frandsen (Eds.), *Kozier & Erb's Fundamentals of Nursing: Concepts, Process, and Practice* (10th ed., pp. 312-327). Australia: Julie Levin Alexander.

Bernabe, E., Sheiham, A., & de Oliveira, C. M. (2008). Condition-Specific Impacts on Quality of Life Attributed to Malocclusion by Adolescents with Normal Occlusion and Class I, II and III Malocclusion. *Angle Orthodontist, 78*(6), 977-982. doi:10.2319/091707-444.1

Bernstein, R. M. (2018). Models of growth for humans and nonhuman primates. In W. Trevathan (Ed.), *The International Encyclopedia of Biological Anthropology*: John Wiley & Sons.

Beunen, G. P., Rogol, A. D., & Malina, R. M. (2006). Indicators of biological maturation and secular changes in biological maturation. *Food and Nutrition Bulletin, 27*(4), S244-S256. doi:Doi 10.1177/15648265060274s508

Binder, H., Sauerbrei, W., & Royston, P. (2013). Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. *Statistics in Medicine, 32*(13), 2262-2277. doi:10.1002/sim.5639

Bishara, S. E. (1981). Longitudinal cephalometric standards from 5 years of age to adulthood. *American Journal of Orthodontics, 79*(1), 35-44.

Bishara, S. E., Jamison, J. E., Peterson, L. C., & DeKock, W. H. (1981). Longitudinal changes in standing height and mandibular parameters between the ages of 8 and 17 years. *American Journal of Orthodontics, 80*(2), 115-135. doi:http://dx.doi.org/10.1016/0002-9416(81)90213-X

Björk, A. (1963). Variations in the Growth Pattern of the Human Mandible: Longitudinal Radiographic Study by the Implant Method. *Journal of Dental Research, 42*(1), 400-411.

Bjork, A., & Helm, S. (1967). Prediction of the age of maximum puberal growth in body height. *Angle Orthodontist, 37*(2), 134-143. doi:10.1043/0003-3219(1967)037<0134:POTAOM>2.0.CO;2

Black, A. C., Harel, O., & McCoach, D. B. (2011). Missing data techniques for multilevel data: implications of model misspecification. *Journal of Applied Statistics, 38*(9), 1845-1865. doi:10.1080/02664763.2010.529882

Bogin, B. (2010). Evolution of Human Growth. In M. P. Muehlenbein (Ed.), *Human Evolutionary Biology* (pp. 379-395). Cambridge: Cambridge University Press.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Chichester : Wiley.

Bollen, K. A., & Curran, P. J. (2006). *Latent Curve Models: A Structural Equation Perspective*. New Jersey: John Wiley & Sons.

Bollhalder, J., Hanggi, M. P., Schatzle, M., Markic, G., Roos, M., & Peltomaki, T. A. (2013). Dentofacial and upper airway characteristics of mild and severe class II division 1 subjects. *European Journal of Orthodontics, 35*(4), 447-453. doi:10.1093/ejo/cjs010

Bose, K. (2007). Concept of human physical growth and development. In Fundamentals of Human Genetics and Human Growth. Retrieved from http://nsdl.niscair.res.in/jspui/handle/123456789/243.

Box, G. E. P., & Tidwell, P. W. (1962). Transformation of the Independent Variables. *Technometrics, 4*(4), 531-550. doi:10.2307/1266288

Boyd, E., Scammon, R. E., Savara, B. S., & Schilke, J. F. (1980). *Origins of the study of human growth*: University of Oregon Health Sciences Center Foundation.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*(3), 345-370. doi:10.1007/bf02294361

Brady, A. P. (2017). Error and discrepancy in radiology: inevitable or avoidable? *Insights into Imaging, 8*(1), 171-182. doi:10.1007/s13244-016-0534-1

Browne, W., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis, 1*(3), 473-513.

Browne, W., & Goldstein, H. (2010). MCMC Sampling for a Multilevel Model With Nonindependent Residuals Within and Between Cluster Units. *35*(4), 453-473. doi:10.3102/1076998609359788

Browne, W. J. (1998). Applying MCMC methods to multi-level models. In: University of Bath.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of Hierarchical Linear Models to Assessing Change. *Psychological Bulletin, 101*(1), 147-158. doi:10.1037/0033-2909.101.1.147

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models for social and behavioural research: Applications and data analysis methods*. Newbury Park: SAGE Publications.

Buis, M. L. (2009). Using and interpreting restricted cubic splines.

Burnham, K., & Anderson, D. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research, 33*(2), 261-304.

Burnham, K. P. (2002). *Model selection and multimodel inference : a practical information-theoretic approach / Kenneth P. Burnham, David R. Anderson* (2nd ed.. ed.). New York: New York : Springer.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference : a practical information-theoretic approach* (2nd ed.. ed.). New York: Springer.

Buschang, P. H., Baume, R. M., & Nass, G. G. (1983). A craniofacial growth maturity gradient for males and females between 4 and 16 years of age. *American Journal of Physical Anthropology, 61*(3), 373-381.

Buschang, P. H., Jacob, H. B., & Demirjian, A. (2013). Female adolescent craniofacial growth spurts: real or fiction? *European Journal of Orthodontics, 35*(6), 819-825. doi:10.1093/ejo/cjs094

Buschang, P. H., Roldan, S. I., & Tadlock, L. P. (2017). Guidelines for assessing the growth and development of orthodontic patients. *Seminars in Orthodontics, 23*(4), 321-335. doi:10.1053/j.sodo.2017.07.001

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

Buschang, P. H., Tanguay, R., Demirjian, A., Lapalme, L., & Goldstein, H. (1988a). Pubertal Growth of the Cephalometric Point Gnathion - Multilevel Models for Boys and Girls. *American Journal of Physical Anthropology, 77*(3), 347-354. doi:DOI 10.1002/ajpa.1330770307

Buschang, P. H., Tanguay, R., Demirjian, A., LaPalme, L., & Goldstein, H. (1989). Modeling longitudinal mandibular growth: percentiles for gnathion from 6 to 15 years of age in girls. *American Journal of Orthodontics and Dentofacial Orthopedics, 95*(1), 60-66.

Buschang, P. H., Tanguay, R., Demirjian, A., Lapalme, L., & Turkewicz, J. (1988b). Mathematical models of longitudinal mandibular growth for children with normal and untreated Class II, division 1 malocclusion. *The European Journal of Orthodontics, 10*(3), 227.

Buschang, P. H., Tanguay, R., Turkewicz, J., Demirjian, A., & La Palme, L. (1986). A polynomial approach to craniofacial growth: description and comparison of adolescent males with normal occlusion and those with untreated Class II malocclusion. *American Journal of Orthodontics and Dentofacial Orthopedics, 90*(5), 437-442.

Busscher, I., Kingma, I., de Bruin, R., Wapstra, F. H., Verkerke, G. J., & Veldhuizen, A. G. (2012). Predicting the peak growth velocity in the individual child: validation of a new growth model. *European Spine Journal, 21*(1), 71-76. doi:10.1007/s00586-011-1845-z

Cameron, N. (2008). The biology of growth. *Nestle Nutrition Workshop Series: Paediatric Programme, 61*, 1-19. doi:10.1159/0000113162

Cameron, N., & Bogin, B. (2012). *Human Growth and Development* (2nd ed.): Academic Press.

Campbell, P. M. (1983). The dilemma of Class III treatment. Early or late? *Angle Orthodontist, 53*(3), 175-191. doi:10.1043/0003-3219(1983)053<0175:TDOCIT>2.0.CO;2

Cao, Z., Hui, L. L., & Wong, M. Y. (2018). New approaches to obtaining individual peak height velocity and age at peak height velocity from the SITAR model. *Computer Methods and Programs in Biomedicine, 163*, 79-85. doi:https://doi.org/10.1016/j.cmpb.2018.05.030

Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine, 19*(9), 1141-1164. doi:Doi 10.1002/(Sici)1097-0258(20000515)19:9<1141::Aid-Sim479>3.0.Co;2-F

Carpenter, J. R., Goldstein, H., & Rasbash, J. (2003). A novel bootstrap procedure for assessing the relationship between class size and achievement. *Journal of the Royal Statistical Society Series C-Applied Statistics, 52*, 431-443. doi:Doi 10.1111/1467-9876.00415

Castaldo, G., & Cerritelli, F. (2015). Craniofacial growth: evolving paradigms. *CRANIO®, 33*(1), 23-31. doi:10.1179/0886963414Z.00000000042

Cavanaugh, J. (2009). Historic growth collections given new life. *American Journal of Orthodontics and Dentofacial Orthopedics, 135*(6), 688. doi:10.1016/j.ajodo.2009.04.004

Centre for Reviews and Dissemination. (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care*: University of York.

Chambers, R., & Chandra, H. (2012). A Random Effect Block Bootstrap for Clustered Data. *Journal of Computational and Graphical Statistics, 22*(2), 452-470. doi:10.1080/10618600.2012.681216

Charlton, C., Rasbash, J., Browne, W., Healy, M., & Cameron, B. (2017). MLwiN Version 3.00. Centre for Multilevel, Modelling University of Bristol.

Chvatal, B. A., Behrents, R. G., Ceen, R. F., & Buschang, P. H. (2005). Development and testing of multilevel models for longitudinal craniofacial growth prediction. *American Journal of Orthodontics and Dentofacial Orthopedics, 128*(1), 45-56. doi:10.1016/j.ajodo.2004.03.035

Cobourne, M. T., & DiBiase, A. T. (2010). *Handbook of Orthodontics*: Elsevier Health Sciences UK.

Coertjens, L., Donche, V., De Maeyer, S., Vanthournout, G., & Van Petegem, P. (2017). To what degree does the missing-data technique influence the estimated growth in learning strategies over time? A tutorial example of sensitivity analysis for longitudinal data. *PloS One, 12*(9). doi:ARTN e0182615

10.1371/journal.pone.0182615

Cole, T., Kuh, D., Johnson, W., Ward, K., Howe, L., Adams, J., *et al.* (2016). Using Super-Imposition by Translation And Rotation (SITAR) to relate pubertal growth to bone health in later life: the Medical Research Council (MRC) National Survey of Health and Development. *International Journal of Epidemiology, 45*(4), 1125-1134. doi:10.1093/ije/dyw134

Cole, T. J. (2003). The secular trend in human physical growth: a biological view. *Economics and Human Biology, 1*(2), 161-168. doi:10.1016/S1570-677x(02)00033-3

Cole, T. J. (2017a). sitar: Super Imposition by Translation and Rotation Growth Curve Analysis. R package version 1.0.10.

Cole, T. J. (2017b). SITAR: Super Imposition by Translation and Rotation Growth Curve Analysis. R package version.1.0.9. Available. https://CRAN.R-project.org/package=sitar.

Cole, T. J. (2018). sitar: Super Imposition by Translation and Rotation Growth Curve Analysis. R package version 1.1.0.

Cole, T. J. (2019). sitar: Super Imposition by Translation and Rotation Growth Curve Analysis. R package version 1.1.1.

Cole, T. J., Donaldson, M. D. C., & Ben-Shlomo, Y. (2010). SITAR—a useful instrument for growth curve analysis. *International Journal of Epidemiology*, *39*(6), 1558-1566. doi:10.1093/ije/dyq115

Cole, T. J., & Mori, H. (2017). Fifty years of child height and weight in Japan and South Korea: Contrasting secular trend patterns analyzed by SITAR. *American Journal of Human Biology*. doi:10.1002/ajhb.23054

Cole, T. J., Pan, H., & Butler, G. E. (2014). A mixed effects model to estimate timing and intensity of pubertal growth from height and secondary sexual characteristics. *Annals of Human Biology*, *41*(1), 76-83. doi:10.3109/03014460.2013.856472

Collins, G. S., Ogundimu, E. O., Cook, J. A., Manach, Y. L., & Altman, D. G. (2016). Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Statistics in Medicine*, *35*(23), 4124-4135. doi:10.1002/sim.6986

Collins, L. M. (2006). Analysis of longitudinal data: the integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology*, *57*, 505-528. doi:10.1146/annurev.psych.57.102904.190146

Corbeil, R. R., & Searle, S. R. (1976). Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model. *Technometrics, 18*(1), 31-38. doi:10.2307/1267913

Cox, N. J. (2008). Speaking Stata: Correlation with confidence, or Fisher's z revisited. *Stata Journal*, *8*(3), 413-439. doi:Doi 10.1177/1536867x0800800307

Crozier, S. R., Johnson, W., Cole, T. J., Macdonald-Wallis, C., Muniz-Terrera, G., Inskip, H. M., *et al.* (2019). A discussion of statistical methods to characterise early growth and its impact on bone mineral content later in childhood. *Annals of Human Biology*, 1-11. doi:10.1080/03014460.2019.1574896

Curran, P. J. (2003). Have Multilevel Models Been Structural Equation Models All Along? *Multivariate Behavioral Research, 38*(4), 529-569. doi:10.1207/s15327906mbr3804_5

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583-619. doi:10.1146/annurev.psych.093008.100356

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. *Journal of cognition and development : official journal of the Cognitive Development Society, 11*(2), 121-136. doi:10.1080/15248371003699969

Curtis, A. C. (2015). Defining adolescence. *Journal of Adolescent and Family Health, 7*(2), 1-39. Retrieved from http://scholar.utc.edu/jafh/vol7/iss2/2

D'Aloisio, A. A., DeRoo, L. A., Baird, D. D., Weinberg, C. R., & Sandler, D. P. (2013). Prenatal and Infant Exposures and Age at Menarche. *Epidemiology, 24*(2), 277-284. doi:10.1097/EDE.0b013e31828062b7

Dahlgren, J. P., Garcia, M. B., & Ehrlen, J. (2011). Nonlinear relationships between vital rates and state variables in demographic models. *Ecology, 92*(5), 1181-1187.

Davidian, M., & Gallant, A. R. (1992). Smooth Nonparametric Maximum-Likelihood-Estimation for Population Pharmacokinetics, with Application to Quinidine. *Journal of Pharmacokinetics and Biopharmaceutics, 20*(5), 529-556. doi:Doi 10.1007/Bf01061470

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge ; New York, NY, USA: Cambridge University Press.

de Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer-Verlag.

De Clerck, H. J., & Proffit, W. R. (2015). Growth modification of the face: A current perspective with emphasis on Class III treatment. *American Journal of Orthodontics and Dentofacial Orthopedics, 148*(1), 37-46. doi:10.1016/j.ajodo.2015.04.017

de Valpine, P. (2012). Frequentist analysis of hierarchical models for population dynamics and demographic data. *Journal of Ornithology, 152*, S393-S408. doi:10.1007/s10336-010-0642-5

Dean, J. A., Jones, J. E., & Vinson, L. A. W. (2016). *McDonald and Avery's Dentistry for the Child and Adolescent* (10th ed.): Elsevier Health Sciences.

Demirjian, A., & Goldstein, H. (1976). New systems for dental maturity based on seven and four teeth. *Annals of Human Biology, 3*(5), 411-421.

Desquilbet, L., & Mariotti, F. (2010). Dose-response analyses using restricted cubic spline functions in public health research. *Statistics in Medicine, 29*(9), 1037-1057. doi:10.1002/sim.3841

DiBiase, A. T., Cobourne, M. T., & Lee, R. T. (2015). The use of functional appliances in contemporary orthodontic practice. *British Dental Journal, 218*(3), 123-128. doi:10.1038/sj.bdj.2015.44

Dimberg, L., Arnrup, K., & Bondemark, L. (2015). The impact of malocclusion on the quality of life among children and adolescents: a systematic review of quantitative studies. *European Journal of Orthodontics, 37*(3), 238-247. doi:10.1093/ejo/cju046

Dolce, C., McGorray, S. P., Brazeau, L., King, G. J., & Wheeler, T. T. (2007). Timing of Class II treatment: skeletal changes comparing 1-phase and 2-phase treatment. *American Journal of Orthodontics and Dentofacial Orthopedics, 132*(4), 481-489. doi:10.1016/j.ajodo.2005.08.046

Duren, D. L., Seselj, M., Froehle, A. W., Nahhas, R. W., & Sherwood, R. J. (2013). Skeletal Growth and the Changing Genetic Landscape during Childhood and Adulthood. *American Journal of Physical Anthropology, 150*(1), 48-57. doi:10.1002/ajpa.22183

Durrleman, S., & Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine, 8*(5), 551-561. doi:10.1002/sim.4780080504

Eager, C., & Roy, J. (2017). *Mixed effects models are sometimes terrible*. Cornell University. Retrieved from http://arxiv.org/abs/1701.04858

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R 2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine, 27*(29), 6137-6157. doi:10.1002/sim.3429

Efron, B. (1979). Bootstrap Methods - Another Look at the Jackknife. *Annals of Statistics, 7*(1), 1-26. doi:DOI 10.1214/aos/1176344552

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Field, C. A., & Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 69*, 369-390. doi:DOI 10.1111/j.1467-9868.2007.00593.x

Fishman, L. S. (1979). Chronological versus skeletal age, an evaluation of craniofacial growth. *Angle Orthodontist, 49*(3), 181-189. doi:10.1043/0003-3219(1979)049<0181:cvsaae>2.0.co;2

Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (2008). *Longitudinal Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Fleming, P. S., & Lee, R. T. (2016). *Orthodontic Functional Appliances: Theory and Practice*: Wiley-Blackwell.

Flores-Mir, C., Nebbe, B., & Major, P. W. (2004). Use of skeletal maturation based on hand-wrist radiographic analysis as a predictor of facial growth: A systematic review. *Angle Orthodontist, 74*(1), 118-124.

Fraser, C., Murray, A., & Burr, J. (2006). Identifying observational studies of surgical interventions in MEDLINE and EMBASE. *BMC Medical Research Methodology, 6*, 41. doi:10.1186/1471-2288-6-41

García-Drago, Gabriel, A., & Arriola-Guillén, L. E. (2014). Duration of the peak of growth in Class I and III subjects using the Baccetti's cervical vertebrae maturation analysis on lateral cephalometric radiographs. *Oral Health Dent Manag, 13*(4), 963-966.

Garn, S. M. (1981). The Growth of Growth *American Journal of Physical Anthropology, 56*, 521-530.

Gasser, T., Gervini, D., & Molinari, L. (2004). Kernel estimation, shape-invariant modelling and structural analysis. In R. C. Hauspie, N. Cameron, & L. Molinari (Eds.), *Methods in Human Growth Research* (pp. 179-204). Cambridge UK.

Gasser, T., Kneip, A., Binding, A., Prader, A., & Molinari, L. (1991). The Dynamics of Linear Growth in Distance, Velocity and Acceleration. *Annals of Human Biology, 18*(3), 187-205. doi:Doi 10.1080/03014469100001522

Gasser, T., Muller, H. G., Kohler, W., Molinari, L., & Prader, A. (1984). Nonparametric Regression-Analysis of Growth-Curves. *Annals of Statistics, 12*(1), 210-229. doi:DOI 10.1214/aos/1176346402

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis, 1*(3), 515-534.

Ghafournia, M., & Hajenourozali Tehrani, M. (2012). Relationship between Bruxism and Malocclusion among Preschool Children in Isfahan. *Journal of Dental Research, Dental Clinics, Dental Prospects, 6*(4), 138-142. doi:10.5681/joddd.2012.028

Ghisletta, P., & Lindenberger, U. (2004). Static and dynamic longitudinal structural analyses of cognitive changes in old age. *Gerontology, 50*(1), 12-16. doi:10.1159/000074383

Ghisletta, P., Renaud, O., Jacot, N., & Courvoisier, D. (2015). Linear Mixed-Effects and Latent Curve Models for Longitudinal Life Course Analyses. In C. Burton-Jeangros, S. Cullati, A. Sacker, & D. Blane (Eds.), *A Life Course Perspective on Health Trajectories and Transitions* (pp. 155-178). Cham: Springer International Publishing.

Gibbons, R. D., Hedeker, D., & DuToit, S. (2010). Advances in analysis of longitudinal data. *Annual Review of Clinical Psychology, 6*, 79-107. doi:10.1146/annurev.clinpsy.032408.153550

Goldstein, H. (1968). Longitudinal Studies and The Measurement of Change. *Journal of the Royal Statistical Society. Series D (The Statistician), 18*(2), 93-117. doi:10.2307/2986775

Goldstein, H. (1979). *The design and analysis of longitudinal studies: Their role in the measurement of change*. London: Academic Press.

Goldstein, H. (1986a). Efficient statistical modelling of longitudinal data. *Annals of Human Biology, 13*(2), 129-141. doi:10.1080/03014468600008271

Goldstein, H. (1986b). Multilevel Mixed Linear-Model Analysis Using Iterative Generalized Least-Squares. *Biometrika, 73*(1), 43-56.

Goldstein, H. (1989a). Efficient prediction models for adult height. In J. M. Tanner (Ed.), *Auxology 88: perspectives in the science of growth and development*. London : Niigata-Shi, Japan: Smith-Gordon ; Nishimura.

Goldstein, H. (1989b). Restricted Unbiased Iterative Generalized Least-Squares Estimation. *Biometrika, 76*(3), 622-623. doi:10.2307/2336130

Goldstein, H. (2011a). *Multilevel statistical models* (4 ed.). Chichester, UK: Wiley.

Goldstein, H. (2011b). *Multilevel statistical models. 4th Edition* (4th ed. ed.). United Kingdom: John Wiley and Sons.

Goldstein, H., Healy, M. J., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine, 13*(16), 1643-1655.

Goldstein, H., Leckie, G., Charlton, C., Tilling, K., & Browne, W. J. (2018). Multilevel growth curve models that incorporate a random coefficient model for the level 1 variance function. *Statistical Methods in Medical Research, 27*(11), 3478-3491. doi:10.1177/0962280217706728

Goldstein, M. S. (1936). Changes in Dimensions and Form of the Face and Head with Age. *American Journal of Physical Anthropology, 22*(1), 37-89. doi:DOI 10.1002/ajpa.1330220104

Grave, K. C., & Brown, T. (1976). Skeletal ossification and the adolescent growth spurt. *American Journal of Orthodontics, 69*(6), 611-619. doi:http://dx.doi.org/10.1016/0002-9416(76)90143-3

Greenland, S. (1995). Dose-Response and Trend Analysis in Epidemiology - Alternatives to Categorical Analysis. *Epidemiology, 6*(4), 356-365. doi:Doi 10.1097/00001648-199507000-00005

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development, 82*(5), 1357-1371. doi:10.1111/j.1467-8624.2011.01630.x

Grippaudo, C., Pantanali, F., Paolantonio, E. G., Saulle, R., La Torre, G., & Deli, R. (2013). Orthodontic treatment timing in growing patients. *European Journal of Paediatric Dentistry, 14*(3), 231-236.

Gumedze, F. N., & Dunne, T. T. (2011). Parameter estimation and inference in the linear mixed model. *Linear Algebra and Its Applications, 435*(8), 1920-1944. doi:10.1016/j.laa.2011.04.015

Gurka, M. J. (2006). Selecting the Best Linear Mixed Model Under REML. *The American Statistician, 60*(1), 19-26. doi:10.1198/000313006X90396

Gurka, M. J., & Edwards, L. J. (2007). Mixed Models. In J. P. M. C.R. Rao & D. C. Rao (Eds.), *Handbook of Statistics* (Vol. Volume 27, pp. 253-280): Elsevier.

Gurri, F. D. (2018). Adolescent growth spurt. In W. Trevathan (Ed.), *The International Encyclopedia of Biological Anthropology*: John Wiley & Sons.

Hägg, U., & Matsson, L. (1985). Dental maturity as an indicator of chronological age: the accuracy and precision of three methods. *European Journal of Orthodontics, 7*(1), 25-34. doi:10.1093/ejo/7.1.25

Hagg, U., & Taranger, J. (1980). Menarche and voice change as indicators of the pubertal growth spurt. *Acta Odontologica Scandinavica, 38*(3), 179-186.

Hägg, U., & Taranger, J. (1982). Maturation indicators and the pubertal growth spurt. *American Journal of Orthodontics, 82*(4), 299-309. doi:http://dx.doi.org/10.1016/0002-9416(82)90464-X

Hannan, E. J., & Quinn, B. G. (1979). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society. Series B (Methodological), 41*(2), 190-195.

Harrell, F. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*: Springer International Publishing.

Harrell Jr, F. E. (2017). rms: Regression Modeling Strategies. R package version 5.1-1. Retrieved from https://CRAN.R-project.org/package=rms

Harring, J. R., & Liu, J. H. (2016). A Comparison of Estimation Methods For Nonlinear Mixed Effects Models Under Model Misspecification and Data Sparseness: A Simulation Study. *Journal of Modern Applied Statistical Methods, 15*(1), 539-569.

Harris, J. E. (1962). A cephalometric analysis of mandibular growth rate. *American Journal of Orthodontics, 48*(3), 161-174. doi:https://doi.org/10.1016/0002-9416(62)90200-2

Hartley, H. O., & Rao, J. N. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika, 54*(1), 93-108.

Harvard Society for the Advancement of Orthodontics. (2007). Advances in orthodontics. . Retrieved from http://www.hsao-online.org/uploads/Advances%202006-07.pdf

Harville, D. A. (1977a). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association, 72*(358), 320-338. doi:10.1080/01621459.1977.10480998

Harville, D. A. (1977b). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems: Rejoinder. *Journal of the American Statistical Association, 72*(358), 339-340. doi:10.2307/2286798

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

Hauspie, R., & Roelants, M. (2012). Adolescent Growth. In N. Cameron & B. Bogin (Eds.), *Human Growth and Development (Second Edition)* (pp. 57-79). Boston: Academic Press.

Hauspie, R. C., Cameron, N., & Molinari, L. (2004). *Methods in Human Growth Research*. Cambridge, UK: Cambridge University Press.

Hellman, M. (1927). Changes in the human face brought about by development. *International Journal of Orthodontia, Oral Surgery and Radiography, 13*(6), 475-516. doi:10.1016/S0099-6963(27)90222-1

Hermanussen, M. (2016). Growth in Childhood and Puberty. In P. Kumanov & A. Agarwal (Eds.), *Puberty: Physiology and Abnormalities* (pp. 65-76). Cham: Springer International Publishing.

Houston, W. J. (1983). The analysis of errors in orthodontic measurements. *American Journal of Orthodontics, 83*(5), 382-390.

Howe, L. D., Tilling, K., Matijasevich, A., Petherick, E. S., Santos, A. C., Fairley, L., *et al.* (2016). Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Statistical Methods in Medical Research, 25*(5), 1854-1874. doi:10.1177/0962280213503925

Hox, J. J. (2010). *Multilevel analysis: techniques and applications* (2nd ed.). New York: Routledge.

Hox, J. J., Moerbeek, M., & Schoot, R. V. D. (2018). *Multilevel analysis: techniques and applications* (3rd ed.). New York: Routledge.

Hox, J. J., & Roberts, J. K. (2011). *Handbook of advanced multilevel analysis*. New York: Routledge.

Hughes-Hallett, D., Lock, P. F., & Gleason, A. M. (2014). *Applied calculus* (Fifth edition. ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Humphrey, L. T. (1998). Growth patterns in the modern human skeleton. *American Journal of Physical Anthropology, 105*(1), 57-72. doi:Doi 10.1002/(Sici)1096-8644(199801)105:1<57::Aid-Ajpa6>3.0.Co;2-A

Hunter, W. S., Baumrind, S., & Moyers, R. E. (1993). An inventory of United States and Canadian growth record sets: Preliminary report. *American Journal of Orthodontics and Dentofacial Orthopedics, 103*(6), 545-555. doi:10.1016/0889-5406(93)70095-6

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and Time Series Model Selection in Small Samples. *Biometrika, 76*(2), 297-307. doi:10.2307/2336663

Ioannidis, J. P. A. (2018). The Proposal to Lower P Value Thresholds to .005. *JAMA, 319*(14), 1429-1430. doi:10.1001/jama.2018.1536

Ioannidis, J. P. A. (2019a). The Importance of Predefined Rules and Prespecified Statistical Analyses Do Not Abandon Significance. *Jama-Journal of the American Medical Association, 321*(21), 2067-2068. doi:10.1001/jama.2019.4582

Ioannidis, J. P. A. (2019b). Retiring statistical significance would give bias a free pass. *Nature, 567*(7749), 461. doi:10.1038/d41586-019-00969-2

Ishak, K. J., Platt, R. W., Joseph, L., Hanley, J. A., & Caro, J. J. (2007). Meta-analysis of longitudinal studies. *Clinical Trials (London, England), 4*(5), 525-539. doi:10.1177/1740774507083567

Jacobson, A., & Jacobson, R. L. (2006). *Radiographic Cephalometry: From Basics to 3-D Imaging*: Quintessence Pub.

Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2017). An R2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics, 44*(6), 1086-1105. doi:10.1080/02664763.2016.1193725

Jamison, J. E., Bishara, S. E., Peterson, L. C., DeKock, W. H., & Kremenak, C. R. (1982). Longitudinal changes in the maxilla and the maxillary-mandibular relationship between 8 and 17 years of age. *American Journal of Orthodontics, 82*(3), 217-230. doi:http://dx.doi.org/10.1016/0002-9416(82)90142-7

Jeelani, W., Fida, M., & Shaikh, A. (2016). Timing of adolescent growth spurt among children with different skeletal classes. *Pakistan Orthodontic Journal, 8*(2), 72-79.

Jelenkovic, A., Sund, R., Hur, Y. M., Yokoyama, Y., Hjelmborg, J. V. B., Moller, S., *et al.* (2016). Genetic and environmental influences on height from infancy to early adulthood: An individual-based pooled analysis of 45 twin cohorts. *Scientific Reports, 6*. doi:ARTN 28496

10.1038/srep28496

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution, 19*(2), 101-108. doi:10.1016/j.tree.2003.10.013

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution, 5*(9), 944-946. doi:10.1111/2041-210X.12225

Johnson, W. (2015). Analytical strategies in human growth research. *American Journal of Human Biology, 27*(1), 69-83. doi:10.1002/ajhb.22589

Johnson, W., Balakrishna, N., & Griffiths, P. L. (2013). Modeling physical growth using mixed effects models. *American Journal of Physical Anthropology, 150*(1), 58-67. doi:10.1002/ajpa.22128

Jones, A. (2004). A review of random effects models in MLwiN (Version 2.0). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.650.9196&rep=rep1&type=pdf

Joshi, N., Hamdan, A. M., & Fakhouri, W. D. (2014). Skeletal Malocclusion: A Developmental Disorder With a Life-Long Morbidity. *Journal of Clinical Medicine Research, 6*(6), 399-408. doi:10.14740/jocmr1905w

Kallunki, J., Bondemark, L., Paulsson, L. J. J. o. D. O. H. C., & Cosmesis. (2018). Outcomes of Early Class II Malocclusion Treatment: A Systematic Review.

Kapust, A. J., Sinclair, P. M., & Turley, P. K. (1998). Cephalometric effects of face mask/expansion therapy in Class III children: a comparison of three age groups. *American Journal of Orthodontics and Dentofacial Orthopedics, 113*(2), 204-212.

Kelly, A., Winer, K. K., Kalkwarf, H., Oberfield, S. E., Lappe, J., Gilsanz, V., *et al.* (2014). Age-based reference ranges for annual height velocity in US children. *Journal of Clinical Endocrinology and Metabolism, 99*(6), 2104-2112. doi:10.1210/jc.2013-4455

Kent, R. L., Reed, R. B., & Moorrees, C. F. A. (1978). Associations in emergence age among permanent teeth. *American Journal of Physical Anthropology, 48*(2), 131-142. doi:10.1002/ajpa.1330480202

Kim, J. H., Viana, M. A., Graber, T. M., Omerza, F. F., & BeGole, E. A. (1999). The effectiveness of protraction face mask therapy: a meta-analysis. *American Journal of Orthodontics and Dentofacial Orthopedics, 115*(6), 675-685.

Kingsbury, B. F. (1924). The significance of the so-called law of cephalocaudal differential growth. *Anatomical Record, 27*(5), 305-321. doi:DOI 10.1002/ar.1090270511

Kluemper, G. T., & Spalding, P. M. (2001). Realities of craniofacial growth modification. *Atlas of the Oral and Maxillofacial Surgery Clinics of North America, 9*(1), 23-51.

Koretsi, V., Zymperdikas, V. F., Papageorgiou, S. N., & Papadopoulos, M. A. (2014). Treatment effects of removable functional appliances in patients with Class II malocclusion: a

systematic review and meta-analysis. *European Journal of Orthodontics*. doi:10.1093/ejo/cju071

Korn, E. L., & Graubard, B. I. (2011). *Analysis of Health Surveys*: Wiley.

Koroluk, L. D. (2016). Adolescent patients. In S. J. Stefanac & S. P. Nesbit (Eds.), *Diagnosis and Treatment Planning in Dentistry (Third Edition)* (pp. 364-380). St. Louis (MO): Elsevier.

Kozak, M., & Piepho, H. P. (2018). What's normal anyway? Residual plots are more telling than significance tests when checking ANOVA assumptions. *Journal of Agronomy and Crop Science, 204*(1), 86-98. doi:10.1111/jac.12220

Kreft, I., & Leeuw, J. d. (1998). *Introducing multilevel modeling*. London: Sage.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*: Academic Press.

Kuc-Michalska, M., & Baccetti, T. (2010). Duration of the pubertal peak in skeletal Class I and Class III subjects. *Angle Orthodontist, 80*(1), 54-57. doi:10.2319/020309-69.1

Kuhn, M., & Wickham, H. (2019). rsample: General Resampling Infrastructure. R package version 0.0.4. Retrieved from https://CRAN.R-project.org/package=rsample

LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained Variance Measures for Multilevel Models. *Organizational Research Methods, 17*(4), 433-451. doi:10.1177/1094428114541701

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*. doi:10.2307/2529876

Lamparski, D. G. (1975). Skeletal age assessment utilizing cervical vertebrae. *American Journal of Orthodontics, 67*(4), 458-459. doi:10.1016/0002-9416(75)90038-X

Laowansiri, U., Behrents, R. G., Araujo, E., Oliver, D. R., & Buschang, P. H. (2013). Maxillary growth and maturation during infancy and early childhood. *Angle Orthodontist, 83*(4), 563-571. doi:10.2319/071312-580.1

Leckie, G., & Charlton, C. (2013a). ‘runmlwin—A Program to Run the MLwiN Multilevel Modelling Software from within Stata. *Journal of Statistical Software, 52*(11), 1-40.

Leckie, G., & Charlton, C. (2013b). runmlwin: A Program to Run the MLwiN Multilevel Modeling Software from within Stata. *Journal of Statistical Software, 52*(11), 1-40.

Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research, 39*(4), 653-686. doi:DOI 10.1207/s15327906mbr3904_4

Lewis, A. B., Roche, A. F., & Wagner, B. (1982). Growth of the Mandible during Pubescence. *Angle Orthodontist, 52*(4), 325-342.

Lewis, A. B., Roche, A. F., & Wagner, B. (1985). Pubertal spurts in cranial base and mandible: comparisons within individuals. *The Angle orthodontist, 55*(1), 17-30.

Lindstrom, M. J., & Vorperian, H. K. (2005). Nonlinear Mixed Effects Models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science*.

Lindstrom, M. L., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics, 46*(3), 673-687.

Liu, X. (2015). *Methods and Applications of Longitudinal Data Analysis*: Elsevier Science.

Liu, X. (2016). Methods for handling missing data. In X. Liu (Ed.), *Methods and Applications of Longitudinal Data Analysis* (pp. 441-473). Oxford: Academic Press.

Long, J., & Ryoo, J. (2010). Using fractional polynomials to model non-linear trends in longitudinal data. *British Journal of Mathematical and Statistical Psychology*, *63*(1), 177-203. doi:10.1348/000711009X431509

Longford, N. T. (1993). *Random coefficient models / Nicholas T. Longford*. Oxford: Oxford : Clarendon Press.

Lux, C. J., Conradt, C., Burden, D., & Komposch, G. (2003). Dental Arch Widths and Mandibular-Maxillary Base Widths in Class II Malocclusions Between Early Mixed and Permanent

Dentitions. *The Angle Orthodontist, 73*(6), 674-685. doi:10.1043/0003-3219(2003)073<0674:DAWAMB>2.0.CO;2

Mallet, A., Mentre, F., Steimer, J. L., & Lokiec, F. (1988). Nonparametric Maximum-Likelihood Estimation for Population Pharmacokinetics, with Application to Cyclosporine. *Journal of Pharmacokinetics and Biopharmaceutics, 16*(3), 311-327. doi:Doi 10.1007/Bf01062140

Mao, J. J., & Nah, H. D. (2004). Growth and development: hereditary and mechanical modulations. *American Journal of Orthodontics and Dentofacial Orthopedics, 125*(6), 676-689. doi:10.1016/S0889540604001908

Marcano Belisario, J. S., Tudor Car, L., Reeves, T. J., Gunn, L. H., & Car, J. (2013). Search strategies to identify observational studies in MEDLINE and EMBASE. *The Cochrane Library*.

Martins-Junior, P. A., Marques, L. S., & Ramos-Jorge, M. L. (2012). Malocclusion: social, functional and emotional influence on children. *Journal of Clinical Pediatric Dentistry, 37*(1), 103-108.

Masood, Y., Masood, M., Zainul, N. N., Araby, N. B., Hussain, S. F., & Newton, T. (2013). Impact of malocclusion on oral health related quality of life in young people. *Health Qual Life Outcomes, 11*, 25. doi:10.1186/1477-7525-11-25

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305-315. doi:10.1016/j.jml.2017.01.001

McArdle, J. J. (2015). Growth Curve Analysis. In J. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (pp. 441-446). Oxford: Elsevier.

McDonald, J. H. (2014). *Handbook of biological statistics* (3rd ed.). Baltimore, Maryland: Sparky House Publishing.

McNamara, J. A., Jr. (1984). A method of cephalometric evaluation. *American Journal of Orthodontics, 86*(6), 449-469. doi:10.1016/S0002-9416(84)90352-X

McNeish, D. (2016). On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(5), 750-773. doi:10.1080/10705511.2016.1186549

McNeish, D., & Matta, T. (2018). Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behavior Research Methods, 50*(4), 1398-1414. doi:10.3758/s13428-017-0976-5

McNeish, D., & Stapleton, L. M. (2016). Modeling Clustered Data with Very Few Clusters. *Multivariate Behavioral Research, 51*(4), 495-518. doi:10.1080/00273171.2016.1167008

McNeish, D., & Wentzel, K. R. (2017). Accommodating Small Sample Sizes in Three-Level Models When the Third Level is Incidental. *Multivariate Behavioral Research, 52*(2), 200-215. doi:10.1080/00273171.2016.1262236

Mehta, P. D., & Neale, M. C. (2005). People are variables too: multilevel structural equations modeling. *Psychological Methods, 10*(3), 259-284. doi:10.1037/1082-989x.10.3.259

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

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107-122. doi:10.1007/BF02294746

Miller, A. J. (2002). *Subset selection in regression* (2nd ed.). Boca Raton: Chapman & Hall/CRC.

Mitani, H. (1977). Occlusal and craniofacial growth changes during puberty. *American Journal of Orthodontics, 72*(1), 76-84. doi:http://dx.doi.org/10.1016/0002-9416(77)90126-9

Mitchell, L. (2013). *An Introduction to Orthodontics* (4th ed.): Oxford University Press.

Moehring, K., & Schmidt-Catran, A. (2013). MLT: Stata module to provide multilevel tools. Boston College Department of Economics: Statistical Software Components S457577. Retrieved from https://EconPapers.repec.org/RePEc:boc:bocode:s457577

Molinari, L., & Gasser, T. (2004). The human growth curve: distance, velocity and acceleration. In R. C. Hauspie, N. Cameron, & L. Molinari (Eds.), *Methods in Human Growth Research* (pp. 27-54). Cambridge UK.

Montasser, M. A. (2019). Craniofacial growth spurt in Class I subjects. *American Journal of Orthodontics and Dentofacial Orthopedics, 155*(4), 473-481. doi:10.1016/j.ajodo.2018.05.013

Moore, R. N., Moyer, B. A., & DuBois, L. M. (1990). Skeletal maturation and craniofacial growth. *American Journal of Orthodontics and Dentofacial Orthopedics, 98*(1), 33-40. doi:http://dx.doi.org/10.1016/0889-5406(90)70029-C

Morgan, G. W. (1987). *Statistical models for biological growth curves.* Sheffield Hallam University, Retrieved from http://shura.shu.ac.uk/20082/1/10697389.pdf

Motulsky, H., & Christopoulos, A. (2004). *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford: Oxford University Press.

Moyers, R. E., Growth, U. o. M. C. f. H., & Development. (1976). *Standards of human occlusal development*: Center for Human Growth and Development, Univ. of Michigan.

Muller, H. G., & Yao, F. (2010). Empirical Dynamics for Longitudinal Data. *Annals of Statistics, 38*(6), 3458-3486. doi:10.1214/09-Aos786

Musekiwa, A., Manda, S. O. M., Mwambi, H. G., & Chen, D.-G. (2016). Meta-Analysis of Effect Sizes Reported at Multiple Time Points Using General Linear Mixed Model. *PloS One, 11*(10), e0164898. doi:10.1371/journal.pone.0164898

Nahhas, R. W., Valiathan, M., & Sherwood, R. J. (2014). Variation in timing, duration, intensity, and direction of adolescent growth in the mandible, maxilla, and cranial base: the Fels longitudinal study. *Anat Rec (Hoboken), 297*(7), 1195-1207. doi:10.1002/ar.22918

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133-142. doi:10.1111/j.2041-210x.2012.00261.x

Nanda, R. S. (1955). The rates of growth of several facial components measured from serial cephalometric roentgenograms. *American Journal of Orthodontics, 41*(9), 658-673. doi:http://dx.doi.org/10.1016/0002-9416(55)90112-3

Nanda, R. S. (1971). Growth Changes in Skeletal-Facial Profile and Their Significance in Orthodontic Diagnosis. *American Journal of Orthodontics, 59*(5), 501-513. doi:Doi 10.1016/0002-9416(71)90085-6

Nanda, S. K. (1983). *The developmental basis of occlusion and malocclusion*: Quintessence Pub. Co.

Nanda, S. K. (1988). Patterns of vertical growth in the face. *American Journal of Orthodontics and Dentofacial Orthopedics, 93*(2), 103-116. doi:http://dx.doi.org/10.1016/0889-5406(88)90287-9

Nanda, S. K. (1992). Differential growth of the female face in the anteroposterior dimension. *The Angle Orthodontist, 62*(1), 23-34.

Nanda, S. K., & Rowe, T. K. (1989). Circumpubertal growth spurt related to vertical dysplasia. *Angle Orthodontist, 59*(2), 113-122. doi:10.1043/0003-3219(1989)059<0113:cgsrtv>2.0.co;2

Nowak, A. J., Christensen, J. R., Mabry, T. R., Townsend, J. A., & Wells, M. (2019). *Pediatric dentistry : infancy through adolescence* (Sixth edition. ed.). Philadelphia, PA: Elsevier.

Ochoa, B. K., & Nanda, R. S. (2004). Comparison of maxillary and mandibular growth. *American Journal of Orthodontics and Dentofacial Orthopedics, 125*(2), 148-159. doi:http://dx.doi.org/10.1016/j.ajodo.2003.03.008

Ong, M. A. (2004). Spectrum of dentofacial deformities: a retrospective survey. *Annals of the Academy of Medicine, Singapore, 33*(2), 239-242.

Oravecz, Z., & Muth, C. (2017). Fitting growth curve models in the Bayesian framework. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-017-1281-0

Orsini, N., & Greenland, S. (2011). A procedure to tabulate and plot results after flexible modeling of a quantitative covariate. *Stata Journal, 11*(1), 1-29.

Perinetti, G., & Contardo, L. (2017). Reliability of Growth Indicators and Efficiency of Functional Treatment for Skeletal Class II Malocclusion: Current Evidence and Controversies. *Biomed Res Int, 2017*, 1367691. doi:10.1155/2017/1367691

Perinetti, G., Primozic, J., Franchi, L., & Contardo, L. (2015). Treatment Effects of Removable Functional Appliances in Pre-Pubertal and Pubertal Class II Patients: A Systematic Review and Meta-Analysis of Controlled Studies. *PloS One, 10*(10), e0141198. doi:10.1371/journal.pone.0141198

Pileski, R. C., Woodside, D. G., & James, G. A. (1973). Relationship of the ulnar sesamoid bone and maximum mandibular growth velocity. *Angle Orthodontist, 43*(2), 162-170. doi:10.1043/0003-3219(1973)043<0162:ROTUSB>2.0.CO;2

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2017). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-134.

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2018). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-137.

Pinheiro, J. C., & Bates, D. M. (1995a). Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics, 4*(1), 12-35. doi:10.2307/1390625

Pinheiro, J. C., & Bates, D. M. (1995b). Model Building for Nonlinear Mixed-Effects Models. Retrieved from https://www.biostat.wisc.edu/content/model-building-nonlinear-mixed-effects-models

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Prahl-Andersen B, & CJ., K. (1979). *A mixed-longitudinal interdisciplinary study of growth and development*. New York: New York : Academic Press.

Prahl-Andersen, B., & Kowalski, C. J. (1973). A mixed longitudinal, interdisciplinary study of the growth and development of Dutch children. *Growth, 37*(3), 281-295.

Preece, M. A., & Baines, M. J. (1978). A new family of mathematical models describing the human growth curve. *Annals of Human Biology, 5*(1), 1-24.

Preedy, V. R. (2011). *Handbook of Growth and Growth Monitoring in Health and Disease*: Springer New York.

Proffit, W. R. (2006). The timing of early treatment: An overview. *American Journal of Orthodontics and Dentofacial Orthopedics, 129*(4), S47-S49. doi:10.1016/j.ajodo.2005.09.014

Proffit, W. R. (2014). Concepts of Growth and Development. In W. R. Proffit, H. W. Fields, & D. M. Sarver (Eds.), *Contemporary Orthodontics* (5th ed., pp. 20-65): Elsevier Health Sciences.

Proffit, W. R., Fields, H. W., & Sarver, D. M. (2014). *Contemporary Orthodontics* (5th ed.): Elsevier Health Sciences.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing. R version 3.1.3. Retrieved from https://www.R-project.org/

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing. R version 3.4.3. Retrieved from https://www.R-project.org/

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). New York: Springer.

Rasbash, J., Steele, F., Browne, W., & Goldstein, H. (2016). A User's Guide to MLwiN, Version 2.36. Centre for Multilevel Modelling, University of Bristol. In.

Rauber, A. (1990). Growth and Development. In H. K. Walker, W. D. Hall, & J. W. Hurst (Eds.), *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd ed.): Boston: Butterworths.

Raudenbush, S. W. (2001). Toward a coherent framework for comparing trajectories of individual change. In L. M. C. A. G. Sayer (Ed.), *New methods for the analysis of change* (pp. 35-64). Washington, DC, US: American Psychological Association.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.

Riddell, C. A., Platt, R. W., Bodnar, L. M., & Hutcheon, J. A. (2017). Classifying Gestational Weight Gain Trajectories Using the SITAR Growth Model. *Paediatric and Perinatal Epidemiology, 31*(2), 116-125. doi:10.1111/ppe.12336

Riolo, M. L. (1974). *An Atlas of craniofacial growth: cephalometric standards from the University School Growth Study, the University of Michigan*. Ann Arbor, Mich.]: Ann Arbor, Mich. : Center for Human Growth and Development. University of Michigan.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14*(5), 465-471. doi:10.1016/0005-1098(78)90005-5

Roche, A. F. (1992). *Growth, Maturation, and Body Composition: The Fels Longitudinal Study 1929–1991*. Cambridge: Cambridge University Press.

Rogol, A. D., Clark, P. A., & Roemmich, J. N. (2000). Growth and pubertal development in children and adolescents: effects of diet and physical activity. *The American Journal of Clinical Nutrition, 72*(2), 521s-528s.

Royston, P. (2017). Model selection for univariable fractional polynomials. *Stata J, 17*(3), 619-629.

Royston, P., & Altman, D. G. (1994). Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 43*(3), 429-467. doi:10.2307/2986270

Salazar-Lazo, R., Arriola-Guillen, L. E., & Flores-Mir, C. (2014). Duration of the peak of adolescent growth spurt in class i and ii malocclusion subjects using a cervical vertebrae maturation analysis. *Acta Odontologica Latinoamericana, 27*(2), 96-101. doi:10.1590/S1852-48342014000200009

San Roman, P., Palma, J. C., Oteo, M. D., & Nevado, E. (2002). Skeletal maturation determined by cervical vertebrae development. *European Journal of Orthodontics, 24*(3), 303-311.

Sanders, J. O., Browne, R. H., McConnell, S. J., Margraf, S. A., Cooney, T. E., & Finegold, D. N. (2007). Maturity assessment and curve progression in girls with idiopathic scoliosis. *Journal of Bone and Joint Surgery (American Volume), 89*(1), 64-73. doi:10.2106/JBJS.F.00067

Sanders, J. O., Qiu, X., Lu, X., Duren, D. L., Liu, R. W., Dang, D., *et al.* (2017). The Uniform Pattern of Growth and Skeletal Maturation during the Human Adolescent Growth Spurt. *Scientific Reports, 7*. doi:https://doi.org/10.1038/s41598-017-16996-w

Sandhu, S. S., Cheema, M. S., & Khehra, H. S. (2016). Comparative effectiveness of pharmacologic and nonpharmacologic interventions for orthodontic pain relief at peak pain intensity: A Bayesian network meta-analysis. *American Journal of Orthodontics and Dentofacial Orthopedics, 150*(1), 13-32. doi:10.1016/j.ajodo.2015.12.025

SAS Institute Inc. (2015). *SAS/STAT® 14.1 User's Guide*. Cary, NC: SAS Institute Inc.

Sassouni, V. (1969). A classification of skeletal facial types. *American Journal of Orthodontics, 55*(2), 109-123. doi:http://dx.doi.org/10.1016/0002-9416(69)90122-5

Savara, B. S., & Steen, J. C. (1978). Timing and sequence of eruption of permanent teeth in a longitudinal sample of children from Oregon. *The Journal of the American Dental Association, 97*(2), 209-214. doi:http://dx.doi.org/10.14219/jada.archive.1978.0277

Savara, B. S., & Tracy, W. E. (1967). Norms of size and annual increments for five anatomical measures of the mandible in boys from three to sixteen years of age. *Archives of Oral Biology, 12*(4), 469-486.

Sayers, A., Baines, M., & Tilling, K. (2013). A new family of mathematical models describing the human growth curve—Erratum: Direct calculation of peak height velocity, age at take-off and associated quantities. *Annals of Human Biology, 40*(3), 298-299. doi:10.3109/03014460.2013.772655

Scammon, R. E. (1930). The Measurement of the Body in Childhood. In Harris J A, Jackson C M, Paterson D G, & Scammon R E (Eds.), *The measurement of man* (pp. 173–215). Minniapolis: University of Minnesota Press.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and Analgesia, 126*(5), 1763-1768. doi:10.1213/Ane.0000000000002864

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics, 6*(2), 461-464.

Sclove, S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333-343. doi:10.1007/BF02294360

Searle, S., Casella, G., & McCulloch, C. (2006). *Variance components*. New Jersey: John Wiley & Sons.

Serroyen, J., Molenberghs, G., Verbeke, G., & Davidian, M. (2009). Non-linear Models for Longitudinal Data. *Am Stat, 63*(4), 378-388. doi:10.1198/tast.2009.07256

Shadish, W. R., Brasil, I. C., Illingworth, D. A., White, K. D., Galindo, R., Nagler, E. D., *et al.* (2009). Using UnGraph to extract data from image files: verification of reliability and validity. *Behavior Research Methods, 41*(1), 177-183. doi:10.3758/brm.41.1.177

Sheiner, L. B., & Beal, S. L. (1980). Evaluation of Methods for Estimating Population Pharmacokinetic Parameters .1. Michaelis-Menten Model - Routine Clinical Pharmacokinetic Data. *Journal of Pharmacokinetics and Biopharmaceutics, 8*(6), 553-571. doi:Doi 10.1007/Bf01060053

Sherar, L. B., Baxter-Jones, A. D. G., & Mirwald, R. L. (2004). Limitations to the use of secondary sex characteristics for gender comparisons. *Annals of Human Biology, 31*(5), 586-593. doi:10.1080/03014460400001222

Shim, J. J., Bogowicz, P., Heo, G., & Lagravere, M. O. (2012). Interrelationship and limitations of conventional radiographic assessments of skeletal maturation. *International Orthodontics, 10*(2), 135-147. doi:10.1016/j.ortho.2012.03.007

Simpkin, A. J., Durban, M., Lawlor, D. A., MacDonald-Wallis, C., May, M. T., Metcalfe, C., *et al.* (2018). Derivative estimation for longitudinal data analysis: Examining features of blood pressure measured repeatedly during pregnancy. *Statistics in Medicine*. doi:10.1002/sim.7694

Simpkin, A. J., Sayers, A., Gilthorpe, M. S., Heron, J., & Tilling, K. (2017). Modelling height in adolescence: a comparison of methods for estimating the age at peak height velocity. *Annals of Human Biology*, 1-8. doi:10.1080/03014460.2017.1391877

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*: Oxford university press.

Singh, I. J., & Savara, B. S. (1966). Norms of size and annual increments of seven anatomical measures of maxillae in girls from three to sixteen years of age. *Angle Orthodontist, 36*(4), 312-324. doi:10.1043/0003-3219(1966)036<0312:nosaai>2.0.co;2

Smith, P. L. (1979). Splines as a Useful and Convenient Statistical Tool. *American Statistician, 33*(2), 57-62. doi:Doi 10.2307/2683222

Smith, S. L., & Buschang, P. H. (2005). Longitudinal models of long bone growth during adolescence. *American Journal of Human Biology, 17*(6), 731-745. doi:10.1002/ajhb.20441

Snijders, T. A. B., & Bosker, R. J. (1994). Modeled Variance in Two-Level Models. *Sociological Methods & Research, 22*(3), 342-363. doi:10.1177/0049124194022003004

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd Ed.): SAGE Publications.

Stagi, S., Cavalli, L., Iurato, C., Seminara, S., Brandi, M. L., & de Martino, M. (2013). Bone metabolism in children and adolescents: main characteristics of the determinants of peak bone mass. *Clinical Cases in Mineral and Bone Metabolism, 10*(3), 172-179.

StataCorp. (2015a). *STATA Programming Reference Manual Release 14*. College Station, TX: StataCorp, LP.

StataCorp. (2015b). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

StataCorp. (2017a). *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LP.

StataCorp. (2017b). *Stata user's guide: release 15*. College Station, Texas: Stata Press.

Steele, R. (2013). Model Selection for Multilevel Models. In *The SAGE Handbook of Multilevel Modeling* (pp. 109-126). London: SAGE Publications Ltd.

Stegmueller, D. (2013). How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches. *American Journal of Political Science, 57*(3), 748-761. doi:10.1111/ajps.12001

Stone, C. J. (1986). Generalized Additive Models: Comment. *Statist. Sci., 1*(3), 312-314. doi:10.1214/ss/1177013607

Stone, C. J., & Koo, C.-Y. (1985). Additive splines in statistics. *Statistical Computing Section, Proceedings of the American Statistical Association.*, 45-58.

Strang, G., & Herman, E. J. (2016a). *Calculus Volume 1*. In Gilbert Strang & E. J. Herman (Eds.). Retrieved from https://d3bxy9euw4e147.cloudfront.net/oscms-prodcms/media/documents/CalculusVolume1-OP.pdf

Strang, G., & Herman, E. J. (2016b). *Calculus Volume 2*. In Gilbert Strang & E. J. Herman (Eds.). Retrieved from https://d3bxy9euw4e147.cloudfront.net/oscms-prodcms/media/documents/CalculusVolume2-OP_7nNwGJD.pdf

Stulp, G., & Barrett, L. (2016). Evolutionary perspectives on human height variation. *Biological Reviews, 91*(1), 206-234. doi:10.1111/brv.12165

Tanner, J. M. (1951). Some notes on the reporting of growth data. *Human Biology, 23*(2), 93-159.

Tanner, J. M. (1981). Growth and Maturation during Adolescence. *Nutrition Reviews, 39*(2), 43-55. doi:10.1111/j.1753-4887.1981.tb06734.x

Tanner, J. M. (1981). *A history of the study of human growth / J.M. Tanner*. Cambridge: Cambridge : Cambridge University Press.

Tanner, J. M., & Cameron, N. (1980). Investigation of the Mid-Growth Spurt in Height, Weight and Limb Circumferences in Single-Year Velocity Data from the London 1966-67 Growth Survey. *Annals of Human Biology, 7*(6), 565-577. doi:Doi 10.1080/03014468000004681

Taranger, J., & Hagg, U. (1980). The Timing and Duration of Adolescent Growth. *Acta Odontologica Scandinavica, 38*(1), 57-67. doi:Doi 10.3109/00016358008997719

Thai, H. T., Mentre, F., Holford, N. H. G., Veyrat-Follet, C., & Comets, E. (2013). A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. *Pharmaceutical Statistics, 12*(3), 129-140. doi:10.1002/pst.1561

The Craniofacial Research Instrumentation Laboratory. (2014). The Mathews Acquisition Group. Retrieved from http://www.cril.org/cril_documents/acqDescription_mathews.htm

Thiesen, G., Pletsch, G., Zastrow, M. D., Valle, C. V. M. d., Valle-Corotti, K. M. d., Patel, M. P., *et al.* (2013). Comparative analysis of the anterior and posterior length and deflection angle

of the cranial base, in individuals with facial Pattern I, II and III. *Dental Press Journal of Orthodontics, 18*, 69-75.

Thilander, B. (1995). Basic mechanisms in craniofacial growth. *Acta Odontologica Scandinavica, 53*(3), 144-151.

Thilander, B., Persson, M., & Adolfsson, U. (2005). Roentgen-cephalometric standards for a Swedish population. A longitudinal study between the ages of 5 and 31 years. *European Journal of Orthodontics, 27*(4), 370-389. doi:10.1093/ejo/cji033

Thiruvenkatachari, B., Harrison, J. E., Worthington, H. V., & O'Brien, K. D. (2013). Orthodontic treatment for prominent upper front teeth (Class II malocclusion) in children. *Cochrane Database Syst Rev*(11), CD003452. doi:10.1002/14651858.CD003452.pub3

Thompson, G. W., & Popovich, F. (1977). A Longitudinal Evaluation of the Burlington Growth Centre Data. *Journal of Dental Research, 56*(3 suppl), C71-C78. doi:10.1177/00220345770560032101l

Thompson, G. W., Popovich, F., & Anderson, D. L. (1976). Maximum Growth Changes in Mandibular Length, Stature and Weight. *Human Biology, 48*(2), 285-293.

Tilling, K., Macdonald-Wallis, C., Lawlor, D. A., Hughes, R. A., & Howe, L. D. (2014). Modelling Childhood Growth Using Fractional Polynomials and Linear Splines. *Annals of Nutrition and Metabolism, 65*(3), 129-138. doi:10.1159/000362695

Tracy, W. E., & Savara, B. S. (1966). Norms of size and annual increments of five anatomical measures of the mandible in girls from 3 to 16 years of age. *Archives of Oral Biology, 11*(6), 587-598.

Tracy, W. E., Savara, B. S., & Brant, J. W. (1965). Relation of height, width and depth of the mandible. *Angle Orthodontist, 35*(4), 269-277. doi:10.1043/0003-3219(1965)035<0269:ROHWAD>2.0.CO;2

Trikalinos, T. A., & Olkin, I. (2012). Meta-analysis of effect sizes reported at multiple time points: a multivariate approach. *Clinical Trials (London, England), 9*(5), 610-620. doi:10.1177/1740774512453218

Turpin, D. L. (2000). Good time for discussion of early treatment. *American Journal of Orthodontics and Dentofacial Orthopedics, 118*(3), 247-247. doi:DOI 10.1067/mod.2000.110506

Ulijaszek, S. J. (2010). Variation in Human Growth Patterns due to Environmental Factors. In M. P. Muehlenbein (Ed.), *Human Evolutionary Biology* (pp. 396-404). Cambridge: Cambridge University Press.

University of Texas School of Public Health. (2015, 10 October 2015). Search filters for case-control studies, cohort studies, cross-sectional studies, clinical trials, epidemiological studies. Retrieved from http://libguides.sph.uth.tmc.edu/search_filters/ovid_medline_filters

Valeggia, C. R., & Núñez-de la Mora, A. (2015). Chapter 21 - Human Reproductive Ecology. In M. P. Muehlenbein (Ed.), *Basics in Human Evolution* (pp. 295-308). Boston: Academic Press.

Vallejo, G., Ato, M., & Valdés, T. (2008). Consequences of Misspecifying the Error Covariance Structure in Linear Mixed Models for Longitudinal Data. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 4*(1), 10-21. doi:10.1027/1016-9040.12.1.10

Vallejo, G., Fernandez, M. P., Livacic-Rojas, P. E., & Tuero-Herrero, E. (2011). Selecting the best unbalanced repeated measures model. *Behavior Research Methods, 43*(1), 18-36. doi:10.3758/s13428-010-0040-1

Vallejo, G., Tuero-Herrero, E., Núñez, J. C., & Rosário, P. (2014). Performance evaluation of recent information criteria for selecting multilevel models in Behavioral and Social Sciences. *International Journal of Clinical and Health Psychology, 14*(1), 48-57. doi:10.1016/S1697-2600(14)70036-5

van der Beek, M. C., Hoeksma, J. B., & Prahl-Andersen, B. (1996). Vertical facial growth and statural growth in girls: a longitudinal comparison. *European Journal of Orthodontics, 18*(6), 549-555.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.

Verzelen, N., Tao, W. W., & Muller, H. G. (2012). Inferring stochastic dynamics from functional data. *Biometrika, 99*(3), 533-550. doi:10.1093/biomet/ass015

Vonesh, E., & Chinchilli, V. M. (1996). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*: CRC Press.

Vonesh, E. F. (2012). *Generalized linear and nonlinear models for correlated data: theory and applications using SAS*. Cary, NC, USA: SAS Institute Inc.

Vonesh, E. F., & Carter, R. L. (1992). Mixed-Effects Nonlinear-Regression for Unbalanced Repeated Measures. *Biometrics, 48*(1), 1-17. doi:Doi 10.2307/2532734

Vonesh, E. F., Chinchilli, V. M., & Pu, K. (1996). Goodness-of-Fit in Generalized Nonlinear Mixed-Effects Models. *Biometrics, 52*(2), 572-587. doi:10.2307/2532896

Waldo, C. M. (1936). Orthodontic research as a component part of a balanced longitudinal study of 100 children. *International Journal of Orthodontia and Oral Surgery, 22*(7), 659-673. doi:http://dx.doi.org/10.1016/S1072-3498(36)80115-0

Wang, J., & Schaalje, G. B. (2009). Model Selection for Linear Mixed Models Using Predictive Criteria. *Communications in Statistics - Simulation and Computation, 38*(4), 788-801. doi:10.1080/03610910802645362

Ware, J. H. (1985). Linear-Models for the Analysis of Longitudinal-Studies. *American Statistician, 39*(2), 95-101. doi:Doi 10.2307/2682803

Weir, C. J., Butcher, I., Assi, V., Lewis, S. C., Murray, G. D., Langhorne, P., *et al.* (2018). Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes:

a systematic review. *BMC Medical Research Methodology, 18*(1), 25. doi:10.1186/s12874-018-0483-0

Wells, J. C. (2017). Worldwide variability in growth and its association with health: Incorporating body composition, developmental plasticity, and intergenerational effects. *American Journal of Human Biology, 29*(2). doi:10.1002/ajhb.22954

Wheeler, T. T., McGorray, S. P., Dolce, C., & King, G. J. (2006). The timing of Class II treatment. *American Journal of Orthodontics and Dentofacial Orthopedics, 129*(4), S66-S70. doi:10.1016/j.ajodo.2005.09.015

White, I. R. (2010). simsum: Analyses of simulation studies including Monte Carlo error. *Stata Journal, 10*(3), 369-385.

Wishart, J. (1938). Growth-Rate Determinations in Nutrition Studies with the Bacon Pig, and Their Analysis. *Biometrika, 30*(1/2), 16-28. doi:10.2307/2332221

Wold, S. (1974). Spline Functions in Data-Analysis. *Technometrics, 16*(1), 1-11. doi:Doi 10.2307/1267485

Woodside, D. G. (1968). *Distance, velocity and relative growth rate standards for mandibular growth for Canadian males and females aged three to twenty years*: Department of Orthodontics, Faculty of Dentistry, University of Toronto.

Wu, W., & West, S. G. (2013). Detecting Misspecification in Mean Structures for Growth Curve Models: Performance of Pseudo R 2 s and Concordance Correlation Coefficients. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(3), 455-478. doi:10.1080/10705511.2013.797829

Wu, W., West, S. G., & Taylor, A. B. (2009). Evaluating Model Fit for Growth Curve Models: Integration of Fit Indices from SEM and MLM Frameworks. *Psychological Methods, 14*(3), 183-201. doi:10.1037/a0015858

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine, 22*(22), 3527-3541. doi:10.1002/sim.1572

Yao, F., Muller, H. G., & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association, 100*(470), 577-590. doi:10.1198/016214504000001745

Yuan, K.-H., Zhang, Z., & Deng, L. (2019). Fit indices for mean structures with growth curve models. *Psychological Methods, 24*(1), 36-53. doi:10.1037/met0000186

Zere, E., Chaudhari, P. K., Sharan, J., Dhingra, K., & Tiwari, N. (2018). Developing Class III malocclusions: challenges and solutions. *Clinical Cosmetic and Investigational Dentistry, 10*, 99-116. doi:10.2147/Ccide.S134303

APPENDIX A. GROWTH CURVE MODEL SELECTION (CHAPTER 5)

A.1. Methodology

Table A-1. Population parameters used in simulation study.

Population parameters	Linear–linear data generation model	Quadratic-linear data generation model	Quadratic-quadratic data generation model
Fixed effect parameters			
Intercept	74.00	68.71	69.65
Linear slope	1.27	2.28	2.09
Quadratic slope		-0.04	-0.03
Random effect parameters			
Intercept variance	37.98	41.63	46.04
Intercept - Linear slope covariance	-0.85	-0.84	-3.81
Linear slope variance	0.19	0.15	1.02
Intercept - Quadratic slope covariance			0.19
Linear slope - Quadratic slope covariance			-0.05
Quadratic slope variance			0.02
Residual variance			
Residual variance	4.48	4.14	3.77

Table A-2. Summary of simulations.

Condition number [a]	Number of Simulations	Data generation model	Balanced data					Unbalanced data [b]				
			Level-2 sample size	Level-1 sample size [c]			N	Level-2 sample size	Level-1 sample size [c]			N [d]
				Avg.	Min	Max			Avg.	Min	Max	
1	500	Linear-linear	1000	20	20	20	20,000	1000	16.0	9.5	20	16009
2	500	Quadratic-linear	1000	20	20	20	20,000	1000	16.0	9	20	15992
3	500	Quadratic-quadratic	1000	20	20	20	20,000	1000	16.0	10	20	16024
4	500	Linear-linear	1000	5	5	5	5,000	1000	4.0	2	5	4004
5	500	Quadratic-linear	1000	5	5	5	5,000	1000	4.0	2	5	4018
6	500	Quadratic-quadratic	1000	5	5	5	5,000	1000	4.0	2	5	3997
7	500	Linear-linear	100	20	20	20	2,000	100	16.1	11	20	1605
8	500	Quadratic-linear	100	20	20	20	2,000	100	16.0	11	20	1604
9	500	Quadratic-quadratic	100	20	20	20	2,000	100	16.1	11	20	1606
10	500	Linear-linear	100	5	5	5	500	100	4.0	2	5	403
11	500	Quadratic-linear	100	5	5	5	500	100	4.0	2	5	405
12	500	Quadratic-quadratic	100	5	5	5	500	100	4.0	2	5	398

[a] Total number of conditions is 24; 12 for the balanced data and 12 for the unbalanced data.
[b] 20% repeated measurements were randomly dropped from within the level-2 units. Thus, data were unbalanced with respect to number of repeated measurements per individuals and time intervals between the measurements as the level-2 sample size was same for balanced and unbalanced data.
[c] Average, minimum and maximum (across 500 simulations) number of repeated measurements per individual.
[d] Total number of observations

A.2. Software

To calculate different versions of information criteria and the prediction criteria included in the simulation study (Chapter 5), I wrote a new Stata program ('gcmfit'). The need to write a new program and working of the 'gcmfit' is described below.

A.2.1. Information criteria

Stata calculates Akaike's information criterion (AIC) and Bayesian information criterion (BIC) using the total number of parameters, and by default, includes the total number of observations when calculating BIC (StataCorp, 2017b). Although user can specify a different sample size for BIC (such as the level-2 sample size or the total number of observations), there is no option to change the number of parameters used in the penalty terms of AIC or the BIC. The 'gcmfit' allows inclusion of the total number of parameters or the number of variance covariance parameters into the penalty terms of AIC and the BIC. For BIC, user can specify the total number of observations or the level-2 sample size.

Though I focused on AIC and BIC in the simulation study (see Chapter 5, Section 5.2.2), the 'gcmfit' also computes other information criteria which are often used for model selection. These are : i) Akaike's information criterion corrected for small sample bias, AICC (Burnham, 2002; Hurvich & Tsai, 1989), ii) consistent Akaike's information criterion, CAIC (Bozdogan, 1987), iii) adjusted Bayesian information criterion ABIC (Rissanen, 1978; Sclove, 1987), and iv) Hannan and Quinn information criterion, HQIC (Hannan & Quinn, 1979). None of these four information criteria (AICC, CAIC, ABIC and HQIC) are computed by the Stata software.

A.2.2. Prediction criteria

No Stata package was available to calculate the prediction criteria for GCMs included in this study. The 'mlt' package written by Moehring and Schmidt-Catran (2013) provides the 'mltrsq'

command to compute variance explained by the models (R^2 statistics) but is restricted to computing R^2 statistics proposed by Bryk and Raudenbush (1992) and Snijders and Bosker (1994). Furthermore, Snijders and Bosker (1994) provide formulation to include random slopes when computing R^2 statistics but 'mltrsq' excludes random slope even when fitted model includes random slope(s).

The 'gcmfit' program computes 10 different prediction criteria which include concordance correlation coefficient (CCC) proposed by Vonesh et al. (1996) and nine R^2 statistics introduced by i) Bryk and Raudenbush (1992), ii) Snijders and Bosker (1994), both random intercept and random slope versions, iii) Vonesh et al. (1996), iv) Xu (2003), v) Edwards et al. (2008), vi) Snijders and Bosker (2012), vii) Nakagawa and Schielzeth (2013), viii) Johnson (2014), and ix) LaHuis et al. (2014).

Out of these 10 prediction criteria (CCC and nine R^2 statistics), I calculated four prediction criteria which include CCC by Vonesh et al. (1996) and three R^2 statistics proposed by Vonesh et al. (1996), Nakagawa and Schielzeth (2013), and Johnson (2014). See Chapter 5 (Section 5.2.2) for reasons behind selecting these four prediction criteria.

A.3. Results

Here I present the performance results for the six data analysis models fitted to the unbalanced data simulated using the linear-linear data generation model (Table A-3), quadratic-linear data generation model (Table A-4), and quadratic-quadratic data generation model (Table A-5). Results for balanced data are reported in Chapter 5 (Section 5.3.1).

Results are displayed as per centage (proportion x 100) of simulated unbalanced datasets in which each fit criterion selected each data analysis model. Out of a total six data analysis models fitted to the data, the true models are linear-linear, quadratic-linear and quadratic-quadratic for Table A-3, Table A-4 and Table A-5, respectively.

Table A-3. Results for unbalanced data generated using the linear-linear model.

Level 2 [a]	Level 1 [b]	Data analysis models	Information criteria						Prediction criteria			
			AICk	AICq	BICkN	BICqN	BICkL2	BICqL2	$CCC_{VC[c]}$	$R^2_{VC[c]}$	$R^2_{JNS[c]}$	$R^2_{NS[c]}$
			(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
1000	20	Linear-linear	**72.8**	0.4	**100**	*0*	**99**	*0*	29.6	29.4	30.4	*0.6*
1000	20	Quadratic-linear	12.6	2.2	0.2	3.4	1	2.4	10.2	10.8	9	4.2
1000	20	Quadratic-quadratic	6.2	7.8	*0*	*0*	*0*	*0*	**47.4**	**47**	**48.4**	9
1000	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	3.6
1000	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**46**
1000	20	Cubic-linear	8.4	**89.6**	*0*	**96.6**	*0*	**97.6**	12.8	12.8	12.2	36.6
1000	5	Linear-linear	**76.2**	*0*	**98.6**	0.2	**98**	*0*	21.2	21.2	24.6	*1.2*
1000	5	Quadratic-linear	13	3.6	*0*	2.4	0.6	3.4	12.6	12.6	11	3
1000	5	Quadratic-quadratic	6.2	7.6	*0*	*0*	*0*	*0*	**51.2**	**51.2**	**52.8**	9.6
1000	5	Quadratic-intercept	*0*	*0*	1.4	0.4	1.2	0.2	*0*	*0*	*0*	4.6
1000	5	Cubic-intercept	*0*	0.2	*0*	11.4	*0*	6.2	*0*	*0*	*0*	**44**
1000	5	Cubic-linear	4.6	**88.6**	*0*	**85.6**	0.2	**90.2**	15	15	11.6	37.6
100	20	Linear-linear	**74**	*0*	**99.2**	*0*	**95.4**	*0*	32.4	33.2	33.6	*0.2*
100	20	Quadratic-linear	12.2	1.8	0.4	2.2	3.4	2	12.4	12.8	12	2.6
100	20	Quadratic-quadratic	6.2	7.2	0.4	0.4	0.4	0.4	**37.8**	**36.4**	**39.6**	7.4
100	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	3.2
100	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**49.4**
100	20	Cubic-linear	7.6	**91**	*0*	**97.4**	0.8	**97.6**	17.4	17.6	14.8	37.2
100	5	Linear-linear	**43.4**	*0.4*	20.8	*0*	29	0.2	28.2	28.4	28.8	*0.6*
100	5	Quadratic-linear	7.2	2.4	0.2	*0*	0.4	0.8	14.2	14	13.6	2.2
100	5	Quadratic-quadratic	5.2	5.6	*0*	*0*	*0*	*0*	**34**	**33.6**	**37**	9.2
100	5	Quadratic-intercept	33.6	5.2	**77.4**	9	**68.2**	8.2	5.2	6	5.2	5.8
100	5	Cubic-intercept	6.4	**49.6**	1.2	**85**	2	**78.8**	*3.2*	*3*	*2.2*	**45.4**
100	5	Cubic-linear	*3.8*	36.4	*0*	5.6	*0*	11.6	15.2	15	13.2	36.8

Note: Highest and lowest values are highlighted using bold and italics fonts, respectively.
[a] Level-2: Number of individuals; [b] Level-1: Average number of repeated measurements per individual.

Table A-4. Results for unbalanced data generated using the quadratic-linear model.

Level 2 [a]	Level 1 [b]	Data analysis models	Information criteria						Prediction criteria			
			AICk	AICq	BICkN	BICqN	BICkL2	BICqL2	$CCC_{VC[c]}$	$R^2_{VC[c]}$	$R^2_{JNS[c]}$	$R^2_{NS[c]}$
			(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
1000	20	Linear-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
1000	20	Quadratic-linear	**75.6**	1	**99.8**	2.2	**99.4**	1.4	37	37.6	37.4	4.2
1000	20	Quadratic-quadratic	9.6	8.4	*0*	*0*	*0*	*0*	**45**	**44.6**	**47**	9
1000	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	4.2
1000	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**44.6**
1000	20	Cubic-linear	14.8	**90.6**	0.2	**97.8**	0.6	**98.6**	18	17.8	15.6	38
1000	5	Linear-linear	**48.2**	*0*	**91.4**	*0*	**90.4**	*0*	12.8	13.2	14.6	*0.4*
1000	5	Quadratic-linear	32	1.6	4.2	1.4	6.8	1.4	19.6	19.8	20.2	2.6
1000	5	Quadratic-quadratic	9.6	9	*0*	*0*	*0*	*0*	**48.2**	**48.2**	**49**	8.4
1000	5	Quadratic-intercept	*0*	*0*	4.4	*0*	2.8	*0*	*0*	*0*	*0*	4
1000	5	Cubic-intercept	*0*	0.2	*0*	13	*0*	7.8	*0*	*0*	*0*	**46**
1000	5	Cubic-linear	10.2	**89.2**	*0*	**85.6**	*0*	**90.8**	19.4	18.8	16.2	38.6
100	20	Linear-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
100	20	Quadratic-linear	**75.4**	1.6	**99**	2.2	**96.6**	2	**41.8**	**42.8**	41	3.4
100	20	Quadratic-quadratic	11	8.2	0.6	0.6	1	1	37.6	35.8	**42**	10.4
100	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	3.8
100	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**44.8**
100	20	Cubic-linear	13.6	**90.2**	0.4	**97.2**	2.4	**97**	20.6	21.4	17	37.6
100	5	Linear-linear	**46.8**	*0.4*	24.2	0.2	35	0.2	24.2	24.8	25.2	*1*
100	5	Quadratic-linear	6.4	3	0.2	0.6	0.4	0.6	14	13.8	12.8	1.8
100	5	Quadratic-quadratic	4.6	5.2	*0*	*0*	*0*	*0*	**39**	**38.4**	**42.2**	8.4
100	5	Quadratic-intercept	31.8	3.4	**74.6**	7.2	**62.4**	6.6	4.8	4.8	4.6	4.2
100	5	Cubic-intercept	6.8	**46**	1	**87.8**	2.2	**80**	*2.6*	*2.6*	*2*	**45.4**
100	5	Cubic-linear	*3.6*	42	*0*	4.2	*0*	12.6	15.4	15.6	13.2	39.2

Note: Highest and lowest values are highlighted using bold and italics fonts, respectively.
[a] Level-2: Number of individuals; [b] Level-1: Average number of repeated measurements per individual.

Table A-5. Results for unbalanced data generated using the quadratic-quadratic model.

Level 2 [a]	Level 1 [b]	Data analysis models	Information criteria						Prediction criteria			
			AICk	AICq	BICkN	BICqN	BICkL2	BICqL2	$CCC_{VC[c]}$	$R^2_{VC[c]}$	$R^2_{JNS[c]}$	$R^2_{NS[c]}$
			(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
1000	20	Linear-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
1000	20	Quadratic-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	3.8
1000	20	Quadratic-quadratic	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	29
1000	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	5.6
1000	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**35.2**
1000	20	Cubic-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	26.8
1000	5	Linear-linear	27.8	*0*	**93.4**	*0*	**90.2**	*0*	9.8	10.2	10	*0.2*
1000	5	Quadratic-linear	20	0.6	6.4	2	9.6	2	17.8	17.8	18.4	2.2
1000	5	Quadratic-quadratic	**42.8**	42	0.2	0.2	0.2	1	**55**	**54.6**	**59.2**	9.6
1000	5	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	4
1000	5	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**43**
1000	5	Cubic-linear	9.4	**57.4**	*0*	**97.8**	*0*	**97**	17.4	17.4	12.4	41
100	20	Linear-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
100	20	Quadratic-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	3
100	20	Quadratic-quadratic	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	32
100	20	Quadratic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	6.4
100	20	Cubic-intercept	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	**33.8**
100	20	Cubic-linear	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	24.8
100	5	Linear-linear	**59.4**	*0.6*	**55.2**	*0*	**63.4**	0.4	26	26.6	28	*1*
100	5	Quadratic-linear	9.6	5	0.8	1.6	1.4	2.2	19	18.8	16.6	3.4
100	5	Quadratic-quadratic	8.2	9	*0*	*0*	*0*	*0.2*	**36.6**	**36**	**39.8**	7.4
100	5	Quadratic-intercept	12.6	2.4	43.6	6.8	33.8	5.6	*0.6*	*0.6*	*0.8*	5.6
100	5	Cubic-intercept	*3.8*	19.8	0.2	**66.4**	1	**52.8**	1.4	1.4	1.2	**43**
100	5	Cubic-linear	6.2	**63**	*0*	25	0.2	38.6	16.4	16.6	13.6	39.6

Note: Highest and lowest values are highlighted using bold and italics fonts, respectively.
[a] Level-2: Number of individuals; [b] Level-1: Average number of repeated measurements per individual.

APPENDIX B. COMPARISON OF LINEAR AND NONLINEAR GROWTH CURVE MODELS (CHAPTER 6)

B.1. Methodology

In Chapter 6 (See Section 6.2.2), I have provided methodological details on fitting covariate-adjusted CP GCM and estimating growth trajectories (distance, velocity and acceleration) and the adolescent growth spurt parameters (age at peak growth velocity, APGV; and peak growth velocity, PGV). Here I give details on covariate-adjusted linear (FP and RCS) and nonlinear (SITAR and PB) GCMs and estimating growth trajectories for these four GCMs. I will not repeat here methodological details on estimating APGV and PGV using derivatives which are same as described in Chapter 6 (see Section 6.2.5 and Section 6.2.6).

B.1.1. Growth curve models

I use the same notation described in Chapter 6 (Section 6.2.2) where y_{ij} denote the growth measurement for individual j ($j = 1, 2, 3, \dots, J$) at occasion i ($i = 1, 2, 3, \dots, n_j$) and age_{ij} is the age in years of the individual at that occasion. The design matrices for study ($\boldsymbol{X_{d_s ij}}$) and class ($\boldsymbol{X_{d_c ij}}$) effects defined in Chapter 6 (Equations (6-3) and (6-4)) are re-written as an Equation (B-1) as:

$$\boldsymbol{X_{d_s ij}} = \begin{cases} X_{d_s 1ij} = 1\text{f study} = \text{Burlington, and 0 otherwise} \\ X_{d_s 2ij} = 1\text{f study} = \text{Denver, and 0 otherwise} \\ X_{d_s 3ij} = 1\text{f study} = \text{Forsyth, and 0 otherwise} \\ X_{d_s 4ij} = 1\text{f study} = \text{Iowa, and 0 otherwise} \\ X_{d_s 5ij} = 1\text{f study} = \text{Mathews, and 0 otherwise} \\ X_{d_s 6ij} = 1\text{f study} = \text{Michigan, and 0 otherwise} \\ X_{d_s 7ij} = 1\text{f study} = \text{Oregon, and 0 otherwise} \end{cases} \qquad (B\text{-}1)$$

$$\boldsymbol{X_{d_c ij}} = \begin{cases} X_{d_c 1ij} = 1 \text{ if Class } = \text{Class II, and 0 otherwise} \\ X_{d_c 2ij} = 1 \text{ if Class } = \text{Class III, and 0 otherwise} \end{cases}$$

Where $d_s 1, .., d_s 7$ are seven dummy variables created for study ($d_s k - 1$ where $d_s k = 8$ i..e., eight growth studies). The Bolton growth study is used as the reference category. Likewise, $d_c 1, d_c 2$ are two dummy variables created for class ($d_c k - 1$ where $d_c k = 3$ i.,e, the number of categories of the class of skeletal jaw relationship). Class I is used as the reference category.

Fractional polynomial (FP)

The standard two level unadjusted (i.e., without covariate) FP GCM is described in Chapter 4 (Section 4.4.2). A two-level covariate-adjusted FP GCM is written as:

$$
\begin{aligned}
y_{ij} &= \beta_0 + \sum_{r=1}^{m_p} \beta_r age_{ij}^{p_r} + u_{0j} \\
&+ \sum_{d_s=1}^{d_s k-1} \beta_{d_s 0} X_{d_s ij} + \sum_{d_s=1}^{d_s k-1} \sum_{r=1}^{m_p} \beta_{d_s r} X_{d_s ij} age_{ij}^{p_r} \\
&+ \sum_{d_c=1}^{d_c k-1} \beta_{d_c 0} X_{d_c ij} + \sum_{d_c=1}^{d_c k-1} \sum_{r=1}^{m_p} \beta_{d_c r} X_{d_c ij} age_{ij}^{p_r} \\
&+ \sum_{s=1}^{m_q} u_{sj} age_{ij}^{p_s} \\
&+ e_{ij}
\end{aligned}
\qquad (B\text{-}2)
$$

where m_p represents the degree of the fixed effects part of the model; m_q represent the degree of the random effects part of the model; and p_r and p_s denote the power(s) of fixed and random parts of the model, respectively.

As powers p_r and p_s can be repeated ($m_p > 1$; $m_q > 1$), model (B-2) with a mixture of zero and non-zero powers with and without repetitions can be written as (B-3):

$$
\begin{aligned}
{}_{ij} = \beta_0 + \sum_{r=1}^{m_p} \beta_r \begin{bmatrix} age_{ij}^{p_r} & if\ p_r \neq 0; c = 1, \\ age_{ij}^{p_r}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_r \neq 0; c > 1, \\ \log(age_{ij})^{p_r} & if\ p_r = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ p_r = 0; c > 1, \end{bmatrix} \\
+ \sum_{d_s=1}^{d_s k-1} \beta_{d_s 0} X_{d_s ij} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c 0} X_{d_c ij} \\
+ \sum_{d_s=1}^{d_s k-1} \sum_{r=1}^{m_p} \beta_{d_s r} X_{d_s ij} \begin{bmatrix} age_{ij}^{p_r} & if\ p_r \neq 0; c = 1, \\ age_{ij}^{p_r}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_r \neq 0; c > 1, \\ \log(age_{ij})^{p_r} & if\ p_r = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ p_r = 0; c > 1, \end{bmatrix} \\
+ \sum_{d_c=1}^{d_c k-1} \sum_{r=1}^{m_p} \beta_{d_c r} X_{d_c ij} \begin{bmatrix} age_{ij}^{p_r} & if\ p_r \neq 0; c = 1, \\ age_{ij}^{p_r}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_r \neq 0; c > 1, \\ \log(age_{ij})^{p_r} & if\ p_r = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ p_r = 0; c > 1, \end{bmatrix} \\
+ \ u_{0j} \ + \sum_{s=1}^{m_q} u_{sj} \begin{bmatrix} age_{ij}^{p_s} & if\ p_s \neq 0; c = 1, \\ age_{ij}^{p_s}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_s \neq 0; c > 1, \\ \log(age_{ij})^{p_s} & if\ p_s = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ p_s = 0; c > 1, \end{bmatrix} \\
+ \ e_{ij}
\end{aligned}
\qquad (B\text{-}3)
$$

where c denotes the count (number of occurrences) of the same powers. When a power appears for the first time (i.e. first occurrence and not repeated), then $c = 1$ and when it is repeated for the first time (i.e. second occurrence), then $c = 2$ and so on (see Chapter 4, Section 4.4.2). The $\beta_0, \dots, \beta_{m_p}$ denote intercept and growth trajectories for reference categories for study (Bolton) and class (Class I) variables. The $\beta_{ds0}, \dots, \beta_{dsm_p}$ capture differences in the growth trajectories between Bolton study and these 7 studies. Similarly, the $\beta_{dc0}, \dots, \beta_{dcm_p}$ model differences in growth trajectories between Class I and dummy coded Class II and Class III. Individual-specific random effects ($u_{0j}, \dots, u_{m_q j}$) and residuals (e_{ij}) are assumed to be mutually independent and follow a normal distribution with mean 0 and respective covariance matrices. Please see Chapter 4 (Section 4.4.2) for further details.

Restricted cubic spline (RCS)

The standard two level unadjusted (for covariate effects) RCS GCM is described in Chapter 4 (Section 4.4.2). The covariate-adjusted RCS GCM with k_p knots $(t_1, \ldots, t_{k_p})$ for the fixed part of the model and k_q knots $(t_1, \ldots, t_{k_q})$ for the random part of the model is written as (B-4):

$$
\begin{aligned}
y_{ij} = {} & \beta_0 + \beta_1 age_{ij} + \sum_{p=2}^{k_p-2} \beta_p . S(age_{ij}) \\
& + \sum_{d_s=1}^{d_s k-1} \beta_{d_s 0} X_{d_s ij} + \sum_{d_s=1}^{d_s k-1} \beta_{d_s 1} X_{d_s ij} age_{ij} + \sum_{d_s=1}^{d_s k-1} \sum_{p=2}^{k_p-2} \beta_{d_s p} X_{d_s ij} . S(age_{ij}) \\
& + \sum_{d_c=1}^{d_c k-1} \beta_{d_c 0} X_{d_c ij} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c 1} X_{d_c ij} age_{ij} + \sum_{d_c=1}^{d_c k-1} \sum_{p=2}^{k_p-2} \beta_{d_c p} X_{d_c ij} . S(age_{ij}) \\
& + u_{0j} + u_{1j} age_{ij} + \sum_{q=2}^{k_q-2} u_{qj} . S(age_{ij}) \\
& + e_{ij}
\end{aligned}
\qquad (B\text{-}4)
$$

In this covariate-adjusted model (B-4), the $\beta_0 + \beta_1 age_{ij} + \sum_{p=2}^{k_p-2} \beta_p . S(age_{ij})$ denote intercept and growth trajectories for Class I and Bolton study estimated by the fixed effects $\beta_0, \beta_1 \ldots, \beta_p$. The $\sum_{d_s=1}^{d_s k-1} \beta_{d_s 0} X_{d_s ij} + \sum_{d_s=1}^{d_s k-1} \beta_{d_s 1} X_{d_s ij} age_{ij} + \sum_{d_s=1}^{d_s k-1} \sum_{p=2}^{k_p-2} \beta_{d_s p} X_{d_s ij} . S(age_{ij})$ part of the model (B-4) denote differences in growth trajectories between Bolton study and 7 other studies coded using dummy variable captured by fixed effects $\beta_{d_s}, \beta_{d_{s1}}, \ldots, \beta_{d_{sp}}$. Similarly, the $\sum_{d_c=1}^{d_c k-1} \beta_{d_c 0} X_{d_c ij} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c 1} X_{d_c ij} age_{ij} + \sum_{d_c=1}^{d_c k-1} \sum_{p=2}^{k_p-2} \beta_{d_c p} X_{d_c ij} . S(age_{ij})$ part of the model represents growth trajectories for the dummy coded Class II and Class III variables and the $\beta_{d_c}, \beta_{d_{c1}}, \ldots, \beta_{d_{cp}}$ denote Class-specific differences in growth trajectories.

The cubic spline function S(.) constructs the restricted cubic spline design matrix using truncated power basis. Full details on the construction of the restricted cubic spline design matrix

including alternative methods (such as B-spline basis) are provided in Chapter 4. Here I briefly outline the key steps which would later help in understanding the differentiation of spline functions (Section B.1.2). Using the truncated power basis "+" function, function, $k_p - 2$, new variables (spline terms) are created for the fixed part of the model as (see Chapter 4, Section 4.4.2):

$$\frac{(age_{ij} - t_p)_+^3}{(t_{k_p} - t_1)^2} - \frac{(age_{ij} - t_{k_p-1})_+^3 (t_{k_p} - t_p)}{(t_{k_p} - t_{k_p-1})(t_{k_p} - t_1)^2} + \frac{(age_{ij} - t_{k_p})_+^3 (t_{k_p-1} - t_p)}{(t_{k_p} - t_{k_p-1})(t_{k_p} - t_1)^2}$$

where "+" function is defined as:

$$S_+ = S \; if \; S > 0$$
$$= 0 \; if \; S \leq 0$$

Similarly, $k_q - 2$ new variables (spline terms) created for the random part of the model:

$$\frac{(age_{ij} - t_q)_+^3}{(t_{k_q} - t_1)^2} - \frac{(age_{ij} - t_{k_q-1})_+^3 (t_{k_q} - t_q)}{(t_{k_q} - t_{k_q-1})(t_{k_q} - t_1)^2} + \frac{(age_{ij} - t_{k_q})_+^3 (t_{k_q-1} - t_q)}{(t_{k_q} - t_{k_q-1})(t_{k_q} - t_1)^2}$$

Super Imposition by Translation and Rotation (SITAR)

Details on a two-level unadjusted SITAR GCM are provided in Chapter 4 (Section 4.2.3). The covariate-adjusted SITAR GCM is written as (B-5):

$$y_{ij} = \alpha_0 + \sum_{d_s=1}^{d_s k-1} \alpha_{d_s 0} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \alpha_{d_c 0} \boldsymbol{X_{d_c j}} + \alpha_j + S\left(\frac{age_{ij} - \left(\beta_0 + \sum_{d_s=1}^{d_s k-1} \beta_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c} \boldsymbol{X_{d_c j}} + \beta_j\right)}{e^{-\left(\gamma_0 + \sum_{d_s=1}^{d_s k-1} \gamma_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c} \boldsymbol{X_{d_c j}} + \gamma_j\right)}} \right) + e_{ij} \quad (B\text{-}5)$$

Here $S(.)$ is the restricted cubic spline function (as described earlier for the RCS GCM) to construct $k_p - 2$ new variables (spline terms) where k_p is the number of knots: The model (B-5) can be written as (B-6):

$$y_{ij} = \alpha_0 + \sum_{d_s=1}^{d_s k-1} \alpha_{d_s 0} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \alpha_{d_c 0} \boldsymbol{X_{d_c j}} + \alpha_j + s_1 \left(\frac{age_{ij} - \left(\beta_0 + \sum_{d_s=1}^{d_s k-1} \beta_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c} \boldsymbol{X_{d_c j}} + \beta_j\right)}{e^{-\left(\gamma_0 + \sum_{d_s=1}^{d_s k-1} \gamma_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c} \boldsymbol{X_{d_c j}} + \gamma_j\right)}} \right) + \sum_{p=2}^{k_p-2} s_p . S\left(\frac{age_{ij} - \left(\beta_0 + \sum_{d_s=1}^{d_s k-1} \beta_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c} \boldsymbol{X_{d_c j}} + \beta_j\right)}{e^{-\left(\gamma_0 + \sum_{d_s=1}^{d_s k-1} \gamma_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c} \boldsymbol{X_{d_c j}} + \gamma_j\right)}} \right) + e_{ij} \quad (B\text{-}6)$$

The $k_p - 2$ spline terms are created as:

$$\frac{\left(\left(\frac{age_{ij} - \left(\beta_0 + \sum_{d_s=1}^{d_s k-1} \beta_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c} \boldsymbol{X_{d_c j}} + \beta_j\right)}{e^{-\left(\gamma_0 + \sum_{d_s=1}^{d_s k-1} \gamma_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c} \boldsymbol{X_{d_c j}} + \gamma_j\right)}}\right) - t_p\right)_+^3}{\left(t_{k_p} - t_1\right)^2}$$

$$- \frac{\left(\left(\frac{age_{ij} - \left(\beta_0 + \sum_{d_s=1}^{d_s k-1} \beta_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c} \boldsymbol{X_{d_c j}} + \beta_j\right)}{e^{-\left(\gamma_0 + \sum_{d_s=1}^{d_s k-1} \gamma_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c} \boldsymbol{X_{d_c j}} + \gamma_j\right)}}\right) - t_{k_p-1}\right)_+^3 \left(t_{k_p} - t_p\right)}{\left(t_{k_p} - t_{k_p-1}\right)\left(t_{k_p} - t_1\right)^2}$$

$$+ \frac{\left(\left(\frac{age_{ij} - \left(\beta_0 + \sum_{d_s=1}^{d_s k-1} \beta_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c} \boldsymbol{X_{d_c j}} + \beta_j\right)}{e^{-\left(\gamma_0 + \sum_{d_s=1}^{d_s k-1} \gamma_{d_s} \boldsymbol{X_{d_s j}} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c} \boldsymbol{X_{d_c j}} + \gamma_j\right)}}\right) - t_{k_p}\right)_+^3 \left(t_{k_p-1} - t_p\right)}{\left(t_{k_p} - t_{k_p-1}\right)\left(t_{k_p} - t_1\right)^2}$$

In addition to the $k_p - 1$ ($s_1, \ldots, s_p$) spline regression coefficients capturing the mean growth trajectories, the Model (B-6) estimates α_0, β_0, γ_0 fixed effects denoting the size (α_0), timing (β_0) and (γ_0) the velocity parameters for the reference categories for study (Bolton) and class (Class I) variables. The differences in the size, timing and the intensity of growth between Bolton study and 7 other studies coded using dummy variable are captured by fixed effects α_{ds_0}, β_{ds}, γ_{ds}. Similarly, the α_{dc_0}, β_{dc}, γ_{dc} fixed effects denote differences in the size, timing and the intensity of growth between Class I and dummy coded Class II and Class III.

Individual-specific variations in the size, timing and the intensity of growth are denoted by random effects α_j, β_j and γ_j. The Model (B-6) estimated level-1 residuals are denoted by e_{ij}. The random effects (α_j, β_j and γ_j.) and residuals (e_{ij}) are assumed to be mutually independent and follow a normal distribution with mean 0 and respective covariance matrices. Please see Chapter 4 (Section 4.2.3) for further details.

Preece-Baines (PB)

Chapter 4 (Section 4.2.3) provides full details on the unadjusted two-level PB GCM. The covariate adjusted PB GCM is written as (B-7):

$$
\begin{aligned}
y_{ij} = & \left(s_{max} + \sum_{d_s=1}^{d_s k-1} s_{max_{d_s}} X_{d_s ij} + \sum_{d_c=1}^{d_c k-1} s_{max_{d_c}} X_{d_c ij} + s_{max_j} \right) \\
& - 2 \left(\left(s_{max} + \sum_{d_s=1}^{d_s k-1} s_{max_{d_s}} X_{d_s ij} + \sum_{d_c=1}^{d_c k-1} s_{max_{d_c}} X_{d_c ij} + s_{max_j} \right) \right. \\
& \left. - \left(s_\theta + \sum_{d_s=1}^{d_s k-1} s_{\theta_{d_s}} X_{d_s ij} + \sum_{d_c=1}^{d_c k-1} s_{\theta_{d_c}} X_{d_c ij} + s_{\theta_j} \right) \right) \\
& / \left(e^{\left(s_0 + \sum_{d_s=1}^{d_s k-1} s_{0_{d_s}} X_{d_s ij} + \sum_{d_c=1}^{d_c k-1} s_{0_{d_c}} X_{d_c ij} + s_{0_j}\right) \left(age - \left(\theta_0 + \sum_{d_s=1}^{d_s k-1} \theta_{d_s} X_{d_s ij} + \sum_{d_c=1}^{d_c k-1} \theta_{d_c} X_{d_c ij} + \theta_j\right)\right)} \right. \\
& \left. + e^{\left(s_1 + \sum_{d_s=1}^{d_s k-1} s_{1_{d_s}} X_{d_s ij} + \sum_{d_c=1}^{d_c k-1} s_{1_{d_c}} X_{d_c ij} + s_{0_j}\right) \left(age - \left(\theta_0 + \sum_{d_s=1}^{d_s k-1} \theta_{d_s} X_{d_s ij} + \sum_{d_c=1}^{d_c k-1} \theta_{d_c} X_{d_c ij} + \theta_j\right)\right)} \right) \\
& + e_{ij}
\end{aligned}
\qquad (B\text{-}7)
$$

where $s_{\max}$, s_θ , s_0 , s_1 and θ five fixed effect parameters describe mean growth trajectory for the reference categories study (Bolton) and class (Class I). The $s_{\max}$ is the maximum adulthood size and s_θ is the size at age θ where parameter θ locates the adolescent growth spurt during puberty. The s_0 and s_1 are growth-rate constants parameters related to prepubertal and pubertal growth velocities.

The differences in growth trajectories between Bolton study and 7 other studies coded using dummy variable are captured by fixed effects $s_{max_{d_s}}$, $s_{\theta_{d_s}}$, $s_{0_{d_s}}$, $s_{1_{d_s}}$ and θ_{d_s}. Similarly, the by fixed effects $s_{max_{d_c}}$, $s_{\theta_{d_c}}$, $s_{0_{d_c}}$, $s_{1_{d_c}}$ and θ_{d_c} denote differences in growth trajectories between Class I and Class II and Class III.

The random effects (s_{max_j}, s_{θ_j}, s_{0_j}, s_{1_j} and θ_j) denote individual-specific growth maximum adulthood size (s_{max_j}), size during the adolescent growth spurt (s_{θ_j}), the prepubertal (s_{0_j}) and pubertal growth (s_{1_j}) velocities and the corresponding to the adolescent growth spurt (θ_j).

The e_{ij} are residuals. It is assumed that the random effects and residuals are mutually independent and follow a normal distribution. Please see Chapter 4 (Section 4.2.3) for further details.

B.1.2. **Distance and derivatives estimation**

Chapter 4 (4.3.2) provides a detailed review of advanced differentiation rules required for estimating derivatives of CP, FP, RCS, SITAR and PB GCMs.

Chapter 6 (Section 6.2.5) described in detail distance and derivatives estimation for the covariate-adjusted CP GCM.

Here I provide equations to estimate population-average and individual-specific distance and derivatives covariate-adjusted FP, RCS, SITAR and PB GCMs. Please see Section B.1.1 for covariate adjusted FP, RCS, SITAR and PB GCMs for whom distance and derivatives estimation are shown below.

Fractional polynomial (FP)

Distance

Population-average, $\boldsymbol{f_{FP_0}}$

$$\beta_0 + \sum_{r=1}^{m_p} \beta_r \begin{cases} age_{ij}^{p_r} & if\ p_r \neq 0; c = 1, \\ age_{ij}^{p_r}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_r \neq 0; c > 1, \\ \log(age_{ij})^{p_r} & if\ \ p_r = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ \ p_r = 0; c > 1, \end{cases}$$
$$+ \sum_{d_s=1}^{d_s k-1} \beta_{d_s 0} X_{d_s ij} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c 0} X_{d_c ij}$$
$$+ \sum_{d_s=1}^{d_s k-1} \sum_{r=1}^{m_p} \beta_{d_s r} X_{d_s ij} \begin{cases} age_{ij}^{p_r} & if\ p_r \neq 0; c = 1, \\ age_{ij}^{p_r}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_r \neq 0; c > 1, \\ \log(age_{ij})^{p_r} & if\ \ p_r = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ \ p_r = 0; c > 1, \end{cases} \quad (B\text{-}8)$$
$$+ \sum_{d_c=1}^{d_c k-1} \sum_{r=1}^{m_p} \beta_{d_c r} X_{d_c ij} \begin{cases} age_{ij}^{p_r} & if\ p_r \neq 0; c = 1, \\ age_{ij}^{p_r}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_r \neq 0; c > 1, \\ \log(age_{ij})^{p_r} & if\ \ p_r = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ \ p_r = 0; c > 1, \end{cases}$$

Individual-specific, $\boldsymbol{f_{FP_j}}$

$$\beta_0 + \sum_{r=1}^{m_p} \beta_r \begin{Bmatrix} age_{ij}^{p_r} & if\ p_r \neq 0; c = 1, \\ age_{ij}^{p_r}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_r \neq 0; c > 1, \\ \log(age_{ij})^{p_r} & if\ p_r = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ p_r = 0; c > 1, \end{Bmatrix}$$

$$+ \sum_{d_s=1}^{d_s k-1} \beta_{d_s 0} X_{d_s ij} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c 0} X_{d_c ij}$$

$$+ \sum_{d_s=1}^{d_s k-1} \sum_{r=1}^{m_p} \beta_{d_s r} X_{d_s ij} \begin{Bmatrix} age_{ij}^{p_r} & if\ p_r \neq 0; c = 1, \\ age_{ij}^{p_r}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_r \neq 0; c > 1, \\ \log(age_{ij})^{p_r} & if\ p_r = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ p_r = 0; c > 1, \end{Bmatrix} \qquad (B\text{-}9)$$

$$+ \sum_{d_c=1}^{d_c k-1} \sum_{r=1}^{m_p} \beta_{d_c r} X_{d_c ij} \begin{Bmatrix} age_{ij}^{p_r} & if\ p_r \neq 0; c = 1, \\ age_{ij}^{p_r}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_r \neq 0; c > 1, \\ \log(age_{ij})^{p_r} & if\ p_r = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ p_r = 0; c > 1, \end{Bmatrix}$$

$$+ \ u_{0j} + \sum_{s=1}^{m_q} u_{sj} \begin{Bmatrix} age_{ij}^{p_s} & if\ p_s \neq 0; c = 1, \\ age_{ij}^{p_s}\left(\log(age_{ij})\right)^{(c-1)} & if\ p_s \neq 0; c > 1, \\ \log(age_{ij})^{p_s} & if\ p_s = 0; c = 1, \\ \log(age_{ij})^{(c-1)} & if\ p_s = 0; c > 1, \end{Bmatrix}$$

First derivatives

Population-average, f'_{FP_0}

$$\sum_{r=1}^{m_p}\beta_r\begin{Bmatrix} p_r(age_{ij})^{(p_r-1)} & if\ p_r\neq 0; c=1,\\ \begin{pmatrix} p_r(age_{ij})^{(p_r-1)}\big(\log(age_{ij})\big)^{(c-1)}\\ +\ age_{ij}^{(p_r-1)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\end{pmatrix} & if\ p_r\neq 0; c>1,\\ p_r\big(\log(age_{ij})\big)^{(p_r-1)} & if\ p_r=0; c=1,\\ age_{ij}^{-1}(for)\big(\log(age_{ij})\big)^{(c-1)} & if\ p_r=0; c>1,\end{Bmatrix}$$

$$+\sum_{d_s=1}^{d_sk-1}\sum_{r=1}^{m_p}\beta_{d_sr}X_{d_sij}\begin{Bmatrix} p_r(age_{ij})^{(p_r-1)} & if\ p_r\neq 0; c=1,\\ \begin{pmatrix} p_r(age_{ij})^{(p_r-1)}\big(\log(age_{ij})\big)^{(c-1)}\\ +\ age_{ij}^{(p_r-1)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\end{pmatrix} & if\ p_r\neq 0; c>1,\\ p_r\big(\log(age_{ij})\big)^{(p_r-1)} & if\ p_r=0; c=1,\\ age_{ij}^{-1}(for)\big(\log(age_{ij})\big)^{(c-1)} & if\ p_r=0; c>1,\end{Bmatrix} \qquad (B\text{-}10)$$

$$+\sum_{d_c=1}^{d_ck-1}\sum_{r=1}^{m_p}\beta_{d_cr}X_{d_cij}\begin{Bmatrix} p_r(age_{ij})^{(p_r-1)} & if\ p_r\neq 0; c=1,\\ \begin{pmatrix} p_r(age_{ij})^{(p_r-1)}\big(\log(age_{ij})\big)^{(c-1)}\\ +\ age_{ij}^{(p_r-1)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\end{pmatrix} & if\ p_r\neq 0; c>1,\\ p_r\big(\log(age_{ij})\big)^{(p_r-1)} & if\ p_r=0; c=1,\\ age_{ij}^{-1}(for)\big(\log(age_{ij})\big)^{(c-1)} & if\ p_r=0; c>1,\end{Bmatrix}$$

Individual-specific, f'_{FP_j}

$$\sum_{r=1}^{m_p}\beta_r\begin{cases} p_r(age_{ij})^{(p_r-1)} & if\ p_r\neq 0; c=1,\\ \begin{pmatrix} p_r(age_{ij})^{(p_r-1)}\big(\log(age_{ij})\big)^{(c-1)}\\ +\ age_{ij}^{(p_r-1)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\end{pmatrix} & if\ p_r\neq 0; c>1,\\ p_r\big(\log(age_{ij})\big)^{(p_r-1)} & if\ p_r=0; c=1,\\ age_{ij}^{-1}(for)\big(\log(age_{ij})\big)^{(c-1)} & if\ p_r=0; c>1,\end{cases}$$

$$+\sum_{d_s=1}^{d_sk-1}\sum_{r=1}^{m_p}\beta_{d_sr}X_{d_sij}\begin{cases} p_r(age_{ij})^{(p_r-1)} & if\ p_r\neq 0; c=1,\\ \begin{pmatrix} p_r(age_{ij})^{(p_r-1)}\big(\log(age_{ij})\big)^{(c-1)}\\ +\ age_{ij}^{(p_r-1)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\end{pmatrix} & if\ p_r\neq 0; c>1,\\ p_r\big(\log(age_{ij})\big)^{(p_r-1)} & if\ p_r=0; c=1,\\ age_{ij}^{-1}(for)\big(\log(age_{ij})\big)^{(c-1)} & if\ p_r=0; c>1,\end{cases} \quad (B\text{-}11)$$

$$+\sum_{d_c=1}^{d_ck-1}\sum_{r=1}^{m_p}\beta_{d_cr}X_{d_cij}\begin{cases} p_r(age_{ij})^{(p_r-1)} & if\ p_r\neq 0; c=1,\\ \begin{pmatrix} p_r(age_{ij})^{(p_r-1)}\big(\log(age_{ij})\big)^{(c-1)}\\ +\ age_{ij}^{(p_r-1)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\end{pmatrix} & if\ p_r\neq 0; c>1,\\ p_r\big(\log(age_{ij})\big)^{(p_r-1)} & if\ p_r=0; c=1,\\ age_{ij}^{-1}(for)\big(\log(age_{ij})\big)^{(c-1)} & if\ p_r=0; c>1,\end{cases}$$

$$+\sum_{s=1}^{m_q}u_{js}\begin{cases} p_s\big(age_{ij}\big)^{(p_s-1)} & if\ p_s\neq 0; c=1,\\ \begin{pmatrix} p_s\big(age_{ij}\big)^{(p_s-1)}\big(\log(age_{ij})\big)^{(c-1)}\\ +\ age_{ij}^{(p_s-1)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\end{pmatrix} & if\ p_s\neq 0; c>1,\\ p_s\big(\log(age_{ij})\big)^{(p_s-1)} & if\ p_s=0; c=1,\\ age_{ij}^{-1}(fos)\big(\log(age_{ij})\big)^{(c-1)} & if\ p_s=0; c>1,\end{cases}$$

Second derivative

Population-average, f''_{FP_0}

$$\sum_{r=1}^{m_p}\beta_r\begin{cases} p_r(p_r-1)(age_{ij})^{(p_r-2)} & if\ p_r\neq 0; c=1,\\ \begin{pmatrix} p_r(p_r-1)(age_{ij})^{(p_r-2)}\big(\log(age_{ij})\big)^{(c-1)}\\ +\,p_r(age_{ij})^{(p_r-2)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\\ +\,age_{ij}^{(p_r-2)}(c^2-3c+2)\big(\log(age_{ij})\big)^{(c-3)}\\ +\,(p_r-1)\big(age_{ij}\big)^{(p_r-2)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\end{pmatrix} & if\ p_r\neq 0; c>1,\\ p_r(p_r-1)\big(\log(age_{ij})\big)^{(p_r-2)} & if\ \ p_r=0; c=1,\\ \begin{pmatrix} age_{ij}^{-2}(c(c-1))\big(\log(age_{ij})\big)^{(c-2)}\\ -\,age_{ij}^{-2}\,(c)\big(\log(age_{ij})\big)^{(c-1)}\end{pmatrix} & if\ \ p_r=0; c>1,\end{cases}$$

$$+\sum_{d_s=1}^{d_sk-1}\sum_{r=1}^{m_p}\beta_{d_sr}X_{d_sij}\begin{cases} p_r(p_r-1)(age_{ij})^{(p_r-2)} & if\ p_r\neq 0; c=1,\\ \begin{pmatrix} p_r(p_r-1)(age_{ij})^{(p_r-2)}\big(\log(age_{ij})\big)^{(c-1)}\\ +\,p_r(age_{ij})^{(p_r-2)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\\ +age_{ij}^{(p_r-2)}(c^2-3c+2)\big(\log(age_{ij})\big)^{(c-3)}\\ +\ (p_r-1)\big(age_{ij}\big)^{(p_r-2)}(c-1)^{(c-2)}\\ \qquad\big(\log(age_{ij})\big)\end{pmatrix} & if\ p_r\neq 0; c>1,\\ p_r(p_r-1)\big(\log(age_{ij})\big)^{(p_r-2)} & if\ \ p_r=0; c=1,\\ \begin{pmatrix} age_{ij}^{-2}(c(c-1))\big(\log(age_{ij})\big)^{(c-2)}\\ -\,age_{ij}^{-2}\,(c)\big(\log(age_{ij})\big)^{(c-1)}\end{pmatrix} & if\ \ p_r=0; c>1,\end{cases} \qquad (B\text{-}12)$$

$$+\sum_{d_c=1}^{d_ck-1}\sum_{r=1}^{m_p}\beta_{d_cr}X_{d_cij}\begin{cases} p_r(p_r-1)(age_{ij})^{(p_r-2)} & if\ p_r\neq 0; c=1,\\ \begin{pmatrix} p_r(p_r-1)(age_{ij})^{(p_r-2)}\big(\log(age_{ij})\big)^{(c-1)}\\ +\,p_r(age_{ij})^{(p_r-2)}(c-1)\big(\log(age_{ij})\big)^{(c-2)}\\ +age_{ij}^{(p_r-2)}(c^2-3c+2)\big(\log(age_{ij})\big)^{(c-3)}\\ +\ (p_r-1)\big(age_{ij}\big)^{(p_r-2)}(c-1)^{(c-2)}\\ \qquad\big(\log(age_{ij})\big)\end{pmatrix} & if\ p_r\neq 0; c>1,\\ p_r(p_r-1)\big(\log(age_{ij})\big)^{(p_r-2)} & if\ \ p_r=0; c=1,\\ \begin{pmatrix} age_{ij}^{-2}(c(c-1))\big(\log(age_{ij})\big)^{(c-2)}\\ -\,age_{ij}^{-2}\,(c)\big(\log(age_{ij})\big)^{(c-1)}\end{pmatrix} & if\ \ p_r=0; c>1,\end{cases}$$

Individual-specific, f''_{FP_j}

$$\sum_{r=1}^{m_p}\beta_r\begin{Bmatrix} p_r(p_r-1)(age_{ij})^{(p_r-2)} & if\ p_r\neq 0; c=1, \\ \begin{pmatrix} p_r(p_r-1)(age_{ij})^{(p_r-2)}\left(\log(age_{ij})\right)^{(c-1)} \\ +p_r(age_{ij})^{(p_r-2)}(c-1)\left(\log(age_{ij})\right)^{(c-2)} \\ +age_{ij}^{(p_r-2)}(c^2-3c+2)\left(\log(age_{ij})\right)^{(c-3)} \\ +(p_r-1)(age_{ij})^{(p_r-2)}(c-1)\left(\log(age_{ij})\right)^{(c-2)} \end{pmatrix} & if\ p_r\neq 0; c>1, \\ p_r(p_r-1)\left(\log(age_{ij})\right)^{(p_r-2)} & if\ \ p_r=0; c=1, \\ \begin{pmatrix} age_{ij}^{-2}(c(c-1))\left(\log(age_{ij})\right)^{(c-2)} \\ -age_{ij}^{-2}(c)\left(\log(age_{ij})\right)^{(c-1)} \end{pmatrix} & if\ \ p_r=0; c>1, \end{Bmatrix}$$

$$+\sum_{d_s=1}^{d_sk-1}\sum_{r=1}^{m_p}\beta_{d_sr}X_{d_sij}\begin{Bmatrix} p_r(p_r-1)(age_{ij})^{(p_r-2)} & if\ p_r\neq 0; c=1, \\ \begin{pmatrix} p_r(p_r-1)(age_{ij})^{(p_r-2)}\left(\log(age_{ij})\right)^{(c-1)} \\ +p_r(age_{ij})^{(p_r-2)}(c-1)\left(\log(age_{ij})\right)^{(c-2)} \\ +age_{ij}^{(p_r-2)}(c^2-3c+2)\left(\log(age_{ij})\right)^{(c-3)} \\ +\ (p_r-1)(age_{ij})^{(p_r-2)}(c-1)^{(c-2)}\left(\log(age_{ij})\right) \end{pmatrix} & if\ p_r\neq 0; c>1, \\ p_r(p_r-1)\left(\log(age_{ij})\right)^{(p_r-2)} & if\ \ p_r=0; c=1, \\ \begin{pmatrix} age_{ij}^{-2}(c(c-1))\left(\log(age_{ij})\right)^{(c-2)} \\ -age_{ij}^{-2}(c)\left(\log(age_{ij})\right)^{(c-1)} \end{pmatrix} & if\ \ p_r=0; c>1, \end{Bmatrix} \qquad (B\text{-}13)$$

$$+\sum_{d_c=1}^{d_ck-1}\sum_{r=1}^{m_p}\beta_{d_cr}X_{d_cij}\begin{Bmatrix} p_r(p_r-1)(age_{ij})^{(p_r-2)} & if\ p_r\neq 0; c=1, \\ \begin{pmatrix} p_r(p_r-1)(age_{ij})^{(p_r-2)}\left(\log(age_{ij})\right)^{(c-1)} \\ +p_r(age_{ij})^{(p_r-2)}(c-1)\left(\log(age_{ij})\right)^{(c-2)} \\ +age_{ij}^{(p_r-2)}(c^2-3c+2)\left(\log(age_{ij})\right)^{(c-3)} \\ +\ (p_r-1)(age_{ij})^{(p_r-2)}(c-1)^{(c-2)}\left(\log(age_{ij})\right) \end{pmatrix} & if\ p_r\neq 0; c>1, \\ p_r(p_r-1)\left(\log(age_{ij})\right)^{(p_r-2)} & if\ \ p_r=0; c=1, \\ \begin{pmatrix} age_{ij}^{-2}(c(c-1))\left(\log(age_{ij})\right)^{(c-2)} \\ -age_{ij}^{-2}(c)\left(\log(age_{ij})\right)^{(c-1)} \end{pmatrix} & if\ \ p_r=0; c>1, \end{Bmatrix}$$

$$+\sum_{s=1}^{m_q} u_{js}\begin{cases} p_s(p_s-1)\left(age_{ij}\right)^{(p_s-2)} & if\ p_s \neq 0; c = 1, \\ \begin{pmatrix} p_s(p_s-1)\left(age_{ij}\right)^{(p_s-2)}\left(\log\left(age_{ij}\right)\right)^{(c-1)} \\ + p_s\left(age_{ij}\right)^{(p_s-2)}(c-1)\left(\log(age_{ij})\right)^{(c-2)} \\ + age_{ij}^{(p_s-2)}(c^2-3c+2)\left(\log(age_{ij})\right)^{(c-3)} \\ + (p_s-1)\left(age_{ij}\right)^{(p_s-2)}(c-1)\left(\log(age_{ij})\right)^{(c-2)} \end{pmatrix} & if\ p_s \neq 0; c > 1, \\ p_s(p_s-1)\left(\log(age_{ij})\right)^{(p_s-2)} & if\ p_s = 0; c = 1, \\ \begin{pmatrix} age_{ij}^{-2}(c(c-1))\left(\log(age_{ij})\right)^{(c-2)} \\ - age_{ij}^{-2}\,(c)\left(\log(age_{ij})\right)^{(c-1)} \end{pmatrix} & if\ \ p_s = 0; c > 1, \end{cases}$$

Restricted cubic spline (RCS)

Distance

Population-average, f_{RCS_0}

$$\beta_0 + \beta_1\{age_{ij}\} + \sum_{p=2}^{k_p-2} \beta_p . S\{(age_{ij})\}$$
$$+ \sum_{d_s=1}^{d_s k-1} \beta_{d_s 0} X_{d_s ij} + \sum_{d_s=1}^{d_s k-1} \beta_{d_{s1}} X_{d_s ij}\{age_{ij}\} + \sum_{d_s=1}^{d_s k-1} \sum_{p=2}^{k_p-2} \beta_{d_{sp}} X_{d_s ij} . S\{(age_{ij})\} \quad (B\text{-}14)$$
$$+ \sum_{d_c=1}^{d_c k-1} \beta_{d_c 0} X_{d_c ij} + \sum_{d_c=1}^{d_c k-1} \beta_{d_{c1}} X_{d_c ij}\{age_{ij}\} + \sum_{d_c=1}^{d_c k-1} \sum_{p=2}^{k_p-2} \beta_{d_{cp}} X_{d_c ij} . S\{(age_{ij})\}$$

Where

$$\sum_{p=2}^{k_p-2} \beta_p . S\{(age_{ij})\}$$
$$= \beta_p \left\{ \frac{(age_{ij} - t_p)_+^3}{(t_{k_p} - t_1)^2} - \frac{(age_{ij} - t_{k_p-1})_+^3 (t_{k_p} - t_p)}{(t_{k_p} - t_{k_p-1})(t_{k_p} - t_1)^2} + \frac{(age_{ij} - t_{k_p})_+^3 (t_{k_p-1} - t_p)}{(t_{k_p} - t_{k_p-1})(t_{k_p} - t_1)^2} \right\}$$

Individual-specific, $\boldsymbol{f_{RCS_j}}$

$$\begin{aligned}
&\beta_0 + \beta_1\{age_{ij}\} + \sum_{p=2}^{k_p-2} \beta_p . S\{(age_{ij})\} \\
&+ \sum_{d_s=1}^{d_s k-1} \beta_{d_s 0} X_{d_s ij} + \sum_{d_s=1}^{d_s k-1} \beta_{d_s 1} X_{d_s ij}\{age_{ij}\} + \sum_{d_s=1}^{d_s k-1} \sum_{p=2}^{k_p-2} \beta_{d_s p} X_{d_s ij} . S\{(age_{ij})\} \\
&+ \sum_{d_c=1}^{d_c k-1} \beta_{d_c 0} X_{d_c ij} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c 1} X_{d_c ij}\{age_{ij}\} + \sum_{d_c=1}^{d_c k-1} \sum_{p=2}^{k_p-2} \beta_{d_c p} X_{d_c ij} . S\{(age_{ij})\} \\
&+ u_{0j} + u_{1j}\{age_{ij}\} + \sum_{q=2}^{k_q-2} u_{qj} . S\{(age_{ij})\}
\end{aligned} \tag{B-15}$$

where

$$\sum_{p=2}^{k_p-2} \beta_p . S\{(age_{ij})\} + \sum_{q=2}^{k_q-2} u_{qj} . S\{(age_{ij})\}$$

$$= \beta_p \left\{ \frac{(age_{ij} - t_p)_+^3}{(t_{k_p} - t_1)^2} - \frac{(age_{ij} - t_{k_p-1})_+^3 (t_{k_p} - t_p)}{(t_{k_p} - t_{k_p-1})(t_{k_p} - t_1)^2} + \frac{(age_{ij} - t_{k_p})_+^3 (t_{k_p-1} - t_p)}{(t_{k_p} - t_{k_p-1})(t_{k_p} - t_1^2)} \right\}$$

$$+ u_{qj} \left\{ \frac{(age_{ij} - t_q)_+^3}{(t_{k_q} - t_1)^2} - \frac{(age_{ij} - t_{k_q-1})_+^3 (t_{k_q} - t_q)}{(t_{k_q} - t_{k_q-1})(t_{k_q} - t_1)^2} + \frac{(age_{ij} - t_{k_q})_+^3 (t_{k_q-1} - t_q)}{(t_{k_q} - t_{k_q-1})(t_{k_q} - t_1)^2} \right\}$$

First derivative

Population-average, $\boldsymbol{f'_{RCS_0}}$

$$\beta_1\{1(age_{ij})^{(1-1)}\} + \sum_{p=2}^{k_p-2} \beta_p . S\{3(age_{ij})^{(3-1)}\}$$
$$+ \sum_{d_s=1}^{d_s k-1} \beta_{d_{s1}} X_{d_s ij}\{1(age_{ij})^{(1-1)}\}$$
$$+ \sum_{d_s=1}^{d_s k-1} \sum_{p=2}^{k_p-2} \beta_{d_{sp}} X_{d_s ij} . S\{3(age_{ij})^{(3-1)}\} \qquad (B\text{-}16)$$
$$+ \sum_{d_c=1}^{d_c k-1} \beta_{d_{c1}} X_{d_c ij}\{1(age_{ij})^{(1-1)}\}$$
$$+ \sum_{d_c=1}^{d_c k-1} \sum_{p=2}^{k_p-2} \beta_{d_{cp}} X_{d_c ij} . S\{3(age_{ij})^{(3-1)}\}$$

Where

$$\sum_{p=2}^{k_p-2} \beta_p . S\{3(age_{ij})^{(3-1)}\}$$

$$= \beta_p \left\{ \frac{3(age_{ij} - t_p)_+^{(3-1)}}{(t_{k_p} - t_1)^2} - \frac{3(age_{ij} - t_{k_p-1})_+^{(3-1)}(t_{k_p} - t_p)}{(t_{k_p} - t_{k_p-1})(t_{k_p} - t_1)^2} + \frac{3(age_{ij} - t_{k_p})_+^{(3-1)}(t_{k_p-1} - t_p)}{(t_{k_p} - t_{k_p-1})(t_{k_p} - t_1)^2} \right\}$$

Individual-specific, f'_{RCS_j}

$$\begin{aligned}\beta_1\{1(age_{ij})^{(1-1)}\} + \sum_{p=2}^{k_p-2}\beta_p.S\{3(age_{ij})^{(3-1)}\} + \sum_{d_s=1}^{d_sk-1}\beta_{d_{s1}}X_{d_sij}\{1(age_{ij})^{(1-1)}\} \\ + \sum_{d_s=1}^{d_sk-1}\sum_{p=2}^{k_p-2}\beta_{d_{sp}}X_{d_sij}.S\{3(age_{ij})^{(3-1)}\} + \sum_{d_c=1}^{d_ck-1}\beta_{d_{c1}}X_{d_cij}\{1(age_{ij})^{(1-1)}\} \\ + \sum_{d_c=1}^{d_ck-1}\sum_{p=2}^{k_p-2}\beta_{d_{cp}}X_{d_cij}.S\{3(age_{ij})^{(3-1)}\} + u_{1j}\{1(age_{ij})^{(1-1)}\} \\ + \sum_{q=2}^{k_q-2}u_{qj}.S\{3(age_{ij})^{(3-1)}\}\end{aligned} \quad (B\text{-}17)$$

Where

$$\sum_{p=2}^{k_p-2}\beta_p.S\{3(age_{ij})^{(3-1)}\} + \sum_{q=2}^{k_q-2}u_{qj}.S\{3(age_{ij})^{(3-1)}\}$$

$$= \beta_p\left\{\frac{3(age_{ij}-t_p)_+^{(3-1)}}{(t_{k_p}-t_1)^2} - \frac{3(age_{ij}-t_{k_p-1})_+^{(3-1)}(t_{k_p}-t_p)}{(t_{k_p}-t_{k_p-1})(t_{k_p}-t_1)^2} + \frac{3(age_{ij}-t_{k_p})_+^{(3-1)}(t_{k_p-1}-t_p)}{(t_{k_p}-t_{k_p-1})(t_{k_p}-t_1)^2}\right\}$$

$$+ u_{qj}\left\{\frac{3(age_{ij}-t_q)_+^{(3-1)}}{(t_{k_q}-t_1)^2} - \frac{3(age_{ij}-t_{k_q-1})_+^{(3-1)}(t_{k_q}-t_q)}{(t_{k_q}-t_{k_q-1})(t_{k_q}-t_1)^2} + \frac{3(age_{ij}-t_{k_q})_+^{(3-1)}(t_{k_q-1}-t_q)}{(t_{k_q}-t_{k_q-1})(t_{k_q}-t_1)^2}\right\}$$

Second derivative

Population-average, $\boldsymbol{f''_{RCS_0}}$

$$\beta_1\{1(1-1)(age_{ij})^{(1-2)}\} + \sum_{p=2}^{k_p-2} \beta_p.S\{3(3-1)(age_{ij})^{(3-2)}\}$$

$$+ \sum_{d_s=1}^{d_sk-1} \beta_{d_{s1}} X_{d_sij}\{1(1-1)(age_{ij})^{(1-2)}\}$$

$$+ \sum_{d_s=1}^{d_sk-1} \sum_{p=2}^{k_p-2} \beta_{d_{sp}} X_{d_sij}.S\{3(3-1)(age_{ij})^{(3-2)}\} \qquad (B\text{-}18)$$

$$+ \sum_{d_c=1}^{d_ck-1} \beta_{d_{c1}} X_{d_cij}\{1(1-1)(age_{ij})^{(1-2)}\}$$

$$+ \sum_{d_c=1}^{d_ck-1} \sum_{p=2}^{k_p-2} \beta_{d_{cp}} X_{d_cij}.S\{3(3-1)(age_{ij})^{(3-2)}\}$$

Where

$$\sum_{p=2}^{k_p-2} \beta_p.S\{3(3-1)(age_{ij})^{(3-2)}\}$$

$$= \beta_p \left\{ \begin{array}{c} \dfrac{3(3-1)\left(age_{ij}-t_p\right)_+^{(3-2)}}{\left(t_{k_p}-t_1\right)^2} - \dfrac{3(3-1)\left(age_{ij}-t_{k_p-1}\right)_+^{(3-2)}\left(t_{k_p}-t_p\right)}{\left(t_{k_p}-t_{k_p-1}\right)\left(t_{k_p}-t_1\right)^2} \\ + \dfrac{3(3-1)\left(age_{ij}-t_{k_p}\right)_+^{(3-2)}\left(t_{k_p-1}-t_p\right)}{\left(t_{k_p}-t_{k_p-1}\right)\left(t_{k_p}-t_1\right)^2} \end{array} \right\}$$

Individual-specific, $\boldsymbol{f''_{RCS_j}}$

$$\beta_1\{1(1-1)(age_{ij})^{(1-2)}\} + \sum_{p=2}^{k_p-2} \beta_p.S\{3(3-1)(age_{ij})^{(3-2)}\}$$

$$+ \sum_{d_s=1}^{d_sk-1} \beta_{d_{s1}} X_{d_sij}\{1(1-1)(age_{ij})^{(1-2)}\}$$

$$+ \sum_{d_s=1}^{d_sk-1} \sum_{p=2}^{k_p-2} \beta_{d_{sp}} X_{d_sij}.S\{3(3-1)(age_{ij})^{(3-2)}\}$$

$$+ \sum_{d_c=1}^{d_ck-1} \beta_{d_{c1}} X_{d_cij}\{1(1-1)(age_{ij})^{(1-2)}\} \qquad (B\text{-}19)$$

$$+ \sum_{d_c=1}^{d_ck-1} \sum_{p=2}^{k_p-2} \beta_{d_{cp}} X_{d_cij}.S\{3(3-1)(age_{ij})^{(3-2)}\}$$

$$+ u_{1j}\{1(1-1)(age_{ij})^{(1-2)}\} + \sum_{q=2}^{k_q-2} u_{qj}.S\{3(3-1)(age_{ij})^{(3-2)}\}$$

Where

$$\sum_{p=2}^{k_p-2} \beta_p.S\{3(3-1)(age_{ij})^{(3-2)}\} + \sum_{q=2}^{k_q-2} u_{qj}.S\{3(3-1)(age_{ij})^{(3-2)}\}$$

$$= \beta_p \left\{ \frac{3(3-1)\left(age_{ij}-t_p\right)_+^{(3-2)}}{\left(t_{k_p}-t_1\right)^2} - \frac{3(3-1)\left(age_{ij}-t_{k_p-1}\right)_+^{(3-2)}\left(t_{k_p}-t_p\right)}{\left(t_{k_p}-t_{k_p-1}\right)\left(t_{k_p}-t_1\right)^2} + \frac{3(3-1)\left(age_{ij}-t_{k_p}\right)_+^{(3-2)}\left(t_{k_p-1}-t_p\right)}{\left(t_{k_p}-t_{k_p-1}\right)\left(t_{k_p}-t_1\right)^2} \right\}$$

$$+ u_{qj} \left\{ \frac{3(3-1)\left(age_{ij}-t_q\right)_+^{3-2}}{\left(t_{k_q}-t_1\right)^2} - \frac{3(3-1)\left(age_{ij}-t_{k_q-1}\right)_+^{3-2}\left(t_{k_q}-t_q\right)}{\left(t_{k_q}-t_{k_q-1}\right)\left(t_{k_q}-t_1\right)^2} + \frac{3(3-1)\left(age_{ij}-t_{k_q}\right)_+^{3-2}\left(t_{k_q-1}-t_q\right)}{\left(t_{k_q}-t_{k_q-1}\right)\left(t_{k_q}-t_1\right)^2} \right\}$$

Super Imposition by Translation and Rotation (SITAR)

Distance

Population-average, f_{SITAR_0}

$$\alpha_{0_{adj}} + s_1\left\{\left(\frac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}}\right)\right\} + \sum_{p=2}^{k_p-2} s_p.S\left\{\left(\frac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}}\right)\right\} \qquad (B\text{-}20)$$

Where $\alpha_{0_{adj}}$, $\beta_{0_{adj}}$ and $\gamma_{0\,adj}$ are:

$$\alpha_{0_{adj}} = \alpha_0 + \sum_{d_s=1}^{d_sk-1} \alpha_{d_s0}\{\boldsymbol{X_{d_sj}}\} + \sum_{d_c=1}^{d_ck-1} \alpha_{d_c0}\{\boldsymbol{X_{d_cij}}\} - \frac{1}{n_j}\sum_{1}^{n_j}\left(\sum_{d_s=1}^{d_sk-1} \alpha_{d_s0}\{\boldsymbol{X_{d_sj}}\} + \sum_{d_c=1}^{d_ck-1} \alpha_{d_c0}\{\boldsymbol{X_{d_cij}}\}\right)$$

$$\beta_{0_{adj}} = \beta_0 + \sum_{d_s=1}^{d_sk-1} \beta_{d_s}\{\boldsymbol{X_{d_sj}}\} + \sum_{d_c=1}^{d_ck-1} \beta_{d_c}\{\boldsymbol{X_{d_cij}}\} - \frac{1}{n_j}\sum_{1}^{n_j}\left(\sum_{d_s=1}^{d_sk-1} \beta_{d_s}\{\boldsymbol{X_{d_sj}}\} + \sum_{d_c=1}^{d_ck-1} \beta_{d_c}\{\boldsymbol{X_{d_cij}}\}\right)$$

$$\gamma_{0_{adj}} = \gamma_0 + \sum_{d_s=1}^{d_sk-1} \gamma_{d_s}\{\boldsymbol{X_{d_sj}}\} + \sum_{d_c=1}^{d_ck-1} \gamma_{d_c}\{\boldsymbol{X_{d_cij}}\} - \frac{1}{n_j}\sum_{1}^{n_j}\left(\sum_{d_s=1}^{d_sk-1} \gamma_{d_s}\{\boldsymbol{X_{d_sj}}\} + \sum_{d_c=1}^{d_ck-1} \gamma_{d_c}\{\boldsymbol{X_{d_cij}}\}\right)$$

; and

$$\sum_{p=2}^{k_p-2} s_p.S\left\{\left(\frac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}}\right)\right\}$$

$$= s_p\left\{\begin{array}{c}\dfrac{\left(\left(\dfrac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}}\right) - t_p\right)_+^3}{\left(t_{k_p} - t_1\right)^2} - \dfrac{\left(\left(\dfrac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}}\right) - t_{k_p-1}\right)_+^3 \left(t_{k_p} - t_p\right)}{\left(\left(t_{k_p} - t_{k_p-1}\right)\left(t_{k_p} - t_1\right)^2\right)} \\ + \dfrac{\left(\left(\dfrac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}}\right) - t_{k_p}\right)_+^3 \left(t_{k_p-1} - t_p\right)}{\left(\left(t_{k_p} - t_{k_p-1}\right)\left(t_{k_p} - t_1\right)^2\right)}\end{array}\right\}$$

Individual-specific, f_{SITAR_j}

$$\alpha_{0_{adj}} + \alpha_j + s_1\left\{\left(\frac{age_{ij} - (\beta_{0_{adj}} + \beta_j)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}}\right)\right\} + \sum_{p=2}^{k_p-2} s_p.S\left\{\left(\frac{age_{ij} - (\beta_{0_{adj}} + \beta_j)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}}\right)\right\} \qquad (B\text{-}21)$$

Where $\alpha_{0_{adj}}$, $\beta_{0_{adj}}$ and $\gamma_{0_{adj}}$ are same as above (B-20); and

$$\sum_{p=2}^{k_p-2} s_p.S\left\{\left(\frac{age_{ij} - (\beta_{0_{adj}} + \beta_j)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}}\right)\right\}$$

$$= s_p\left\{\begin{array}{c}\dfrac{\left(\left(\dfrac{age_{ij} - (\beta_{0_{adj}} + \beta_j)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}}\right) - t_p\right)_+^3}{\left(t_{k_p} - t_1\right)^2} - \dfrac{\left(\left(\dfrac{age_{ij} - (\beta_{0_{adj}} + \beta_j)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}}\right) - t_{k_p-1}\right)_+^3 \left(t_{k_p} - t_p\right)}{\left(\left(t_{k_p} - t_{k_p-1}\right)\left(t_{k_p} - t_1\right)^2\right)} \\ + \dfrac{\left(\left(\dfrac{age_{ij} - (\beta_{0_{adj}} + \beta_j)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}}\right) - t_{k_p}\right)_+^3 \left(t_{k_p-1} - t_p\right)}{\left(\left(t_{k_p} - t_{k_p-1}\right)\left(t_{k_p} - t_1\right)^2\right)}\end{array}\right\}$$

First derivative

Population-average, f'_{SITAR_0}

$$\left[s_1 \left\{ 1 \left(\frac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right)^{(1-1)} \right\} + \sum_{p=2}^{k_p-2} s_p . S \left\{ 3 \left(\frac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right)^{(3-1)} \right\} \right] e^{\gamma_{0adj}} \qquad (B\text{-}22)$$

Where $\beta_{0_{adj}}$ and $\gamma_{0_{adj}}$ are:

$$\beta_{0_{adj}} = \beta_0 + \sum_{d_s=1}^{d_s k-1} \beta_{d_s} \{ \boldsymbol{X_{d_s j}} \} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c} \{ \boldsymbol{X_{d_c ij}} \} - \frac{1}{n_j} \sum_{1}^{n_j} \left(\sum_{d_s=1}^{d_s k-1} \beta_{d_s} \{ \boldsymbol{X_{d_s j}} \} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c} \{ \boldsymbol{X_{d_c ij}} \} \right)$$

$$\gamma_{0_{adj}} = \gamma_0 + \sum_{d_s=1}^{d_s k-1} \gamma_{d_s} \{ \boldsymbol{X_{d_s j}} \} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c} \{ \boldsymbol{X_{d_c ij}} \} - \frac{1}{n_j} \sum_{1}^{n_j} \left(\sum_{d_s=1}^{d_s k-1} \gamma_{d_s} \{ \boldsymbol{X_{d_s j}} \} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c} \{ \boldsymbol{X_{d_c ij}} \} \right)$$

; and

$$\sum_{p=2}^{k_p-2} s_p . S \left\{ 3 \left(\frac{age_{ij} - \beta_{0_{adj_0}}}{e^{-\gamma_{0adj}}} \right)^{(3-1)} \right\}$$

$$= s_p \left\{ \begin{array}{c} \dfrac{3 \left(\left(\dfrac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right) - t_p \right)_+^{(3-1)}}{\left(t_{k_p} - t_1 \right)^2} - \dfrac{3 \left(\left(\dfrac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right) - t_{k_p-1} \right)_+^{(3-1)} \left(t_{k_p} - t_p \right)}{\left(\left(t_{k_p} - t_{k_p-1} \right) \left(t_{k_p} - t_1 \right)^2 \right)} \\ + \dfrac{3 \left(\left(\dfrac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right) - t_{k_p} \right)_+^{(3-1)} \left(t_{k_p-1} - t_p \right)}{\left(\left(t_{k_p} - t_{k_p-1} \right) \left(t_{k_p} - t_1 \right)^2 \right)} \end{array} \right\}$$

Individual-specific, $\boldsymbol{f'_{SITAR_j}}$

$$\left[s_1\left\{1\left(\frac{age_{ij}-\left(\beta_{0_{adj}}+\beta_j\right)}{e^{-\left(\gamma_{0_{adj}}+\gamma_j\right)}}\right)^{(1-1)}\right\} + \sum_{p=2}^{k_p-2} s_p.S\left\{3\left(\frac{age_{ij}-\left(\beta_{0_{adj}}+\beta_j\right)}{e^{-\left(\gamma_{0_{adj}}+\gamma_j\right)}}\right)^{(3-1)}\right\}\right] e^{\left(\gamma_{0_{adj}}+\gamma_j\right)} \quad (B\text{-}23)$$

Where $\beta_{0_{adj}}$ and $\gamma_{0_{adj}}$ are:

$$\beta_{0_{adj}} = \beta_0 + \sum_{d_s=1}^{d_s k-1}\beta_{d_s}\{\boldsymbol{X_{d_s j}}\} + \sum_{d_c=1}^{d_c k-1}\beta_{d_c}\{\boldsymbol{X_{d_c ij}}\} - \frac{1}{n_j}\sum_{1}^{n_j}\left(\sum_{d_s=1}^{d_s k-1}\beta_{d_s}\{\boldsymbol{X_{d_s j}}\} + \sum_{d_c=1}^{d_c k-1}\beta_{d_c}\{\boldsymbol{X_{d_c ij}}\}\right)$$

$$\gamma_{0_{adj}} = \gamma_0 + \sum_{d_s=1}^{d_s k-1}\gamma_{d_s}\{\boldsymbol{X_{d_s j}}\} + \sum_{d_c=1}^{d_c k-1}\gamma_{d_c}\{\boldsymbol{X_{d_c ij}}\} - \frac{1}{n_j}\sum_{1}^{n_j}\left(\sum_{d_s=1}^{d_s k-1}\gamma_{d_s}\{\boldsymbol{X_{d_s j}}\} + \sum_{d_c=1}^{d_c k-1}\gamma_{d_c}\{\boldsymbol{X_{d_c ij}}\}\right)$$

; and

$$\sum_{p=2}^{k_p-2} s_p.S\left\{3\left(\frac{age_{ij}-\left(\beta_{0_{adj}}+\beta_j\right)}{e^{-\left(\gamma_{0_{adj}}+\gamma_j\right)}}\right)^{(3-1)}\right\}$$

$$\left\{\begin{array}{l} \dfrac{3\left(\left(\dfrac{age_{ij}-\left(\beta_{0_{adj}}+\beta_j\right)}{e^{-\left(\gamma_{0_{adj}}+\gamma_j\right)}}\right)-t_p\right)_+^{(3-1)}}{\left(t_{k_p}-t_1\right)^2} \\ -\dfrac{3\left(\left(\dfrac{age_{ij}-\left(\beta_{0_{adj}}+\beta_j\right)}{e^{-\left(\gamma_{0_{adj}}+\gamma_j\right)}}\right)-t_{k_p-1}\right)_+^{(3-1)}\left(t_{k_p}-t_p\right)}{\left(\left(t_{k_p}-t_{k_p-1}\right)\left(t_{k_p}-t_1\right)^2\right)} \\ +\dfrac{3\left(\left(\dfrac{age_{ij}-\left(\beta_{0_{adj}}+\beta_j\right)}{e^{-\left(\gamma_{0_{adj}}+\gamma_j\right)}}\right)-t_{k_p}\right)_+^{(3-1)}\left(t_{k_p-1}-t_p\right)}{\left(\left(t_{k_p}-t_{k_p-1}\right)\left(t_{k_p}-t_1\right)^2\right)} \end{array}\right\}$$

Second derivative

Population-average, f''_{SITAR_0}

$$\left[s_1 \left\{ 1(1-1) \left(\frac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right)^{(1-2)} \right\} + \sum_{p=2}^{k_p-2} s_p . S \left\{ 3(3-1) \left(\frac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right)^{(3-2)} \right\} \right] e^{\left(\gamma_{0_{adj}}\right)^2} \quad (B\text{-}24)$$

Where $\beta_{0_{adj}}$ and $\gamma_{0_{adj}}$ are:

$$\beta_{0_{adj}} = \beta_0 + \sum_{d_s=1}^{d_s k-1} \beta_{d_s} \{ \boldsymbol{X_{d_s j}} \} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c} \{ \boldsymbol{X_{d_c ij}} \} - \frac{1}{n_j} \sum_{1}^{n_j} \left(\sum_{d_s=1}^{d_s k-1} \beta_{d_s} \{ \boldsymbol{X_{d_s j}} \} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c} \{ \boldsymbol{X_{d_c ij}} \} \right)$$

$$\gamma_{0_{adj}} = \gamma_0 + \sum_{d_s=1}^{d_s k-1} \gamma_{d_s} \{ \boldsymbol{X_{d_s j}} \} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c} \{ \boldsymbol{X_{d_c ij}} \} - \frac{1}{n_j} \sum_{1}^{n_j} \left(\sum_{d_s=1}^{d_s k-1} \gamma_{d_s} \{ \boldsymbol{X_{d_s j}} \} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c} \{ \boldsymbol{X_{d_c ij}} \} \right)$$

; and

$$\sum_{p=2}^{k_p-2} s_p . S \left\{ 3(3-1) \left(\frac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right)^{(3-2)} \right\}$$

$$= s_p \left\{ \begin{array}{l} \dfrac{3(3-1) \left(\left(\dfrac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right) - t_p \right)_+^{(3-2)}}{\left(t_{k_p} - t_1 \right)^2} \\ - \dfrac{3(3-1) \left(\left(\dfrac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right) - t_{k_p-1} \right)_+^{(3-2)} \left(t_{k_p} - t_p \right)}{\left(\left(t_{k_p} - t_{k_p-1} \right) \left(t_{k_p} - t_1 \right)^2 \right)} \\ + \dfrac{3(3-1) \left(\left(\dfrac{age_{ij} - \beta_{0_{adj}}}{e^{-\gamma_{0adj}}} \right) - t_{k_p} \right)_+^{(3-2)} \left(t_{k_p-1} - t_p \right)}{\left(\left(t_{k_p} - t_{k_p-1} \right) \left(t_{k_p} - t_1 \right)^2 \right)} \end{array} \right\}$$

Individual-specific, $\boldsymbol{f''_{SITAR_0}}$

$$\left[s_1 \left\{ 1(1-1) \left(\frac{age_{ij} - (\beta_{0_{adj}} + \beta_j)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}} \right)^{(1-2)} \right\} + \sum_{p=2}^{k_p-2} s_p . S \left\{ 3(3-1) \left(\frac{age_{ij} - \left(\beta_{0_{adj}} + \beta_j\right)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}} \right)^{(3-2)} \right\} \right] e^{\left(\gamma_{0_{adj}} + \gamma_j\right)^2} \quad (B\text{-}25)$$

Where $\beta_{0_{adj}}$ and $\gamma_{0_{adj}}$ are:

$$\beta_{0_{adj}} = \beta_0 + \sum_{d_s=1}^{d_s k-1} \beta_{d_s}\{\boldsymbol{X_{d_s j}}\} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c}\{\boldsymbol{X_{d_c ij}}\} - \frac{1}{n_j} \sum_{1}^{n_j} \left(\sum_{d_s=1}^{d_s k-1} \beta_{d_s}\{\boldsymbol{X_{d_s j}}\} + \sum_{d_c=1}^{d_c k-1} \beta_{d_c}\{\boldsymbol{X_{d_c ij}}\} \right)$$

$$\gamma_{0_{adj}} = \gamma_0 + \sum_{d_s=1}^{d_s k-1} \gamma_{d_s}\{\boldsymbol{X_{d_s j}}\} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c}\{\boldsymbol{X_{d_c ij}}\} - \frac{1}{n_j} \sum_{1}^{n_j} \left(\sum_{d_s=1}^{d_s k-1} \gamma_{d_s}\{\boldsymbol{X_{d_s j}}\} + \sum_{d_c=1}^{d_c k-1} \gamma_{d_c}\{\boldsymbol{X_{d_c ij}}\} \right)$$

; and

$$\sum_{p=2}^{k_p-2} s_p . S \left\{ 3(3-1) \left(\frac{age_{ij} - \left(\beta_{0_{adj}} + \beta_j\right)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}} \right)^{(3-2)} \right\}$$

$$= s_p \left\{ \begin{array}{l} \dfrac{3(3-1) \left(\left(\dfrac{age_{ij} - \left(\beta_{0_{adj}} + \beta_j\right)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}} \right) - t_p \right)_+^{(3-2)}}{\left(t_{k_p} - t_1\right)^2} \\ - \dfrac{3(3-1) \left(\left(\dfrac{age_{ij} - \left(\beta_{0_{adj}} + \beta_j\right)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}} \right) - t_{k_p-1} \right)_+^{(3-2)} \left(t_{k_p} - t_p\right)}{\left(\left(t_{k_p} - t_{k_p-1}\right) \left(t_{k_p} - t_1\right)^2 \right)} \\ + \dfrac{3(3-1) \left(\left(\dfrac{age_{ij} - \left(\beta_{0_{adj}} + \beta_j\right)}{e^{-(\gamma_{0_{adj}} + \gamma_j)}} \right) - t_{k_p} \right)_+^{(3-2)} \left(t_{k_p-1} - t_p\right)}{\left(\left(t_{k_p} - t_{k_p-1}\right) \left(t_{k_p} - t_1\right)^2 \right)} \end{array} \right\}$$

Preece-Baines (PB)

Distance

Population-average, $\boldsymbol{f_{PB_0}}$

$$s_{max_{adj}} - \frac{2\left(s_{max_{adj}} - s_{\theta_{adj}}\right)}{e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}} \tag{B-26}$$

Where $s_{max_{adj}}$, $s_{\theta_{adj}}$, θ_{adj}, $s_{0_{adj}}$ and $s_{1_{adj}}$ are:

$$s_{max_{adj}} = s_{max} + \sum_{d_s=1}^{d_s k-1}\left\{s_{max_{d_s}}\right\}X_{d_s ij} + \sum_{d_c=1}^{d_c k-1}\left\{s_{max_{d_c}}\right\}X_{d_c ij}$$

$$s_{\theta_{adj}} = s_{\theta} + \sum_{d_s=1}^{d_s k-1}\left\{s_{\theta_{d_s}}\right\}X_{d_s ij} + \sum_{d_c=1}^{d_c k-1}\left\{s_{\theta_{d_c}}\right\}X_{d_c ij}$$

$$\theta_{adj} = \theta + \sum_{d_s=1}^{d_s k-1}\{\theta_{d_s}\}X_{d_s ij} + \sum_{d_c=1}^{d_c k-1}\{\theta_{d_c}\}X_{d_c ij}$$

$$s_{0_{adj}} = s_0 + \sum_{d_s=1}^{d_s k-1}\left\{s_{0_{d_s}}\right\}X_{d_s ij} + \sum_{d_c=1}^{d_c k-1}\left\{s_{0_{d_c}}\right\}X_{d_c ij}$$

$$s_{1_{adj}} = s_1 + \sum_{d_s=1}^{d_s k-1}\left\{s_{1_{d_s}}\right\}X_{d_s ij} + \sum_{d_c=1}^{d_c k-1}\left\{s_{1_{d_c}}\right\}X_{d_c ij}$$

Individual-specific, $\boldsymbol{f_{PB_j}}$

$$\left(s_{max_{adj}} + s_{max_j}\right) - \frac{2\left(\left(s_{max_{adj}} + s_{max_j}\right) - \left(s_{\theta_{adj}} + s_{\theta_j}\right)\right)}{e^{\left(s_{0_{adj}}+s_{0_j}\right)\left(age_{ij}-\left(\theta_{adj}+\theta_j\right)\right)} + e^{\left(s_{1_{adj}}+s_{1_j}\right)\left(age_{ij}-\left(\theta_{adj}+\theta_j\right)\right)}} \tag{B-27}$$

Where $s_{max_{adj}}$, $s_{\theta_{adj}}$, θ_{adj}, $s_{0_{adj}}$ and $s_{1_{adj}}$ are same as (B-26).

First derivative

Population-average, $\boldsymbol{f'_{PB_0}}$

$$\frac{2\left(s_{max_{adj}} - s_{\theta_{adj}}\right)}{e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}} \times \left(\frac{s_{0_{adj}} e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + s_{1_{adj}} e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}}{e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}} \right) \quad (B\text{-}28)$$

Where $s_{max_{adj}}$, $s_{\theta_{adj}}$, θ_{adj}, $s_{0_{adj}}$ and $s_{1_{adj}}$ are same as (B-26).

Individual-specific, $\boldsymbol{f'_{PB_j}}$

$$\frac{2\left((s_{max_{adj}} + s_{max_j}) - (s_{\theta_{adj}} + s_{\theta_j})\right)}{e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}} \times \left(\frac{(s_{0_{adj}} + s_{0_j}) e^{\left(s_{0_{adj}}+s_{0_j}\right)(age_{ij}-\theta_{adj})} + (s_{1_{adj}} + s_{1_j}) e^{\left(s_{1_{adj}}+s_{1_j}\right)(age_{ij}-\theta_{adj})}}{e^{\left(s_{0_{adj}}+s_{0_j}\right)(age_{ij}-\theta_{adj})} + e^{\left(s_{1_{adj}}+s_{1_j}\right)(age_{ij}-\theta_{adj})}} \right) \quad (B\text{-}29)$$

Where $s_{max_{adj}}$, $s_{\theta_{adj}}$, θ_{adj}, $s_{0_{adj}}$ and $s_{1_{adj}}$ are same as (B-26).

Second derivative

Population-average, f''_{PB_0}

$$\frac{2\left(s_{max_{adj}} - s_{\theta_{adj}}\right)}{e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}} \times \left\{ -2\left(\frac{s_{0_{adj}} e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + s_{1_{adj}} e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}}{e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}} \right)^2 + \left(s_{0_{adj}} + s_{1_{adj}}\right)\left(\frac{s_{0_{adj}} e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + s_{1_{adj}} e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}}{e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}} \right) - \left(s_{0_{adj}} \times s_{1_{adj}}\right) \right\} \quad (B\text{-}30)$$

Where $s_{max_{adj}}$, $s_{\theta_{adj}}$, θ_{adj}, $s_{0_{adj}}$ and $s_{1_{adj}}$ are same as (B-26).

Individual-specific, f''_{PBj}

$$\frac{2\left((s_{max_{adj}} + s_{max_j}) - (s_{\theta_{adj}} + s_{\theta_j})\right)}{e^{s_{0_{adj}}(age_{ij}-\theta_{adj})} + e^{s_{1_{adj}}(age_{ij}-\theta_{adj})}} \times \left\{ -2\left(\frac{(s_{0_{adj}} + s_{0_j}) e^{(s_{0_{adj}}+s_{0_j})(age_{ij}-\theta_{adj})} + (s_{1_{adj}} + s_{1_j}) e^{(s_{1_{adj}}+s_{1_j})(age_{ij}-\theta_{adj})}}{e^{\left(s_{0_{adj}}+s_{0_j}\right)(age_{ij}-\theta_{adj})} + e^{\left(s_{1_{adj}}+s_{1_j}\right)(age_{ij}-\theta_{adj})}} \right)^2 + \left((s_{0_{adj}} + s_{0_j}) + (s_{1_{adj}} + s_{1_j})\right) \times \left(\frac{(s_{0_{adj}} + s_{0_j}) e^{\left(s_{0_{adj}}+s_{0_j}\right)(age_{ij}-\theta_{adj})} + (s_{1_{adj}} + s_{1_j}) e^{\left(s_{1_{adj}}+s_{1_j}\right)(age_{ij}-\theta_{adj})}}{e^{\left(s_{0_{adj}}+s_{0_j}\right)(age_{ij}-\theta_{adj})} + e^{\left(s_{1_{adj}}+s_{1_j}\right)(age_{ij}-\theta_{adj})}} \right) - (s_{0_{adj}} + s_{0_j}) \times (s_{1_{adj}} + s_{1_j}) \right\} \quad (B\text{-}31)$$

Where $s_{max_{adj}}$, $s_{\theta_{adj}}$, θ_{adj}, $s_{0_{adj}}$ and $s_{1_{adj}}$ are same as (B-26).

B.2. Software

Chapter 4 reviewed software available for fitting linear and nonlinear GCMs, derivative estimation and to estimate APGV and the PGV. As stated earlier in Chapter 4, I work primarily within the Stata software and call required specialized MLwiN (for CP, FP and RCS GCMs) and R (for SITAR GCM) software from within the Stata. While I use available 'runmlwin' command to call MLwiN software, I have written my own command to call R software.

As no software package was available to estimate covariate-adjusted derivatives, I wrote a set of post-estimation Stata programs to implement my methodology of estimating covariate-adjusted derivatives (velocity and acceleration) and the adolescent growth spurt parameters (APGV and PGV). As a package, these programs also estimate covariate-adjusted distance.

As post-estimation programs required key information on model fitting, I wrote another set of programs which create required design matrix for fitting CP, FP, RCS and SITAR GCMs and pass on key information to post-estimation Stata programs such as the degree of conventional polynomials (CP), degree and powers of fractional polynomials (FP), and the number and location of knots (RCS and SITAR). The information required to estimate derivative for the PB GCM was directly recovered from the list returned by Stata's 'menl' command.

To estimate adolescent growth spurt parameters (APGV and PGV), a wrote another Stata program which utilize the covariate adjusted first and second derivatives estimated by my post-estimation programs for the CP, FP, RCS and SITAR GCMs.

Lastly, to implement nonparametric bootstrapping of linear GCMs (CP, RCS and SITAR) I adapted and modified Stata's 'bootstrap' prefix command to be compatible with 'runmlwin' command. As Stata's 'bootstrap' prefix command is not allowed with its 'menl' command, which is I used to fit PB GCM, I wrote a new program which is based on 'bootstrap' prefix command. For

the SITAR GCM, I used recently introduced 'apgv_se' function by Tim Cole (2019) for his 'sitar' R package.

Below I describe in detail working of software used for model fitting, estimating distance and derivatives, estimation of adolescent growth spurt parameters, bootstrapping and summarising results presented in Chapter 6.

B.2.1. **Model fitting**

Linear growth curve models

The linear GCMs (CP, FP and RCS) were fitted using the MLwiN software (version 3.01) (Charlton et al., 2017) calling it from within Stata software (version 15.1) (StataCorp, 2017a) via the user-written 'runmlwin' command (Leckie & Charlton, 2013b). All three GCMs were estimated using the restricted iterative generalized least squares (RIGLS) method as implemented in the MLwiN software. The RIGLS is equivalent to the restricted maximum likelihood (REML) method of estimation (Goldstein, 1989b). To fit multiple GCMs (e.g., different degree of polynomials for CP and FP, and number of knots for the RCS) to find best-fitting CP, FP and RCS, I wrote and implemented custom loops in a Stata.

To pass on the key model fitting information to post-estimation program (see below) for distance and derivatives estimation, I wrote three commands ('gencp', 'genfp' and 'gensp') to generate conventional polynomials (CP), fractional polynomials (FP) and restricted cubic spline (RCS).

The 'gencp' generates conventional polynomials using use-specified degree of polynomials (e.g., cp1, cp2 and so on) and whether to centre age (mean centring or any other value) or to use the original scale of age. The 'gencp' leaves behind the key information such as the name of polynomials generated (e.g., cp1, cp2 and so on), the name of the predictor (i.e., age) and the value used for centring the predictor.

The ‘genfp’ and ‘gensp’ are based on Stata’s ‘fp generate’ and ‘mkspline’ commands, respectively. User can specify all options available for ‘fp generate’ and ‘mkspline’ as both these commands work the same way as ‘fp generate’ and ‘mkspline’ with the exception that they leave behind key information used to generate fractional polynomials and restricted cubic splines.

The ‘genfp’ stores information the name of fractional polynomials generated (e.g., fp1, fp2 and so on), the name of the predictor (i.e., age), powers used to generate fractional polynomials, the value used for centring the predictor as well as well as scaling factors (scale a and scale b) used to scale the predictor before generating fractional polynomials.[10]

The ‘gensp’ leaves information on the name of restricted cubic spline terms generated (e.g., sp1, sp2 and so on), the name of the predictor (i.e., age), the number of knots, location of knots, the value used for centring the predictor etc.

Nonlinear growth curve models

To fit PB GCM, I used a recently introduced ‘menl’ command in Stata software (version 15 onwards). The ‘menl’ implements linearization method (Lindstrom & Bates, 1990; Pinheiro & Bates, 1995a). The REML method was used for estimating the PB GCM. Chapter 4 provides details on the linearization method as well as the REML method. To get better initial values and to enhance the chances of model convergence, I followed the following strategy to fit the PB GCM.

In the first step, I fit a fixed-effect PB GCM using initial values recommended by Hauspie et al. (2004, p. 216), which are as follows. The parameter dmax (size at adulthood) is set as the sex-specific maximum value for adult size (age 18 years in my study). The starting value for dtheta (size at peak growth velocity) is calculated as 90 per cent of the dmax (0.9 × dmax). For theta (age at peak growth velocity), authors (Hauspie et al., 2004) recommend a rough estimate of 12 years for

[10] See https://www.stata.com/manuals/rfp.pdf for details.

females, and 14 years for males. I followed the same initial values for theta while fitting PB GCM to male (14 years) and female data (12 years). However, I tried different age ranges for theta (11-13 years for females; and 12-14 years for males) if model did not converge. For height, the starting values for parameter s0 (pre-adolescent growth rate) and s1 (adolescent growth rate) are usually set to 0.1 and 1.2, respectively. I followed the same initial values for s0 and s1 parameters. In case model failed to converge, I tried a range of 0.08 to 0.12 for s0, and 0.9 to 1.5 for s1.

In the second step, I used parameters estimated by the above fitted fixed-effect PB GCM as the initial values for regression parameters for the random-effect model. To improve convergence and to speed-up the estimation, I added random effects (Dmax, Dtheta, Theta, S1 and S0) in a sequence using a custom loop written to fit models of increasing complexity wherein the fixed and the random effect parameter estimated from a preceding simpler model (e.g. a model with only Dmax and Dtheta as random effects) are used as initial values for the next complex model of increased complexity (e.g. a model with Dmax, Dtheta and Theta as random effects) and so on.

The SITAR GCM was fitted using the well-documented 'sitar' R package (version 1.0.10) (Cole, 2017a). The 'sitar' package calls the 'nlme' package (Pinheiro et al., 2017) to fit nonlinear mixed effect models. Like Stata's 'menl' command, the 'nlme' implements the linearization method for model estimation (Lindstrom & Bates, 1990; Pinheiro & Bates, 1995a). I used the REML method to estimate the SITAR GCM.

As reported earlier in Chapter 4 (see Section 4.2.3) and explained in Chapter 6 (see Section 6.2.8), I constructed restricted cubic splines design matrix using truncated spline basis instead of B-spline basis matrix as implemented in the 'sitar' package. Therefore, I modified the 'sitar' package which now calls the 'rms' R package written by Harrell Jr (2017) instead of the 'splines' package (R Core Team, 2017). The 'rms' construct restricted cubic spline functions using truncated power

basis matrices similar to Stata's 'mkspline' command. The 'ns' function in the 'splines' R package generates B-spline basis matrix for a restricted cubic spline (see Chapter 6, Section 6.2.8).

To fit the SITAR GCM from within the Stata, I wrote two programs. The first program, the 'sitarR', writes R script and then calls the R software (version 3.4.3) (R Core Team, 2017) via Stata's 'shell' command to execute the R script. The second program 'getsitarR' imports back results of the fitted SITAR GCM from R into the Stata. The results imported in the Stata environment include fixed (regression coefficients) effects, individual-specific random effects, variance-covariance matrix, level-1 residuals and the information used in constructing the restricted cubic splines such the name of restricted cubic spline terms (e.g., sp1, sp2 and so on), the name of the predictor (i.e., age), the number of knots, location of knots, the value used for centring the predictor etc..).

From this point onwards i.e. after getting back results of the fitted SITAR GCM from R into the Stata, estimation of derivatives as well as APGV and the PGV using post-estimation programs (see below) is same as for other four GCMs (CP, FP, RCS and PB) which are fitted directly within the Stata.

B.2.2. **Distance and derivatives estimation**

To implement covariate-adjusted growth trajectories (distance, velocity and acceleration) estimation as described in Chapter 6 (Section 6.2.5), I wrote a set of Stata programs. The main program, the 'predictgcm', calls five subprograms i.e., each for the five GCMs (CP, FP, RCS, SITAR and PB) to estimate population-average and individual-specific growth trajectories for CP, FP, RCS, SITAR and PB GCMs.

The 'predictgcm' program works as follows. The syntax of the main program allows options such as the model for which distance and derivatives to be estimated (e.g. CP, FP etc), covariates (e.g. study and class) and whether to estimate unadjusted or covariate adjusted population-average

and/or individual-specific growth trajectories. The 'predictgcm' then pass on the use-specified options to relevant subprogram (depending on the model option) which then estimates growth trajectories by using Stata's 'predictnl' command.

Each subprogram first automatically recovers the information stored by the 'gencp', 'genfp' and 'gensp' (see B.2.1) and then use this information to construct exact same conventional polynomials (CP), fractional polynomials (FP), or restricted cubic splines (RCS and SITAR) design matrix as was used for fitting these GCMs. For the PB GCM, required information such as predictor (age) and value used for centring the predictor (not relevant here as I fit model using original metric of age) are directly used from the list returned by Stata's 'menl' command. For linear GCMs (CP, FP and RCS) programs are compatible with 'runmlwin' and Stata's 'mixed' command.

B.2.3. **Timing and intensity of the adolescent growth spurt**

To estimate timing (APGV) and intensity (PGV) of the adolescent growth spurt, I used the quadratic function method (see Chapter 6, Section 6.2.6; and Chapter 4 Section 4.4.2).

To implement the quadratic function method for estimating APGV and PGV, I wrote a new Stata program, the 'getpeak'. The 'getpeak' internally calls another subprogram, the 'iage', which prepares the dataset required for the quadratic function method (see Chapter 4, Section 4.4.2 for details).

The 'getpeak' estimates APGV and the PGV using the first and second derivatives estimated by the 'predictgcm' program. The population-average APGV and PGV are estimated using the population-average derivatives whereas individual-specific APGV and the PGV estimates are estimated using the individual-specific derivatives.

B.2.4. **Bootstrapped confidence intervals**

Linear growth curve models

I used Stata's machinery to implement nonparametric bootstrapping linear GCMs (CP, FP and RCS). Stata's 'bootstrap' prefix with 'mixed' command (which fits the linear mixed effects models) implements nonparametric bootstrapping (by cluster with replacement) (StataCorp, 2017a). The key options such as the 'idcluster' and the 'newgrouped' maintain the hierarchal structure of the data while resampling (StataCorp, 2017a). For resampling and model fitting, the 'bootstrap' prefix internally calls the '_loop_bs' command which loops over the requested number of replications.

As I fit linear GCMs using the 'runmlwin' command (Leckie & Charlton, 2013b) and use my custom-written programs such 'predictgcm' and 'getpeak' to estimate growth trajectories (distance and derivatives) and the adolescent growth spurt parameters (APGV and PGV), I modified the 'bootstrap' prefix in the following ways. First, I made it compatible with the 'runmlwin' command. Second, I customized the internal '_loop_bs' command, which is called by the 'bootstrap' prefix, to run my programs ('predictgcm' and 'getpeak') at each replication and save growth trajectories and the adolescent growth spurt parameters estimated by these programs. Below I explain changes I made to the 'bootstrap' prefix and the '_loop_bs' commands.

An initial trial run of using 'bootstrap' prefix with 'runmlwin' showed errors as Stata was not accepting the 'runmlwin' returned variance-covariance matrix names. To overcome the problem of for name conflicts, I wrote my new prefix command, 'bootstrap_runmlwin' which works exactly same as Stata's 'bootstrap' prefix except that it accepts the variance-covariance matrix names returned by the 'runmlwin' command.

I then added an option 'EXEcutecommands' to the 'bootstrap_runmlwin' prefix command. The 'EXEcutecommands' accepts string argument comprised of valid post-estimation Stata

commands (e.g., 'predict') or use-written commands (e.g., 'predictgcm'). Any number of commands separated by semicolon ';' can be combined as a string.

Based on Stata's '_loop_bs' command, I wrote '_loop_bs_runmlwin' command which is exactly same as the '_loop_bs' except that it receives the string comprised of valid post-estimation commands passed to it by the 'bootstrap_runmlwin' prefix (via 'EXEcutecommands' option) and then execute each command at each iteration of resampling by splitting the string at the semicolon ';' location. At each iteration, data are resampled, the GCM is fitted and each of the post-estimation commands specified by the user are executed. The resample data along with the newly created variable(s) such predictions on the resample data are saved and appended.

As an example, to estimate population-average and individual-specific growth trajectories (distance, velocity and acceleration) and the adolescent growth spurt parameters (APGV and PGV) for 1000 bootstrapps, I use 'bootstrap_runmlwin' prefix to the 'runmlwin' command as follows: bootstrap_runmlwin, seed(1234) exe(predictgcm, options; getpeak, options) rep(1000) cluster(id) idcluster(newid) saving(bootdata ,every(1) replace) : runmlwin…..

In the above example, the 'bootstrap_runmlwin' pass on the string comprised of the post-estimation commands ("predictgcm, options; getpeak, options") to the internal '_loop_bs_runmlwin' command which then resamples the data, fits the runmlwin model (CP, FP or RCS) and then splits the string "predictgcm, options; getpeak, options" at ';' location and first executes the 'predictgcm' program and then 'getpeak'. The saving option with 'every (1)' sub-option saves bootstrapped data along with growth trajectories (distance, velocity, acceleration) and the adolescent growth spurt parameters (APGV and PGV) estimated for each iteration. The 2.5th and 97.5th percentiles of APGV and the PGV are used as 95 per cent confidence intervals (see Chapter 6, Section 6.2.7).

Nonlinear growth curve models

I used Stata's 'menl' command for fitting the PB GCM. Unlike the 'mixed' command, Stata's 'bootstrap' prefix is not allowed with the 'menl' command. Therefore, I wrote my own bootstrap program to implement nonparametric bootstrap for the PB GCM. The basic structure of the program is same as described above.

Recently, Tim Cole (2019) added a function 'apv_se' to his 'sitar' R package (version 1.1.1) for estimating bootstrapped standard error for the APGV and the PGV. For resampling, the 'apv_se' function calls the 'bootstrap' function from the 'rsample' R package (Kuhn & Wickham, 2019).

The 'sitar' package reference manual does not mention whether 'apv_se' implements parametric, semiparametric or nonparametric bootstrapping. However, the 'rsample' package reference manual states that the 'bootstrap' function resample data with replacement (Kuhn & Wickham, 2019, p. 4).

The 'apv_se' function return the standard deviation of bootstrapped estimates as standard error of APGV and the PGV. The CIs can be computed by using the standard error. However, I made no distributional assumptions and therefore report percentile-based CIs (see Chapter 6, Section 6.2.7). For this, I extracted bootstrapped estimates of APGV and PGV (for each iteration) and then computed 2.5th and 97.5th percentiles as CIs for APGV and the PGV.

B.2.5. **Summarizing and presenting results**

Except for estimating the Spearman's rank correlation between individual-specific APGV and the PGV, I use Stata's built-in commands to plot figures, to get summary statistics (mean, standard deviation, median, interquartile range etc.) and to compute 2.5th and 97.5th percentiles of bootstrapped APGV and the PGV as 95% CIs.

The Spearman's rank correlation coefficient is a Pearson correlation coefficient calculated with the ranks of the variables instead of their actual values (Schober et al., 2018). Stata's 'spearman' command for computing Spearman's rank correlation does not provide confidence intervals.

I utilized a user-written command 'corrci' to compute Spearman's rank correlation and associated 95% CIs (Cox, 2008). The 'corrci' calculates Pearson correlations and CIs using Fisher's z transform. I run 'corrci' command on rank transformed APGV and PGV. The rank transformation was done by using the 'rank' function of Stata' 'egen' command.

B.3. Results

B.3.1. Sensitivity analysis

Effect of different spline basis on the fit of the SITAR model

Comparison of SITAR GCM fitted to the upper and lower jaw length and total face height using the truncated spline basis and the B-spline basis to construct the cubic spline design matrix is shown in Table B-1.

Table B-1. Comparison of superimposition by translation and rotation (SITAR) growth curve model applied to male and female data using B-spline basis and the truncated spline basis.

Outcome	SITAR model fit using the truncated spline basis to construct the cubic spline design matrix.					SITAR model fit using the B-spline basis to construct the cubic spline design matrix.				
	Knots	AIC	RSD	APGV	PGV	Knots	AIC	RSD	APGV	PGV
Male										
Upper jaw length (COPAD)	4	3774.42	1.01	12.62	2.88	4	3769.01	1.01	12.57	2.92
Lower jaw length (COPOD)	7	4336.22	1.23	13.35	4.76	7	4356.40	1.23	13.32	4.56
Total face height (TFHNP)	7	4335.07	1.23	13.43	4.75	7	4366.32	1.24	13.36	4.61
Female										
Upper jaw length (COPAD)	4	3399.04	0.79	11.14	2.21	4	3394.90	0.79	11.13	2.26
Lower jaw length (COPOD)	5	4275.69	1.02	12.16	3.49	5	4300.73	1.02	12.21	4.06
Total face height (TFHNP)	5	4114.11	0.93	12.43	3.28	5	4115.25	0.93	12.37	3.33

Knots: number of knots

AIC Akaike information criterion; RSD residual standard deviation in millimetres

APGV age at peak growth velocity in years; PGV peak growth velocity in mm/year

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Except for the upper jaw length data for both males and females, the AIC estimates are lower for the model using the truncated spline basis instead of the B-spline basis. These differences are substantial (based on AIC values) for lower jaw length and total face height for data males, and lower jaw length data for females. The AIC values are almost same for the total face height data for females.

For upper jaw length data for both sexes, the AIC for model using the truncated spline basis is higher (around 5 AIC units) than model which uses the B-spline basis for the construct the cubic spline design matrix. The difference in APGV estimated by these two approaches is less than 0.07 year for any of three outcomes for both sexes. The difference in PGV is slightly greater especially for the lower jaw length for females (0.57 mm/year). For all other outcomes, the difference is less than or equal to 0.20 mm/year.

Optimizing the SITAR model

Results show that the transformations of age and/or outcomes have no influence on the number of knots for the best-fitting SITAR GCM. The AIC and the RSD estimates for the best-fitting SITAR GCMs with and without log and/or square root transformation of age and each of the three outcome for males (Table B-2) and females (Table B-3) show that the SITAR GCM fits better to the jaw growth data with original scales of age and each of the three outcome i.e., upper jaw length, lower jaw length and total face height.

Table B-2. Comparison of superimposition by translation and rotation (SITAR) growth curve model applied to male data with and without transformation of age and outcomes.

Age	Upper jaw length (COPAD)				Lower jaw length (COPOD)				Total face height (TFHNP)			
	Outcome	Knots	AIC	RSD	Outcome	Knots	AIC	RSD	Outcome	Knots	AIC	RSD
original	original*	4	3774.42	1.01	original*	7	4336.22	1.23	original*	7	4335.07	1.23
original	sqrt	7	3851.90	0.05	sqrt	7	4435.14	0.06	sqrt	7	4422.50	0.06
original	log				log				log			
sqrt	original	7	3800.21	1.01	original	7	4376.51	1.25	original	7	4338.60	1.20
log	original				original				original			
log	log				log				log			
sqrt	sqrt				sqrt				sqrt			
sqrt	log				log				log			
log	sqrt				sqrt				sqrt			

Missing rows indicate model did not converge.

Transformation: original: untransformed; sqrt: square root transformation; log: log transformation

AIC Akaike information criterion; RSD residual standard deviation in millimetres on transformed outcome scale

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Knots: number of knots

Table B-3. Comparison of superimposition by translation and rotation (SITAR) growth curve model applied to female data with and without transformation of age and outcomes.

Age	Upper jaw length (COPAD)				Lower jaw length (COPOD)				Total face height (TFHNP)			
	Outcome	Knots	AIC	RSD	Outcome	Knots	AIC	RSD	Outcome	Knots	AIC	RSD
original	original*	4	3399.04	3399.04	original*	5	4275.69	1.02	original*	5	4114.11	0.93
original	sqrt	4	3486.61	3486.61	sqrt	5	4458.40	0.05	sqrt	5	4207.33	0.04
original	log				log				log			
sqrt	original	4	3437.13	3437.13	original	5	4306.30	1.02	original	5	4213.14	1.12
log	original				original				original			
log	log				log				log			
sqrt	sqrt				sqrt				sqrt			
sqrt	log				log				log			
log	sqrt				sqrt				sqrt			

Missing rows indicate model did not converge.

Transformation: original: untransformed; sqrt: square root transformation; log: log transformation

AIC Akaike information criterion; RSD residual standard deviation in millimetres on transformed outcome scale

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Knots: number of knots

Table B-4 and Table B-5 summarize the APGV and the PGV estimated by the SITAR model with and without transformation of age and outcomes for males and females. For both males and females, the difference in APGV is less than 0.15 year for all three outcomes. The difference in PGV is less than 0.25 mm/year for all three outcomes for both sexes.

Table B-4. Comparison of age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by the superimposition by translation and rotation (SITAR) growth curve model applied to male data with and without transformation of age and outcomes.

Age	Upper jaw length (COPAD)				Lower jaw length (COPOD)				Total face height (TFHNP)			
	Outcome	Knots	APGV	PGV	Outcome	Knots	APGV	PGV	Outcome	Knots	APGV	PGV
original*	original*	4	12.62	2.88	original*	7	13.35	4.76	original*	7	13.43	4.75
original	sqrt	7	12.66	2.89	sqrt	7	13.41	4.85	sqrt	7	13.39	4.89
original	log				log				log			
sqrt	original	7	12.55	2.90	original	7	13.39	4.74	original	7	13.20	4.93
log	original				original				original			
log	log				log				log			
sqrt	sqrt				sqrt				sqrt			
sqrt	log				log				log			
log	sqrt				sqrt				sqrt			

Missing rows indicate model did not converge. * model used for data analysis

Transformation: original: untransformed; sqrt: square root transformation; log: log transformation

APGV age at peak growth velocity in years; PGV peak growth velocity in mm/year

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Knots: number of knots

Table B-5. Comparison of age at peak growth velocity (APGV) and peak growth velocity (PGV) estimated by the superimposition by translation and rotation (SITAR) growth curve model applied to female data with and without transformation of age and outcomes.

Age	Upper jaw length (COPAD)				Lower jaw length (COPOD)				Total face height (TFHNP)			
	Outcome	Knots	APGV	PGV	Outcome	Knots	APGV	PGV	Outcome	Knots	APGV	PGV
original*	original*	4	11.14	2.21	original*	5	12.16	3.49	original*	5	12.43	3.28
original	sqrt	4	11.26	2.24	sqrt	5	12.27	6.00	sqrt	5	12.46	3.373
original	log				log				log			
sqrt	original	4	11.04	2.30	original	5	12.19	3.64	original	5	12.26	2.382
log	original				original				original			
log	log				log				log			
sqrt	sqrt				sqrt				sqrt			
sqrt	log				log				log			
log	sqrt				sqrt				sqrt			

Missing rows indicate model did not converge. * model used for data analysis

Transformation: original: untransformed; sqrt: square root transformation; log: log transformation

APGV age at peak growth velocity in years; PGV peak growth velocity in mm/year

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Knots: number of knots

Effect of age scale on the fit of FP model

Comparisons of fit statistics and growth parameters estimated for third-degree FP GCM applied to the data with original age and re-scaled age [(age+0)/10] are shown Table B-6. Interestingly, powers chosen for original age and re-scaled age differ. While power selected for lower jaw length and total face height for males with original age were 1 2 3, powers selected for re-scaled age were 3 3 3 for both outcomes.

For males, model fitted to all three outcomes using re-scaled age provided better fit than the original age (AIC differences larger than 15 units). For females, the model fit was almost identical for upper and lower jaw length as the AIC difference is less than 1.5 units. For total face height, however, FP GCM fitted using re-scaled age provided better fit (AIC 4165.04) than the original age scale (AIC 4186.18). Irrespective of the outcome and the sex, the APGV and the PGV estimated by model fitted to the data with original age were lower than model applied to the data with re-scaled age.

Table B-6. Comparison of fractional polynomial (FP) growth curve model applied to the male and female data with original age and re-scaled age.

Outcome	FP model fit to re-scaled age [(age+0)/10]						FP model fit to the original age scale [(age+0)/1]					
	Degree	Powers	AIC	RSD	APGV	PGV	Degree	Powers	AIC	RSD	APGV	PGV
Male												
Upper jaw length (COPAD)	3	1 3 3	3838.47	1.05	12.05	2.20	3	1 1 3	3854.84	1.06	11.73	2.18
Lower jaw length (COPOD)	3	3 3 3	4433.16	1.39	12.67	2.56	3	1 2 3	4454.53	1.42	12.36	2.50
Total face height (TFHNP)	3	3 3 3	4425.89	1.42	12.57	2.52	3	1 2 3	4446.79	1.44	12.27	2.47
Female												
Upper jaw length (COPAD)	3	1 1 1	3430.38	0.83	10.25	1.61	3	0 0.5 2	3431.65	0.83	10.19	1.62
Lower jaw length (COPOD)	3	3 3 3	4330.46	1.10	11.94	2.44	3	2 2 3	4329.06	1.14	11.70	2.43
Total face height (TFHNP)	3	3 3 3	4165.04	1.02	12.17	2.10	3	0.5 2 3	4186.18	1.05	11.70	2.08

Degree: degree of FP model; Powers: fractional polynomial powers

AIC: Akaike information criterion; RSD: residual standard deviation in millimetres

APGV age at peak growth velocity in years; PGV peak growth velocity in mm/year

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

B.3.2. **Fit to the data**

Akaike information criterion (AIC) and residual standard deviation (RSD)

The AIC and the RSD estimated by GCMs fitted to the upper and lower jaw length and total face height with different number of polynomial terms (CP and FP) and number of knots (RCS and SITAR) for males and females are shown below. The PB GCM failed to converge for any of the three outcomes for both males and females.

Conventional polynomial (CP)

Table B-7. Fit of conventional polynomial (CP) growth curve model applied to male and female data.

	Degree	Upper jaw length (COPAD)		Lower jaw length (COPOD)		Total face height (TFHNP)	
		AIC	RSD	AIC	RSD	AIC	RSD
Male	1	4159.11	1.45	4614.98	1.80	4579.36	1.78
	2	4132.20	1.34	4563.07	1.62	4540.99	1.62
	3	3844.78	1.06	4454.53	1.42	4446.81	1.44
	4	-	-	4361.73	1.35	4361.83	1.36
	5	-	-	-	-	-	-
Female	1	3932.49	1.21	4760.69	1.59	4526.15	1.40
	2	3608.40	0.94	-	-	-	-
	3	3439.85	0.83	4335.36	1.12	4180.73	1.04
	4	-	-	-	-	-	-
	5	-	-	-	-	-	-

Note: Missing row(s) indicate that model did not converge.

Degree: Number of conventional polynomial terms

AIC Akaike information criterion; RSD residual standard deviation (millimetres)

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Fractional polynomial (FP)

Table B-8. Fit of fractional polynomial (FP) growth curve model applied to male and female data.

	Degree	Upper jaw length (COPAD)		Lower jaw length (COPOD)		Total face height (TFHNP)	
		AIC	RSD	AIC	RSD	AIC	RSD
Male	1	4159.11	1.45	4614.98	1.80	4579.37	1.78
	2	3984.14	1.21	4476.93	1.52	4459.23	1.51
	3	3838.47	1.05	4433.16	1.39	4425.89	1.42
Female	1	3725.02	1.08	4722.86	1.56	4526.15	1.40
	2	3519.54	0.91	4577.36	1.37	4334.61	1.17
	3	3430.38	0.83	4330.46	1.10	4165.04	1.02

Note: Missing row(s) indicate that model did not converge.

Degree: Number of fractional polynomial terms.

AIC Akaike information criterion; RSD residual standard deviation (millimetres)

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Fractional Polynomial powers:

Degree	Male			Female		
	Upper jaw length (COPAD)	Lower jaw length (COPOD)	Total face height (TFHNP)	Upper jaw length (COPAD)	Lower jaw length (COPOD)	Total face height (TFHNP)
1	1	1	1	-.5	.5	1
2	3 3	3 3	3 3	2 3	-2 -2	3 3
3	1 3 3	3 3 3	3 3 3	1 1 1	3 3 3	3 3 3

Restricted cubic spline (RCS)

Table B-9. Fit of restricted cubic spline (RCS) growth curve model applied to male and female data.

	Knots	Upper jaw length (COPAD)		Lower jaw length (COPOD)		Total face height (TFHNP)	
		AIC	RSD	AIC	RSD	AIC	RSD
Male	3	4137.33	1.35	4562.74	1.62	4538.50	1.62
	4	3826.71	1.03	4489.04	1.46	4481.44	1.47
	5	3762.95	0.94	4399.04	1.26	4404.25	1.29
	6	3754.55	0.90	4348.36	1.12	4352.82	1.14
	7	-	-	4320.65	1.03	4312.34	1.01
Female	3	3612.50	0.94	-	-	-	-
	4	3388.79	0.77	4331.44	1.10	4202.03	1.03
	5	3379.19	0.73	4245.63	0.95	4100.77	0.89
	6	-	-	-	-	-	-
	7	-	-	-	-	-	-

Note: Missing row(s) indicate that model did not converge.

Knots: Number of knots used for the construction of spline design matrix

AIC Akaike information criterion; RSD residual standard deviation (millimetres)

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Superimposition by translation and rotation (SITAR)

Table B-10. Fit of superimposition by translation and rotation (SITAR) growth curve model applied to male and female data.

	Knots	Upper jaw length (COPAD)		Lower jaw length (COPOD)		Total face height (TFHNP)	
		AIC	RSD	AIC	RSD	AIC	RSD
Male	3	-	-	-	-	-	-
	4	3774.42	1.01	-	-	-	-
	5	3783.79	1.02	4855.37	1.24	4452.79	1.19
	6	-	-	4361.29	1.25	4347.01	1.21
	7	-	-	4336.22	1.23	4335.07	1.23
Female	3	-	-	-	-	-	-
	4	3399.04	0.79	4316.77	1.03	-	-
	5	3391.98	0.79	4275.69	1.02	4114.11	0.93
	6	3397.65	0.76	4291.64	1.00	4109.93	0.93
	7	-	-	4342.68	1.01	4140.50	0.93

Note: Missing row(s) indicate that model did not converge.

Knots: Number of knots used for the construction of spline design matrix

AIC Akaike information criterion; RSD residual standard deviation (millimetres)

COPAD condyle–point A measurement in millimetres; COPOD condyle–pogonion measurement in millimetres; TFHNP total face height measurement in millimetres.

Extent of agreement between the observed outcome and the model predictions

The difference between observed and model predicted outcomes for the upper and lower jaw length and total face height for males and females are shown below.

Upper jaw length (COPAD)

Table B-11 Fit of linear and nonlinear growth curve models applied to the upper jaw length (COPAD) data for males.

Age (year)	Obs. mean (SD)	CP		FP		RCS		SITAR	
		Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)
7-8	82.23 (3.10)	82.43 (2.57)	-0.20 (-0.92 to 0.51)	82.41 (2.59)	-0.17 (-0.89 to 0.54)	82.21 (2.63)	0.03 (-0.79 to 0.84)	82.30 (2.73)	-0.07 (-0.90 to 0.76)
8-9	83.40 (3.08)	83.07 (2.66)	0.33 (-0.38 to 1.04)	83.10 (2.66)	0.30 (-0.41 to 1.01)	83.33 (2.70)	0.08 (-0.74 to 0.89)	83.42 (2.65)	-0.02 (-0.83 to 0.80)
9-10	84.79 (3.25)	84.56 (2.84)	0.24 (-0.48 to 0.95)	84.58 (2.83)	0.22 (-0.50 to 0.93)	84.80 (2.88)	-0.01 (-0.83 to 0.81)	84.83 (2.76)	-0.04 (-0.84 to 0.76)
10-11	86.21 (3.21)	86.34 (2.92)	-0.13 (-0.85 to 0.59)	86.33 (2.92)	-0.12 (-0.84 to 0.60)	86.24 (2.88)	-0.03 (-0.84 to 0.79)	86.24 (2.84)	-0.03 (-0.84 to 0.78)
11-12	87.72 (3.20)	88.02 (2.97)	-0.30 (-1.02 to 0.43)	87.99 (2.97)	-0.27 (-1.00 to 0.46)	87.69 (2.93)	0.03 (-0.80 to 0.85)	87.66 (2.99)	0.06 (-0.78 to 0.90)
12-13	90.12 (3.49)	90.18 (3.01)	-0.07 (-0.83 to 0.70)	90.16 (3.01)	-0.04 (-0.80 to 0.73)	89.98 (3.03)	0.13 (-0.73 to 1.00)	90.04 (3.11)	0.07 (-0.80 to 0.95)
13-14	92.97 (4.45)	92.58 (3.67)	0.39 (-0.58 to 1.36)	92.58 (3.67)	0.39 (-0.58 to 1.36)	92.81 (3.80)	0.15 (-0.95 to 1.26)	92.99 (3.87)	-0.02 (-1.14 to 1.09)
14-15	94.29 (4.39)	94.19 (3.58)	0.10 (-0.95 to 1.16)	94.22 (3.59)	0.07 (-0.98 to 1.13)	94.49 (3.64)	-0.20 (-1.38 to 0.99)	94.60 (3.50)	-0.31 (-1.48 to 0.86)
15-16	95.67 (3.29)	95.82 (3.27)	-0.15 (-1.26 to 0.96)	95.85 (3.27)	-0.19 (-1.30 to 0.92)	95.75 (3.29)	-0.08 (-1.38 to 1.21)	96.07 (3.14)	-0.40 (-1.68 to 0.88)
16-17	95.74 (3.76)	96.02 (3.14)	-0.28 (-1.25 to 0.69)	96.04 (3.14)	-0.30 (-1.27 to 0.67)	95.77 (3.04)	-0.03 (-1.11 to 1.05)	95.81 (2.97)	-0.07 (-1.15 to 1.00)
17-18	96.19 (3.33)	96.66 (3.35)	-0.47 (-1.79 to 0.86)	96.54 (3.33)	-0.35 (-1.68 to 0.97)	96.88 (3.07)	-0.69 (-2.18 to 0.79)	96.70 (3.02)	-0.51 (-2.01 to 0.99)

Age (year): Growth periods; Obs. mean (SD): observed mean and standard deviation; Pred. mean (SD): predicted mean and standard deviation

Mean difference (95% CI): difference between the observed and predicted means (Obs. mean - Pred. mean) and the limits within which 95% of the differences between observed and predicted values lie

CP Conventional polynomial; FP Fractional polynomial; RCS Restricted cubic spline; SITAR Superimposition by translation and rotation

COPAD condyle-point-A measurement in millimetres

Table B-12. Fit of linear and nonlinear growth curve models applied to the upper jaw length (COPAD) data for females.

Age (year)	Obs. mean (SD)	CP		FP		RCS		SITAR	
		Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)
7-8	81.43 (2.62)	81.33 (2.23)	0.10 (-0.48 to 0.69)	81.41 (2.22)	0.03 (-0.56 to 0.61)	81.38 (2.25)	0.05 (-0.54 to 0.64)	81.43 (2.10)	0.01 (-0.57 to 0.59)
8-9	82.47 (2.74)	82.40 (2.20)	0.07 (-0.50 to 0.65)	82.33 (2.19)	0.14 (-0.43 to 0.72)	82.42 (2.25)	0.05 (-0.53 to 0.62)	82.46 (2.22)	0.01 (-0.58 to 0.59)
9-10	83.44 (2.52)	83.61 (2.10)	-0.17 (-0.70 to 0.37)	83.55 (2.11)	-0.11 (-0.65 to 0.43)	83.50 (2.08)	-0.06 (-0.59 to 0.48)	83.40 (2.22)	0.04 (-0.51 to 0.59)
10-11	84.92 (2.55)	85.03 (2.13)	-0.11 (-0.63 to 0.42)	85.04 (2.15)	-0.11 (-0.64 to 0.41)	84.91 (2.11)	0.02 (-0.51 to 0.54)	84.83 (2.22)	0.10 (-0.44 to 0.63)
11-12	86.80 (2.64)	86.92 (2.10)	-0.12 (-0.68 to 0.44)	86.97 (2.12)	-0.17 (-0.73 to 0.39)	86.94 (2.11)	-0.15 (-0.71 to 0.41)	87.05 (2.13)	-0.25 (-0.81 to 0.31)
12-13	88.08 (2.97)	87.85 (2.24)	0.22 (-0.39 to 0.84)	87.90 (2.24)	0.17 (-0.44 to 0.79)	88.02 (2.28)	0.05 (-0.57 to 0.67)	88.23 (2.28)	-0.16 (-0.77 to 0.46)
13-14	89.35 (3.08)	89.23 (2.56)	0.12 (-0.59 to 0.84)	89.23 (2.55)	0.12 (-0.60 to 0.83)	89.37 (2.61)	-0.02 (-0.74 to 0.70)	89.43 (2.61)	-0.08 (-0.80 to 0.64)
14-15	89.98 (3.04)	89.93 (2.39)	0.05 (-0.65 to 0.74)	89.90 (2.36)	0.08 (-0.61 to 0.77)	89.88 (2.40)	0.10 (-0.59 to 0.79)	89.78 (2.44)	0.20 (-0.50 to 0.91)
15-16	90.29 (3.55)	90.41 (2.91)	-0.12 (-1.42 to 1.19)	90.36 (2.88)	-0.07 (-1.37 to 1.23)	90.21 (2.85)	0.08 (-1.22 to 1.38)	90.06 (2.89)	0.23 (-1.09 to 1.54)
16-17	90.35 (3.16)	90.51 (2.66)	-0.16 (-0.90 to 0.59)	90.48 (2.65)	-0.13 (-0.88 to 0.61)	90.36 (2.60)	-0.01 (-0.75 to 0.73)	90.31 (2.64)	0.04 (-0.71 to 0.78)
17-18	90.49 (2.91)	90.24 (2.15)	0.25 (-0.85 to 1.35)	90.44 (2.17)	0.05 (-1.06 to 1.15)	90.80 (2.21)	-0.31 (-1.42 to 0.80)	91.01 (2.25)	-0.52 (-1.63 to 0.60)

Age (year): Growth periods; Obs. mean (SD): observed mean and standard deviation; Pred. mean (SD): predicted mean and standard deviation

Mean difference (95% CI): difference between the observed and predicted means (Obs. mean - Pred. mean) and the limits within which 95% of the differences between observed and predicted values lie

CP Conventional polynomial; FP Fractional polynomial; RCS Restricted cubic spline; SITAR Superimposition by translation and rotation

COPAD condyle-point-A measurement in millimetres

Lower jaw length (COPOD)

Table B-13. Fit of linear and nonlinear growth curve models applied to the lower jaw length (COPOD) data for males.

Age (year)	Obs. mean (SD)	CP		FP		RCS		SITAR	
		Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)
7-8	102.84 (3.64)	102.89 (3.12)	-0.05 (-1.01 to 0.91)	103.28 (2.95)	-0.44 (-1.37 to 0.50)	103.03 (3.17)	-0.19 (-1.04 to 0.66)	103.03 (2.85)	-0.19 (-1.02 to 0.64)
8-9	104.92 (3.40)	104.99 (2.78)	-0.07 (-0.94 to 0.79)	104.74 (2.84)	0.18 (-0.69 to 1.05)	104.84 (2.87)	0.08 (-0.70 to 0.86)	104.87 (2.80)	0.04 (-0.73 to 0.82)
9-10	106.81 (3.31)	106.78 (2.78)	0.03 (-0.78 to 0.85)	106.49 (2.88)	0.32 (-0.51 to 1.14)	106.77 (2.88)	0.04 (-0.69 to 0.77)	106.93 (2.89)	-0.12 (-0.85 to 0.61)
10-11	108.69 (3.29)	108.62 (2.83)	0.07 (-0.75 to 0.89)	108.55 (2.86)	0.14 (-0.68 to 0.96)	108.80 (2.90)	-0.11 (-0.84 to 0.62)	109.02 (2.93)	-0.33 (-1.06 to 0.40)
11-12	110.54 (3.67)	110.49 (3.12)	0.04 (-0.87 to 0.96)	110.68 (3.05)	-0.14 (-1.05 to 0.77)	110.63 (3.10)	-0.10 (-0.91 to 0.71)	110.57 (3.12)	-0.03 (-0.85 to 0.78)
12-13	112.40 (3.98)	112.66 (3.24)	-0.26 (-1.23 to 0.70)	112.91 (3.17)	-0.52 (-1.47 to 0.44)	112.43 (3.33)	-0.03 (-0.90 to 0.83)	112.53 (3.34)	-0.13 (-1.00 to 0.74)
13-14	115.40 (4.87)	115.61 (4.01)	-0.21 (-1.41 to 0.99)	115.72 (4.00)	-0.32 (-1.52 to 0.88)	115.38 (4.24)	0.02 (-1.06 to 1.10)	115.84 (4.49)	-0.44 (-1.55 to 0.67)
14-15	118.12 (5.02)	118.17 (4.03)	-0.04 (-1.39 to 1.30)	118.04 (4.04)	0.08 (-1.27 to 1.42)	118.36 (4.16)	-0.24 (-1.44 to 0.97)	118.49 (4.48)	-0.37 (-1.60 to 0.87)
15-16	121.55 (6.00)	120.31 (4.01)	1.24 (-0.75 to 3.23)	120.03 (3.87)	1.52 (-0.45 to 3.49)	120.55 (4.07)	1.00 (-0.82 to 2.82)	121.02 (4.12)	0.54 (-1.30 to 2.38)
16-17	122.06 (5.67)	122.42 (3.81)	-0.36 (-1.88 to 1.16)	122.27 (3.64)	-0.21 (-1.71 to 1.29)	122.22 (3.78)	-0.15 (-1.54 to 1.23)	122.35 (3.73)	-0.29 (-1.68 to 1.11)
17-18	123.77 (6.02)	122.52 (3.80)	1.26 (-1.06 to 3.57)	123.08 (4.14)	0.69 (-1.69 to 3.08)	122.80 (4.17)	0.97 (-1.19 to 3.12)	122.90 (3.88)	0.87 (-1.30 to 3.05)

Age (year): Growth periods; Obs. mean (SD): observed mean and standard deviation; Pred. mean (SD): predicted mean and standard deviation

Mean difference (95% CI): difference between the observed and predicted means (Obs. mean - Pred. mean) and the limits within which 95% of the differences between observed and predicted values lie

CP Conventional polynomial; FP Fractional polynomial; RCS Restricted cubic spline; SITAR Superimposition by translation and rotation

COPOD condyle-pogonion measurement in millimetres

Table B-14. Fit of linear and nonlinear growth curve models applied to the lower jaw length (COPOD) data for females.

Age (year)	Obs. mean (SD)	CP		FP		RCS		SITAR	
		Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)
7-8	98.87 (4.76)	98.98 (4.54)	-0.12 (-1.22 to 0.99)	98.90 (4.54)	-0.03 (-1.13 to 1.07)	98.82 (4.56)	0.04 (-1.06 to 1.15)	98.79 (4.51)	0.08 (-1.02 to 1.18)
8-9	100.54 (5.03)	100.03 (4.73)	0.51 (-0.59 to 1.61)	100.11 (4.73)	0.43 (-0.67 to 1.54)	100.31 (4.79)	0.24 (-0.87 to 1.34)	100.37 (4.72)	0.17 (-0.94 to 1.28)
9-10	101.51 (5.16)	101.42 (4.78)	0.09 (-1.04 to 1.22)	101.49 (4.79)	0.02 (-1.11 to 1.15)	101.54 (4.77)	-0.03 (-1.16 to 1.10)	101.58 (4.83)	-0.08 (-1.22 to 1.06)
10-11	102.97 (5.36)	103.31 (5.10)	-0.34 (-1.48 to 0.80)	103.30 (5.11)	-0.33 (-1.47 to 0.81)	103.12 (5.09)	-0.14 (-1.29 to 1.00)	103.06 (5.10)	-0.09 (-1.24 to 1.06)
11-12	106.25 (5.16)	106.40 (4.70)	-0.15 (-1.28 to 0.98)	106.32 (4.70)	-0.07 (-1.20 to 1.06)	106.11 (4.75)	0.14 (-0.99 to 1.27)	106.26 (4.83)	-0.01 (-1.15 to 1.13)
12-13	108.06 (5.90)	107.87 (5.42)	0.20 (-1.09 to 1.48)	107.79 (5.43)	0.27 (-1.01 to 1.56)	107.87 (5.55)	0.20 (-1.10 to 1.49)	108.05 (5.68)	0.01 (-1.29 to 1.31)
13-14	110.51 (6.14)	110.46 (5.68)	0.04 (-1.43 to 1.52)	110.46 (5.68)	0.04 (-1.43 to 1.52)	110.79 (5.78)	-0.28 (-1.76 to 1.21)	110.84 (5.91)	-0.33 (-1.83 to 1.17)
14-15	112.03 (6.11)	112.23 (5.57)	-0.19 (-1.64 to 1.25)	112.30 (5.56)	-0.27 (-1.71 to 1.17)	112.48 (5.58)	-0.44 (-1.89 to 1.00)	112.35 (5.54)	-0.31 (-1.76 to 1.14)
15-16	113.91 (6.74)	113.76 (6.10)	0.15 (-2.41 to 2.71)	113.87 (6.08)	0.04 (-2.51 to 2.60)	113.63 (6.09)	0.29 (-2.27 to 2.84)	113.70 (5.96)	0.22 (-2.37 to 2.80)
16-17	114.30 (6.11)	114.58 (5.62)	-0.28 (-1.76 to 1.20)	114.61 (5.62)	-0.31 (-1.80 to 1.17)	114.28 (5.57)	0.02 (-1.45 to 1.50)	114.47 (5.40)	-0.17 (-1.64 to 1.30)
17-18	114.60 (5.41)	113.71 (4.56)	0.89 (-1.23 to 3.00)	113.38 (4.60)	1.22 (-0.90 to 3.34)	114.25 (4.82)	0.35 (-1.80 to 2.49)	114.78 (4.81)	-0.18 (-2.33 to 1.96)

Age (year): Growth periods; Obs. mean (SD): observed mean and standard deviation; Pred. mean (SD): predicted mean and standard deviation

Mean difference (95% CI): difference between the observed and predicted means (Obs. mean - Pred. mean) and the limits within which 95% of the differences between observed and predicted values lie

CP Conventional polynomial; FP Fractional polynomial; RCS Restricted cubic spline; SITAR Superimposition by translation and rotation

COPOD condyle-pogonion measurement in millimetres

Total face height (TFHNP)

Table B-15. Fit of linear and nonlinear growth curve models applied to the total face height (TFHNP) data for males.

Age (year)	Obs. mean (SD)	CP		FP		RCS		SITAR	
		Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)
7-8	106.31 (4.13)	106.38 (3.65)	-0.07 (-1.17 to 1.03)	106.70 (3.44)	-0.39 (-1.47 to 0.68)	106.51 (3.73)	-0.20 (-1.17 to 0.77)	106.36 (3.52)	-0.05 (-1.02 to 0.91)
8-9	108.37 (3.95)	108.43 (3.38)	-0.06 (-1.08 to 0.97)	108.18 (3.46)	0.19 (-0.84 to 1.22)	108.28 (3.45)	0.09 (-0.82 to 1.00)	108.28 (3.47)	0.10 (-0.82 to 1.02)
9-10	110.16 (4.11)	110.14 (3.62)	0.02 (-1.01 to 1.05)	109.89 (3.74)	0.27 (-0.77 to 1.32)	110.09 (3.67)	0.07 (-0.84 to 0.98)	110.25 (3.67)	-0.09 (-1.00 to 0.82)
10-11	112.06 (4.09)	111.97 (3.67)	0.10 (-0.94 to 1.13)	111.91 (3.71)	0.15 (-0.89 to 1.20)	112.16 (3.76)	-0.10 (-1.02 to 0.82)	112.36 (3.74)	-0.29 (-1.21 to 0.62)
11-12	113.89 (4.57)	113.79 (3.95)	0.10 (-1.05 to 1.25)	113.96 (3.89)	-0.07 (-1.21 to 1.07)	113.97 (4.00)	-0.08 (-1.10 to 0.94)	113.90 (3.96)	-0.01 (-1.03 to 1.02)
12-13	115.82 (4.61)	115.96 (4.12)	-0.14 (-1.31 to 1.04)	116.20 (4.03)	-0.37 (-1.54 to 0.79)	115.74 (4.22)	0.08 (-0.95 to 1.12)	115.88 (4.05)	-0.05 (-1.10 to 0.99)
13-14	118.55 (5.49)	118.75 (4.91)	-0.20 (-1.61 to 1.22)	118.86 (4.85)	-0.31 (-1.71 to 1.10)	118.45 (5.02)	0.10 (-1.14 to 1.35)	118.96 (5.23)	-0.41 (-1.70 to 0.89)
14-15	121.31 (5.27)	121.42 (4.81)	-0.11 (-1.63 to 1.40)	121.32 (4.74)	-0.02 (-1.52 to 1.49)	121.53 (4.82)	-0.23 (-1.54 to 1.09)	121.73 (5.07)	-0.42 (-1.80 to 0.95)
15-16	124.73 (6.38)	123.23 (4.66)	1.49 (-0.70 to 3.69)	122.97 (4.49)	1.76 (-0.41 to 3.92)	123.47 (4.79)	1.26 (-0.73 to 3.24)	124.06 (4.70)	0.66 (-1.36 to 2.69)
16-17	125.32 (5.96)	125.70 (4.62)	-0.37 (-2.07 to 1.32)	125.51 (4.41)	-0.18 (-1.85 to 1.49)	125.53 (4.60)	-0.21 (-1.72 to 1.31)	125.64 (4.48)	-0.32 (-1.86 to 1.23)
17-18	126.99 (6.40)	125.56 (4.44)	1.43 (-1.13 to 3.98)	126.18 (4.58)	0.81 (-1.77 to 3.39)	125.90 (4.86)	1.09 (-1.26 to 3.43)	126.14 (4.42)	0.85 (-1.53 to 3.22)

Age (year): Growth periods; Obs. mean (SD): observed mean and standard deviation; Pred. mean (SD): predicted mean and standard deviation

Mean difference (95% CI): difference between the observed and predicted means (Obs. mean - Pred. mean) and the limits within which 95% of the differences between observed and predicted values lie

CP Conventional polynomial; FP Fractional polynomial; RCS Restricted cubic spline; SITAR Superimposition by translation and rotation

TFHNP total face height measurement in millimetres

Table B-16. Fit of linear and nonlinear growth curve models applied to the total face height (TFHNP) data for females.

Age (year)	Obs. mean (SD)	CP		FP		RCS		SITAR	
		Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)	Pred. mean (SD)	Mean difference (95% CI)
7-8	102.38 (5.22)	102.52 (4.93)	-0.14 (-1.35 to 1.07)	102.46 (4.91)	-0.07 (-1.28 to 1.14)	102.28 (4.92)	0.10 (-1.11 to 1.31)	102.37 (4.91)	0.01 (-1.21 to 1.23)
8-9	103.81 (5.38)	103.35 (5.01)	0.46 (-0.72 to 1.63)	103.42 (5.02)	0.38 (-0.79 to 1.56)	103.63 (5.06)	0.18 (-1.00 to 1.36)	103.81 (5.05)	-0.00 (-1.19 to 1.19)
9-10	104.54 (5.35)	104.37 (4.98)	0.18 (-1.00 to 1.35)	104.42 (4.99)	0.12 (-1.05 to 1.29)	104.58 (4.97)	-0.04 (-1.21 to 1.13)	104.71 (5.04)	-0.17 (-1.35 to 1.01)
10-11	105.67 (5.52)	105.88 (5.21)	-0.20 (-1.38 to 0.97)	105.86 (5.22)	-0.19 (-1.36 to 0.99)	105.73 (5.20)	-0.06 (-1.23 to 1.12)	105.81 (5.28)	-0.13 (-1.32 to 1.06)
11-12	108.30 (5.29)	108.58 (4.84)	-0.28 (-1.44 to 0.88)	108.50 (4.83)	-0.20 (-1.36 to 0.96)	108.22 (4.85)	0.08 (-1.09 to 1.24)	108.36 (4.86)	-0.05 (-1.22 to 1.11)
12-13	109.80 (5.88)	109.66 (5.40)	0.14 (-1.15 to 1.42)	109.59 (5.40)	0.21 (-1.08 to 1.49)	109.56 (5.49)	0.24 (-1.05 to 1.52)	109.67 (5.53)	0.13 (-1.16 to 1.43)
13-14	112.04 (6.22)	112.03 (5.71)	0.00 (-1.49 to 1.50)	112.04 (5.70)	0.00 (-1.49 to 1.49)	112.31 (5.79)	-0.28 (-1.77 to 1.22)	112.47 (5.91)	-0.43 (-1.94 to 1.08)
14-15	113.65 (6.12)	113.67 (5.55)	-0.01 (-1.45 to 1.43)	113.74 (5.54)	-0.09 (-1.53 to 1.35)	113.96 (5.57)	-0.31 (-1.75 to 1.13)	114.07 (5.49)	-0.41 (-1.85 to 1.03)
15-16	115.96 (7.00)	115.54 (6.05)	0.42 (-2.22 to 3.05)	115.66 (6.04)	0.30 (-2.33 to 2.93)	115.47 (6.06)	0.49 (-2.14 to 3.13)	115.85 (6.00)	0.11 (-2.59 to 2.82)
16-17	115.97 (6.22)	116.27 (5.55)	-0.30 (-1.79 to 1.20)	116.31 (5.55)	-0.35 (-1.84 to 1.15)	116.00 (5.52)	-0.04 (-1.53 to 1.46)	116.18 (5.41)	-0.21 (-1.70 to 1.28)
17-18	116.40 (5.78)	115.71 (4.64)	0.69 (-1.54 to 2.92)	115.40 (4.62)	1.00 (-1.23 to 3.23)	116.01 (4.92)	0.39 (-1.87 to 2.65)	116.32 (4.99)	0.09 (-2.18 to 2.36)

Age (year): Growth periods; Obs. mean (SD): observed mean and standard deviation; Pred. mean (SD): predicted mean and standard deviation

Mean difference (95% CI): difference between the observed and predicted means (Obs. mean - Pred. mean) and the limits within which 95% of the differences between observed and predicted values lie

CP Conventional polynomial; FP Fractional polynomial; RCS Restricted cubic spline; SITAR Superimposition by translation and rotation

TFHNP total face height measurement in millimetres

B.3.3. **Model assumptions**

Upper jaw length (COPAD)

Males

Normality

Random effects

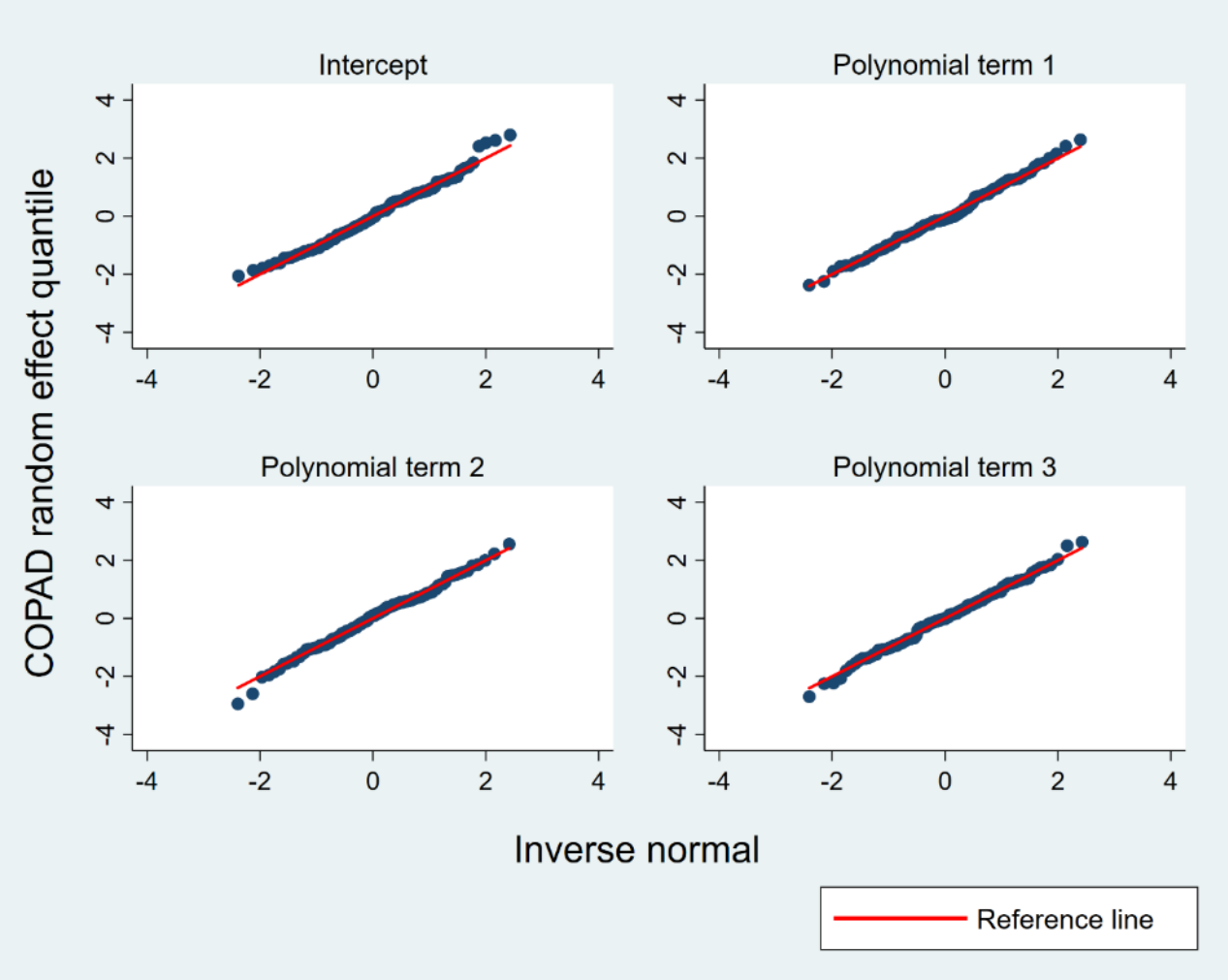


Figure B-1. Normal QQ-plot of random effects for conventional polynomial (CP) growth curve model applied to the upper jaw length (COPAD) data for males.

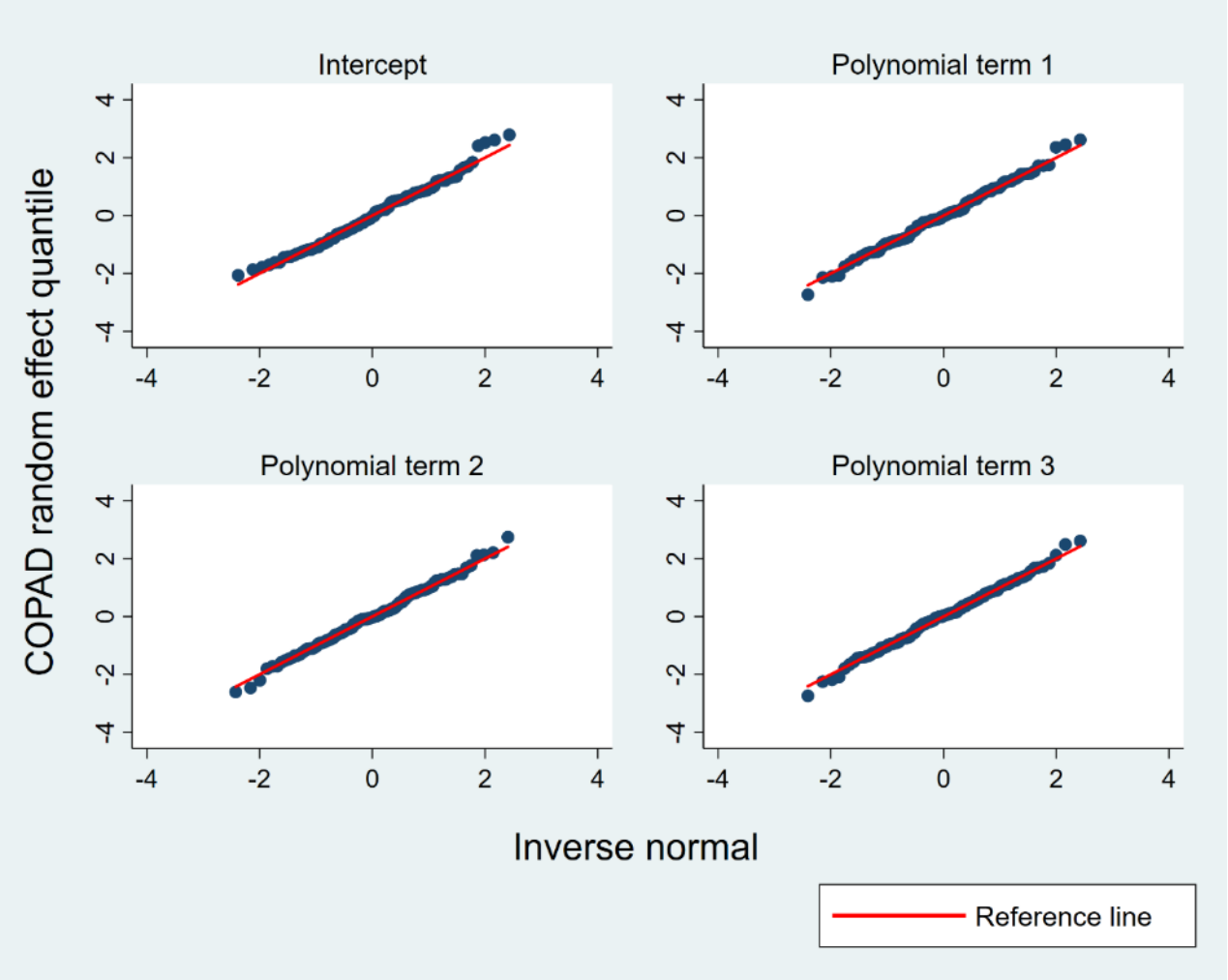


Figure B-2. Normal QQ-plot of random effects for fractional polynomial (FP) growth curve model applied to the upper jaw length (COPAD) data for males.

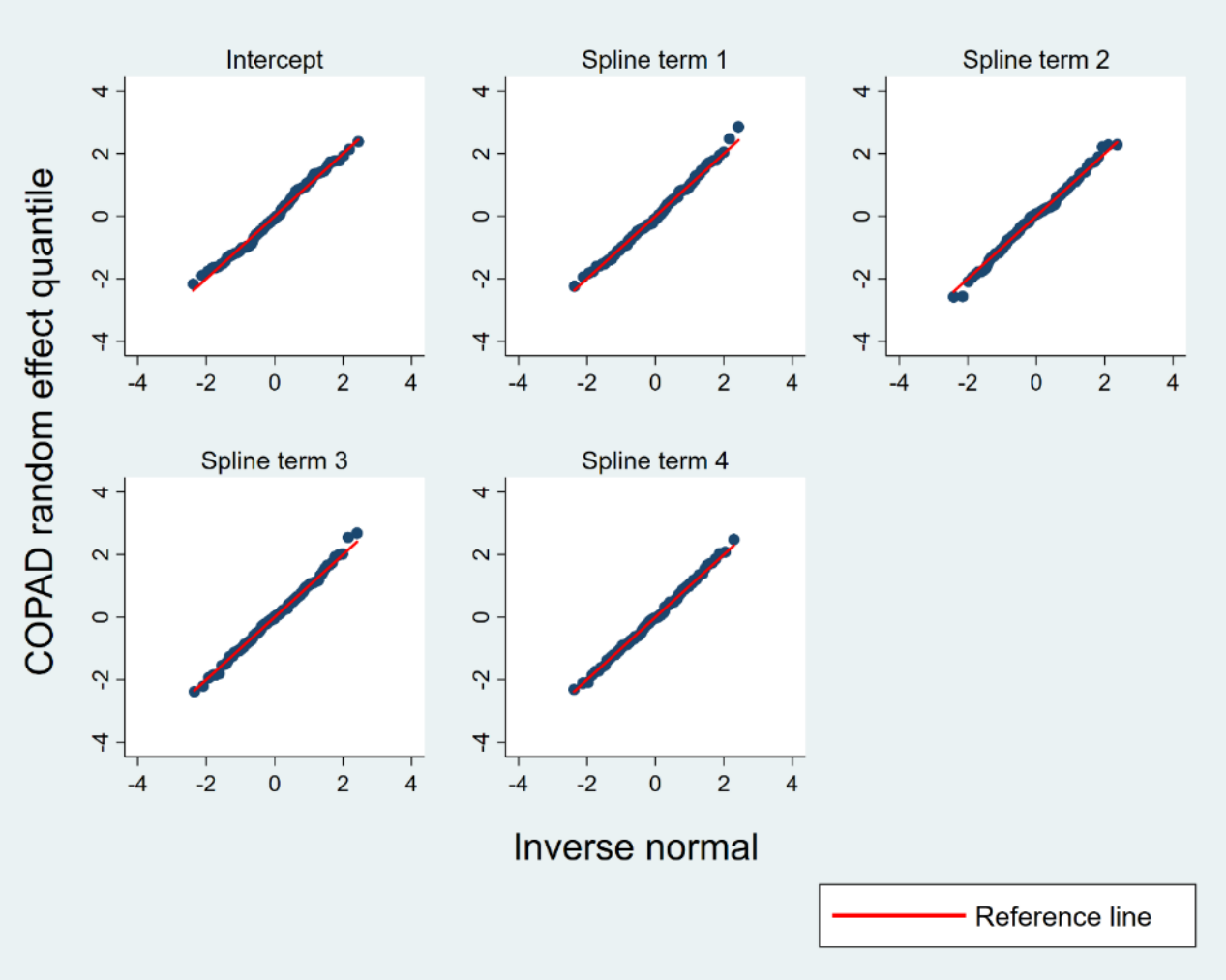


Figure B-3. Normal QQ-plot of random effects for restricted cubic spline (RCS) growth curve model applied to the upper jaw length (COPAD) data for males.

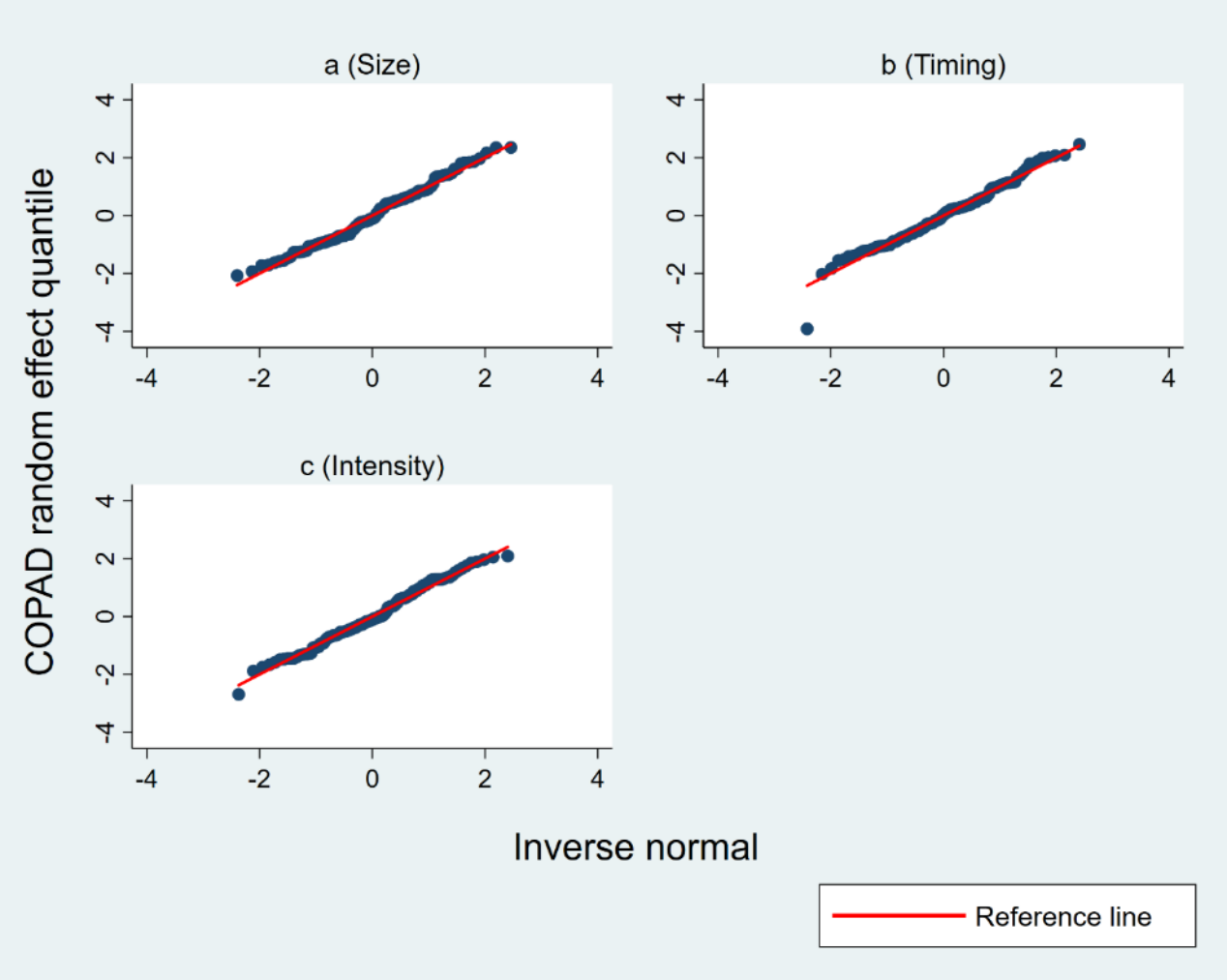


Figure B-4. Normal QQ-plot of random effects for superimposition by translation and rotation (SITAR) growth curve model applied to the upper jaw length (COPAD) data for males.

Level-1 residuals

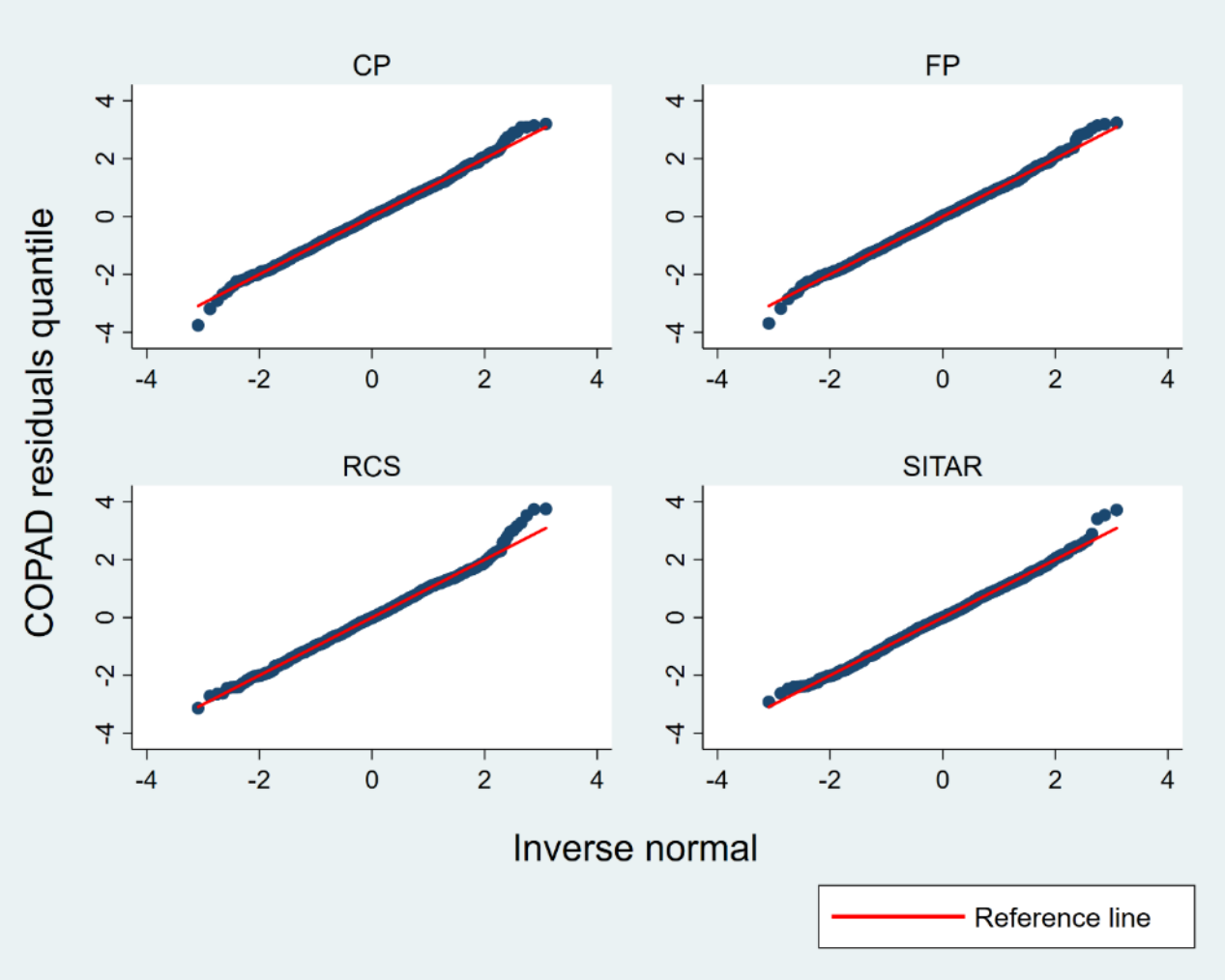


Figure B-5. Normal QQ-plot of level-1 residuals for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the upper jaw length (COPAD) data for males.

Homoscedasticity

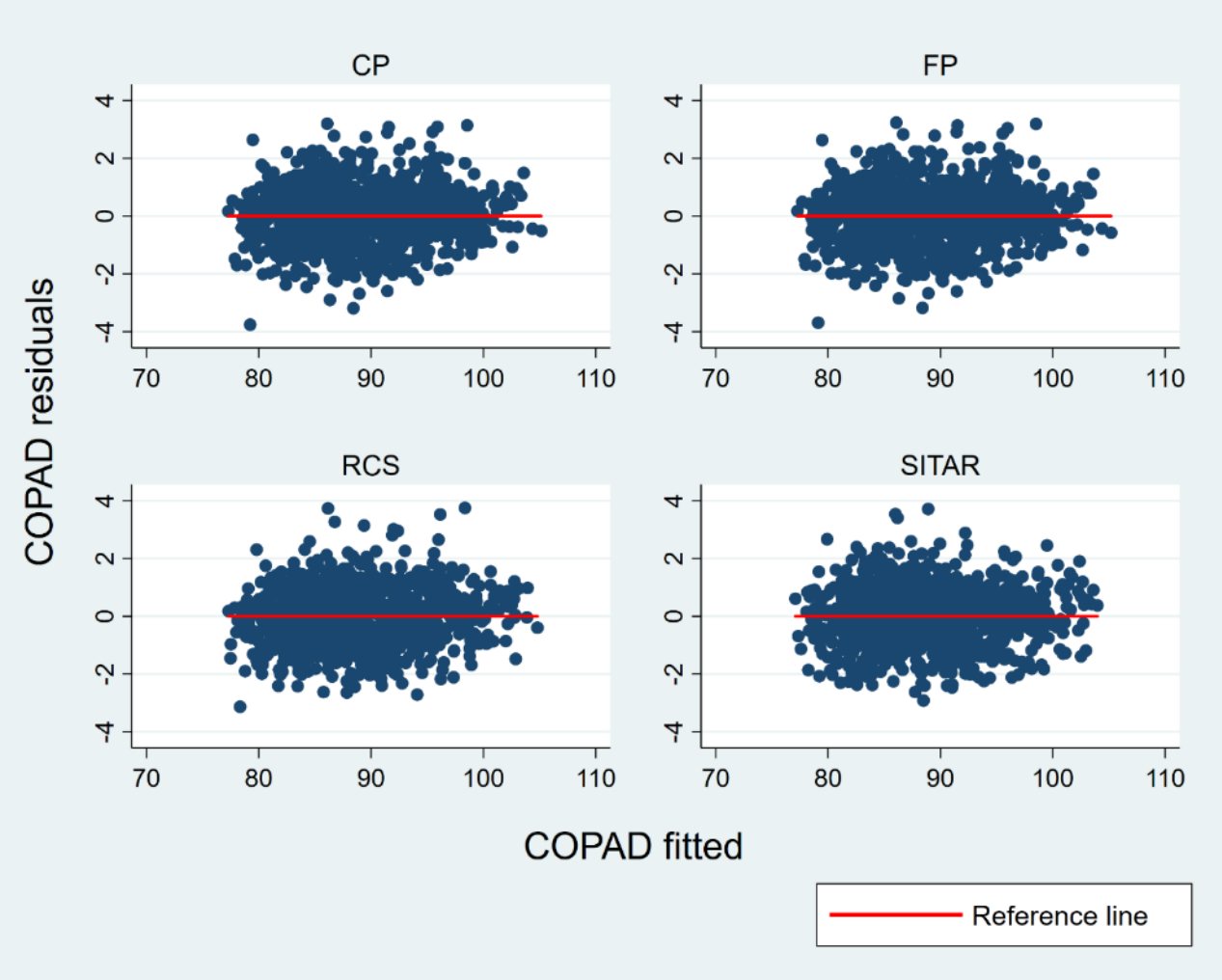


Figure B-6. Level-1 residuals versus fitted plot for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the upper jaw length (COPAD) data for males.

Females

Normality

Random effects

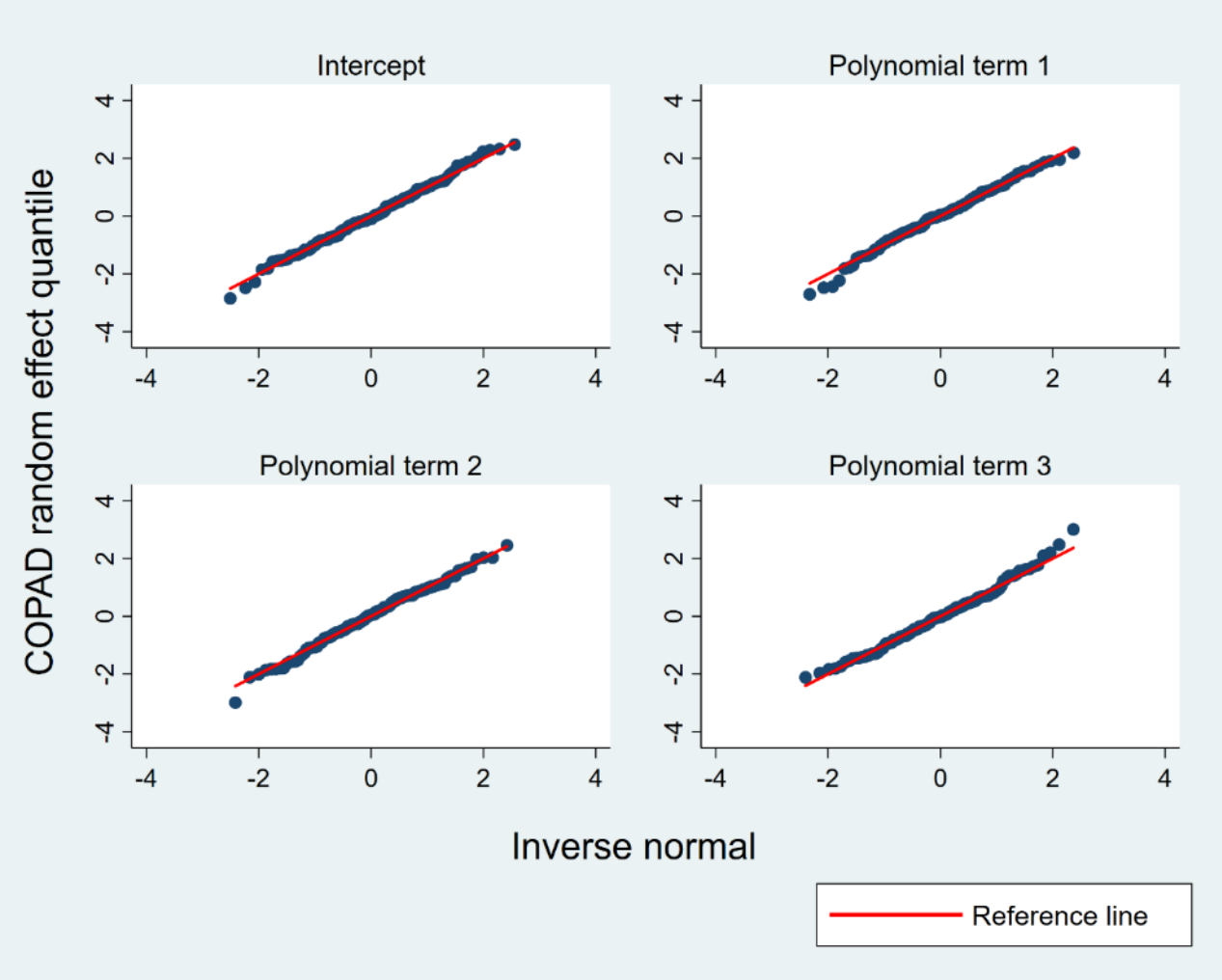


Figure B-7. Normal QQ-plot of random effects for conventional polynomial (CP) growth curve model applied to the upper jaw length (COPAD) data for females.

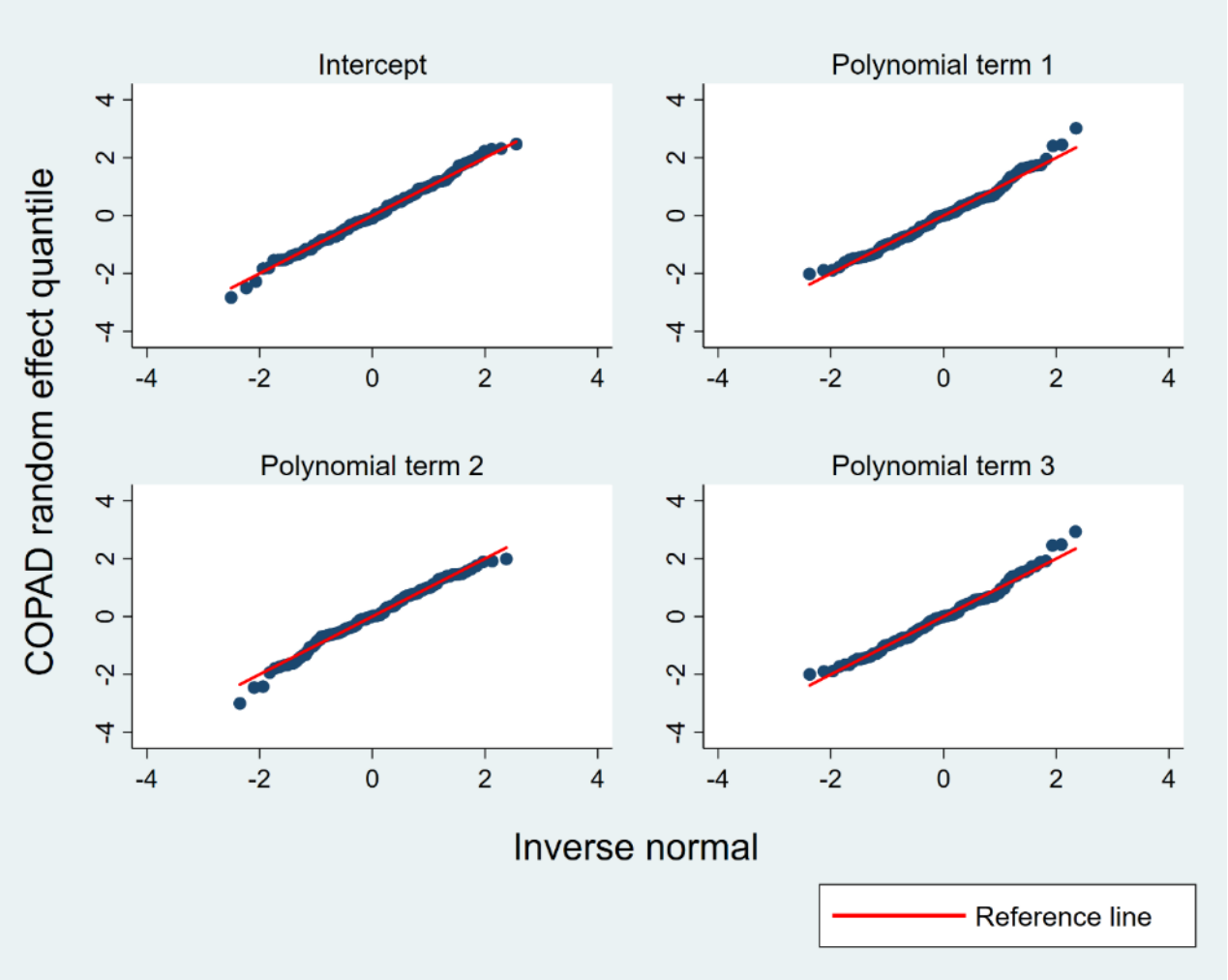


Figure B-8. Normal QQ-plot of random effects for fractional polynomial (FP) growth curve model applied to the upper jaw length (COPAD) data for females.

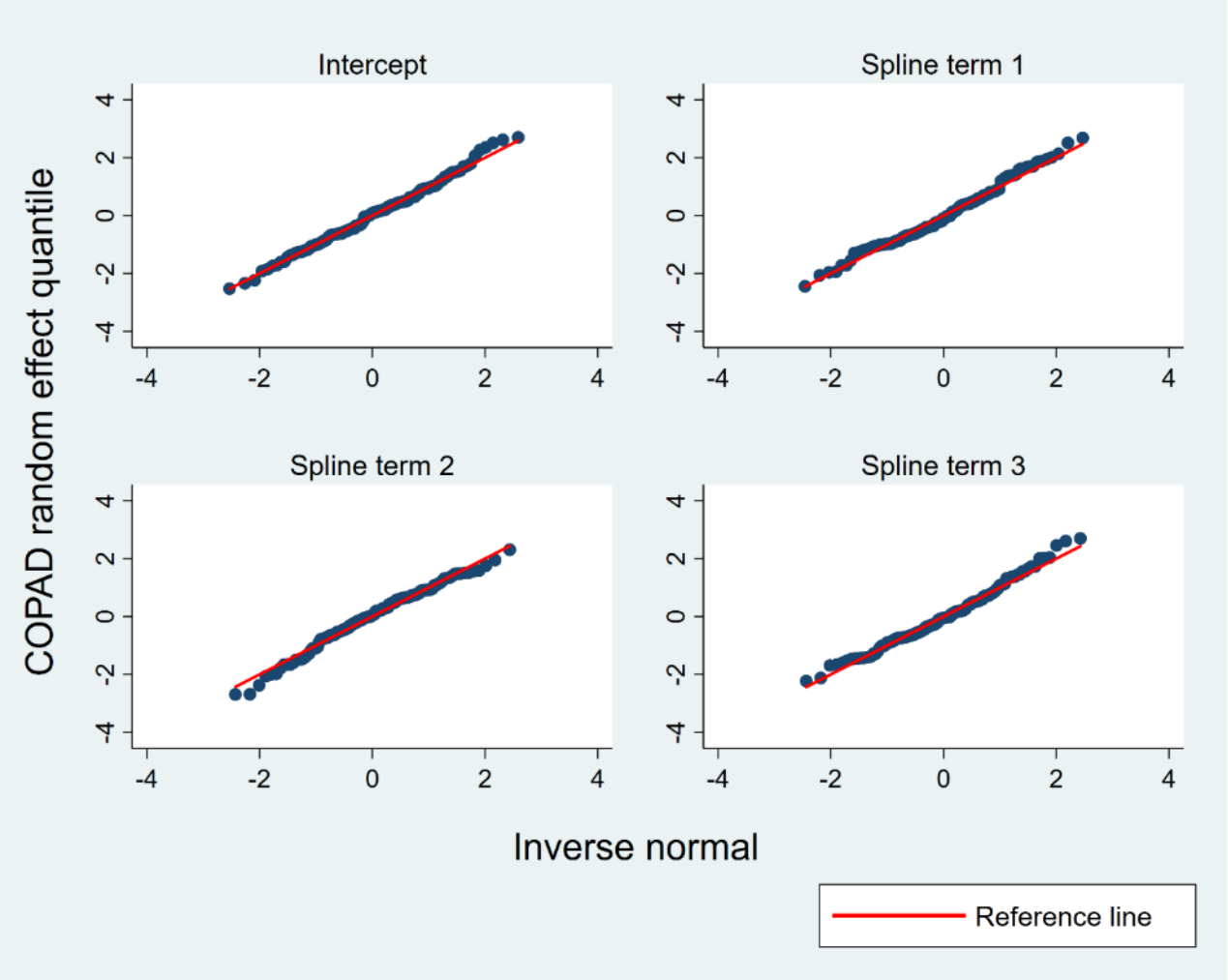


Figure B-9. Normal QQ-plot of random effects for restricted cubic spline (RCS) growth curve model applied to the upper jaw length (COPAD) data for females.

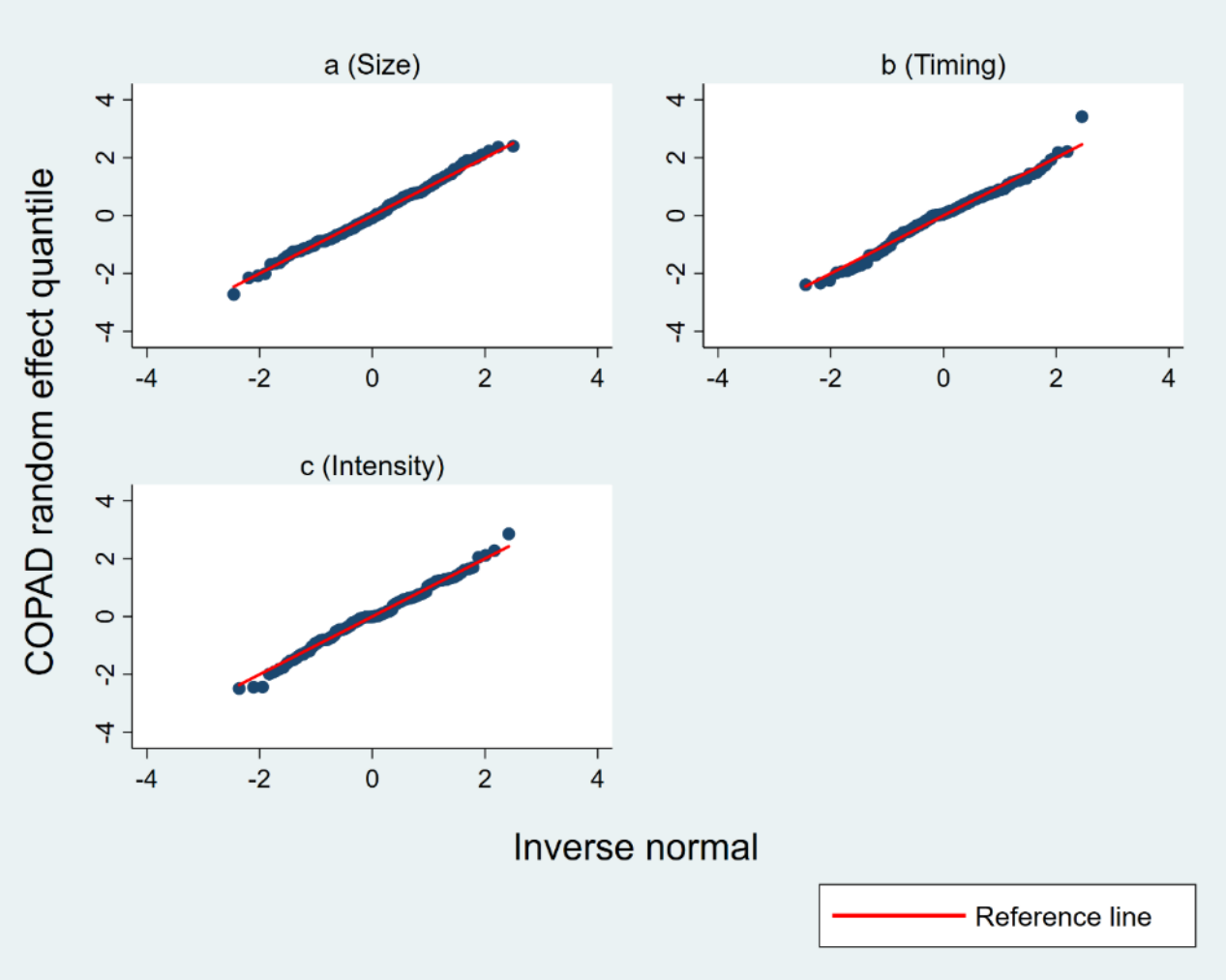


Figure B-10. Normal QQ-plot of random effects for superimposition by translation and rotation (SITAR) growth curve model applied to the upper jaw length (COPAD) data for females.

Level-1 residuals

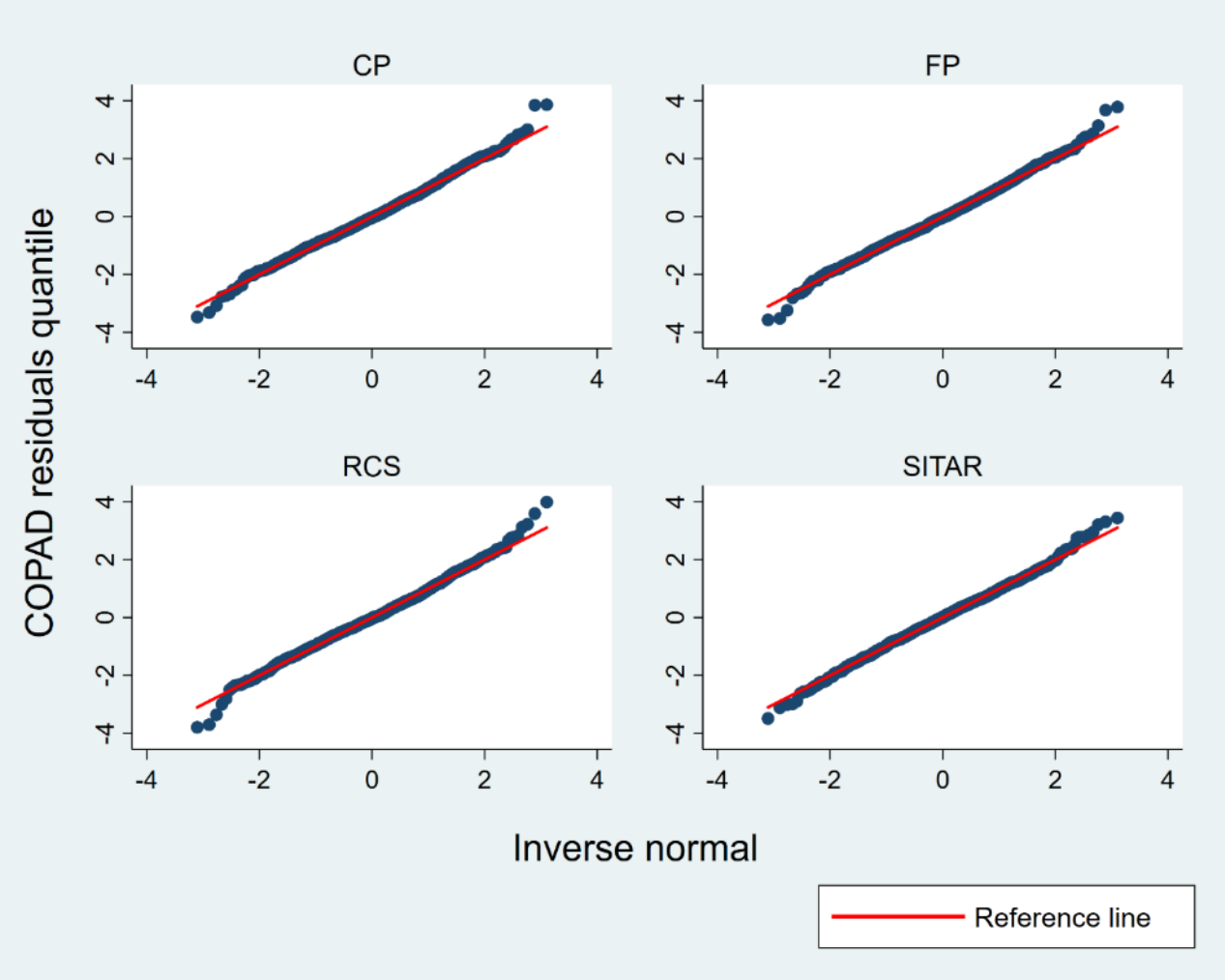


Figure B-11. Normal QQ-plot of level-1 residuals for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the upper jaw length (COPAD) data for females.

Homoscedasticity

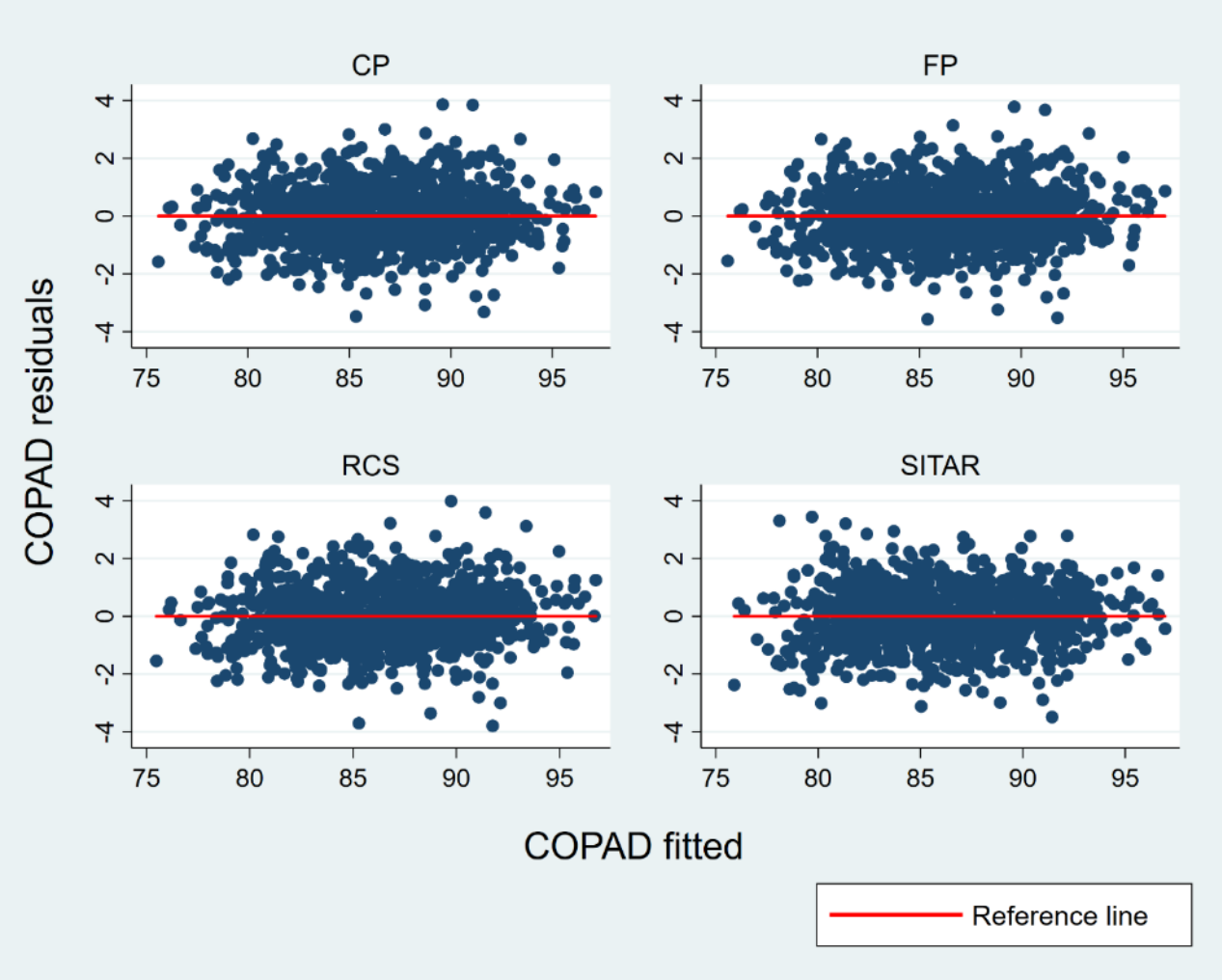


Figure B-12. Level-1 residuals versus fitted plot for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the upper jaw length (COPAD) data for females.

Lower jaw length (COPOD)

Males

Normality

Random effects

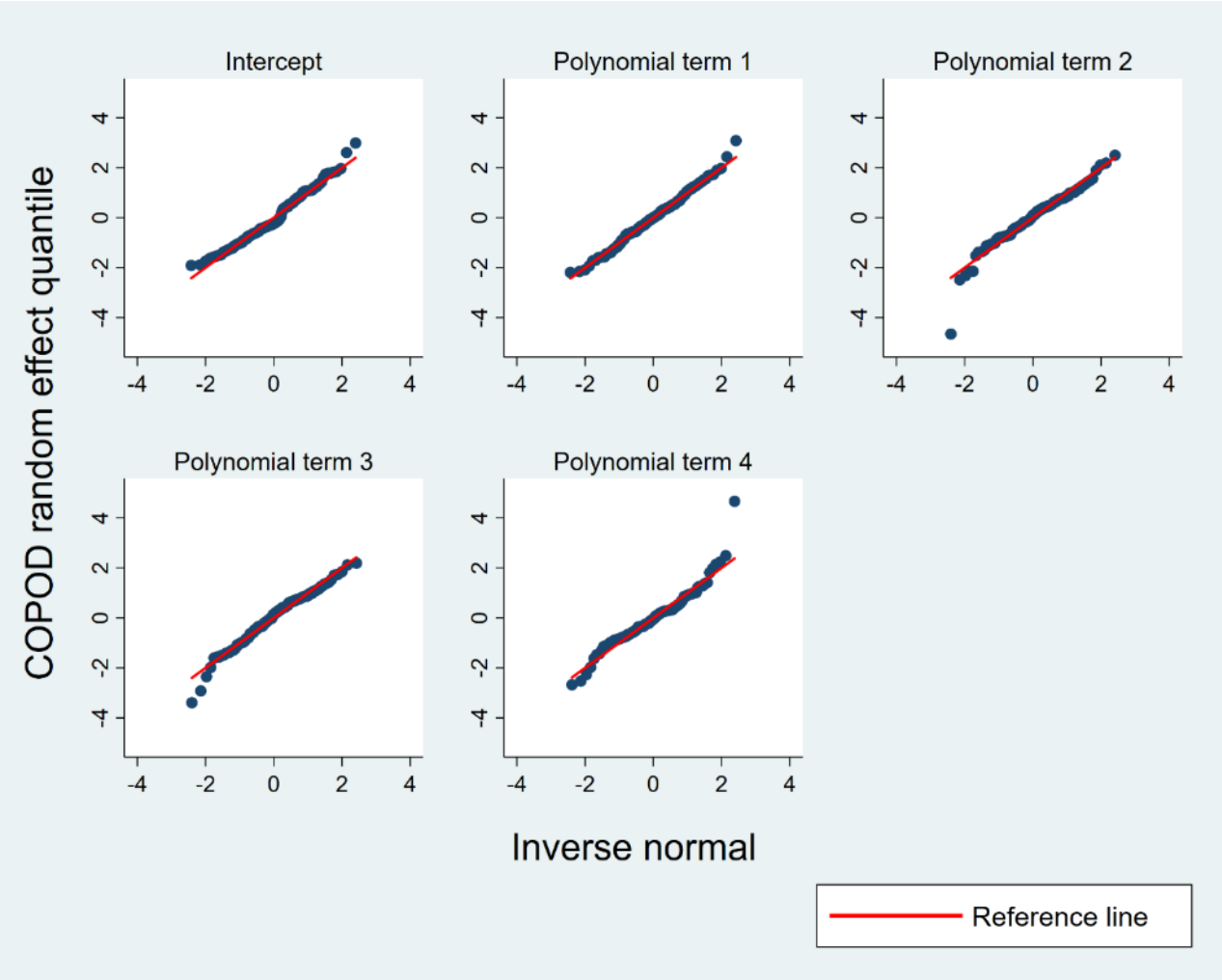


Figure B-13. Normal QQ-plot of random effects for conventional polynomial (CP) growth curve model applied to the lower jaw length (COPOD) data for males.

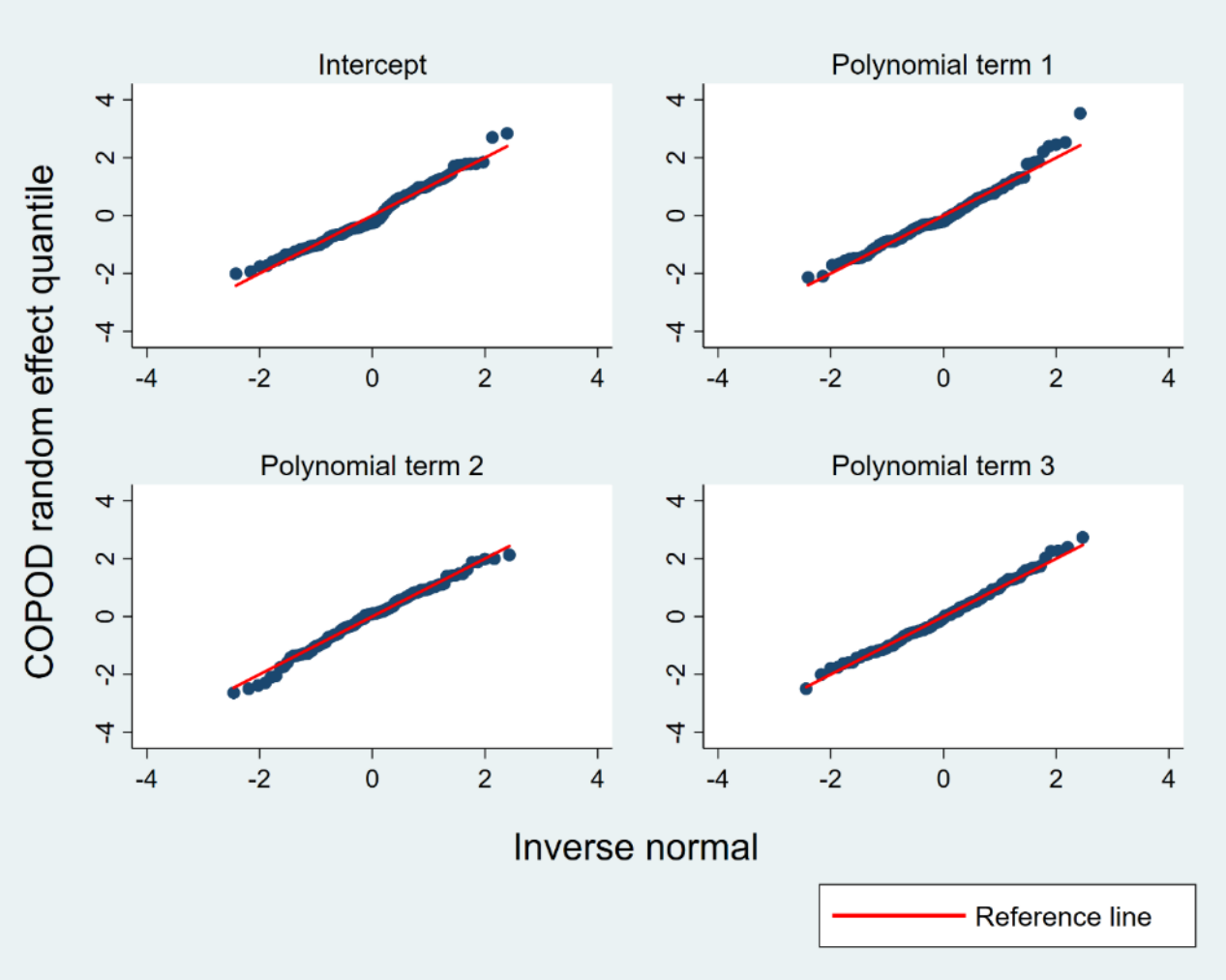


Figure B-14. Normal QQ-plot of random effects for fractional polynomial (FP) growth curve model applied to the lower jaw length (COPOD) data for males.

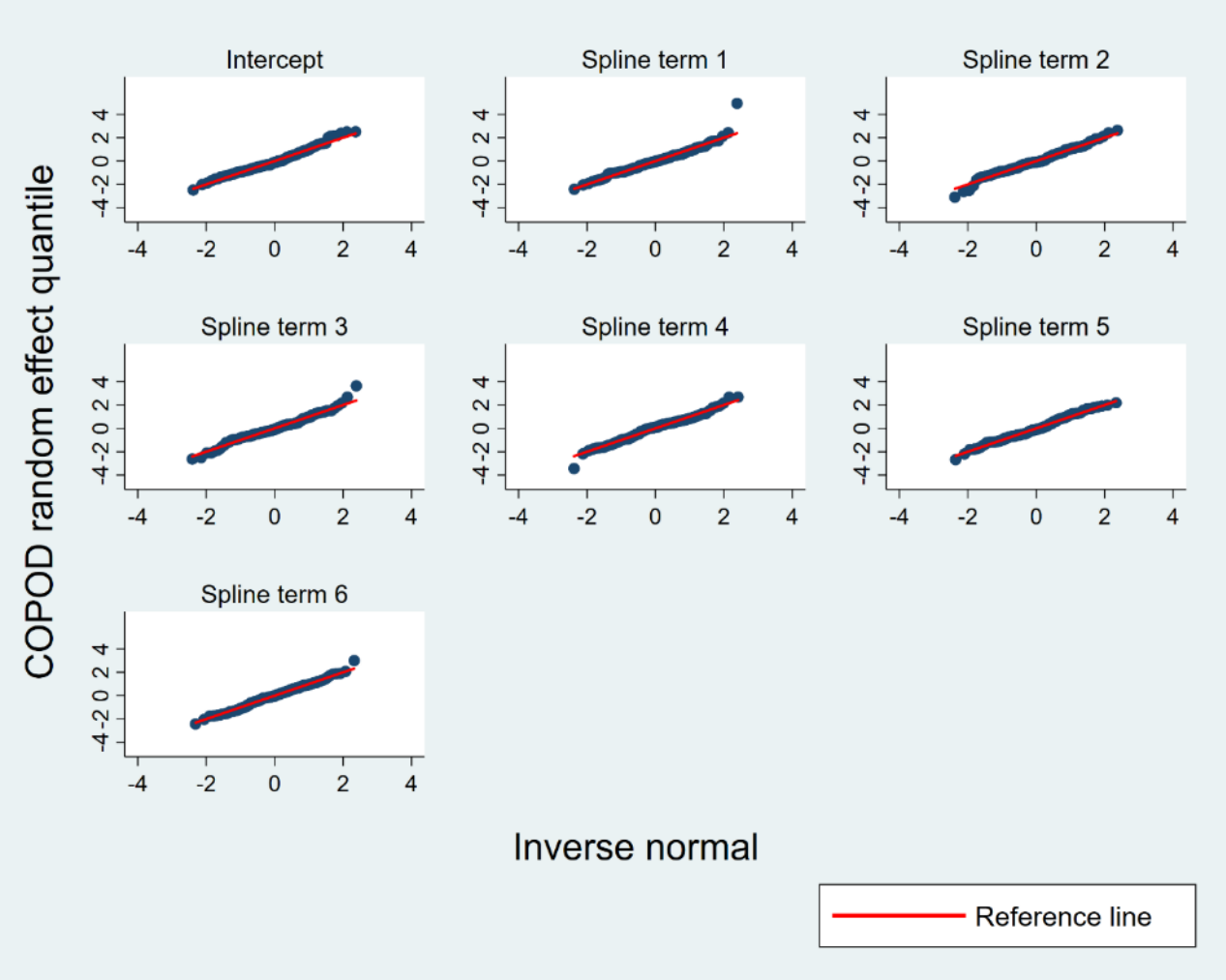


Figure B-15. Normal QQ-plot of random effects for restricted cubic spline (RCS) growth curve model applied to the lower jaw length (COPOD) data for males.

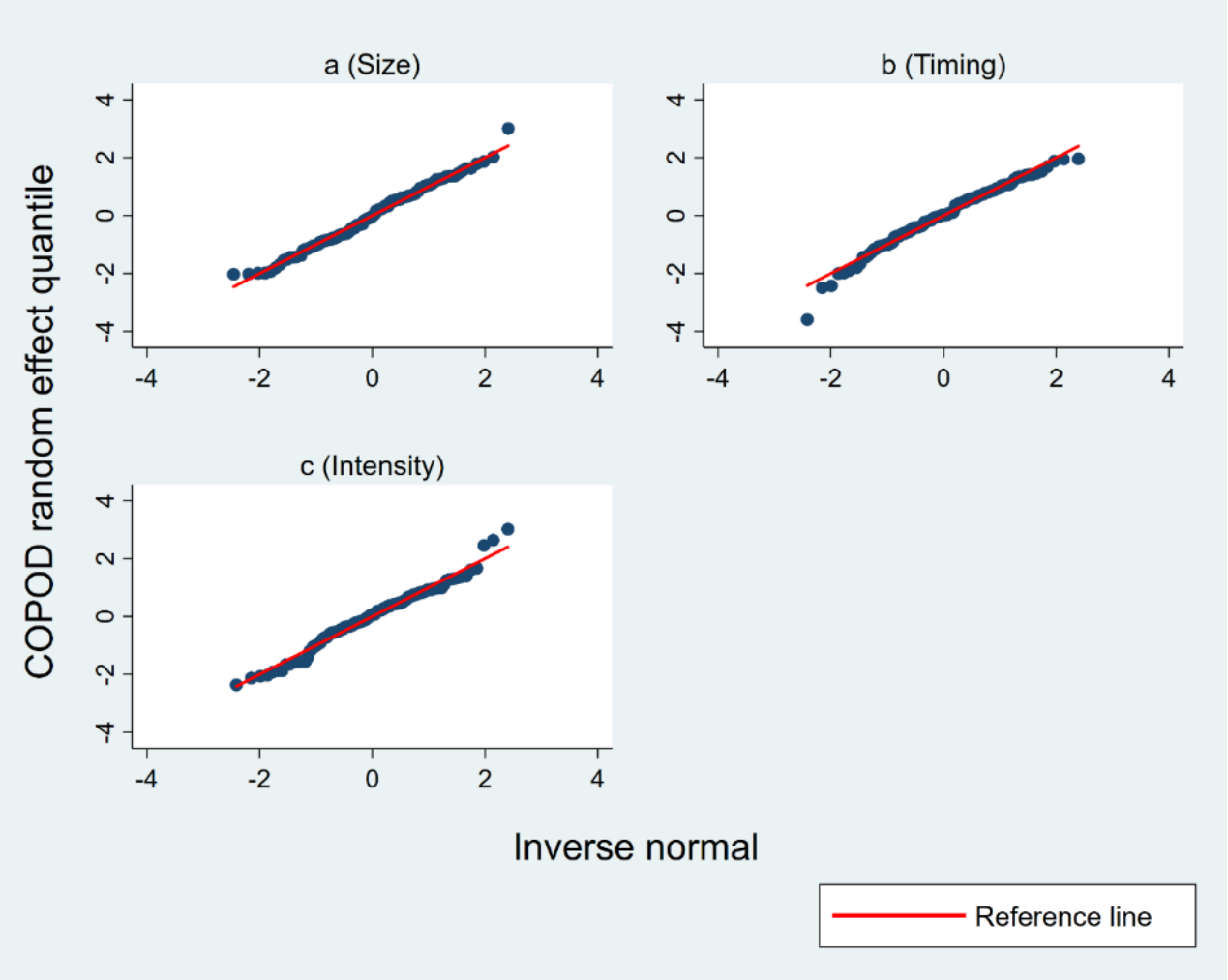


Figure B-16. Normal QQ-plot of random effects for superimposition by translation and rotation (SITAR) growth curve model applied to the lower jaw length (COPOD) data for males.

Level-1 residuals

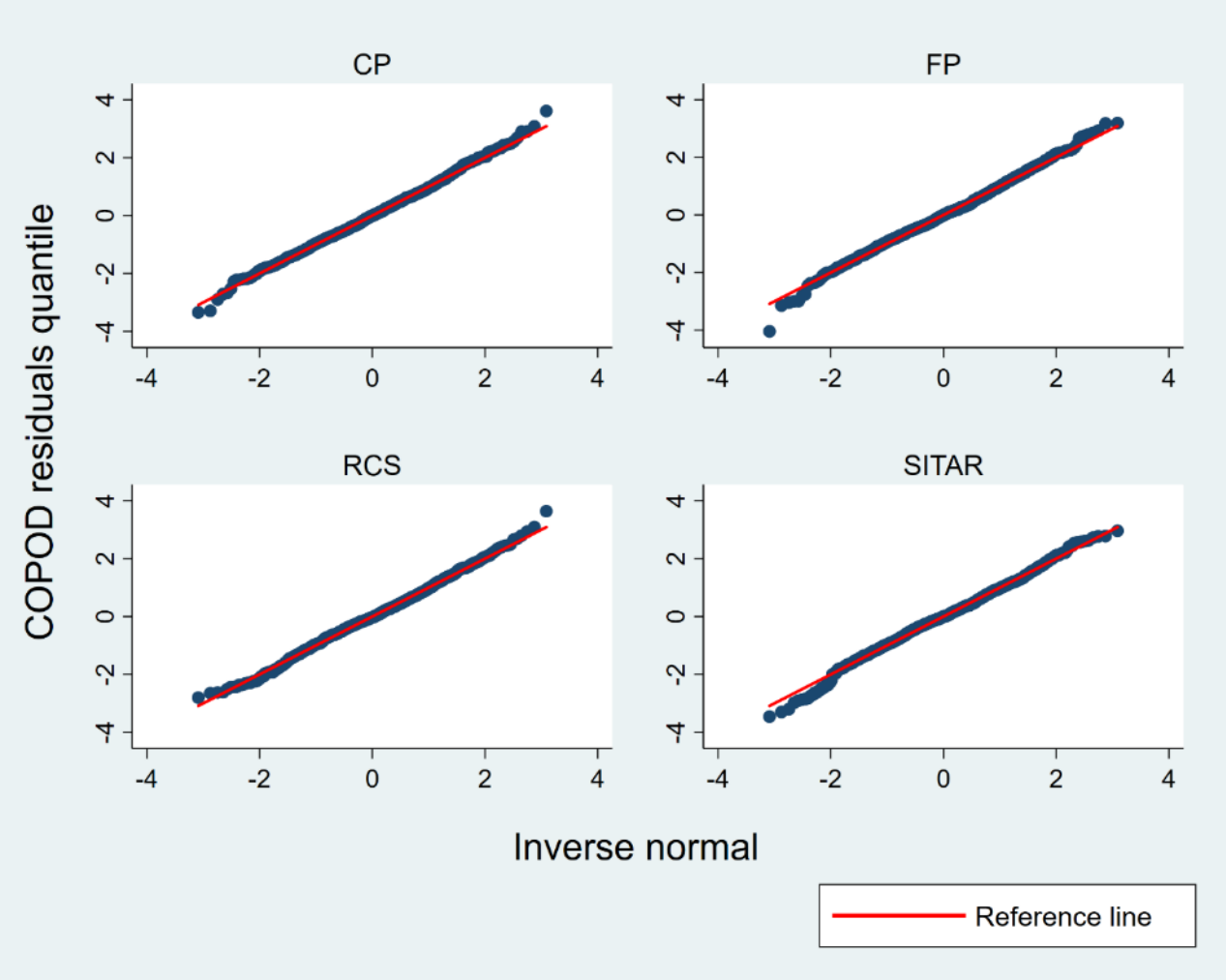


Figure B-17. Normal QQ-plot of level-1 residuals for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the to lower jaw length (COPOD) data for males.

Homoscedasticity

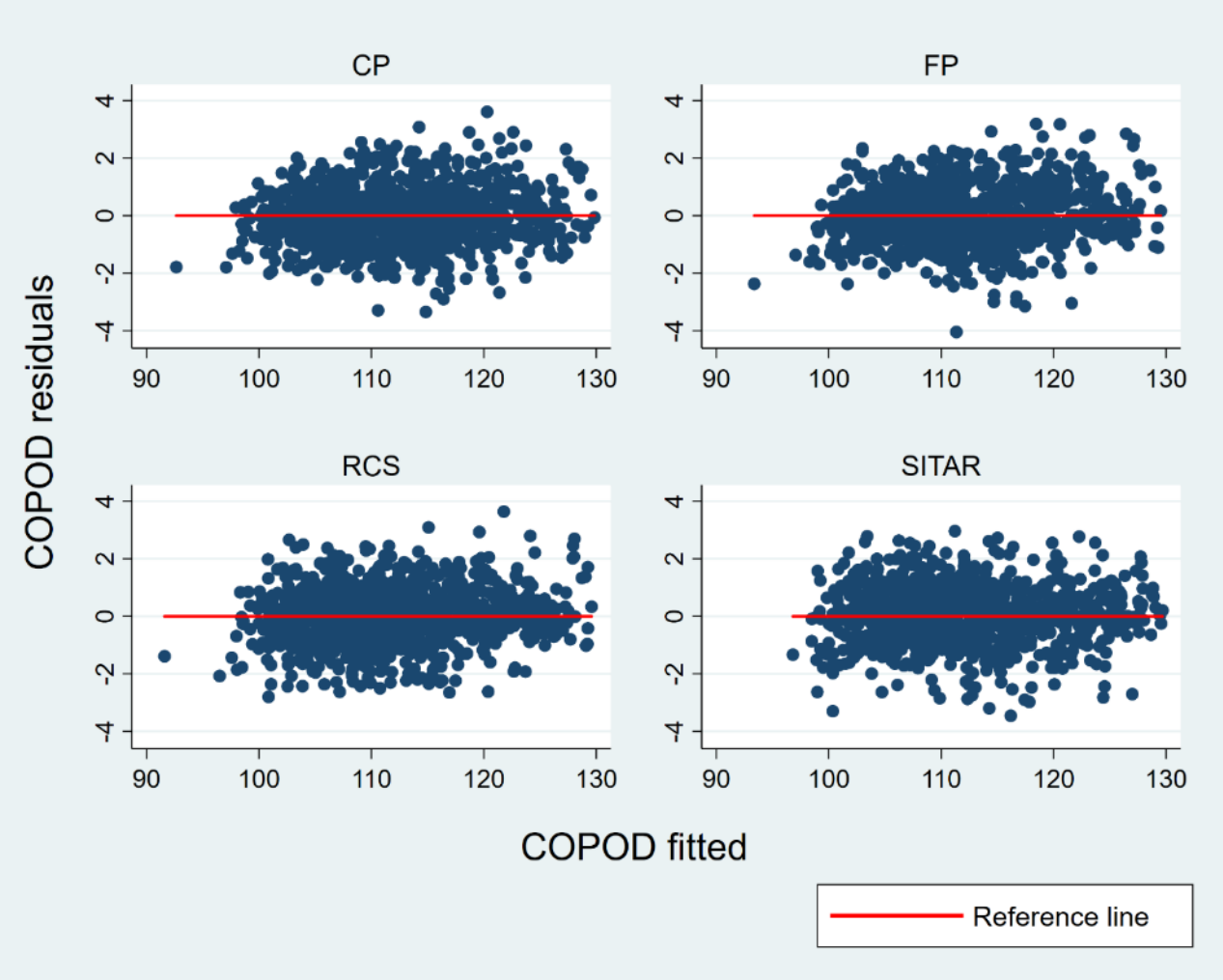


Figure B-18. Level-1 residuals versus fitted plot for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the lower jaw length (COPOD) data for males.

Female

Normality

Random effects

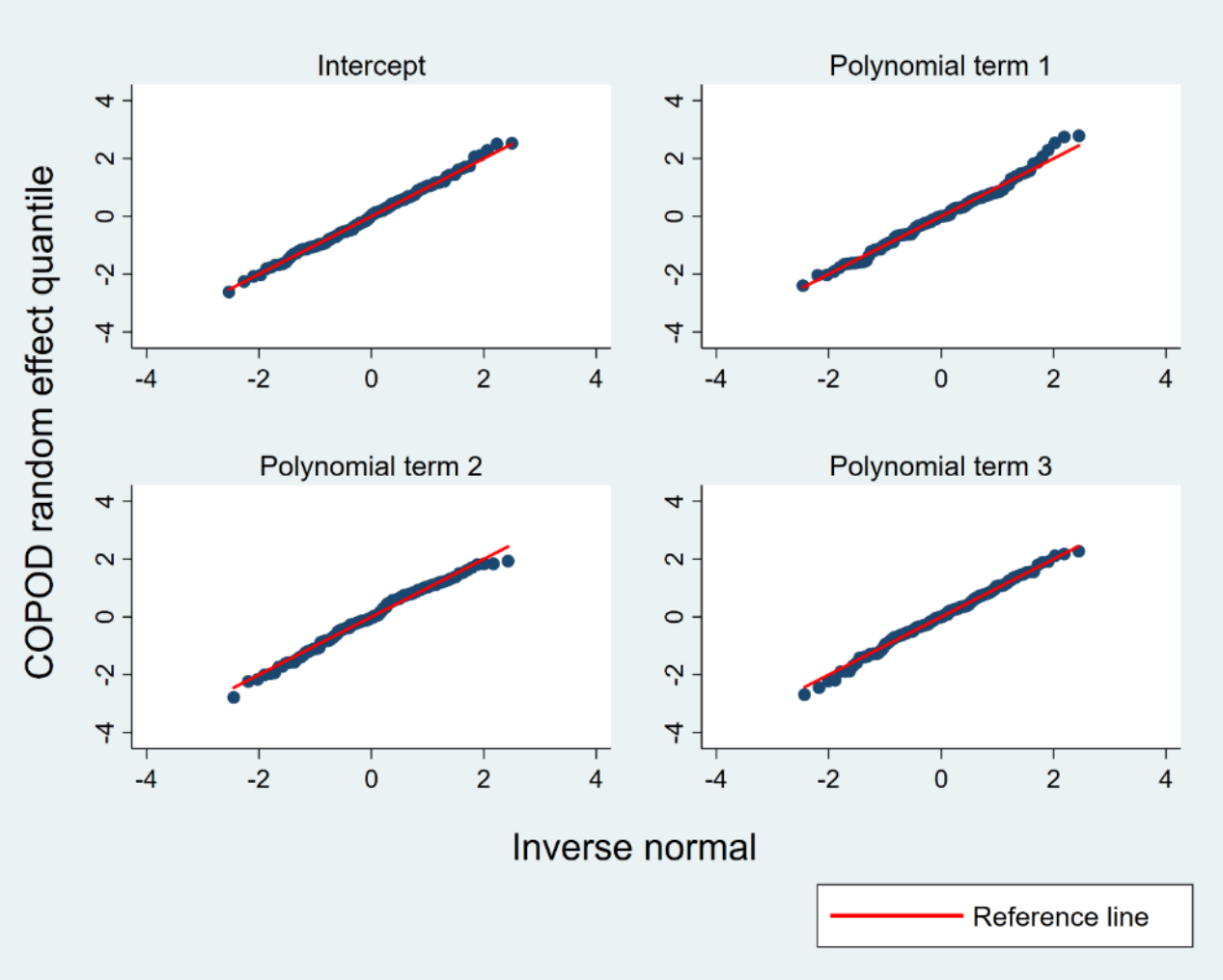


Figure B-19. Normal QQ-plot of random effects for conventional polynomial (CP) growth curve model applied to the lower jaw length (COPOD) data for females.

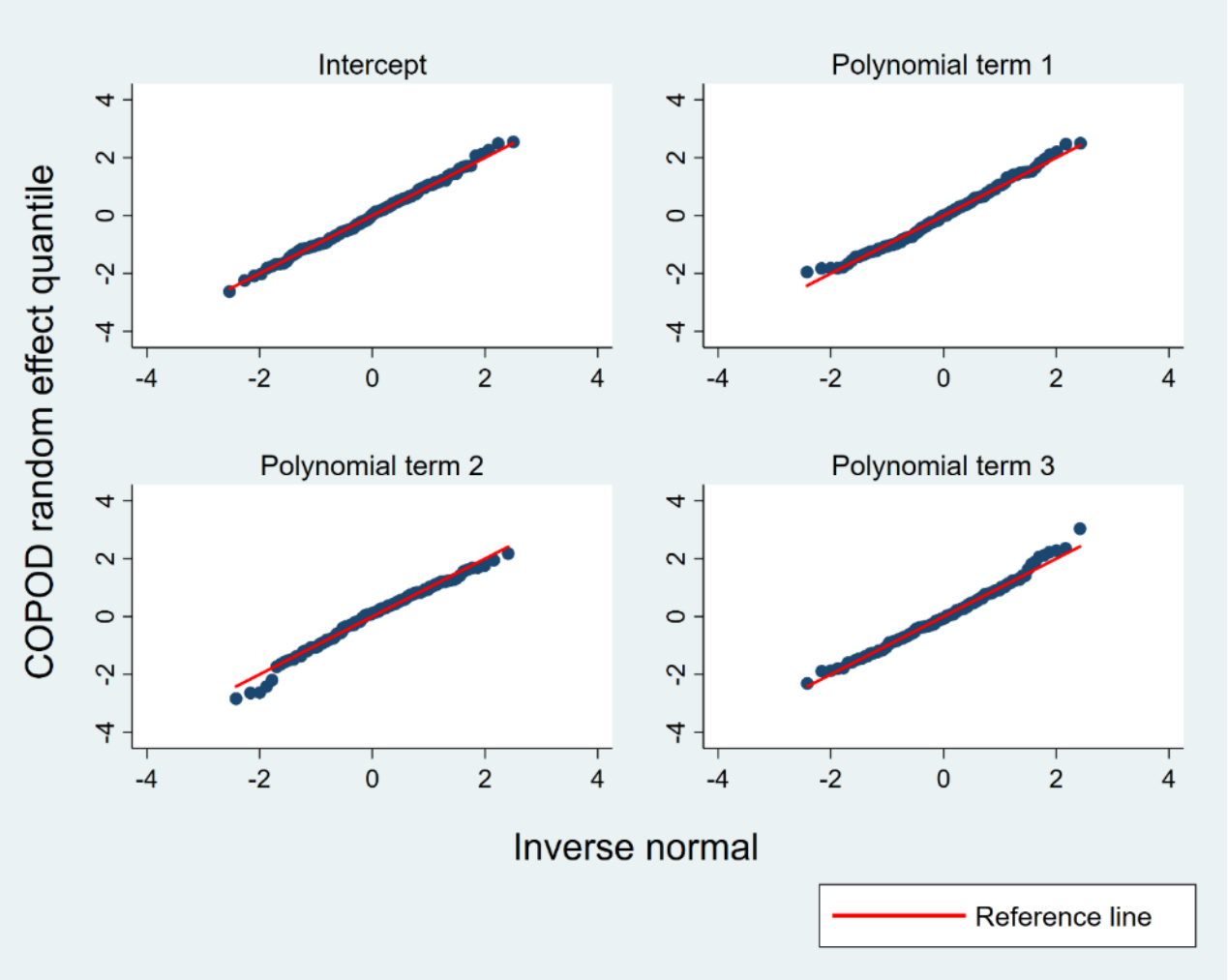


Figure B-20. Normal QQ-plot of random effects for fractional polynomial (FP) growth curve model applied to the lower jaw length (COPOD) data for females.

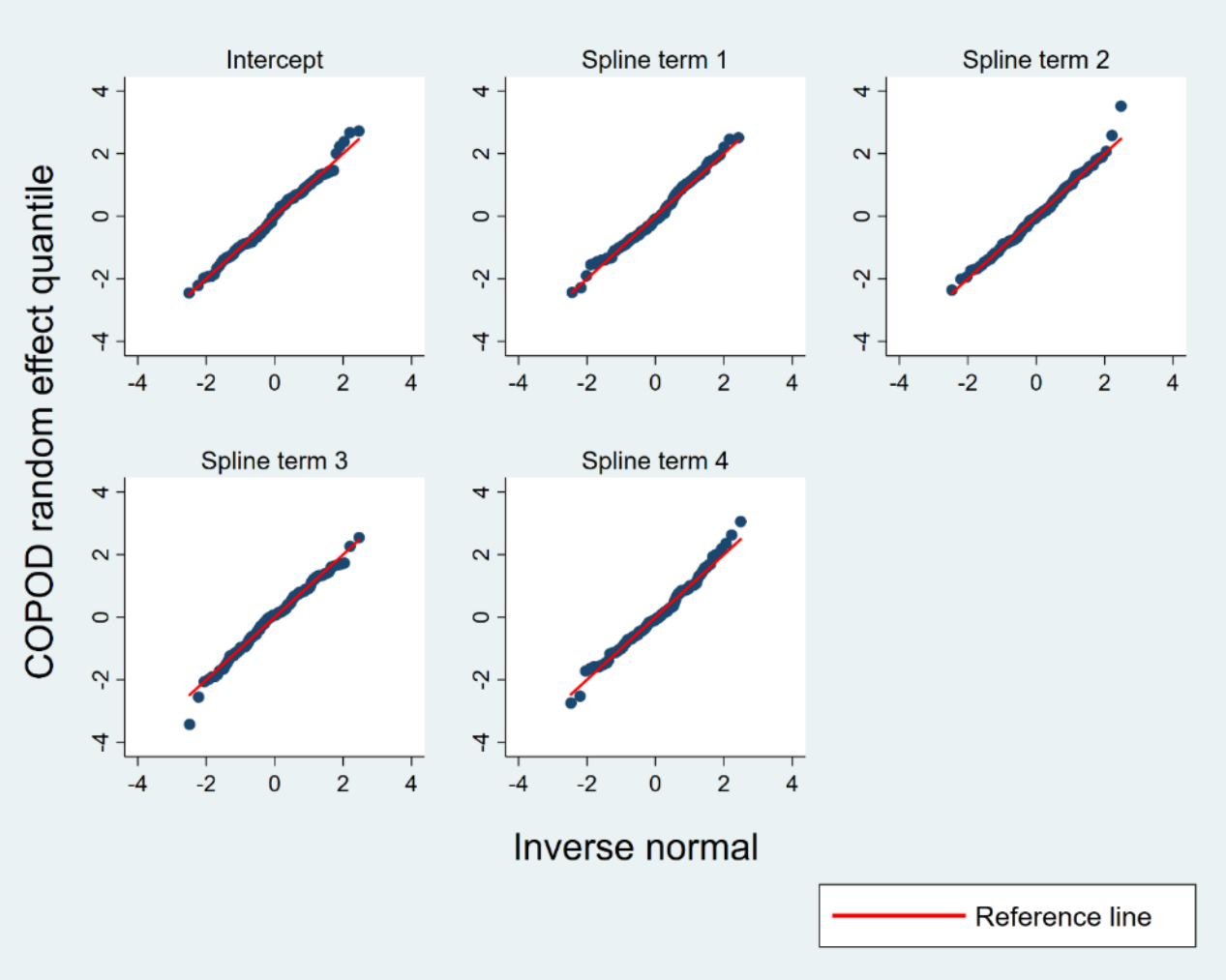


Figure B-21. Normal QQ-plot of random effects for restricted cubic spline (RCS) growth curve model applied to the lower jaw length (COPOD) data for females.

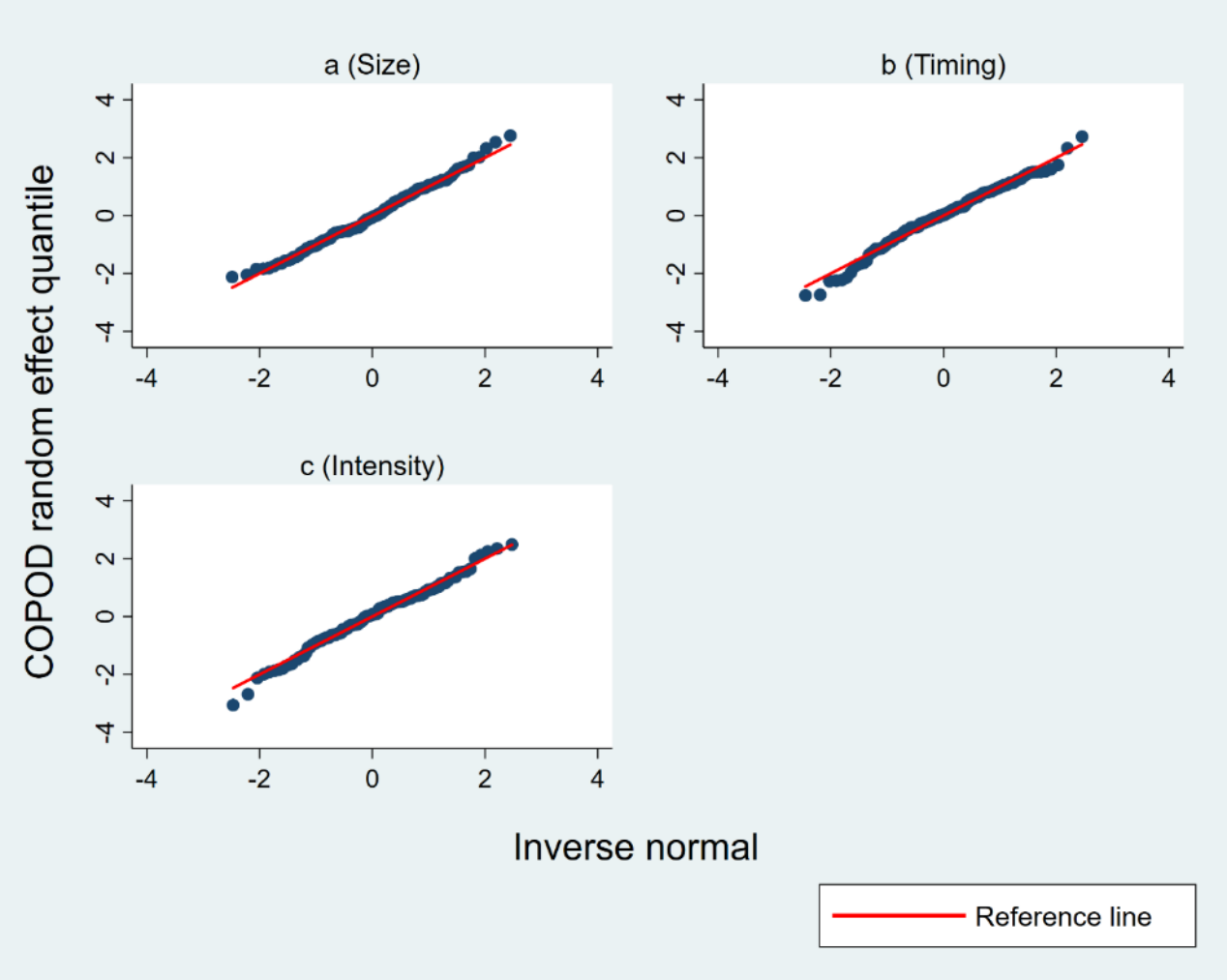


Figure B-22. Normal QQ-plot of random effects for superimposition by translation and rotation (SITAR) growth curve model applied to the lower jaw length (COPOD) data for females.

Level-1 residuals

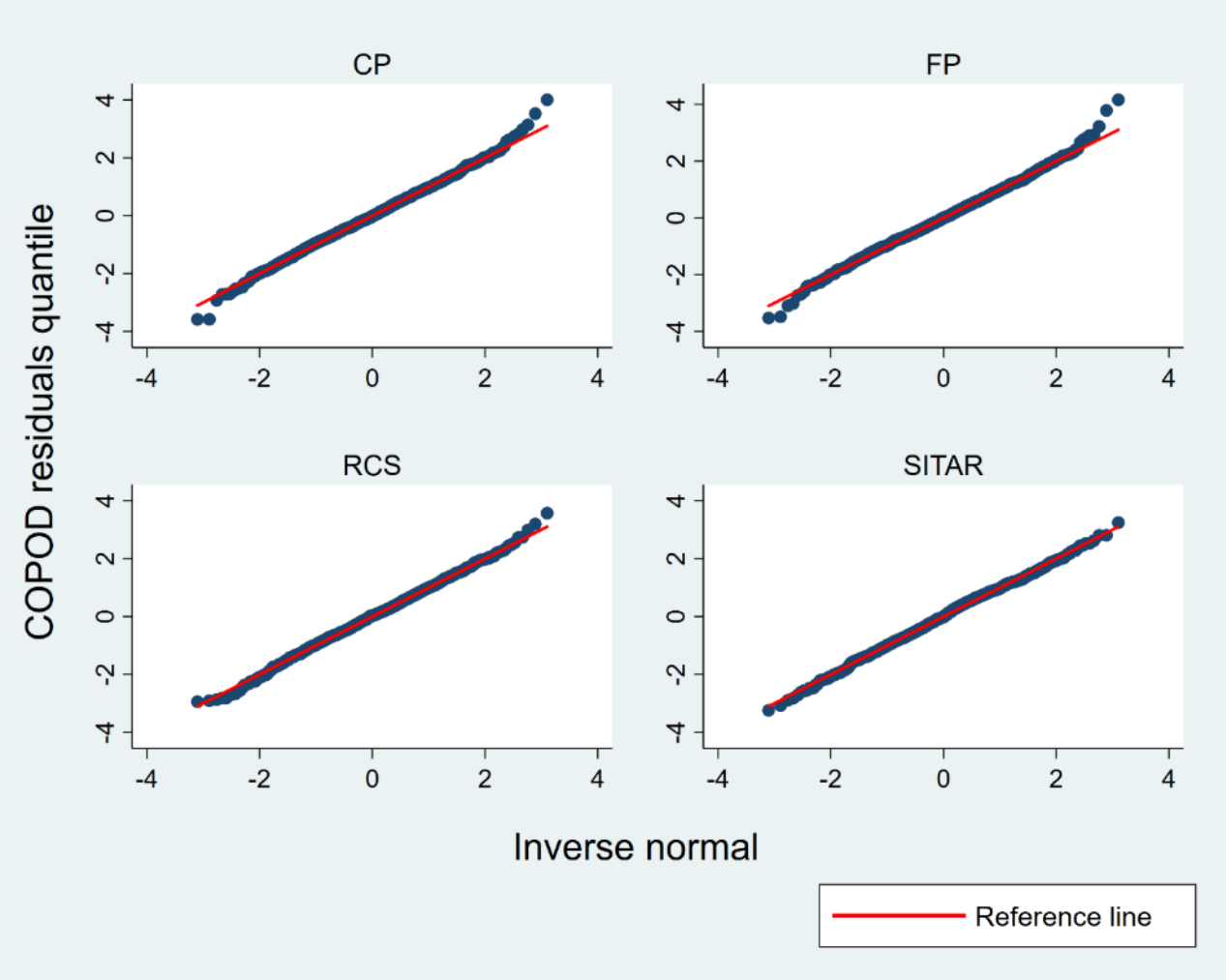


Figure B-23. Normal QQ-plot of level-1 residuals for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the lower jaw length (COPOD) data for females.

Homoscedasticity

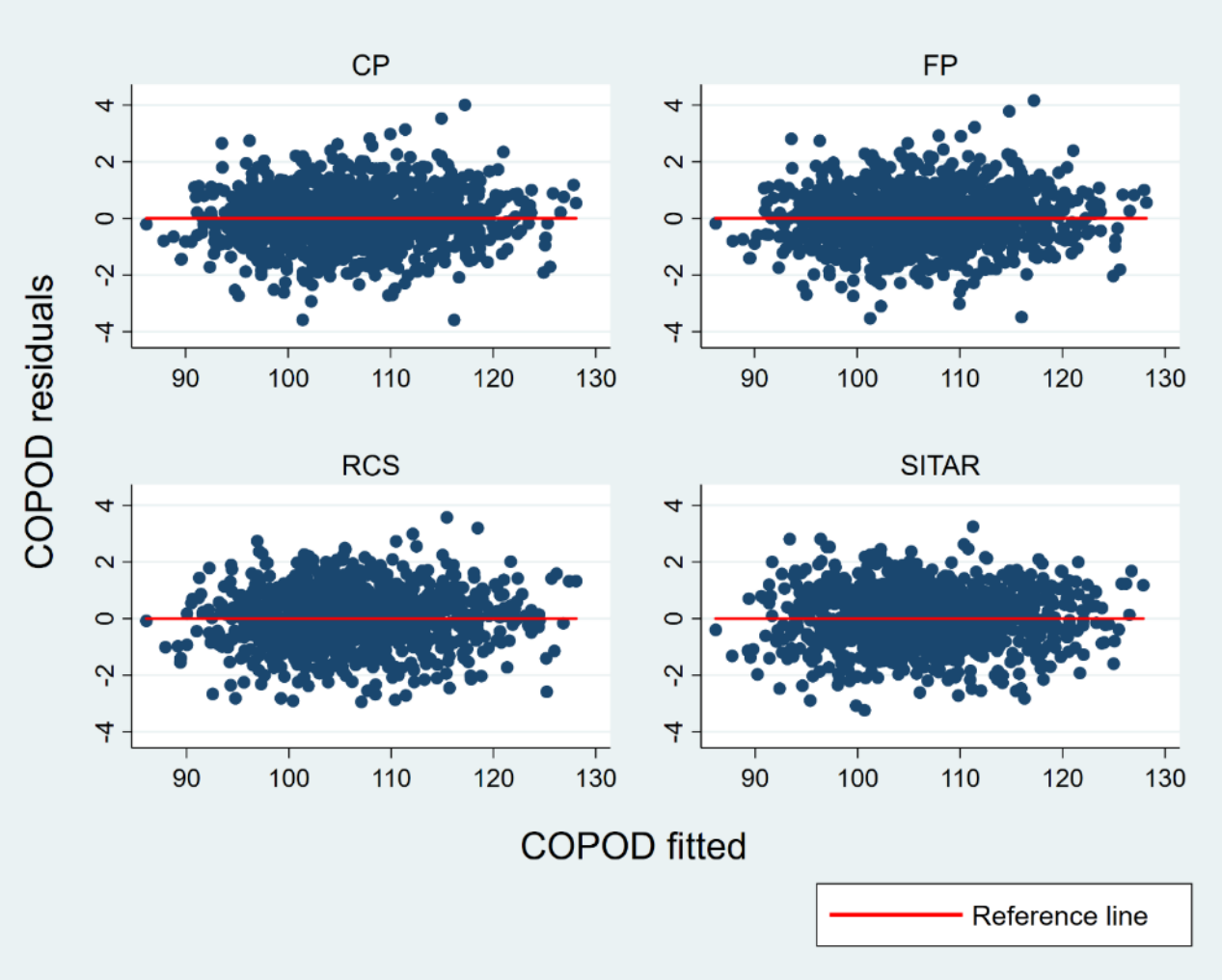


Figure B-24. Level-1 residuals versus fitted plot for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve model applied to the lower jaw length (COPOD) data for females.

Total face height (TFHNP)

Male

Normality

Random effects

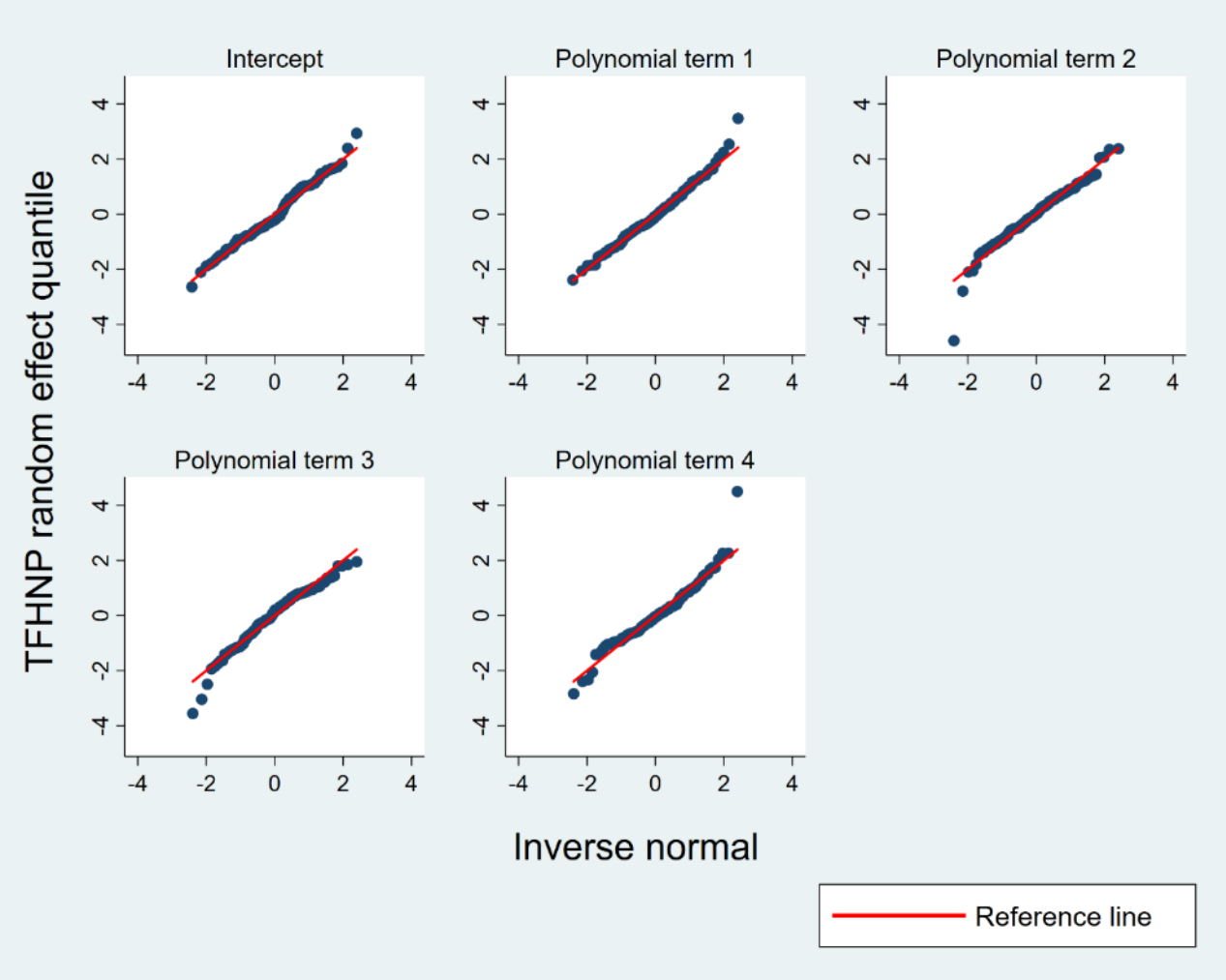


Figure B-25. Normal QQ-plot of random effects for conventional polynomial (CP) growth curve model applied to the total face height (TFHNP) data for males.

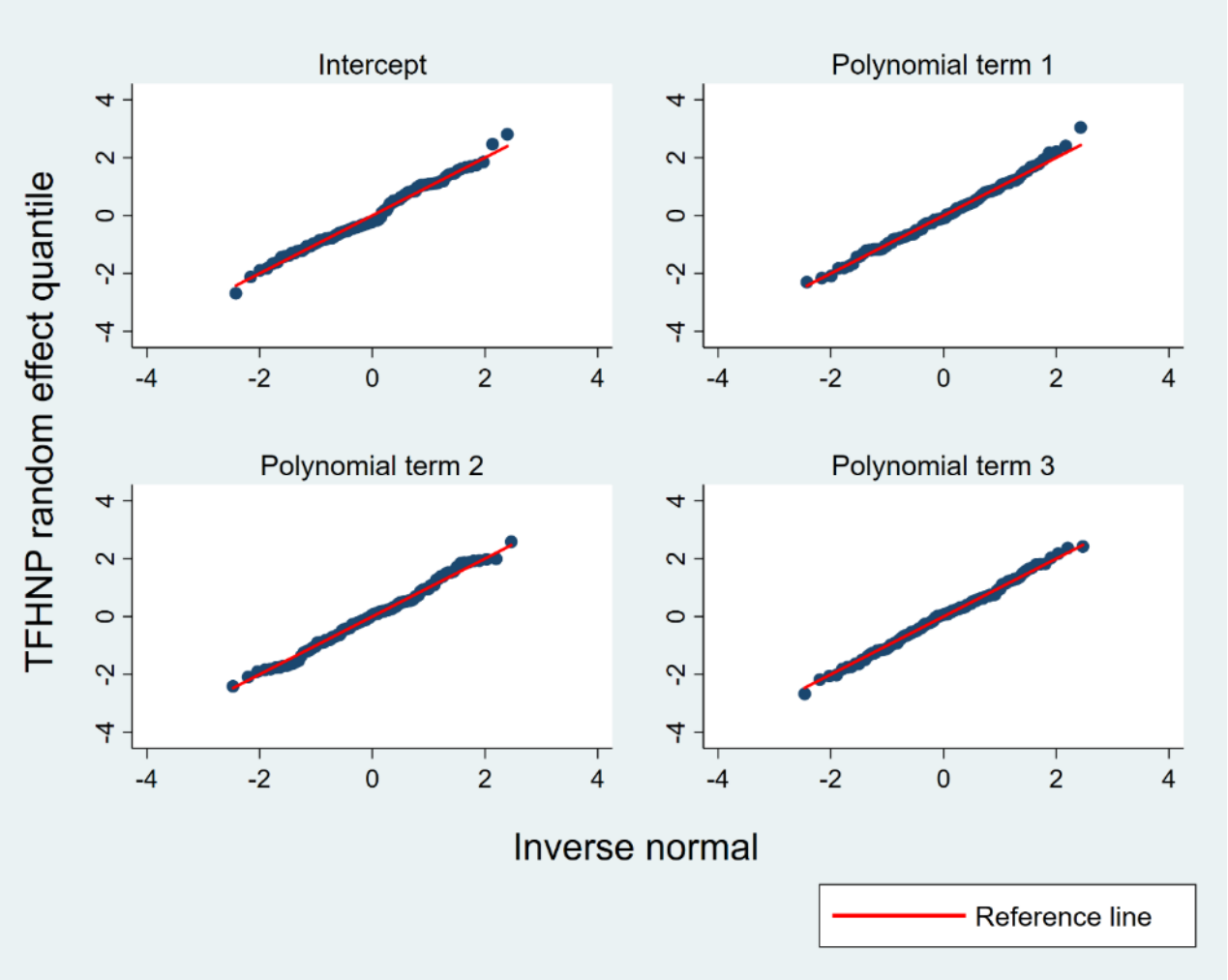


Figure B-26. Normal QQ-plot of random effects for fractional polynomial (FP) growth curve model applied to the total face height (TFHNP) data for males.

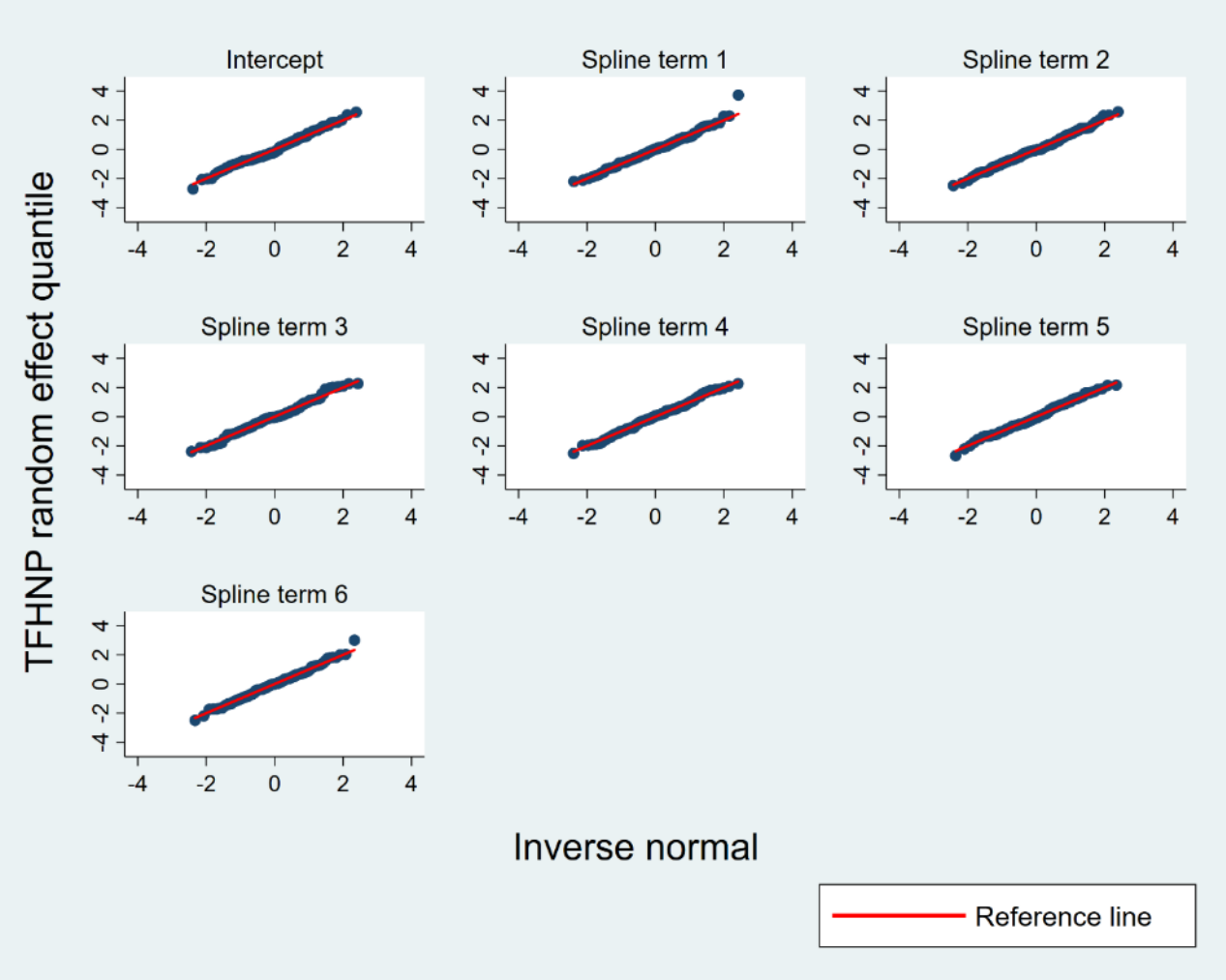


Figure B-27. Normal QQ-plot of random effects for restricted cubic spline (RCS) growth curve model applied to the total face height (TFHNP) data for males.

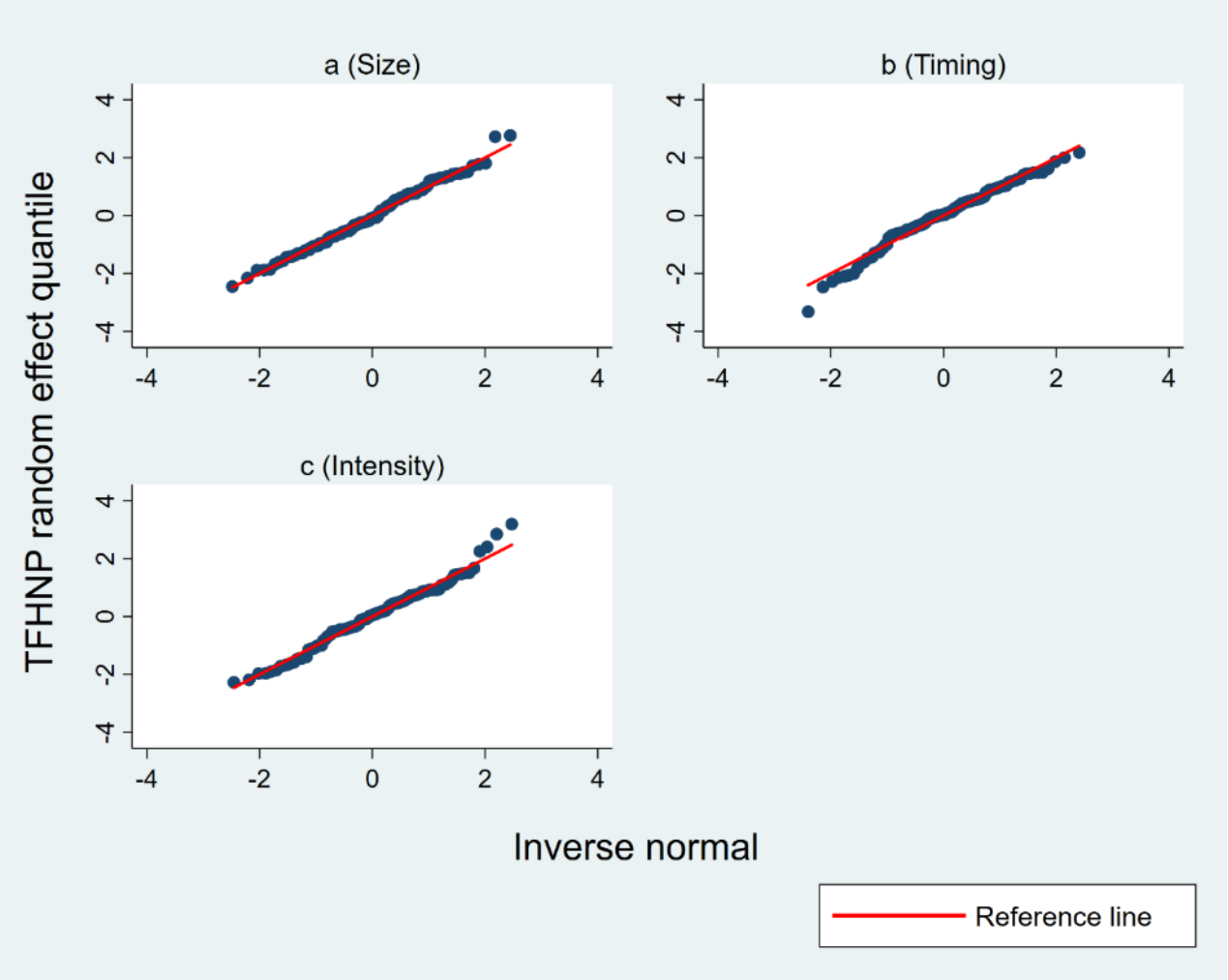


Figure B-28. Normal QQ-plot of random effects for superimposition by translation and rotation (SITAR) growth curve model applied to the total face height (TFHNP) data for males.

Level-1 residuals

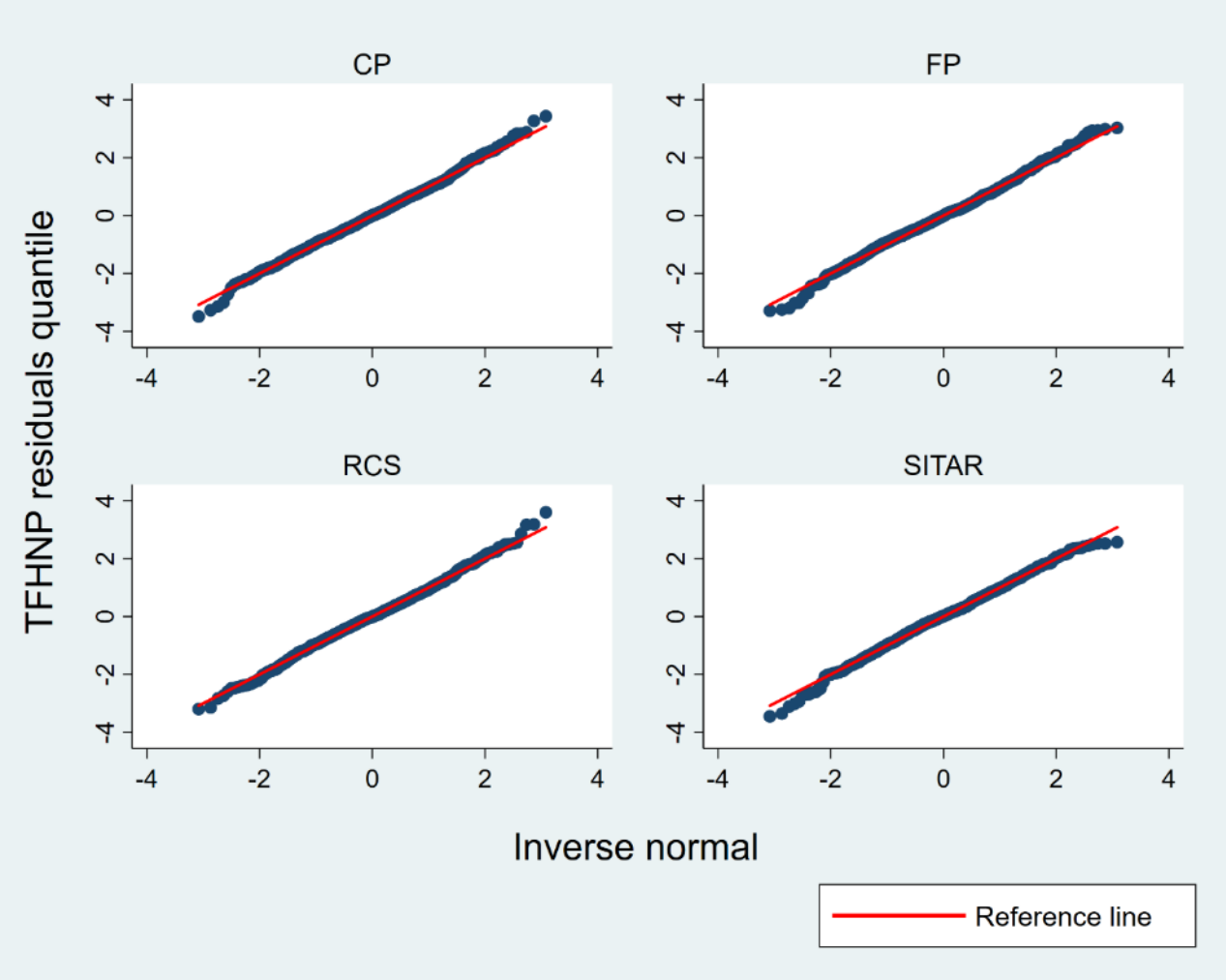


Figure B-29. Normal QQ-plot of level-1 residuals for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the total face height (TFHNP) data for males.

Homoscedasticity

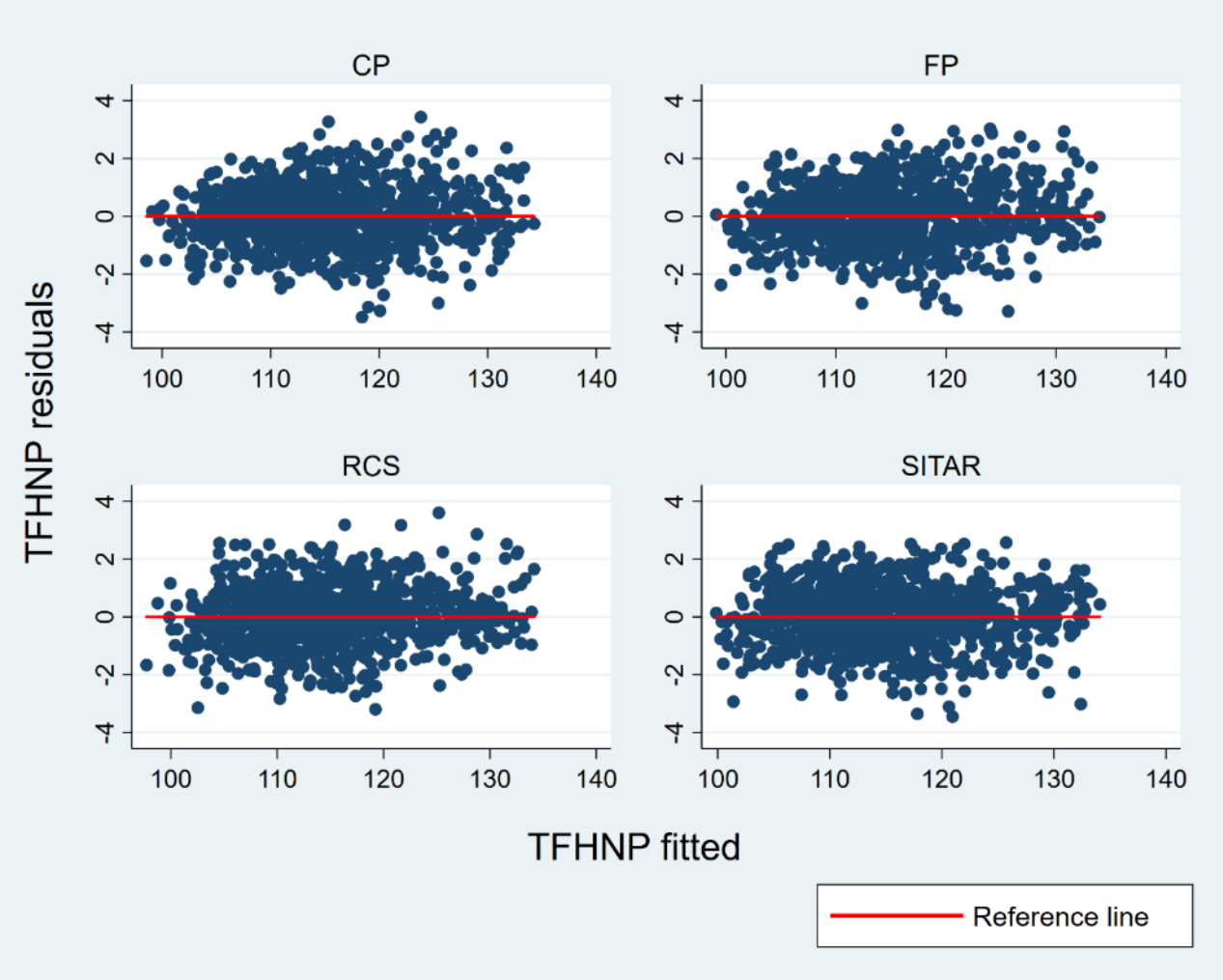


Figure B-30. Level-1 residuals versus fitted plot for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the total face height (TFHNP) data for males.

Female

Normality

Random effects

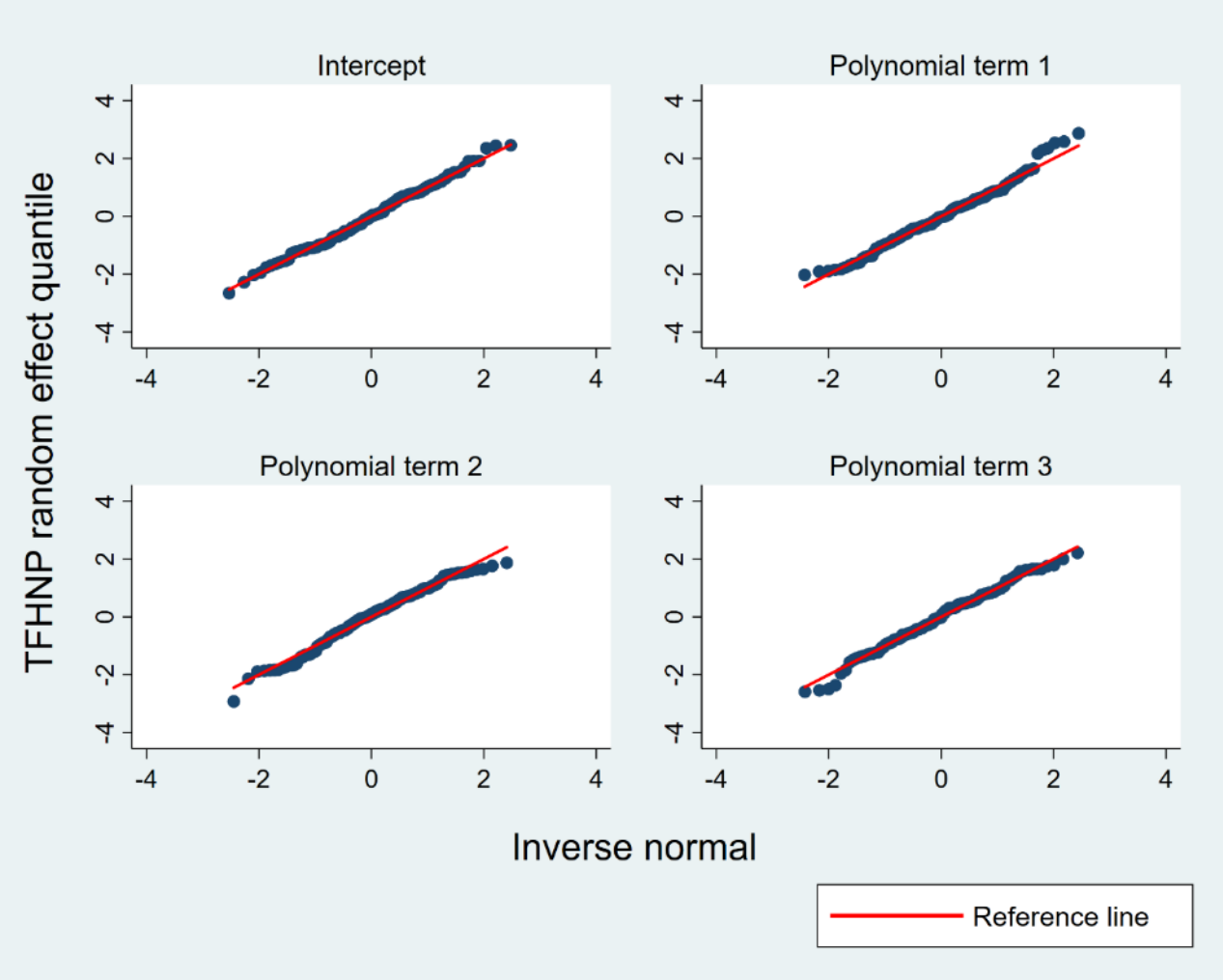


Figure B-31. Normal QQ-plot of random effects for conventional polynomial (CP) growth curve model applied to the total face height (TFHNP) data for females.

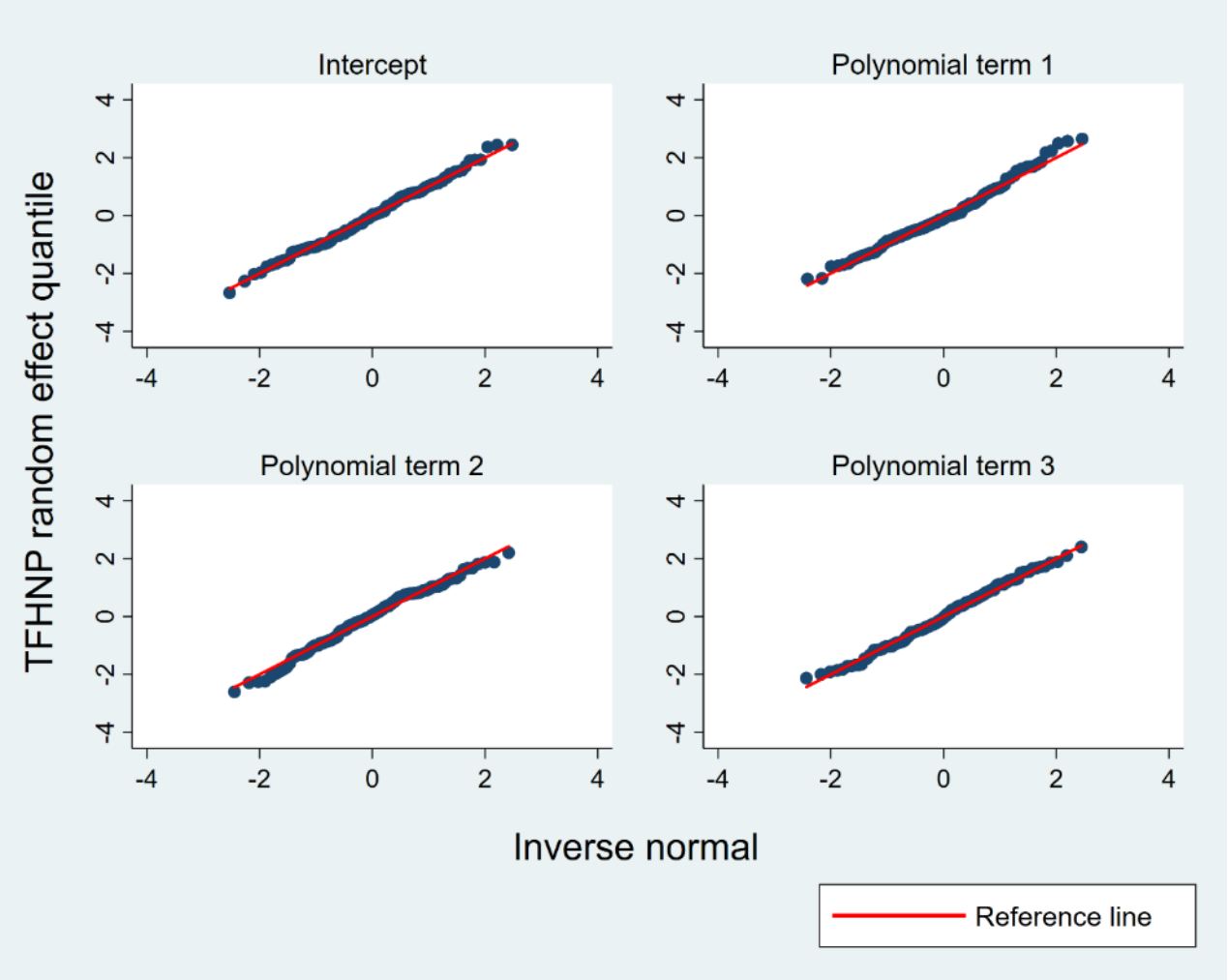


Figure B-32. Normal QQ-plot of random effects for fractional polynomial (FP) growth curve model applied to the total face height (TFHNP) data for females.

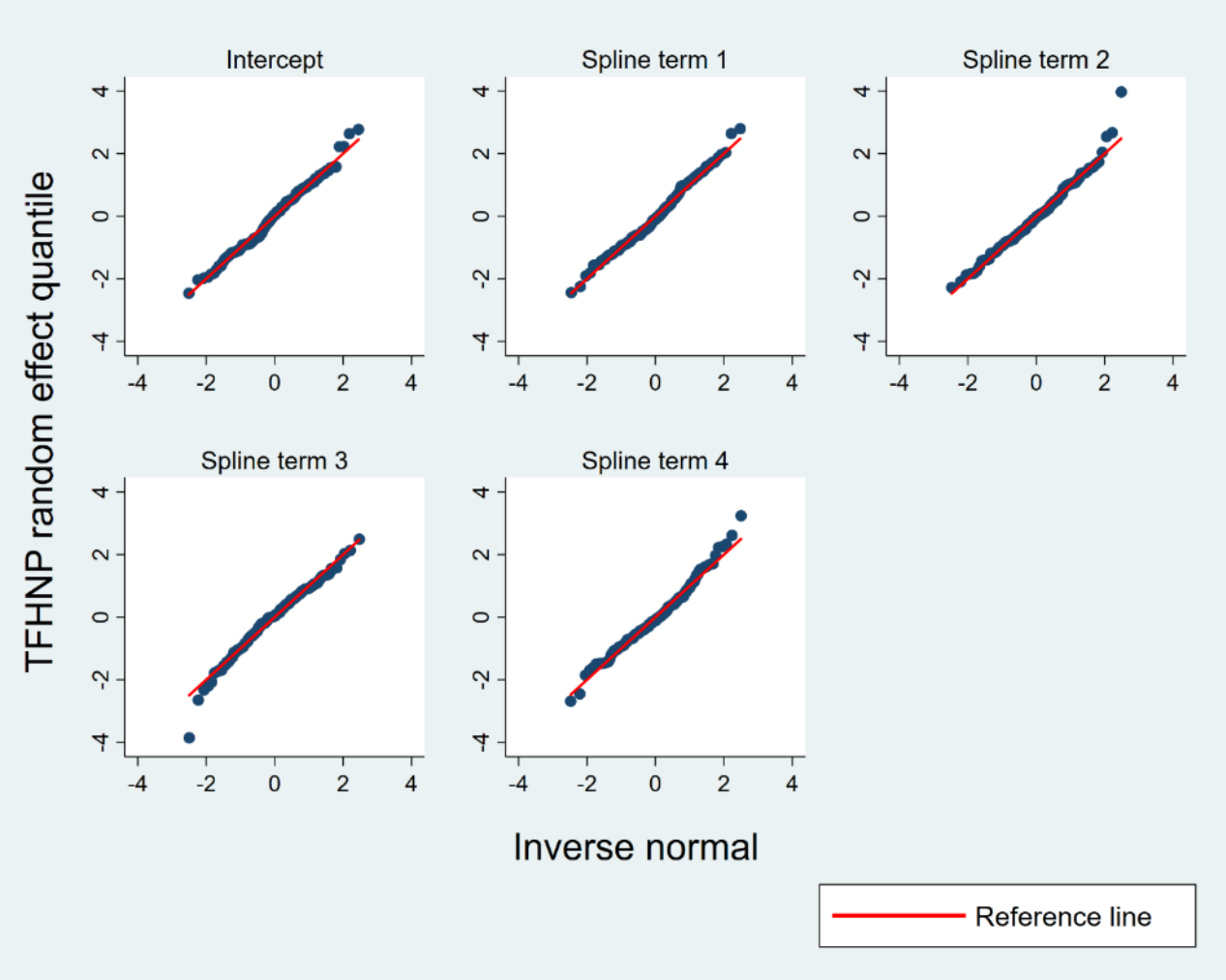


Figure B-33. Normal QQ-plot of random effects for restricted cubic spline (RCS) growth curve model applied to the total face height (TFHNP) data for females.

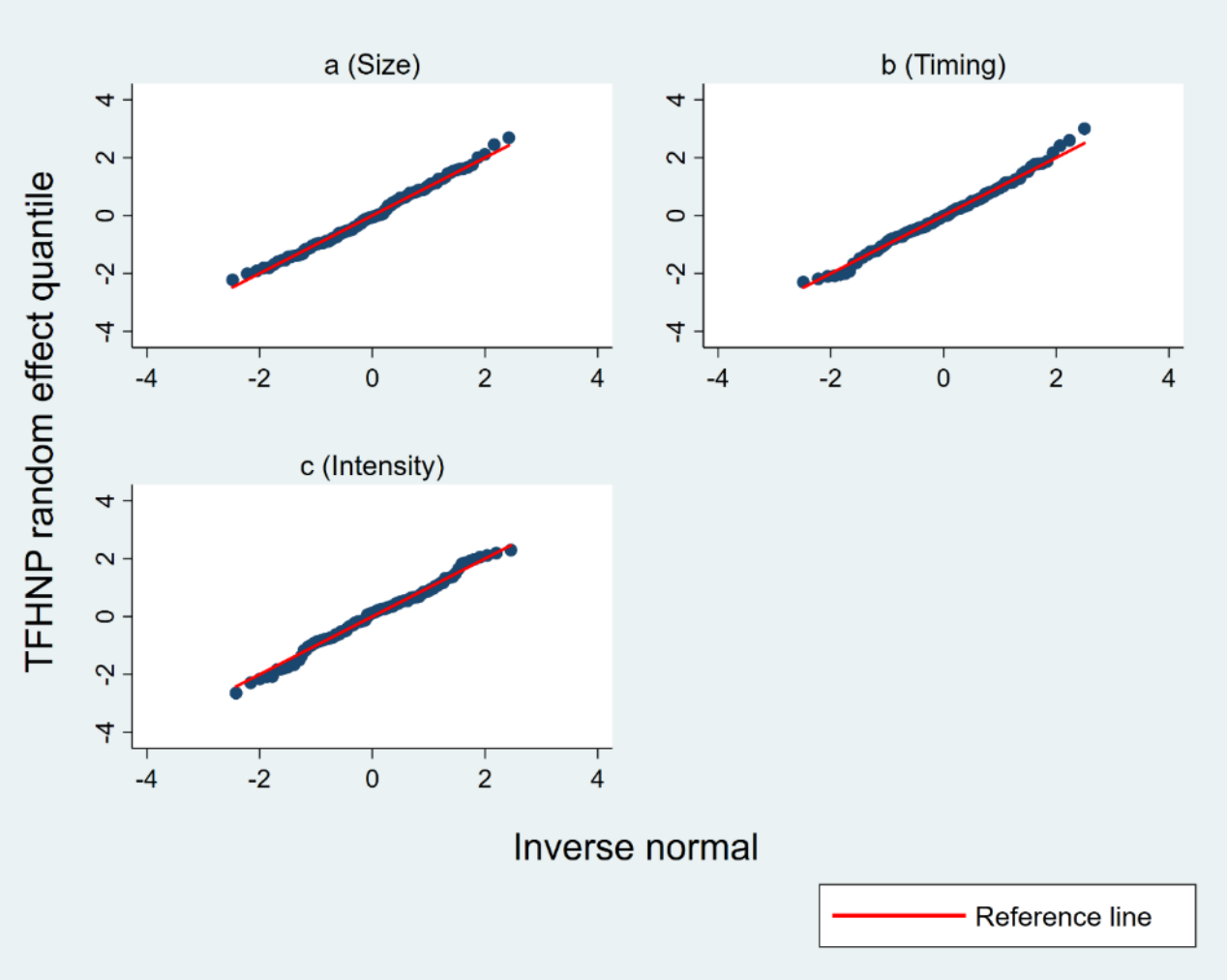


Figure B-34. Normal QQ-plot of random effects for superimposition by translation and rotation (SITAR) growth curve model applied to the total face height (TFHNP) data for females.

Level-1 residuals

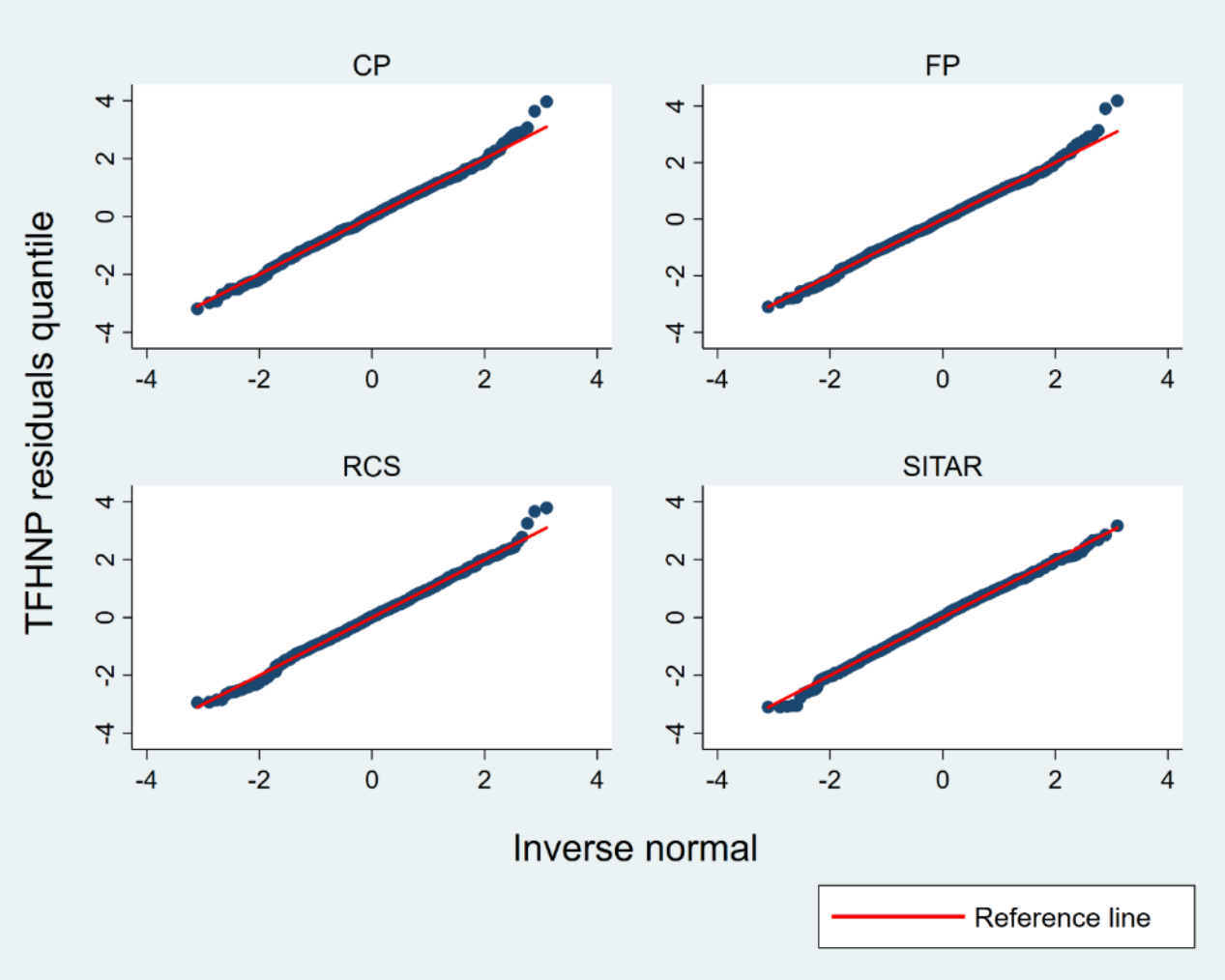


Figure B-35. Normal QQ-plot of level-1 residuals for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the total face height (TFHNP) data for females.

Homoscedasticity

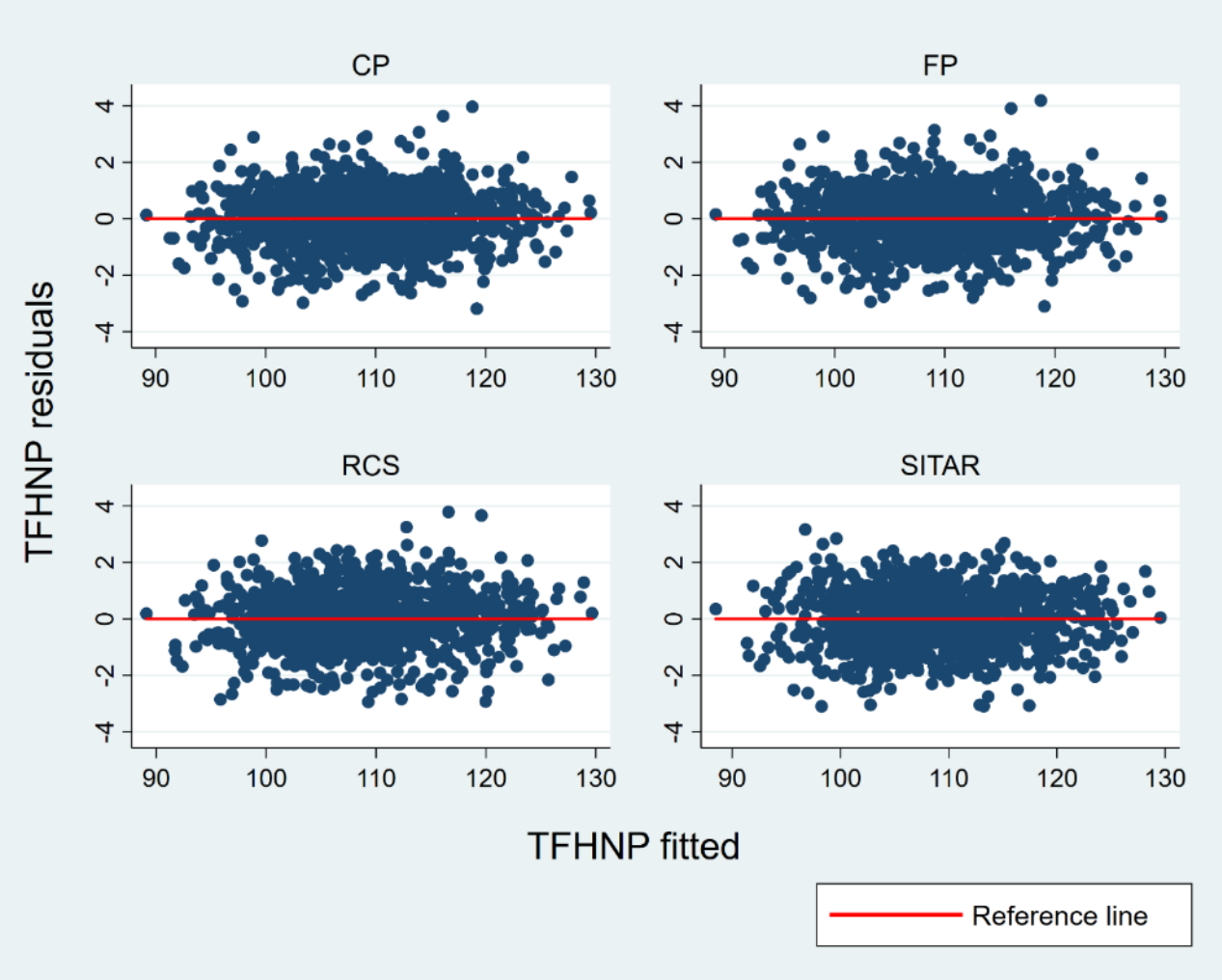


Figure B-36. Level-1 residuals versus fitted plot for conventional polynomial (CP), fractional polynomial (FP), restricted cubic spline (RCS) and superimposition by translation and rotation (SITAR) growth curve models applied to the total face height (TFHNP) data for females.

B.3.4. **Growth acceleration**

Upper jaw length (COPAD)

Male

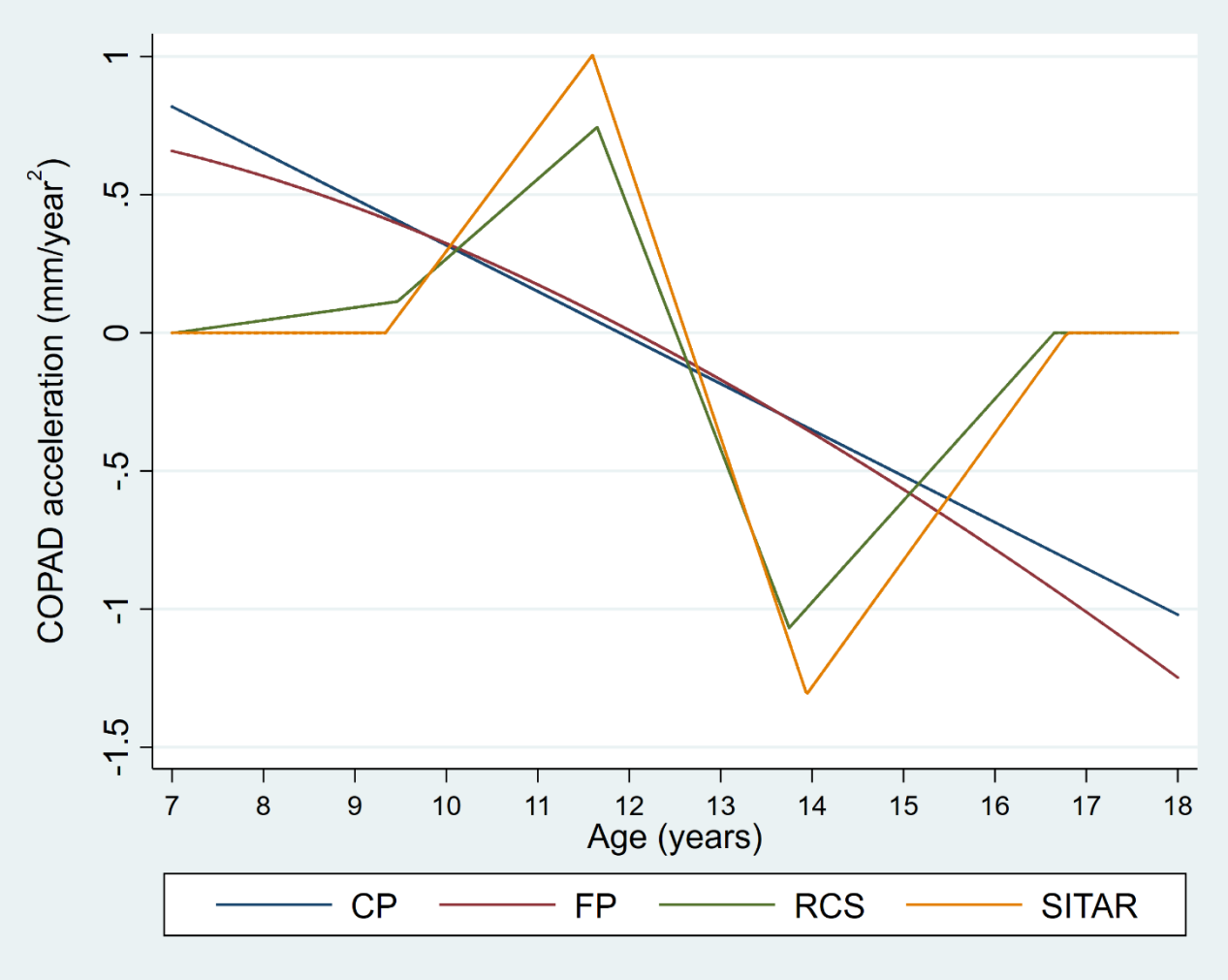


Figure B-37. Population-average growth acceleration curves for the upper jaw length (COPAD) for males.

Female

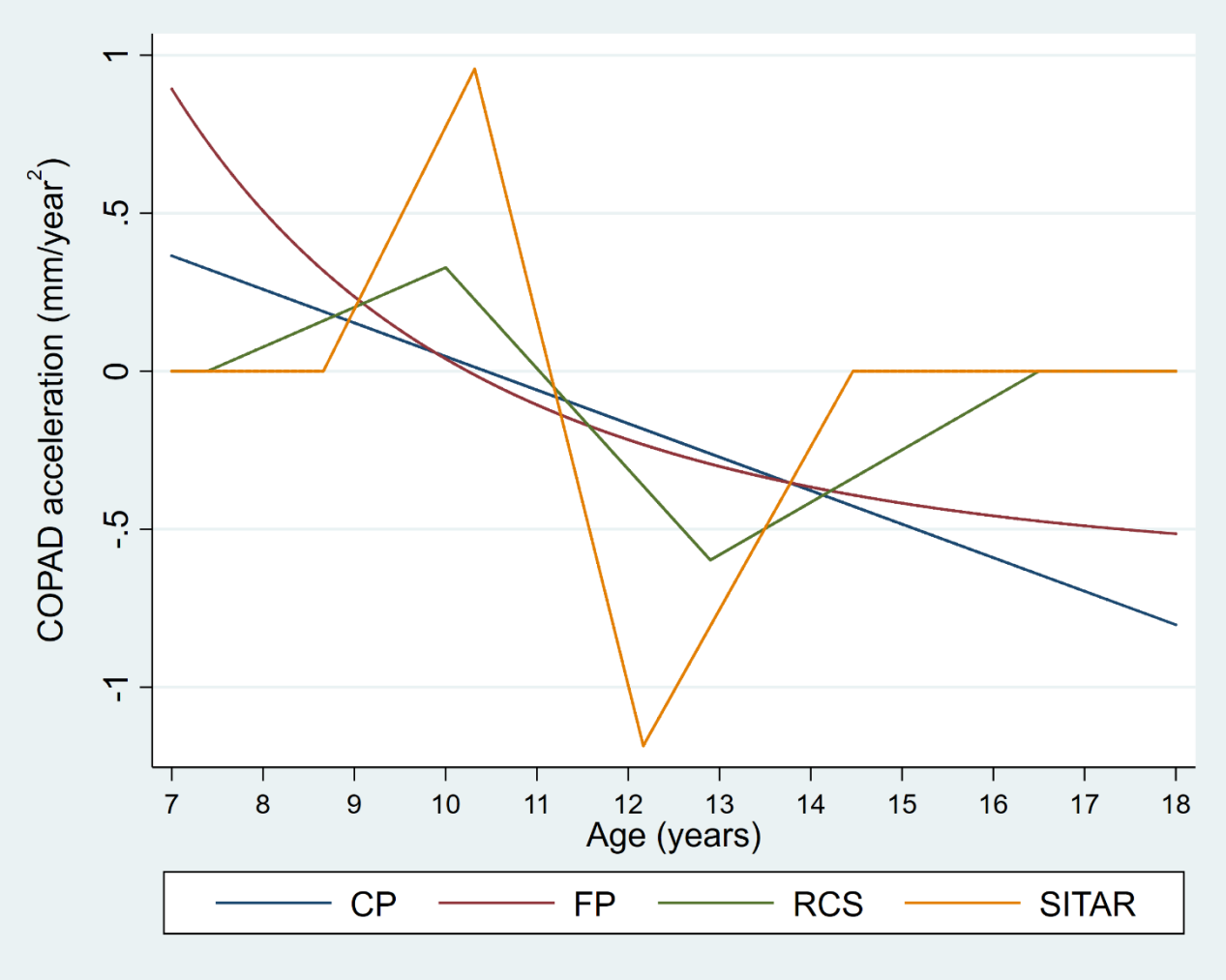


Figure B-38. Population-average growth acceleration curves for the upper jaw length (COPAD) for females.

Lower jaw length (COPOD)

Male

Figure B-39. Population-average growth acceleration curves for the lower jaw length (COPOD) for males.

Female

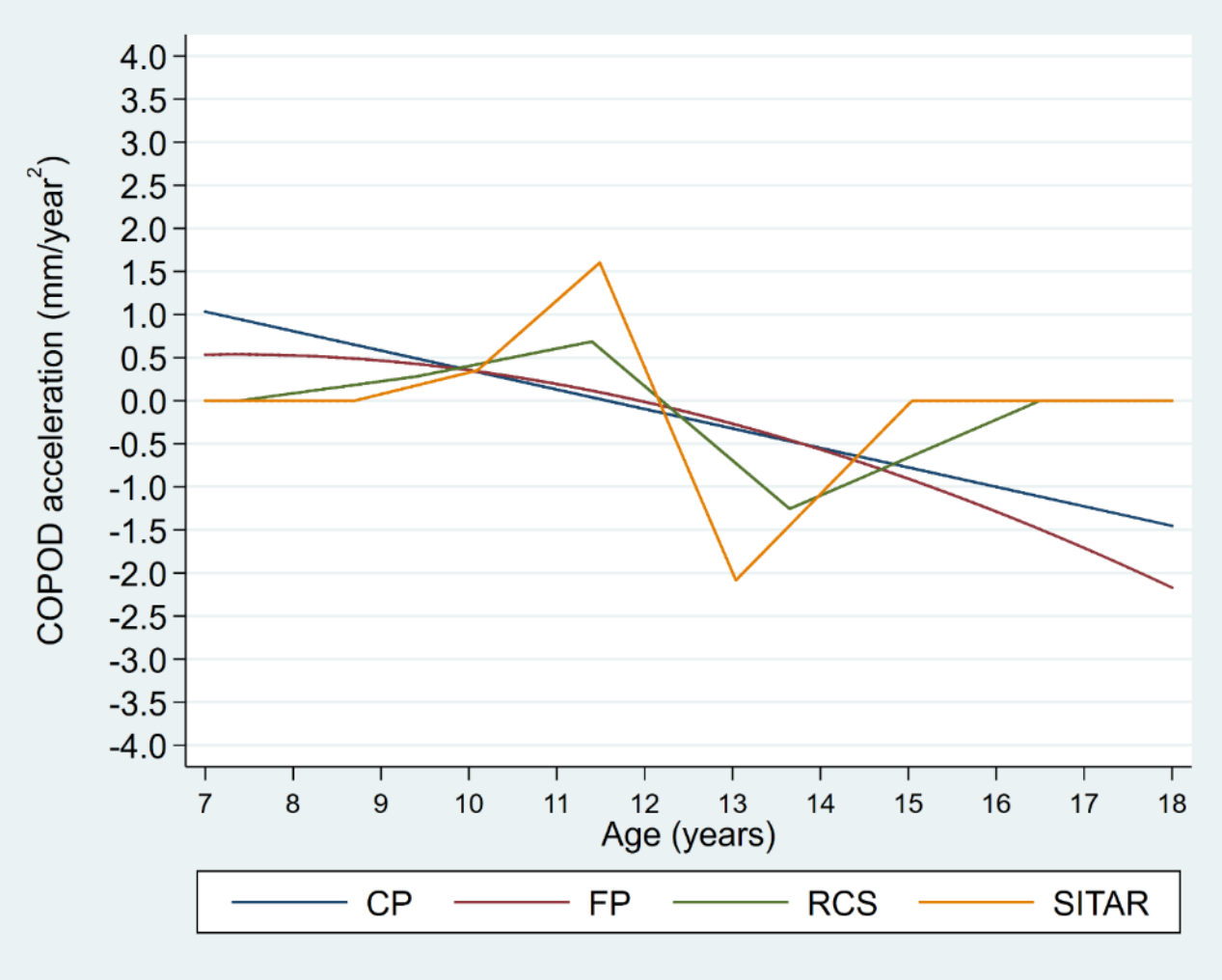


Figure B-40. Population-average growth acceleration curves for the lower jaw length (COPOD) for females.

Total face height (TFHNP)

Male

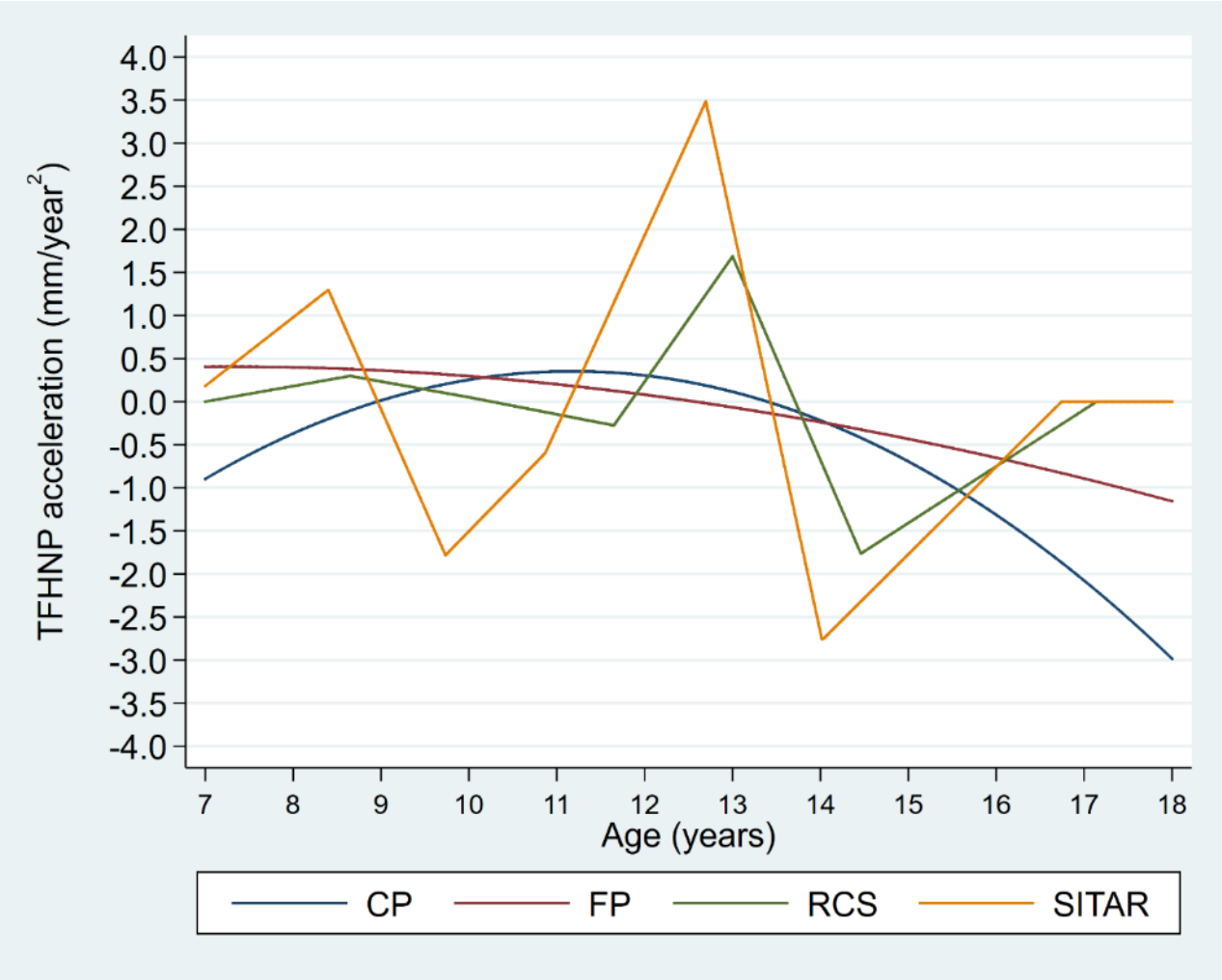


Figure B-41. Population-average acceleration curves for the total face height (TFHNP) for males.

Female

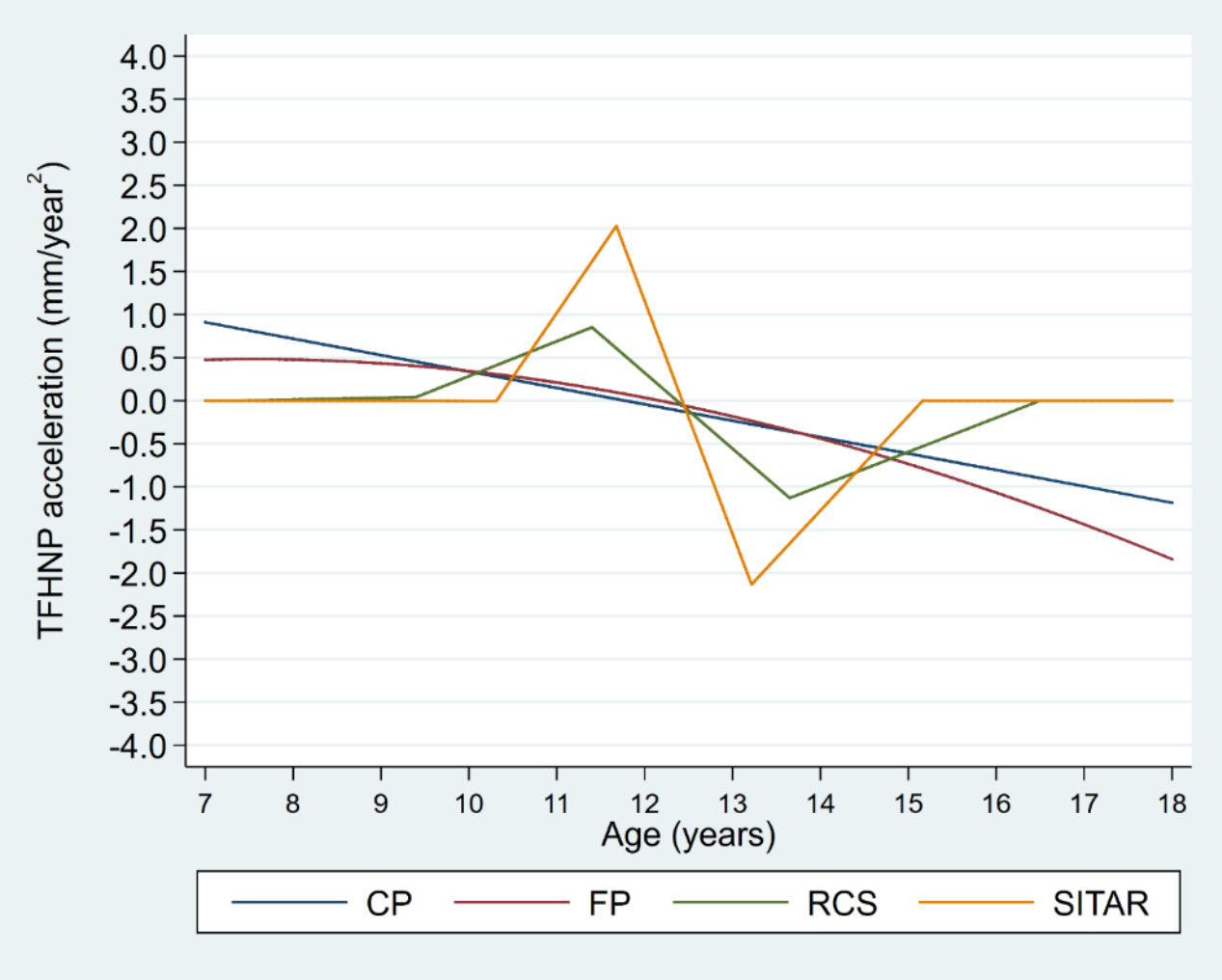


Figure B-42. Population-average growth acceleration curves for the total face height (TFHNP) for females.

APPENDIX C. SEX DIFFERENCES IN JAW GROWTH DURING ADOLESCENCE (CHAPTER 7)

C.1. Software

To compute and test class differences in growth trajectories (distance, velocity and acceleration) adjusted for study (growth study) effects, of a new postestimation Stata program ('covdiffs') was written. This because no program was available, in either Stata (StataCorp, 2015a) , SAS (SAS Institute Inc., 2015) or R (R Core Team, 2015) software , to compute and test differences in covariate growth parameters (distance, velocity and acceleration) estimated using the restricted cubic spline function.

Orsini and Greenland (2011) wrote a postestimation Stata program 'xblc' for a simple linear regression based for the restricted cubic spline model. The 'xblc' computes, at specified values, point estimate and confidence intervals for differences in predictions (i.e., distance) between a continuous covariate and a single reference value (e.g., year 2 year – year 1 year; year 3 year – year 1 year etc.). The 'xblc' does not provide a utility to adjust difference in distance as only one covariate is allowed. The 'xblc' is limited to computing differences in distance.

The 'covdiffs' computes adjusted (e.g., growth study) point estimates and confidence intervals at specified values of a continuous or categorical covariate (e.g., class). Further, as a single reference value limits the utility of the program as it does not allow computing pairwise differences for a nominal covariate with more than two levels (e.g., class), the 'covdiffs' updates the reference category to compute all possible pairwise differences (e.g., Class II vs Class I, Class III vs Class I, Class III vs Class II). The 'covdiffs' computes differences in distance, velocity and acceleration.

The 'covdiffs' can be used after fitting the RCS model using Stata's 'mixed' command (StataCorp, 2017b) or the user written 'runmlwin' command (Leckie & Charlton, 2013a). The 'covdiffs' can be easily extended to compute and test covariate differences in distance, velocity and acceleration for other linear GCMs such as CP and FP.

C.2. Results

C.2.1. **Distance**

Here I display numerical results (as Tables) for class differences in the upper jaw length, the lower jaw length, and the total face height for males and females. The corresponding graphical results have already been presented in Chapter 7 (see Section 7.3).

Upper jaw length (COPAD)

Table C-1. Class differences in the upper jaw length (COPAD) distance between seven and 18 years of age for males.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	81.33	82.24	79.71	0.92 (-0.30 to 2.13)	-1.61 (-3.77 to 0.54)	-2.53 (-4.46 to -0.61)
8	82.46	83.59	80.89	1.14 (-0.07 to 2.34)	-1.56 (-3.75 to 0.63)	-2.70 (-4.67 to -0.73)
9	83.67	84.96	82.15	1.29 (0.00 to 2.59)	-1.52 (-3.90 to 0.86)	-2.81 (-4.95 to -0.67)
10	85.04	86.37	83.55	1.33 (0.03 to 2.63)	-1.49 (-3.87 to 0.89)	-2.82 (-4.96 to -0.68)
11	86.71	88.06	85.11	1.34 (0.02 to 2.66)	-1.61 (-4.00 to 0.79)	-2.95 (-5.10 to -0.80)
12	88.86	90.40	86.78	1.54 (0.12 to 2.96)	-2.08 (-4.65 to 0.50)	-3.62 (-5.93 to -1.31)
13	91.30	93.27	88.42	1.97 (0.42 to 3.51)	-2.89 (-5.69 to -0.08)	-4.86 (-7.37 to -2.34)
14	93.34	95.71	89.69	2.37 (0.73 to 4.01)	-3.65 (-6.65 to -0.66)	-6.03 (-8.71 to -3.34)
15	94.55	97.16	90.42	2.60 (0.99 to 4.21)	-4.14 (-7.08 to -1.20)	-6.74 (-9.38 to -4.10)
16	95.21	97.92	90.79	2.70 (1.14 to 4.26)	-4.42 (-7.27 to -1.57)	-7.12 (-9.69 to -4.56)
17	95.65	98.41	91.02	2.75 (1.11 to 4.39)	-4.63 (-7.65 to -1.60)	-7.38 (-10.10 to -4.66)
18	96.08	98.88	91.25	2.80 (0.92 to 4.68)	-4.83 (-8.32 to -1.34)	-7.63 (-10.78 to -4.47)

COPAD condyle-point-A measurement in millimetres

Table C-2. Class differences in the upper jaw length (COPAD) distance between seven and 18 years of age for females.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	80.74	81.44	77.89	0.70 (-0.46 to 1.86)	-2.85 (-4.70 to -0.99)	-3.55 (-5.41 to -1.69)
8	81.81	82.60	79.04	0.79 (-0.20 to 1.77)	-2.78 (-4.34 to -1.21)	-3.56 (-5.14 to -1.99)
9	82.97	83.86	80.22	0.89 (-0.01 to 1.78)	-2.75 (-4.18 to -1.32)	-3.64 (-5.08 to -2.20)
10	84.32	85.35	81.49	1.02 (0.14 to 1.90)	-2.84 (-4.24 to -1.43)	-3.86 (-5.28 to -2.44)
11	85.93	87.14	82.86	1.21 (0.32 to 2.10)	-3.07 (-4.50 to -1.65)	-4.28 (-5.72 to -2.84)
12	87.54	88.95	84.16	1.41 (0.46 to 2.35)	-3.39 (-4.89 to -1.88)	-4.79 (-6.31 to -3.28)
13	88.86	90.44	85.19	1.58 (0.56 to 2.60)	-3.67 (-5.30 to -2.04)	-5.25 (-6.89 to -3.61)
14	89.66	91.36	85.82	1.70 (0.64 to 2.77)	-3.84 (-5.54 to -2.14)	-5.54 (-7.26 to -3.83)
15	90.06	91.84	86.13	1.78 (0.69 to 2.87)	-3.93 (-5.67 to -2.19)	-5.71 (-7.46 to -3.96)
16	90.22	92.04	86.25	1.83 (0.70 to 2.95)	-3.96 (-5.76 to -2.17)	-5.79 (-7.59 to -3.99)
17	90.29	92.15	86.31	1.86 (0.67 to 3.05)	-3.98 (-5.88 to -2.08)	-5.85 (-7.76 to -3.93)
18	90.36	92.26	86.36	1.90 (0.61 to 3.20)	-4.00 (-6.06 to -1.93)	-5.90 (-7.98 to -3.82)

COPAD condyle-point-A measurement in millimetres

Lower jaw length (COPOD)

Table C-3. Class differences in the lower jaw length (COPOD) distance between seven and 18 years of age for males.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	104.01	101.14	105.38	-2.87 (-4.60 to -1.13)	1.37 (-1.78 to 4.51)	4.23 (1.40 to 7.06)
8	104.93	103.46	107.24	-1.47 (-2.83 to -0.10)	2.31 (-0.27 to 4.90)	3.78 (1.44 to 6.13)
9	106.58	105.45	109.36	-1.13 (-2.41 to 0.15)	2.77 (0.38 to 5.17)	3.90 (1.74 to 6.07)
10	108.97	107.10	111.94	-1.86 (-3.22 to -0.51)	2.97 (0.33 to 5.61)	4.83 (2.43 to 7.24)
11	111.13	108.82	114.99	-2.31 (-3.75 to -0.88)	3.86 (1.15 to 6.56)	6.17 (3.73 to 8.62)
12	112.83	110.84	117.96	-1.99 (-3.56 to -0.42)	5.13 (2.20 to 8.06)	7.12 (4.47 to 9.77)
13	114.92	113.32	121.50	-1.61 (-3.39 to 0.18)	6.58 (3.18 to 9.98)	8.18 (5.10 to 11.27)
14	118.71	116.09	127.08	-2.62 (-4.45 to -0.79)	8.37 (4.90 to 11.84)	10.99 (7.85 to 14.13)
15	122.00	118.12	131.69	-3.88 (-5.82 to -1.94)	9.69 (5.95 to 13.43)	13.57 (10.17 to 16.98)
16	123.70	119.06	133.91	-4.64 (-6.61 to -2.67)	10.21 (6.41 to 14.02)	14.85 (11.39 to 18.32)
17	124.52	119.40	134.81	-5.12 (-7.50 to -2.73)	10.30 (5.76 to 14.83)	15.41 (11.28 to 19.54)
18	125.14	119.62	135.43	-5.53 (-8.77 to -2.28)	10.28 (4.13 to 16.43)	15.81 (10.21 to 21.41)

COPOD condyle-pogonion measurement in millimetres

Table C-4. Class differences in the lower jaw length (COPOD) distance between seven and 18 years of age for females.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	98.38	97.73	100.65	-0.66 (-2.81 to 1.50)	2.27 (-1.17 to 5.71)	2.92 (-0.53 to 6.38)
8	99.78	98.90	102.38	-0.88 (-2.90 to 1.14)	2.60 (-0.62 to 5.82)	3.48 (0.25 to 6.72)
9	101.23	100.21	104.05	-1.02 (-3.02 to 0.97)	2.82 (-0.36 to 6.01)	3.84 (0.64 to 7.05)
10	102.84	101.89	105.63	-0.96 (-2.97 to 1.05)	2.79 (-0.42 to 5.99)	3.75 (0.52 to 6.97)
11	104.84	103.99	107.57	-0.85 (-2.97 to 1.26)	2.72 (-0.64 to 6.09)	3.58 (0.18 to 6.97)
12	107.51	106.45	110.63	-1.06 (-3.28 to 1.17)	3.12 (-0.42 to 6.66)	4.18 (0.61 to 7.74)
13	110.42	108.87	114.33	-1.54 (-3.82 to 0.73)	3.92 (0.28 to 7.55)	5.46 (1.80 to 9.12)
14	112.63	110.64	117.26	-2.00 (-4.32 to 0.33)	4.62 (0.91 to 8.34)	6.62 (2.88 to 10.36)
15	113.74	111.50	118.74	-2.24 (-4.57 to 0.08)	5.00 (1.28 to 8.72)	7.24 (3.50 to 10.99)
16	114.15	111.79	119.31	-2.36 (-4.67 to -0.05)	5.16 (1.47 to 8.84)	7.52 (3.80 to 11.23)
17	114.32	111.89	119.56	-2.43 (-4.75 to -0.10)	5.24 (1.53 to 8.95)	7.67 (3.93 to 11.40)
18	114.48	111.98	119.80	-2.49 (-4.87 to -0.11)	5.32 (1.52 to 9.12)	7.81 (3.97 to 11.65)

COPOD condyle-pogonion measurement in millimetres

Total face height (TFHNP)

Table C-5. Class differences in the total face height (TFHNP) distance between seven and 18 years of age for males.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	107.03	104.79	110.68	-2.25 (-4.21 to -0.28)	3.64 (0.06 to 7.22)	5.89 (2.67 to 9.10)
8	107.88	106.95	112.33	-0.92 (-2.57 to 0.73)	4.45 (1.36 to 7.54)	5.37 (2.58 to 8.16)
9	109.56	108.82	114.05	-0.74 (-2.33 to 0.84)	4.49 (1.55 to 7.43)	5.23 (2.58 to 7.88)
10	112.10	110.39	116.24	-1.71 (-3.42 to -0.01)	4.13 (0.89 to 7.38)	5.85 (2.90 to 8.79)
11	114.50	112.10	119.24	-2.41 (-4.24 to -0.57)	4.74 (1.34 to 8.13)	7.14 (4.09 to 10.20)
12	116.36	114.07	121.24	-2.29 (-4.23 to -0.34)	4.88 (1.27 to 8.49)	7.17 (3.92 to 10.42)
13	118.43	116.45	123.07	-1.98 (-4.08 to 0.11)	4.64 (0.57 to 8.72)	6.62 (2.92 to 10.33)
14	122.06	119.21	128.34	-2.85 (-4.99 to -0.71)	6.28 (2.15 to 10.41)	9.13 (5.37 to 12.89)
15	125.25	121.23	133.32	-4.02 (-6.31 to -1.73)	8.07 (3.55 to 12.59)	12.09 (7.96 to 16.23)
16	126.99	122.12	135.88	-4.87 (-7.24 to -2.51)	8.89 (4.19 to 13.60)	13.76 (9.46 to 18.07)
17	127.92	122.38	137.08	-5.54 (-8.26 to -2.81)	9.17 (3.80 to 14.53)	14.70 (9.79 to 19.61)
18	128.67	122.51	137.99	-6.16 (-9.65 to -2.67)	9.32 (2.48 to 16.17)	15.48 (9.22 to 21.75)

TFHNP total face height measurement in millimetres

Table C-6. Class differences in the total face height (TFHNP) distance between seven and 18 years of age for females.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	101.80	101.26	104.39	-0.54 (-2.81 to 1.73)	2.59 (-1.02 to 6.20)	3.13 (-0.50 to 6.75)
8	103.13	102.20	106.04	-0.92 (-3.05 to 1.20)	2.91 (-0.46 to 6.29)	3.84 (0.44 to 7.23)
9	104.42	103.25	107.54	-1.17 (-3.25 to 0.90)	3.12 (-0.19 to 6.43)	4.29 (0.95 to 7.62)
10	105.64	104.56	108.67	-1.09 (-3.15 to 0.98)	3.03 (-0.27 to 6.32)	4.11 (0.80 to 7.43)
11	107.10	106.23	109.97	-0.87 (-2.99 to 1.25)	2.87 (-0.51 to 6.25)	3.74 (0.34 to 7.14)
12	109.32	108.32	112.47	-1.00 (-3.21 to 1.21)	3.15 (-0.35 to 6.66)	4.15 (0.62 to 7.69)
13	111.97	110.48	115.81	-1.49 (-3.76 to 0.77)	3.84 (0.23 to 7.45)	5.33 (1.69 to 8.97)
14	114.09	112.08	118.57	-2.01 (-4.32 to 0.31)	4.48 (0.78 to 8.18)	6.49 (2.76 to 10.22)
15	115.22	112.87	120.06	-2.35 (-4.67 to -0.04)	4.84 (1.14 to 8.54)	7.20 (3.47 to 10.92)
16	115.74	113.15	120.75	-2.59 (-4.88 to -0.30)	5.02 (1.35 to 8.68)	7.61 (3.91 to 11.30)
17	116.03	113.25	121.16	-2.79 (-5.08 to -0.49)	5.13 (1.45 to 8.82)	7.92 (4.20 to 11.63)
18	116.32	113.34	121.56	-2.98 (-5.33 to -0.63)	5.24 (1.46 to 9.02)	8.22 (4.41 to 12.04)

TFHNP total face height measurement in millimetres

C.2.2. **Growth velocity**

Here I display numerical results (as Tables) for class differences in the upper jaw length, the lower jaw length, and the total face height for males and females. The corresponding graphical results have already been presented in Chapter 7 (see Section 7.3).

Upper jaw length (COPAD)

Table C-7. Class differences in the upper jaw length (COPAD) growth velocity between seven and 18 years of age for males.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	1.12	1.35	1.17	0.23 (-0.29 to 0.74)	0.05 (-0.89 to 1.00)	-0.17 (-1.03 to 0.68)
8	1.16	1.35	1.21	0.20 (-0.24 to 0.63)	0.05 (-0.74 to 0.84)	-0.15 (-0.86 to 0.57)
9	1.28	1.38	1.32	0.10 (-0.13 to 0.34)	0.04 (-0.40 to 0.48)	-0.07 (-0.46 to 0.33)
10	1.49	1.48	1.48	-0.01 (-0.37 to 0.35)	-0.01 (-0.74 to 0.73)	0.00 (-0.68 to 0.68)
11	1.88	1.95	1.62	0.07 (-0.33 to 0.47)	-0.26 (-1.07 to 0.55)	-0.33 (-1.07 to 0.41)
12	2.40	2.75	1.71	0.35 (0.01 to 0.70)	-0.69 (-1.40 to -0.01)	-1.04 (-1.69 to -0.39)
13	2.36	2.82	1.51	0.46 (0.08 to 0.84)	-0.85 (-1.64 to -0.07)	-1.31 (-2.04 to -0.59)
14	1.61	1.92	0.98	0.32 (0.08 to 0.55)	-0.62 (-1.11 to -0.14)	-0.94 (-1.39 to -0.49)
15	0.88	1.03	0.52	0.15 (-0.22 to 0.52)	-0.36 (-1.07 to 0.35)	-0.52 (-1.16 to 0.13)
16	0.49	0.56	0.27	0.06 (-0.44 to 0.57)	-0.23 (-1.19 to 0.74)	-0.29 (-1.17 to 0.59)
17	0.42	0.47	0.22	0.05 (-0.48 to 0.58)	-0.20 (-1.22 to 0.82)	-0.25 (-1.17 to 0.68)
18	0.42	0.47	0.22	0.05 (-0.48 to 0.58)	-0.20 (-1.22 to 0.82)	-0.25 (-1.17 to 0.68)

COPAD condyle-point-A measurement in millimetres

Table C-8. Class differences in the upper jaw length (COPAD) growth velocity between seven and 18 years of age for females.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	1.07	1.16	1.15	0.09 (-0.24 to 0.41)	0.07 (-0.47 to 0.61)	-0.01 (-0.55 to 0.53)
8	1.10	1.19	1.16	0.09 (-0.22 to 0.40)	0.06 (-0.45 to 0.57)	-0.03 (-0.54 to 0.48)
9	1.23	1.35	1.22	0.11 (-0.11 to 0.34)	-0.02 (-0.39 to 0.35)	-0.13 (-0.50 to 0.24)
10	1.50	1.66	1.33	0.16 (0.05 to 0.28)	-0.16 (-0.35 to -0.02)	-0.33 (-0.52 to -0.14)
11	1.66	1.86	1.37	0.20 (0.02 to 0.38)	-0.29 (-0.59 to -0.01)	-0.49 (-0.79 to -0.19)
12	1.52	1.72	1.20	0.20 (0.01 to 0.40)	-0.31 (-0.66 to -0.03)	-0.52 (-0.86 to -0.16)
13	1.06	1.21	0.83	0.15 (0.00 to 0.29)	-0.23 (-0.47 to 0.01)	-0.38 (-0.62 to -0.14)
14	0.57	0.67	0.45	0.09 (-0.04 to 0.23)	-0.13 (-0.35 to 0.10)	-0.22 (-0.44 to 0.00)
15	0.25	0.31	0.20	0.06 (-0.13 to 0.24)	-0.06 (-0.35 to 0.24)	-0.11 (-0.42 to 0.19)
16	0.09	0.13	0.07	0.04 (-0.18 to 0.26)	-0.02 (-0.37 to 0.33)	-0.06 (-0.42 to 0.29)
17	0.07	0.11	0.05	0.04 (-0.18 to 0.26)	-0.02 (-0.37 to 0.34)	-0.05 (-0.42 to 0.31)
18	0.07	0.11	0.05	0.04 (-0.18 to 0.26)	-0.02 (-0.37 to 0.34)	-0.05 (-0.42 to 0.31)

COPAD condyle-point-A measurement in millimetres

Lower jaw length (COPOD)

Table C-9. Class differences in the lower length (COPOD) growth velocity between seven and 18 years of age for males.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	0.79	2.38	1.83	1.58 (0.64 to 2.53)	1.03 (-0.62 to 2.69)	-0.55 (-2.04 to 0.94)
8	1.17	2.21	1.95	1.04 (0.39 to 1.69)	0.78 (-0.31 to 1.87)	-0.26 (-1.23 to 0.71)
9	2.18	1.76	2.32	-0.42 (-1.03 to 0.19)	0.14 (-1.29 to 1.57)	0.56 (-0.79 to 1.91)
10	2.38	1.62	2.86	-0.76 (-1.26 to -0.26)	0.48 (-0.55 to 1.51)	1.24 (0.29 to 2.19)
11	1.93	1.84	3.12	-0.10 (-0.83 to 0.64)	1.19 (-0.41 to 2.79)	1.29 (-0.20 to 2.77)
12	1.56	2.24	2.87	0.67 (-0.22 to 1.56)	1.30 (-0.56 to 3.16)	0.63 (-1.08 to 2.34)
13	3.01	2.73	4.69	-0.28 (-0.92 to 0.36)	1.67 (0.37 to 2.98)	1.96 (0.76 to 3.16)
14	4.02	2.60	5.75	-1.42 (-2.40 to -0.45)	1.73 (-0.33 to 3.79)	3.15 (1.25 to 5.05)
15	2.40	1.41	3.27	-0.99 (-1.62 to -0.35)	0.87 (-0.42 to 2.16)	1.86 (0.67 to 3.04)
16	1.13	0.56	1.37	-0.58 (-1.60 to 0.45)	0.24 (-1.73 to 2.20)	0.81 (-0.98 to 2.61)
17	0.63	0.22	0.62	-0.41 (-1.66 to 0.84)	-0.01 (-2.42 to 2.40)	0.40 (-1.79 to 2.60)
18	0.63	0.21	0.61	-0.41 (-1.67 to 0.84)	-0.01 (-2.43 to 2.40)	0.40 (-1.81 to 2.60)

COPOD condyle-pogonion measurement in millimetres

Table C-10. Class differences in the lower length (COPOD) growth velocity between seven and 18 years of age for females.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	1.39	1.16	1.73	-0.23 (-0.77 to 0.30)	0.34 (-0.55 to 1.23)	0.57 (-0.31 to 1.45)
8	1.41	1.20	1.72	-0.21 (-0.70 to 0.29)	0.31 (-0.51 to 1.12)	0.51 (-0.30 to 1.32)
9	1.51	1.46	1.62	-0.05 (-0.36 to 0.25)	0.11 (-0.40 to 0.62)	0.16 (-0.34 to 0.67)
10	1.75	1.90	1.62	0.15 (-0.34 to 0.63)	-0.13 (-0.93 to 0.67)	-0.28 (-1.09 to 0.53)
11	2.30	2.30	2.40	0.00 (-0.45 to 0.45)	0.10 (-0.64 to 0.84)	0.09 (-0.65 to 0.84)
12	2.94	2.54	3.61	-0.40 (-0.74 to -0.06)	0.67 (0.10 to 1.24)	1.07 (0.50 to 1.65)
13	2.72	2.19	3.55	-0.52 (-0.92 to -0.12)	0.83 (0.17 to 1.50)	1.36 (0.69 to 2.03)
14	1.62	1.27	2.15	-0.34 (-0.58 to -0.11)	0.53 (0.16 to 0.91)	0.88 (0.50 to 1.26)
15	0.68	0.52	0.93	-0.17 (-0.41 to 0.08)	0.24 (-0.15 to 0.64)	0.41 (0.01 to 0.81)
16	0.21	0.14	0.31	-0.08 (-0.40 to 0.25)	0.10 (-0.43 to 0.62)	0.17 (-0.36 to 0.71)
17	0.16	0.09	0.24	-0.07 (-0.40 to 0.27)	0.08 (-0.47 to 0.63)	0.14 (-0.41 to 0.70)
18	0.16	0.09	0.24	-0.07 (-0.40 to 0.27)	0.08 (-0.47 to 0.63)	0.14 (-0.41 to 0.70)

COPOD condyle-pogonion measurement in millimetres

Total face height (TFHNP

Table C-11. Class differences in the total face height (TFHNP) growth velocity between seven and 18 years of age for males.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	0.70	2.22	1.64	1.52 (0.60 to 2.43)	0.94 (-0.67 to 2.55)	-0.58 (-2.03 to 0.87)
8	1.13	2.06	1.67	0.93 (0.30 to 1.57)	0.54 (-0.53 to 1.62)	-0.39 (-1.35 to 0.57)
9	2.28	1.65	1.82	-0.63 (-1.27 to 0.01)	-0.46 (-1.97 to 1.05)	0.17 (-1.25 to 1.59)
10	2.60	1.59	2.72	-1.01 (-1.54 to -0.47)	0.12 (-0.98 to 1.22)	1.13 (0.12 to 2.14)
11	2.16	1.83	2.88	-0.33 (-1.07 to 0.42)	0.72 (-0.95 to 2.39)	1.05 (-0.52 to 2.61)
12	1.64	2.14	1.12	0.50 (-0.44 to 1.44)	-0.52 (-2.63 to 1.59)	-1.02 (-2.98 to 0.94)
13	2.91	2.67	3.54	-0.24 (-0.94 to 0.47)	0.64 (-0.87 to 2.15)	0.88 (-0.53 to 2.28)
14	3.86	2.61	6.01	-1.25 (-2.30 to -0.20)	2.16 (-0.23 to 4.54)	3.41 (1.18 to 5.64)
15	2.38	1.38	3.63	-1.00 (-1.64 to -0.35)	1.25 (-0.18 to 2.69)	2.25 (0.91 to 3.59)
16	1.21	0.49	1.68	-0.73 (-1.73 to 0.27)	0.47 (-1.53 to 2.46)	1.20 (-0.64 to 3.03)
17	0.76	0.14	0.92	-0.62 (-1.85 to 0.60)	0.16 (-2.28 to 2.60)	0.79 (-1.46 to 3.03)
18	0.75	0.13	0.91	-0.62 (-1.85 to 0.61)	0.16 (-2.29 to 2.61)	0.78 (-1.47 to 3.03)

TFHNP total face height measurement in millimetres

Table C-12. Class differences in the total face height (TFHNP) growth velocity between seven and 18 years of age for females.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	1.33	0.94	1.66	-0.39 (-0.91 to 0.12)	0.33 (-0.52 to 1.19)	0.73 (-0.12 to 1.58)
8	1.32	0.97	1.62	-0.35 (-0.83 to 0.12)	0.30 (-0.49 to 1.08)	0.65 (-0.13 to 1.43)
9	1.25	1.15	1.33	-0.10 (-0.38 to 0.17)	0.08 (-0.38 to 0.53)	0.18 (-0.28 to 0.64)
10	1.25	1.48	1.04	0.23 (-0.17 to 0.64)	-0.21 (-0.88 to 0.46)	-0.44 (-1.12 to 0.23)
11	1.77	1.88	1.76	0.11 (-0.28 to 0.50)	-0.01 (-0.65 to 0.64)	-0.12 (-0.77 to 0.53)
12	2.59	2.22	3.15	-0.37 (-0.69 to -0.05)	0.56 (0.01 to 1.10)	0.92 (0.37 to 1.47)
13	2.54	1.98	3.28	-0.56 (-0.92 to -0.20)	0.74 (0.12 to 1.35)	1.30 (0.68 to 1.92)
14	1.59	1.16	2.08	-0.43 (-0.64 to -0.21)	0.49 (0.14 to 0.85)	0.92 (0.56 to 1.28)
15	0.76	0.48	1.00	-0.28 (-0.52 to -0.03)	0.25 (-0.16 to 0.66)	0.53 (0.11 to 0.94)
16	0.34	0.14	0.47	-0.20 (-0.52 to 0.12)	0.13 (-0.41 to 0.66)	0.33 (-0.22 to 0.88)
17	0.29	0.09	0.40	-0.19 (-0.52 to 0.14)	0.11 (-0.44 to 0.67)	0.31 (-0.26 to 0.87)
18	0.29	0.09	0.40	-0.19 (-0.52 to 0.14)	0.11 (-0.44 to 0.67)	0.31 (-0.26 to 0.87)

TFHNP total face height measurement in millimetres

C.2.3. **Growth acceleration**

Here I first display graphs showing for class differences in the upper jaw length, the lower jaw length, and the total face height for males and females, and then the numerical results (as Tables).

Upper jaw length (COPAD)

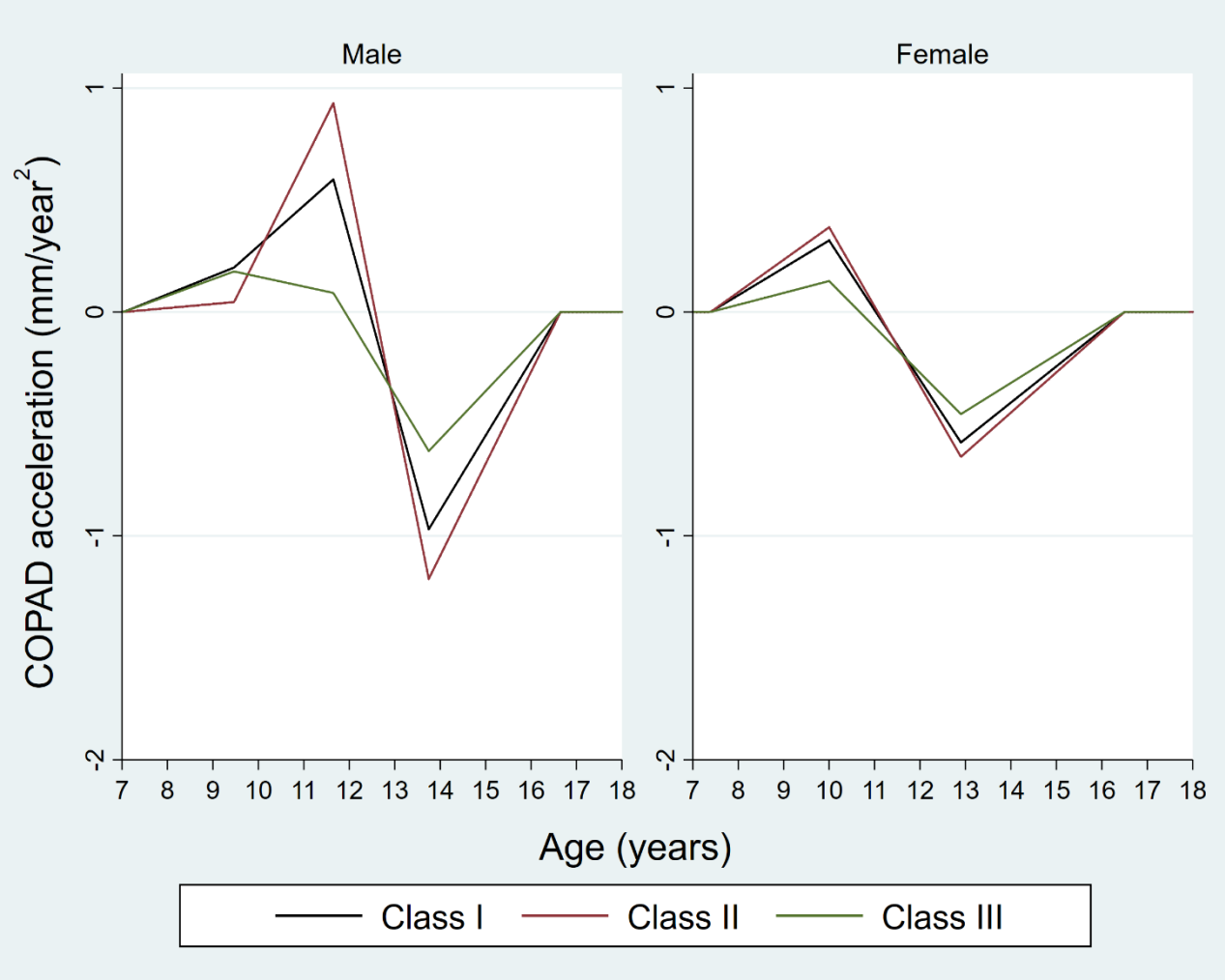


Figure C-1. Class-specific growth acceleration curves for the upper jaw length (COPAD) for males and females.

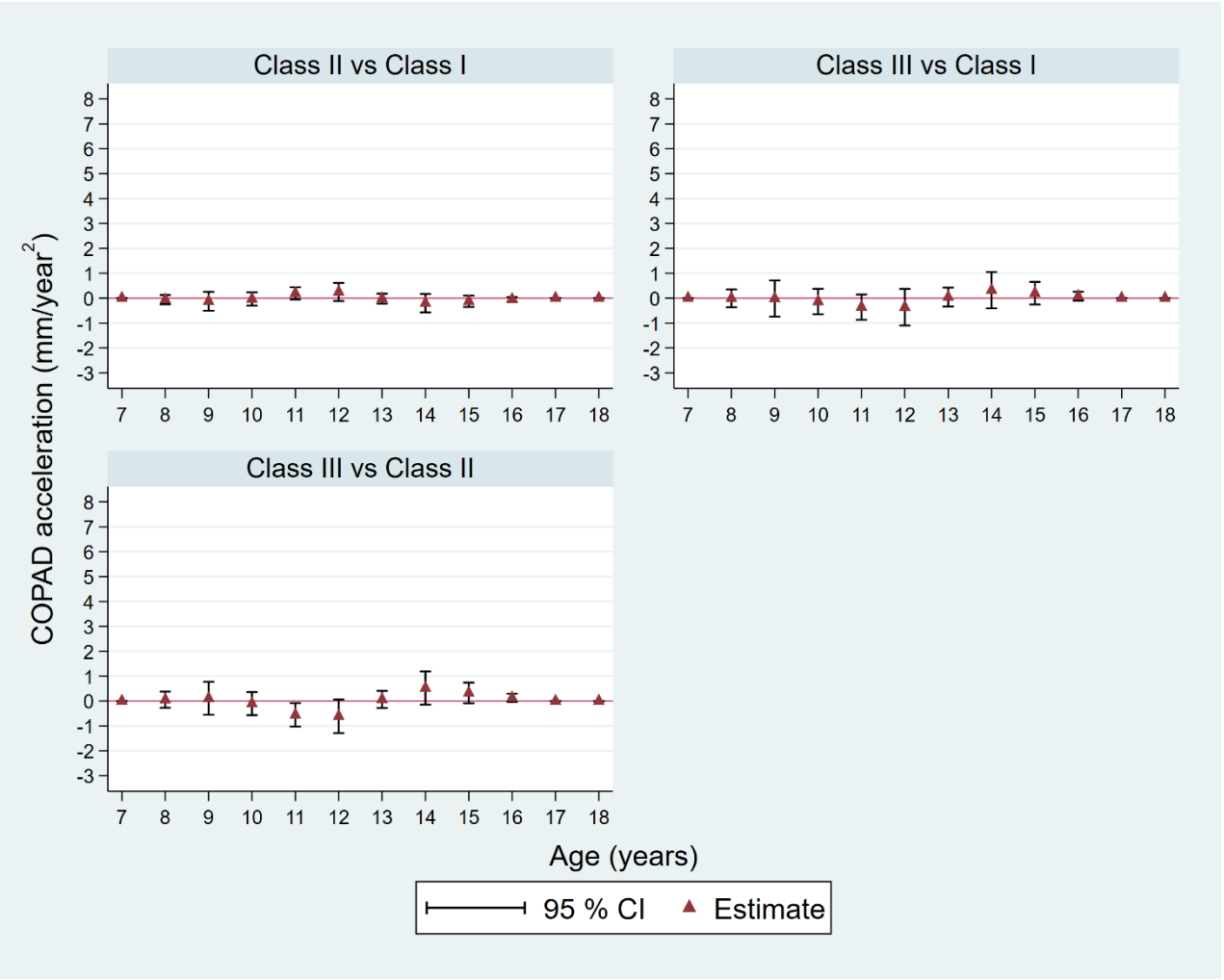


Figure C-2. Class differences in the upper jaw length (COPAD) growth acceleration between seven and 18 years of age for males.

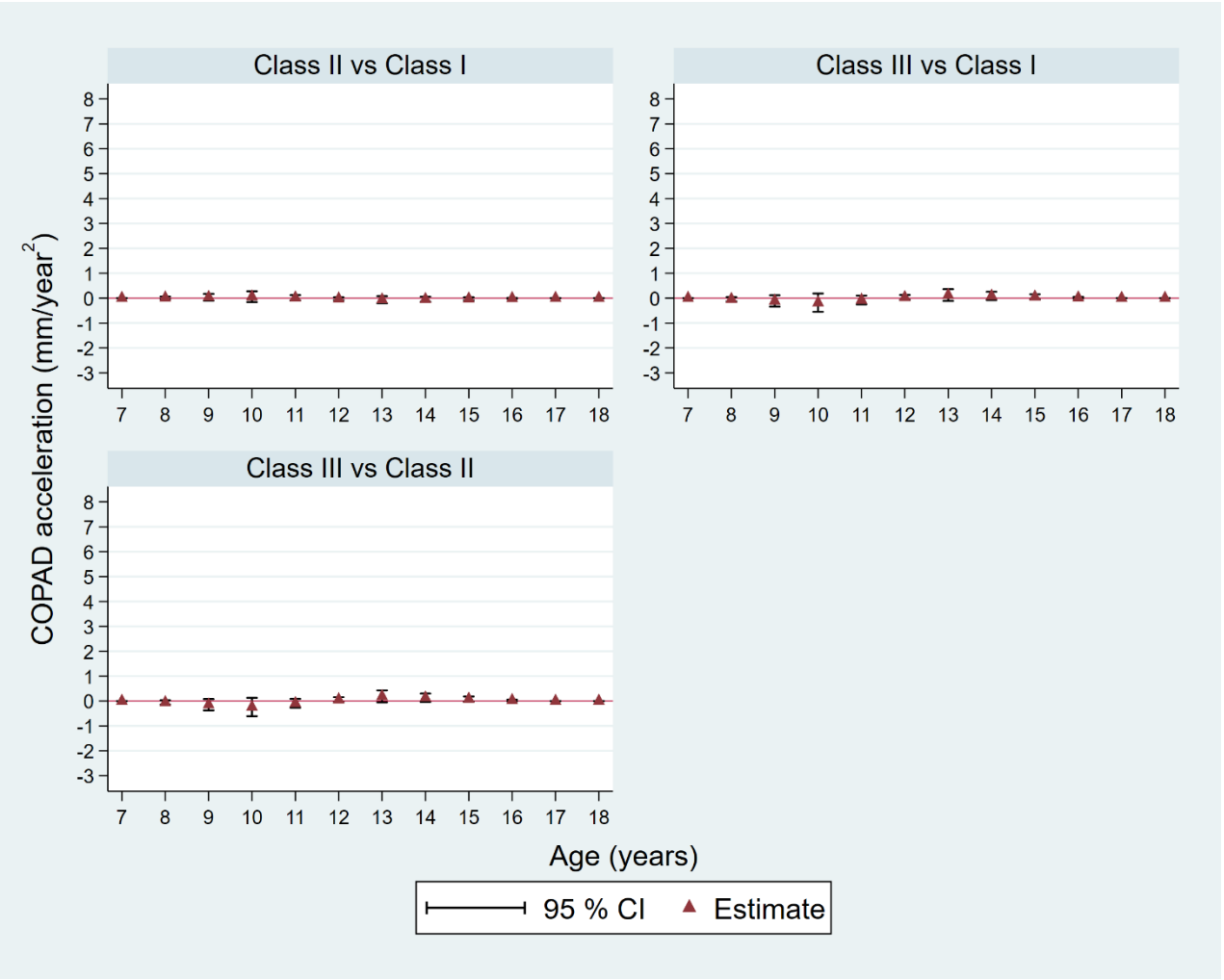


Figure C-3. Class differences in the upper jaw length (COPAD) growth acceleration between seven and 18 years of age for females.

Table C-13. Class differences in the upper jaw length (COPAD) growth acceleration between seven and 18 years of age for males.

Estimate	Estimated difference (95% confidence intervals)

Age (years)	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)
8	0.08	0.02	0.07	-0.06 (-0.25 to 0.12)	-0.01 (-0.36 to 0.35)	0.05 (-0.27 to 0.38)
9	0.16	0.04	0.15	-0.12 (-0.50 to 0.25)	-0.01 (-0.74 to 0.71)	0.11 (-0.55 to 0.77)
10	0.30	0.26	0.16	-0.03 (-0.30 to 0.24)	-0.14 (-0.65 to 0.37)	-0.10 (-0.57 to 0.36)
11	0.48	0.68	0.12	0.19 (-0.05 to 0.44)	-0.36 (-0.87 to 0.15)	-0.56 (-1.02 to -0.09)
12	0.34	0.59	-0.03	0.25 (-0.12 to 0.61)	-0.36 (-1.10 to 0.37)	-0.61 (-1.29 to 0.07)
13	-0.41	-0.43	-0.37	-0.02 (-0.22 to 0.18)	0.04 (-0.34 to 0.42)	0.06 (-0.28 to 0.41)
14	-0.89	-1.10	-0.57	-0.20 (-0.57 to 0.17)	0.32 (-0.41 to 1.05)	0.52 (-0.15 to 1.19)
15	-0.56	-0.68	-0.36	-0.13 (-0.36 to 0.10)	0.20 (-0.26 to 0.65)	0.32 (-0.09 to 0.74)
16	-0.22	-0.27	-0.14	-0.05 (-0.14 to 0.04)	0.08 (-0.10 to 0.26)	0.13 (-0.04 to 0.29)
17	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)
18	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)

COPAD condyle-point-A measurement in millimetres

Table C-14. Class differences in the upper jaw length (COPAD) growth acceleration between seven and 18 years of age for females.

	Estimate	Estimated difference (95% confidence intervals)

Age (years)	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)
8	0.08	0.09	0.03	0.01 (-0.04 to 0.07)	-0.04 (-0.13 to 0.04)	-0.06 (-0.14 to 0.03)
9	0.20	0.24	0.09	0.04 (-0.10 to 0.17)	-0.11 (-0.34 to 0.12)	-0.15 (-0.38 to 0.08)
10	0.32	0.38	0.14	0.06 (-0.16 to 0.28)	-0.18 (-0.55 to 0.19)	-0.24 (-0.61 to 0.13)
11	0.01	0.03	-0.06	0.02 (-0.09 to 0.12)	-0.08 (-0.25 to 0.10)	-0.09 (-0.27 to 0.09)
12	-0.30	-0.33	-0.27	-0.03 (-0.09 to 0.03)	0.03 (-0.07 to 0.13)	0.06 (-0.04 to 0.16)
13	-0.57	-0.63	-0.44	-0.06 (-0.21 to 0.08)	0.12 (-0.11 to 0.36)	0.19 (-0.05 to 0.43)
14	-0.40	-0.45	-0.32	-0.04 (-0.15 to 0.06)	0.09 (-0.08 to 0.26)	0.13 (-0.04 to 0.30)
15	-0.24	-0.27	-0.19	-0.03 (-0.09 to 0.03)	0.05 (-0.05 to 0.15)	0.08 (-0.02 to 0.18)
16	-0.08	-0.09	-0.06	-0.01 (-0.03 to 0.01)	0.02 (-0.02 to 0.05)	0.03 (-0.01 to 0.06)
17	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)
18	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)

COPAD condyle-point-A measurement in millimetres

Lower jaw length (COPOD)

Figure C-4. Class-specific growth acceleration curves for the lower jaw length (COPOD) for males and females.

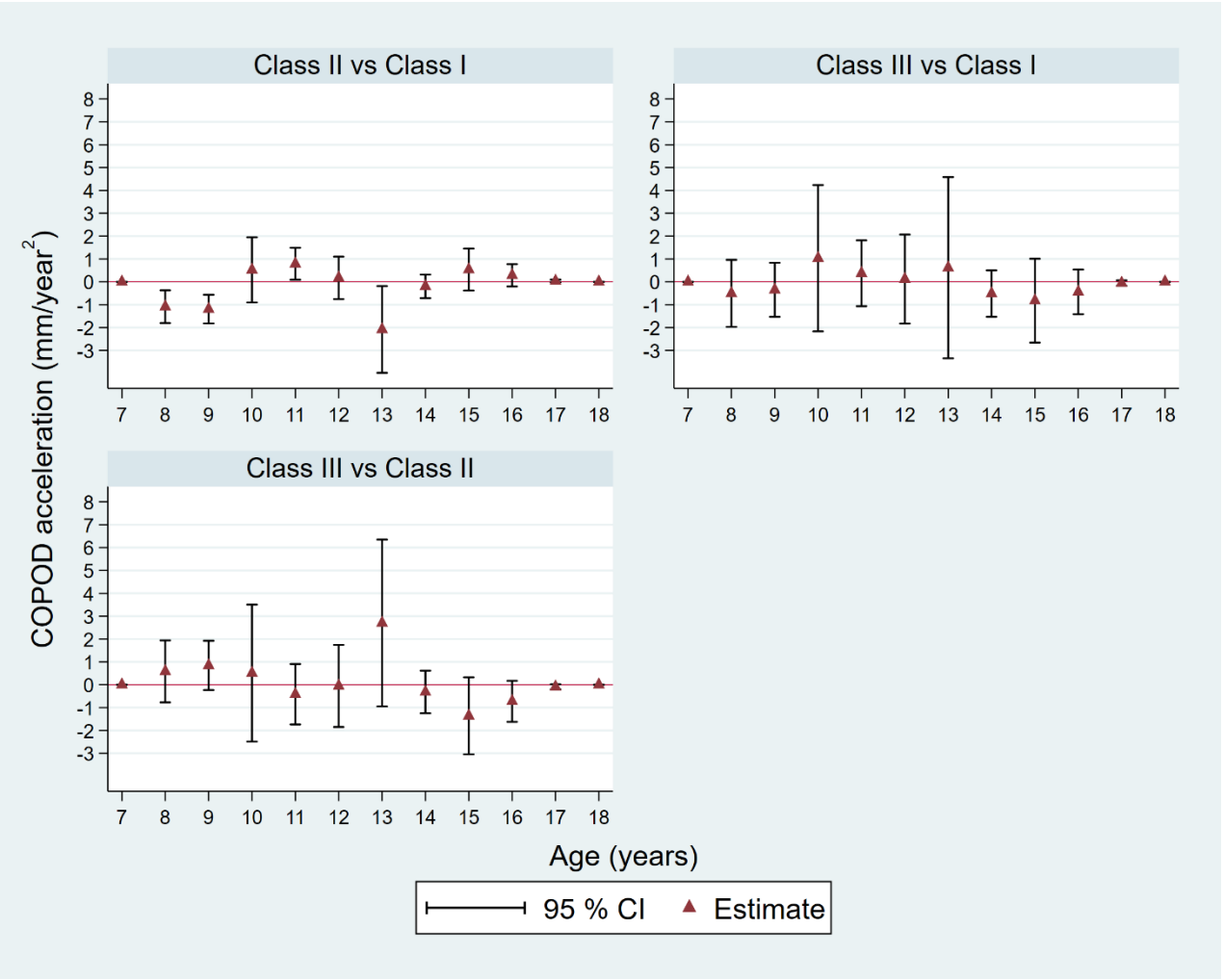


Figure C-5. Class differences in the lower jaw length (COPOD) growth acceleration between seven and 18 years of age for males.

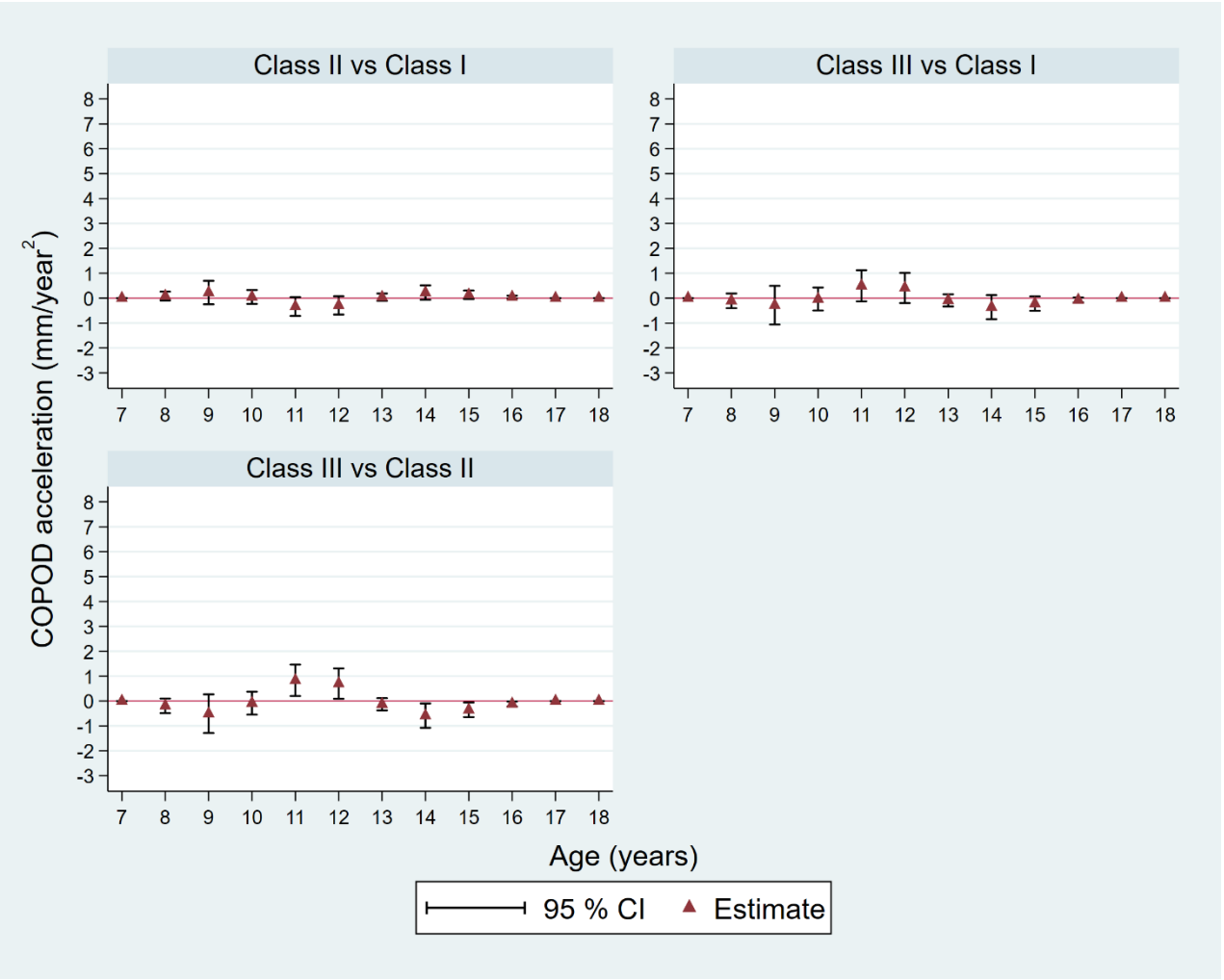


Figure C-6. Class differences in the lower jaw length (COPOD) growth acceleration between seven and 18 years of age for females.

Table C-15. Class differences in the lower length (COPOD) growth acceleration between seven and 18 years of age for males.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)
8	0.75	-0.33	0.25	-1.09 (-1.81 to -0.37)	-0.50 (-1.97 to 0.96)	0.58 (-0.77 to 1.94)
9	0.81	-0.38	0.47	-1.19 (-1.82 to -0.57)	-0.35 (-1.53 to 0.83)	0.85 (-0.23 to 1.92)
10	-0.41	0.11	0.62	0.52 (-0.90 to 1.95)	1.03 (-2.17 to 4.23)	0.51 (-2.49 to 3.51)
11	-0.48	0.31	-0.11	0.79 (0.09 to 1.49)	0.37 (-1.07 to 1.81)	-0.42 (-1.74 to 0.91)
12	0.29	0.46	0.41	0.17 (-0.75 to 1.10)	0.12 (-1.82 to 2.07)	-0.05 (-1.85 to 1.74)
13	2.61	0.53	3.23	-2.08 (-3.98 to -0.18)	0.62 (-3.35 to 4.58)	2.70 (-0.95 to 6.35)
14	-0.60	-0.79	-1.11	-0.20 (-0.71 to 0.32)	-0.51 (-1.53 to 0.50)	-0.32 (-1.25 to 0.61)
15	-1.65	-1.12	-2.48	0.54 (-0.38 to 1.45)	-0.82 (-2.66 to 1.01)	-1.36 (-3.05 to 0.32)
16	-0.88	-0.60	-1.32	0.29 (-0.20 to 0.78)	-0.44 (-1.42 to 0.54)	-0.73 (-1.62 to 0.17)
17	-0.11	-0.07	-0.17	0.04 (-0.03 to 0.10)	-0.05 (-0.18 to 0.07)	-0.09 (-0.20 to 0.02)
18	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)

COPOD condyle-pogonion measurement in millimetres

Table C-16. Class differences in the lower length (COPOD) growth acceleration between seven and 18 years of age for females.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)
8	0.06	0.14	-0.05	0.08 (-0.09 to 0.26)	-0.11 (-0.40 to 0.19)	-0.19 (-0.49 to 0.10)
9	0.15	0.37	-0.13	0.22 (-0.25 to 0.69)	-0.28 (-1.06 to 0.49)	-0.51 (-1.28 to 0.27)
10	0.38	0.43	0.35	0.05 (-0.23 to 0.33)	-0.03 (-0.50 to 0.43)	-0.08 (-0.54 to 0.38)
11	0.71	0.37	1.21	-0.34 (-0.72 to 0.04)	0.50 (-0.13 to 1.12)	0.83 (0.20 to 1.46)
12	0.26	-0.03	0.67	-0.29 (-0.66 to 0.08)	0.41 (-0.20 to 1.01)	0.70 (0.09 to 1.31)
13	-0.71	-0.67	-0.79	0.04 (-0.11 to 0.19)	-0.09 (-0.33 to 0.15)	-0.13 (-0.38 to 0.12)
14	-1.17	-0.95	-1.53	0.22 (-0.07 to 0.51)	-0.36 (-0.85 to 0.12)	-0.59 (-1.07 to -0.10)
15	-0.70	-0.57	-0.92	0.13 (-0.04 to 0.31)	-0.22 (-0.51 to 0.07)	-0.35 (-0.64 to -0.06)
16	-0.23	-0.19	-0.31	0.04 (-0.01 to 0.10)	-0.07 (-0.17 to 0.02)	-0.12 (-0.21 to -0.02)
17	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)
18	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)

COPOD condyle-pogonion measurement in millimetres

Total face height (TFHNP)

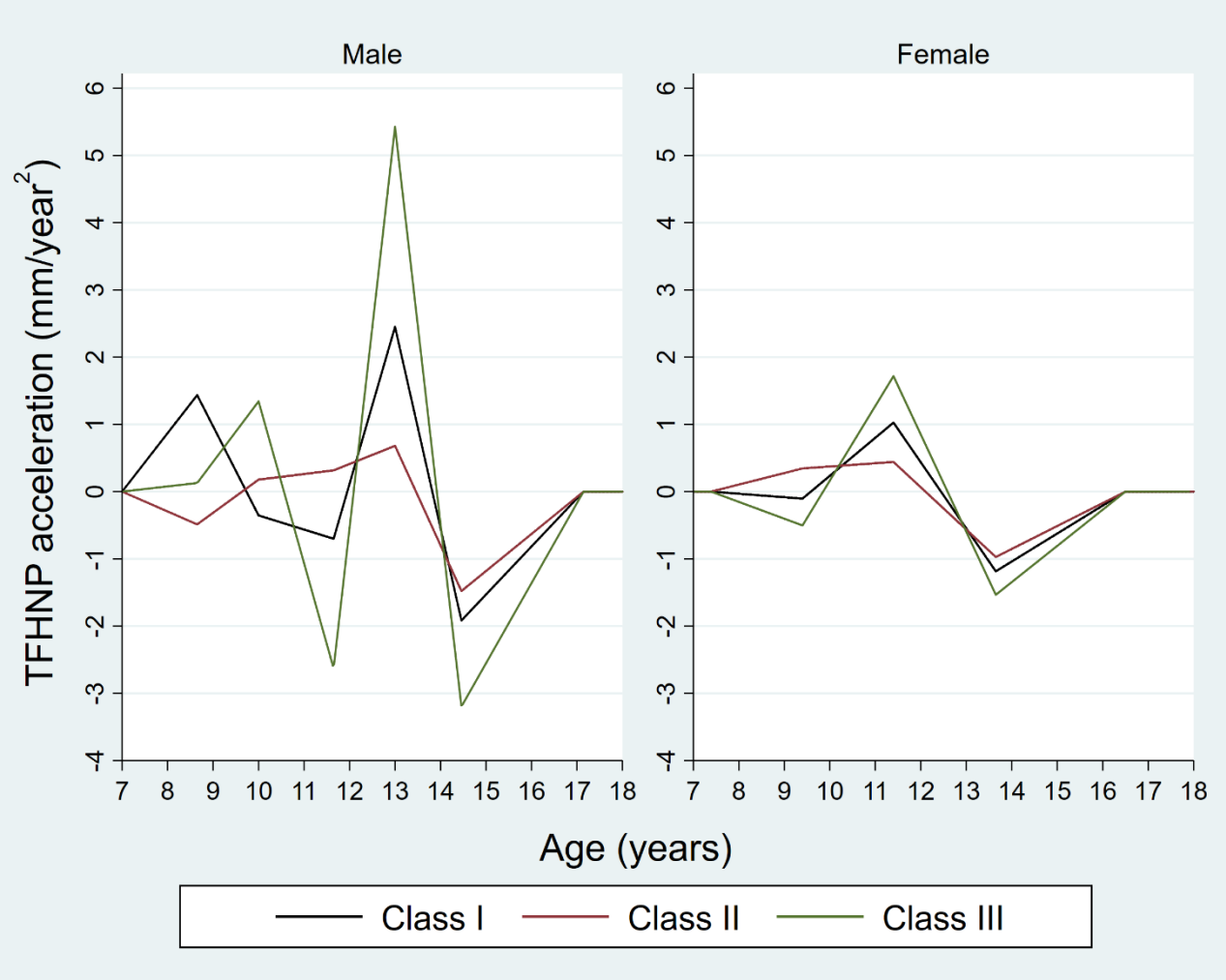


Figure C-7. Class-specific growth acceleration curves for the total face height (TFHNP) for males and females.

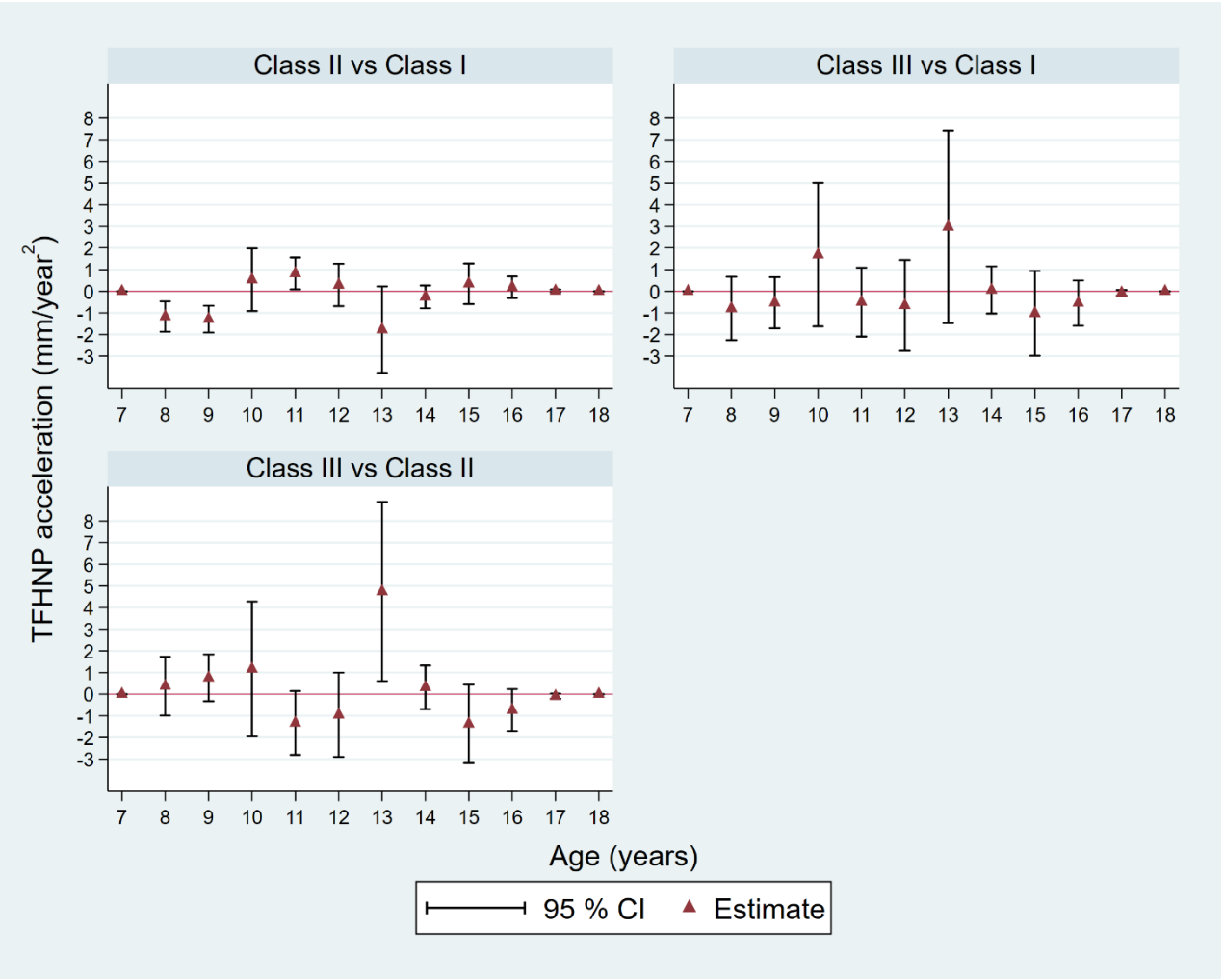


Figure C-8. Class differences in the total face height (TFHNP) growth acceleration between seven and 18 years of age for males.

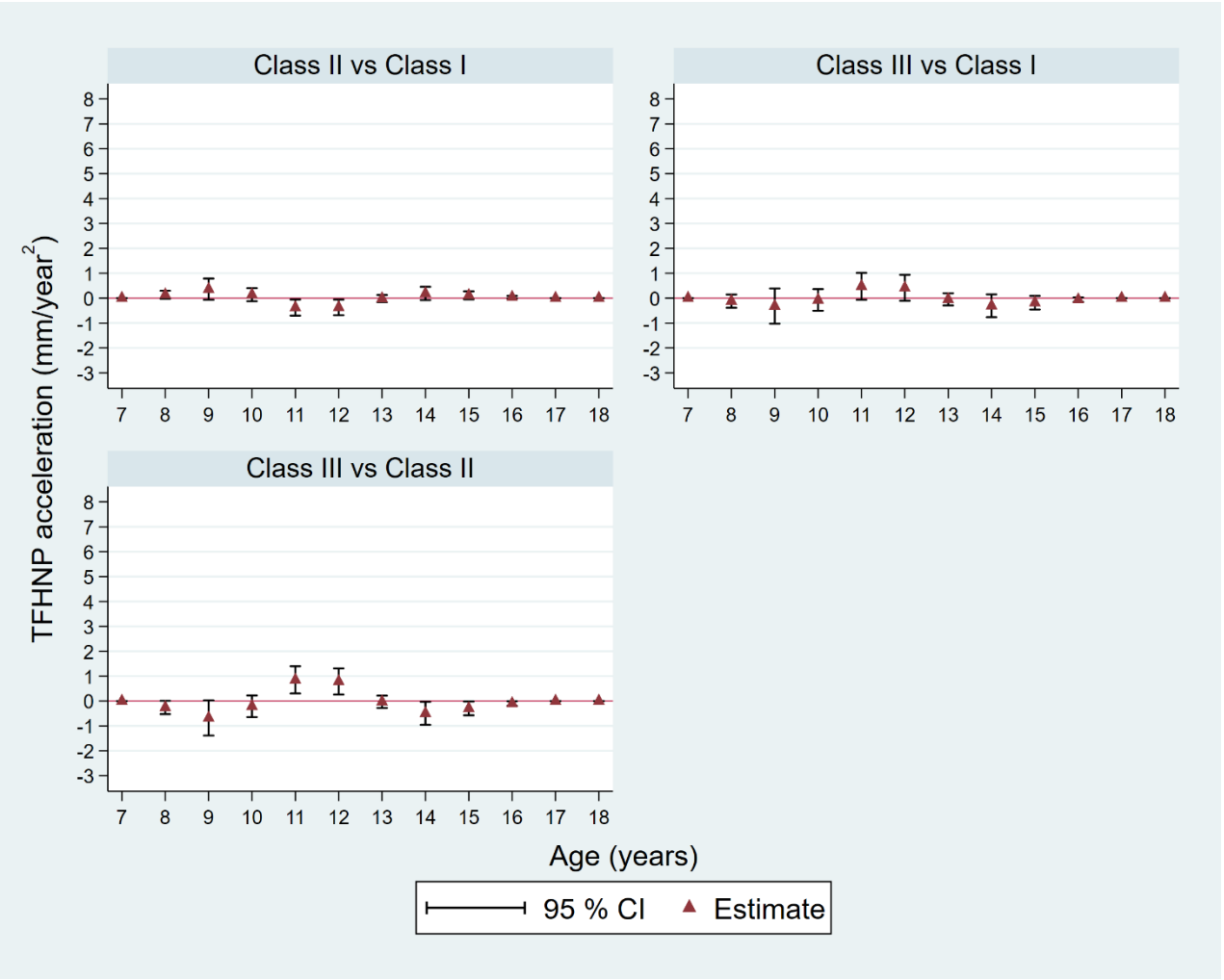


Figure C-9. Class differences in the total face height (TFHNP) growth acceleration between seven and 18 years of age for females.

Table C-17. Class differences in the total face height (TFHNP) growth acceleration between seven and 18 years of age for males.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)
8	0.86	-0.31	0.06	-1.17 (-1.87 to -0.46)	-0.79 (-2.26 to 0.67)	0.37 (-0.99 to 1.73)
9	0.96	-0.33	0.43	-1.29 (-1.91 to -0.67)	-0.53 (-1.71 to 0.65)	0.76 (-0.32 to 1.84)
10	-0.32	0.21	1.37	0.53 (-0.91 to 1.98)	1.70 (-1.62 to 5.01)	1.16 (-1.95 to 4.28)
11	-0.56	0.27	-1.06	0.83 (0.09 to 1.56)	-0.50 (-2.09 to 1.09)	-1.33 (-2.80 to 0.15)
12	0.10	0.40	-0.55	0.29 (-0.68 to 1.27)	-0.66 (-2.75 to 1.44)	-0.95 (-2.90 to 1.00)
13	2.43	0.66	5.41	-1.77 (-3.77 to 0.23)	2.97 (-1.47 to 7.42)	4.74 (0.61 to 8.88)
14	-0.53	-0.79	-0.47	-0.26 (-0.79 to 0.27)	0.06 (-1.03 to 1.15)	0.32 (-0.69 to 1.33)
15	-1.51	-1.17	-2.54	0.35 (-0.59 to 1.29)	-1.02 (-2.98 to 0.94)	-1.37 (-3.18 to 0.44)
16	-0.81	-0.62	-1.35	0.19 (-0.31 to 0.69)	-0.55 (-1.59 to 0.50)	-0.73 (-1.70 to 0.23)
17	-0.10	-0.08	-0.17	0.02 (-0.04 to 0.09)	-0.07 (-0.20 to 0.06)	-0.09 (-0.21 to 0.03)
18	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)

TFHNP total face height measurement in millimetres

Table C-18. Class differences in the total face height (TFHNP) growth acceleration between seven and 18 years of age for females.

Age (years)	Estimate			Estimated difference (95% confidence intervals)		
	Class I	Class II	Class III	Class II Vs Class I (Class II - Class I)	Class III Vs Class I (Class III - Class I)	Class III Vs Class II (Class III - Class II)
7	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)
8	-0.04	0.10	-0.16	0.14 (-0.03 to 0.30)	-0.12 (-0.39 to 0.15)	-0.26 (-0.53 to 0.01)
9	-0.09	0.27	-0.41	0.36 (-0.07 to 0.79)	-0.32 (-1.03 to 0.39)	-0.68 (-1.39 to 0.03)
10	0.23	0.37	0.16	0.14 (-0.12 to 0.40)	-0.07 (-0.51 to 0.36)	-0.21 (-0.64 to 0.22)
11	0.81	0.43	1.29	-0.38 (-0.70 to -0.06)	0.47 (-0.07 to 1.01)	0.85 (0.31 to 1.40)
12	0.45	0.08	0.86	-0.37 (-0.69 to -0.06)	0.41 (-0.11 to 0.94)	0.79 (0.26 to 1.31)
13	-0.54	-0.56	-0.59	-0.02 (-0.16 to 0.13)	-0.05 (-0.29 to 0.19)	-0.03 (-0.28 to 0.22)
14	-1.04	-0.86	-1.35	0.19 (-0.08 to 0.46)	-0.31 (-0.76 to 0.15)	-0.49 (-0.96 to -0.03)
15	-0.63	-0.51	-0.81	0.11 (-0.05 to 0.27)	-0.18 (-0.46 to 0.09)	-0.30 (-0.57 to -0.02)
16	-0.21	-0.17	-0.27	0.04 (-0.02 to 0.09)	-0.06 (-0.15 to 0.03)	-0.10 (-0.19 to -0.01)
17	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)
18	0.00	0.00	0.00	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)	0.00 (0.00 to 0.00)

TFHNP total face height measurement in millimetres

APPENDIX D. DATA (CHAPTER 3)

D.1. American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection project

D.1.1. Historical background

Between 1929 to 1982, various longitudinal growth studies were conducted to collect growth records of craniofacial structures such as jaw sizes and face height (Boyd et al., 1980; J. M. Tanner, 1981). In 1988, the National Institute of Dental Research sponsored a survey to identify and locate existing longitudinal records collections in the United States and Canada (Baumrind & Curry, 2015). In this survey, Hunter et al. (1993) identified 12 relevant longitudinal growth studies providing a valuable collection of records related to growth process which includes radiographic images (such as lateral and cephalograms) of normal, healthy children and adolescents with varied growth patterns such as Class I, Class II and Class III skeletal jaw relationships (Baumrind & Curry, 2015; Hunter et al., 1993).

Each of the 12 growth studies was independent and pursued its own sampling and data collection strategies. Taken together, these different and complementary strategies have produced a rich longitudinal record of craniofacial development among children who did not receive orthodontic treatment. The records have been gathered, catalogued, and studied over a period of decades (Cavanaugh, 2009; Hunter et al., 1993). The contributing collections, working individually, have produced most of the information that is available in the contemporary literature on the craniofacial growth in untreated children (Baumrind & Curry, 2015).

The idea of merging these irreplaceable cephalometric images and associated numeric data had been a dream of clinicians and researchers for many years. However, the huge volume of database available from each growth study coupled with limited technology available at that time made it almost an impossible task to merge these collections. As a result, various proposal were submitted to the American Association of Orthodontists Foundation (AAOF) aimed at preserving these growth collections at a number of different institutes across USA (Cavanaugh, 2009). The AAOF is the charitable arm of the American Association of Orthodontists (AAO) which is primarily responsible for advancement of orthodontic specialty by supporting education and research. The Planning and Awards Review Committee (PARC) of the AAOF plays an important role in making recommendation to the AAOF for allocation of funds (Cavanaugh, 2009).

In January 2005, while considering various individual proposals, the PARC made an important recommendation to the AAOF for adopting a more global approach in preserving longitudinal growth collection data (Cavanaugh, 2009). The objective was to protect, digitize, and disseminate the orthodontic collections so that they would be preserved and available to clinicians and researchers (Cavanaugh, 2009). The AAOF accepted this recommendation and in its next two meetings (held in May 2005 and January 2006), criteria for the records to be preserved and digitized were established, and issues related to how the information would be made available on the Internet were discussed (Cavanaugh, 2009). This initiated the process of developing a collaborate proposal to preserve and merge longitudinal craniofacial growth data.

In November 2007, Mark Hans and his associates at Case Western Reserve University, USA invited representatives of each historic growth collection to a meeting (Cavanaugh, 2009). In 2008, a meeting was held where representatives from interested institutions discussed the potential for developing a shared virtual resource of longitudinal craniofacial growth records (Baumrind & Curry, 2015). The meeting led to the formation of a consortium among the separate collections which

submitted a single collaborated proposal to AAOF for testing the feasibility of constructing a sharable image base and database (Baumrind & Curry, 2015; Cavanaugh, 2009). The proposal was immediately accepted and the AAOF awarded this group $120,000 to begin work (Cavanaugh, 2009).

Nine of the 12 collections identified in the study of Hunter et al. (1993) collaborated in this consortium to share their individual databases (see Section D.1.2 for details). Three collections which were identified by Hunter et al. (1993) but did not participate in the consortium were (see Section D.1.3 for details): 1) Krogman Philadelphia Growth Study (University of Pennsylvania, Pennsylvania USA) 2) Meharry Growth Study (Meharry University, Nashville USA) and Montreal Growth Study (University of Montreal, Montreal Canada).

The Craniofacial Research Instrumentation Laboratory (CRIL) at the University of the Pacific was the designated site for the development of a prototype sharable database and web site to which each participating collection would contribute materials for a representative subset of cases (Baumrind & Curry, 2015). The initial test period which lasted for one and a half year (from June 2009 to December 2010), successfully demonstrated the feasibility and validity of the collaboration. Based on the findings of initial report, the AAOF approved and authorized a full-scale project aimed at the collection of longitudinal records with an emphasis on lateral cephalograms (Baumrind & Curry, 2015). The main data collection period lasted for three years between 2011 and 2014 (Baumrind & Curry, 2015).

D.1.2. **Longitudinal growth studies participating in the project**

Table D-1 belwo summarize nine longitudnal growth studies which participated in the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection project.

Table D-1. Data characterstics and the enrollment periods for studies participating in the the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection project.

Growth study	Participants	Enrolment period
Bolton-Brush	4,309 individuals (birth to adolescence; nearly equal number of male and female). American-born Anglo-Saxon (few black).	1929-1959
Burlington	1258 individuals (3-12-year age; nearly equal number male and female). Caucasian and Anglo-Saxon children.	1952-1971
Denver	313 individuals (few months after birth to 18 years age; 158 females and 155 males). European Caucasians.	1932-1967
Fels	1200 individuals (first month after birth to 18 years age; 158 females and 155 males). European Caucasians.	1931-1982
Forsyth	414 Twin pairs (4-13 years age; 226 male and 178 female). European Caucasians.	1959-1970
Iowa	183 individuals (3 to 5 years age; 91 females and 92 males). European Caucasians.	1946-1960
Mathews	36 individuals (7 to 18 years;13 male 23 female). European Caucasians.	1967-1979
Michigan	721 individuals (3 to 18 years;13 male 23 female). European Caucasians.	1953-1968
Oregon	357 individuals (3 to 18 years;153 male 204 female). European Caucasians.	1950s-mid 1970s

Bolton-Brush growth study

The Bolton-Brush growth study comprises the world's most extensive source of longitudinal human growth data (Behrents & Broadbent, 1984; Boyd et al., 1980; J. M. Tanner, 1981). The foundation of this historic growth study, first of its kind, was laid down in 1926 by Dr T Wingate Todd (Department of Anatomy) by setting up a 'Health Inquiry' or 'Brush Inquiry' to examine physical and mental growth of normal healthy children (6-14 years age) of Ohio, Cleveland. Financial support was provided by the Brush Foundation (Behrents & Broadbent, 1984).

In 1929, Dr B. Holly Broadbent, Sr., an orthodontist and a research fellow in the department of anatomy, developed radiographic cephalometric technique and started an independent but coordinated study in conjunction with the Brush Inquiry to study the growth and development of the face of growing children (Behrents & Broadbent, 1984). Since the financial aid for this project was provided by the Bolton fund, the project was known as 'Bolton Study' (Behrents & Broadbent, 1984).

The recruitment of well-born and healthy children continued for next 30 years (till 1959). Most participants were American-born children of Anglo-Saxon (few black children). However, selection was not based on any race or gender criteria. The only selection criteria was well born and healthy children (Behrents & Broadbent, 1984). Various public and private schools in and around Cleveland participated in this historic and famous longitudinal growth study. There was no systematic bias involved with participants' recruitment (Behrents & Broadbent, 1984). Indeed the selection factors may be fortuitous in terms of applying the findings to present day children (Behrents & Broadbent, 1984).

A total of 4,309 participants (nearly equal number of male and female) were enrolled. Children were examined periodically, every three months in babyhood (birth to first year of age); every six months until 5 years and then every year through adolescence (Behrents & Broadbent,

1984). Lateral and frontal cephalograms as well as hand wrist radiographs and dental study casts were taken. As all were enrolled in the Brush study, therefore, data were also collected for various physical, mental and general health parameters (Behrents & Broadbent, 1984). However, nearly half of the total participants (47 %) were examined only once, therefore did not provide longitudinal growth data (Hunter et al., 1993). As per the survey report submitted by Hunter et al. (1993), longitudinal growth data with at least 10 longitudinal measurement is available from 850 participants.

In the early 1980s and 2000s, two recall efforts took place to assess the present health of the research participants that were studied during their childhood (Behrents & Broadbent, 1984). Over 100 of the original participants were located and invited for participation in this extended study[11].

Burlington growth study

In 1952, Dr Robert E. Moyers, an orthodontist, initiated the Burlington growth study at the Burlington Growth Centre, faculty of dentistry, University of Toronto, Canada (Thompson & Popovich, 1977). Since Dr Moyers moved to United States in 1953 (University of Michigan), much of the data collection was undertaken by Dr Frank Popovich, Professor, and Past Director of Burlington Growth Centre.

The enrolment period lasted for almost 20 years between 1952 and 1971 (Thompson & Popovich, 1977). The initial age groups selected for enrolment in this study were 3, 6, 8, 10, and 12 years (Thompson & Popovich, 1977). The sample of 1258 children enrolled for this study represented approximately 90% of the Burlington children in these age groups at that time. Almost all children were of Northern European ancestry and predominant racial group was Caucasian and

[11] https://dental.case.edu/boltonbrush/

mostly Anglo Saxon[12] (Hunter et al., 1993). In addition to dental records, other general information like height and weight etc. was also recorded.

Children of age groups 8 (N=219), 10 (N=217), and 12 (N=215) years were examined only at one time points, therefore no serial cephalometric data is available. For 3 years age group (N=312), records were taken annually from three to 30 years of age. However, this group was labelled as serial experimental group because treatment, if required, was provided at the Burlington Growth Centre. The 6 years age group constituted the control serial group (N=295) where no treatment was provided at the Burlington Growth Centre. This control group was examined periodically to collect longitudinal records at the ages of 6, 9, 12, 14, 16 and 20 years[13] (Hunter et al., 1993; Thompson & Popovich, 1977). However, few children in control group received orthodontic treatment at private orthodontic centres. Therefore, only 168 (98 females; 68 males) qualified for inclusion into the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection project[14]. Like Bolton-Brush study, Burlington Growth Centre also recalled individuals who participated in the original study. The original three-year-old sample group constituted the 40 years sample in this recall study.

Denver growth study

In 1932, Dr Albert H. Ketcham, an orthodontist, started the Denver growth study at the Child Research Council program, University of Colorado School of Medicine (Waldo, 1936). After demise of Dr Albert H. Ketcham in December 1935, Dr Charles M. Waldo (another orthodontist) got actively involved in taking serial records of children who come from homes of various social and economic positions in the community, representing roughly a cross-section of the population. The children were of European Caucasians origin (Waldo, 1936).

[12] http://www.aaoflegacycollection.org/aaof_collection.html?id=UTBurlington
[13] http://www.aaoflegacycollection.org/aaof_collection.html?id=UTBurlington
[14] http://www.aaoflegacycollection.org/aaof_collection.html?id=UTBurlington

There was no attempt to select special cases for this group. Only selection criteria was that children should be healthy (Waldo, 1936). Children were first seen at the age of two to four weeks, and then followed periodically to collect longitudinal records (Waldo, 1936). When study ended in 1967, the growth data from total 313 untreated children (158 females and 155 males) was collected[15]. In addition to dental records which included lateral and frontal cephalogram, hand wrist films and study cast, data was also collected for height and weight (Waldo, 1936). At present, the collection is housed at the department of orthodontics, college of density, University of Oklahoma, USA[16].

Fels growth study

The Fels Research Institute was founded in 1929 with an objective to collect longitudinal growth data from normal healthy children (Roche, 1992). Arthur E. Morgan, who was president of Antioch College in Yellow Springs, was responsible for setting up this project by getting financial support from his friend Samuel S. Fels (Roche, 1992). Dr Lester W. Sontag was appointed as the first director of Fels Research Institute.

The examinations of participants began in 1930 and longitudinal data was collected for more than 1200 participants. Most of the participants participated in the study since birth (Roche, 1992). All participants provided data for overall general physical and mental health. Out of a total 1200 participants, nearly 400 participants provided lateral cephalogram images (Hunter et al., 1993). This is a randomly ascertained cohort as participants were not selected for any specific feature or trait (Hunter et al., 1993; Roche, 1992).

Study participants generally lived in or near southwest Ohio (Indiana, Kentucky, West Virginia) and were mostly European Caucasians. Data collection was scheduled at 1, 3, 6, 9, and 12 months of age, and then at every 6-month intervals until the age 18 years[17]. Participants continued

[15] http://www.aaoflegacycollection.org/aaof_collection.html?id=UOKDenver
[16] http://www.aaoflegacycollection.org/aaof_collection.html?id=UOKDenver
[17] http://www.aaoflegacycollection.org/aaof_collection.html?id=WSUFels

life-long examinations in the study, and some have over 30 lateral cephalographs spanning up to 45 years in the life of an individual. Cranial radiography (cephalograms) of Fels growth study terminated in 1982[18].

Forsyth growth study

Dr Coenraad F. A. Moorrees, an orthodontist at the Forsyth Dental Centre in Boston and Professor of orthodontics at the Harvard School of Dental Medicine was responsible for setting up this historic longitudinal twin study at Boston, Massachusetts (Harvard Society for the Advancement of Orthodontics, 2007). Study began in 1959 and continued until 1970s. This study was unique in a sense that it enrolled only twins and their families for longitudinal growth data collections. The sample is entirely of European extraction (Hunter et al., 1993).

The participants enrolled in this longitudinal study consisted of 414 (226 male and 178 female) monozygotic and dizygotic Caucasian twin pairs living in the Greater Boston area. They were ascertained at various ages from 4 to 13 years; 228 pairs enrolled between 4 and 6 years, 124 pairs between 7 and 10 years, and the remaining 62 pairs between 11 and 13 years of age. An average of nine annually spaced observations were obtained for each set of twins (Kent et al., 1978).

The longitudinal records consisted of dental records, anthropometric measurements, demographic and dietary information. Dental records consisted of lateral and frontal cephalograms, intra-oral radiographs, hand-wrist films and study casts. For lateral cephalograms, complete longitudinal records between 6 and 16 years are available with thee age groups, from 6 to 10 years, 10 to 16 years, and 6 to 16 years. Records are also available between four and 18 years of age for a few participants (Harvard Society for the Advancement of Orthodontics, 2007; Hunter et al., 1993).

[18] http://www.aaoflegacycollection.org/aaof_collection.html?id=WSUFels

Iowa growth study

In 1946, the Iowa growth study was initiated at the University of Iowa under the direction of two orthodontists, Dr Howard V. Meredith and Dr L. B. Higley. The study was sponsored jointly by the Iowa Child Welfare Research Station and the College of Dentistry[19].

Although study was originally called as 'Facial Growth Study', numerous trunk and limb measurements were obtained in addition to those taken of the head and face. Height, weight, dietary information, and medical history data was also obtained. While most of the data (general physical and dental study casts) was collected semi-annually, lateral and frontal cephalograms were obtained at three-month intervals until age five and twice yearly after the children reached the fifth birthday. After age 12, records were made annually[20].

The Iowa Facial Growth Study is a true longitudinal study. Unlike other mixed longitudinal growth studies where investigators decide to add individuals while the study was ongoing or enrolled participants as cohort (e.g. grouping of children), the Iowa Facial Growth Study is unique because the same individuals were followed throughout its duration[21].

The study collected data for 183 Caucasians (92 males and 91 females) participants. As few participants were enrolled as early as three years of age, and all were included in the study by the age of five years. Ninety-seven percent of the participants were of Northwest European ancestry. Records ceased to be collected in 1960. The collection is presently housed in Iowa Facial Growth Study Laboratory within the Department of Orthodontics, University of Iowa College of Dentistry[22].

[19] http://www.aaoflegacycollection.org/aaof_collection.html?id=UIOWAGrowth
[20] http://www.aaoflegacycollection.org/aaof_collection.html?id=UIOWAGrowth
[21] http://www.aaoflegacycollection.org/aaof_collection.html?id=UIOWAGrowth
[22] http://www.aaoflegacycollection.org/aaof_collection.html?id=UIOWAGrowth

Mathews growth study

In 1967, Dr J. Rodney Mathews, an orthodontist, established the Mathews growth study at the department of orthodontics, School of Dentistry, University of California San Francisco. Longitudinal growth data was collected over next 12 years (till 1979)[23]. Total 36 participants (seven to 18 years age) were enrolled in this study.(Hunter et al., 1993; The Craniofacial Research Instrumentation Laboratory, 2014). Participants were primarily of Northern European origin (Hunter et al., 1993).

Lateral and fontal cephalograms as well as other radiographic image were collected at annual intervals. The full collection consists of approximately 1000 radiographic images and currently housed in the Craniofacial Research Instrumentation Laboratory (CRIL) at the Arthur A. Dugoni School of Dentistry, University of Pacific, USA[24].

Michigan growth study

In 1930, Dean Willard Oslon and Prof. Byron O. Hughes started longitudinal growth study at the School of Education, University of Michigan (Moyers et al., 1976; Riolo, 1974). The Michigan growth study is comprised of annual records taken on students enrolled in the laboratory housed in the School of Education, Ann Arbor campus. The participants were primarily of Northern European ancestry and were followed from age three to 18 years. Although data collection began in 1930s which included medical, dental, psychological, and anthropologic records, the collection of lateral and frontal cephalograms began in 1953 and continued until 1968 (Moyers et al., 1976; Riolo, 1974). Records were taken annually around birthdays (Moyers et al., 1976; Riolo, 1974).

A total 721 participants with almost equal number of males and females were enrolled. However, cephalogram images were collected only for those participants who enter the study after

[23] http://www.aaoflegacycollection.org/aaof_collection.html?id=UOPMathews
[24] http://www.aaoflegacycollection.org/aaof_collection.html?id=UOPMathews

1953. Therefore, the longitudinal cephalometric records are available only for limited sample (about half the total number of participants) constituting about 3,266 repeated measurements[25].

Oregon growth study

In early 1950s, Dr Bhim Sen Savara, professor and chairman of the Child Study Clinic, University of Oregon Health Sciences established Oregon growth study. Longitudinal growth data were collected till mid 1970's (Garn, 1981; Savara & Steen, 1978). A total 357 participants (204 females and 153 males) with age ranging from three to 18 years were enrolled in this study. The sample was entirely of Northern European extraction (Hunter et al., 1993).

Foe approximately one third of the sample, records collection began at three years of age. For remaining two-third, data collected started between four and 12 years of age. More than half the sample has records up to 18 years of age (Hunter et al., 1993). Nearly one third participants (118) had orthodontic treatment (Hunter et al., 1993). The records collected include lateral and frontal cephalogram, dental study casts, hand-wrist radiographs, and general medical records such as height and weight[26] (Hunter et al., 1993).

[25] http://www.aaoflegacycollection.org/aaof_collection.html?id=UMICHGrowth
[26] http://www.aaoflegacycollection.org/aaof_collection.html?id=UOGrowth

D.1.3. Longitudinal growth studies which did not participate in the project

There are five longitudinal growth studies which are not part of the AAOF Craniofacial Growth Legacy Collection project. Three of these five studies were identified by Hunter et al. (1993) as they were conducted across USA and Canada but did not participate in the project. These studies are: i) Krogman Philadelphia growth study (University of Pennsylvania, Pennsylvania USA), ii) Meharry growth study (Meharry University, Nashville USA), and iii) Montreal growth study (University of Montreal, Montreal, Canada). There were no specific reasons why these three growth studies did not participate in the project.[27]

Two other longitudinal growth studies, which are not part of the project were not included in the survey report of Hunter et al. (1993) as these were not conducted in the USA or Canada, are the Belfast growth study (Queen's University Belfast, Belfast Northern Ireland) and the Nijmegen growth study (Radboud University Nijmegen, Nijmegen Netherland).

The Krogman Philadelphia study collected cephalograms (lateral and lateral) as well as hand-wrist radiographs of 750 healthy children. Longitudinal records were collected at four to six time points. Majority of participants were white, and a few were black. The data collection started between the ages of 12 and 15 years and ended at the age of approximately 18 years. The records were gathered by Dr Wilton M. Krogman and Dr Sol Katz (Hunter et al., 1993).

The Meharry Black sample provides longitudinal records (lateral and frontal cephalograms, hand-wrist radiographs) for approximately 160 American black participants who were enrolled at the age of six and then followed every 6 months up to 14 of age. About 100 of these participants were included in extended study period where lateral cephalograms and other dental records were

[27] Personal communication, Dr Sheldon Baumrind (25/10/2016), Director of the American Association of Orthodontists Foundation (AAOF) Craniofacial Growth Legacy Collection project.

taken annually till the age of 20 years. The records were collected under the supervision of Dr Elisha Richardson, the project director (Hunter et al., 1993).

The Montreal Growth Study is a mixed longitudinal sample with two principle cohorts: six to 15 years of age with at least 50 participants for each sex at each age; and 10 to 19 years of age with at least 30 males and 20 females at each age. Because of the overlap of cohorts, there are more than 100 participants of each sex at each annual time point in the age interval between 10 and 14 years. The study sample consisted entirely of French-Canadian participants who were of European ancestry. In addition to the dental records (lateral and frontal cephalograms, hand-wrist radiographs), height and weight records were also obtained. The record were collected under the direction of Dr Arto Demirjian who served as the director (Hunter et al., 1993).

The Belfast Growth Study (1962-1972) was conducted by Dr C. P. Adams with 300 children. Each child was examined 21 times beginning at the age of five years (first visit) up to the age of 15 years (21st visit). Lateral and frontal cephalograms were taken annually and dental study casts made every six months (Adams, 1972; Lux et al., 2003). Therefore, there were 10 longitudinal cephalometric records available from each participant.

The Nijmegen growth study (Radboud University Nijmegen, Nijmegen, Netherland) was conducted by Dr B. Prahl-Andersen B and Dr C. J Kowalski. Total 486 normal Dutch children (232 males and 254 females) were followed for five years (1970–1975) to collect longitudinal growth data, covering a total age range of four to 14 years. Cephalometric radiographs and study casts were made twice a year for each participant. Apart from dental records (cephalograms and dental casts etc.), information was also gathered for other general medical/psychological such as height and weight (Prahl-Andersen B & CJ., 1979; Prahl-Andersen & Kowalski, 1973).

D.2. Data available

Data available from the AAOF Craniofacial Growth Legacy Collection website[28] includes demographics, lateral cephalogram images and data derived from lateral cephalogram images in the form of cephalometric measurements and Cartesian coordinates. The website allows orthodontists to directly download all the data available from the collection. At the time of latest update of the website (July 2016), the collection comprised serial lateral cephalograms of 762 untreated individuals (394 males and 368 females).

Initial scrutiny of data revealed duplicate records for 10 individuals and missing lateral cephalograms for two other individuals (see Figure D-1). This was confirmed by Sean Curry[29], the technical director of the AAOF Craniofacial Growth Legacy Collection website. Thus, data available comprises 9,707 lateral cephalograms of 750 individuals.

Out of the total sample of 750 individuals, measurements of upper and lower length and total face height for 129 individuals are directly available from the AAOF Craniofacial Growth Legacy Collection website. These measurements were calculated from the anatomic landmarks and Cartesian coordinates (horizontal and vertical coordinates, x-y) from the lateral cephalograms. The website provides additional Cartesian coordinates data (without measurements) for the upper and lower jaw length and total face height of 236 individuals. The reason for these two sets of data is that the Cartesian coordinates with and without measurements were computed independently at different universities for research purposes and then pooled together for the AAOF Craniofacial Growth Legacy Collection database.[30]

[28] https://www.aaoflegacycollection.org/aaof_home.html
[29] Personal communication, Sean Curry, 7/9/2016.
[30] Personal communication, Dr Sheldon Baumrind, 25/10/2016.

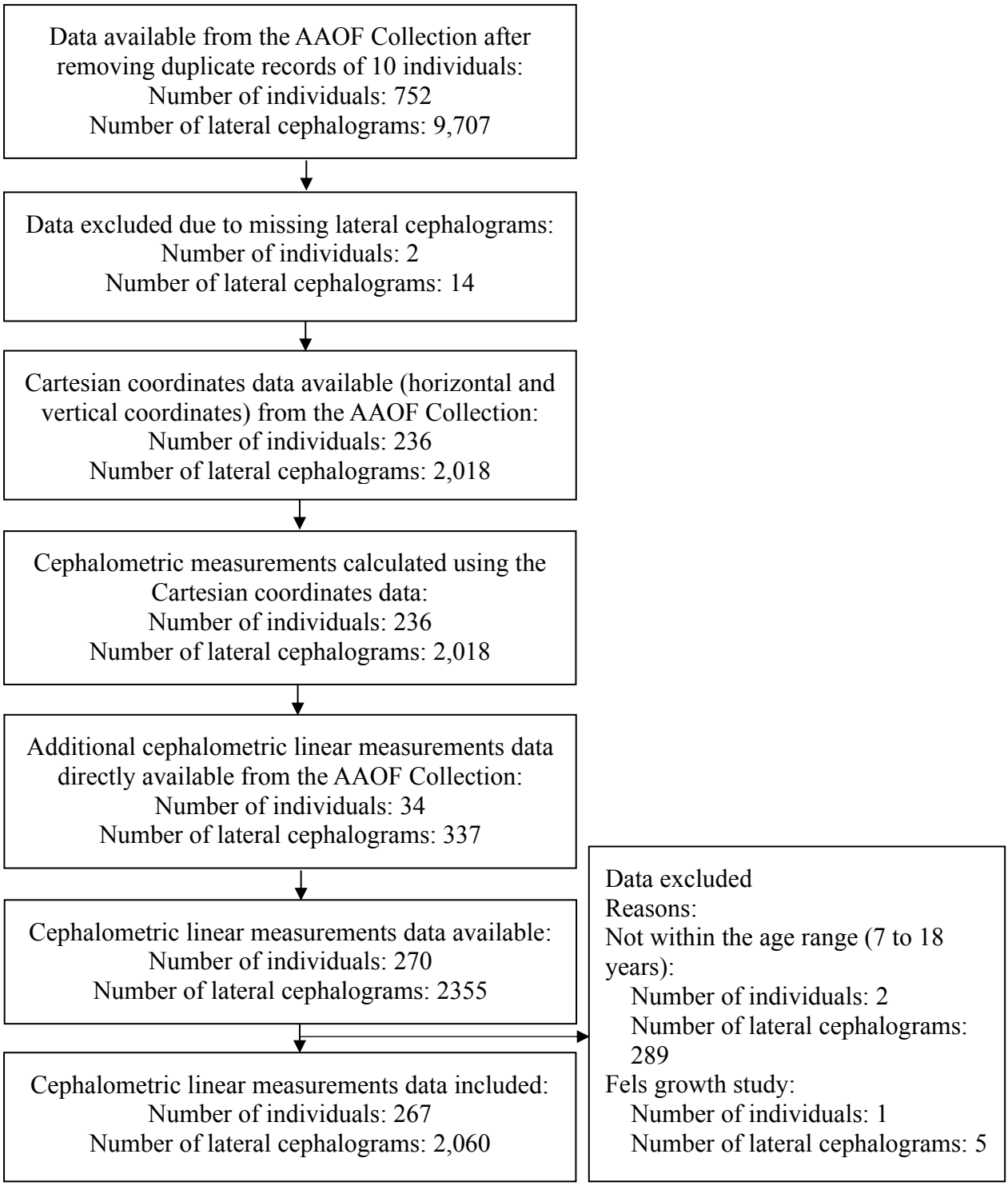


Figure D-1. Flow diagram showing process of data inclusion for analysis.

To best use the data, upper and lower jaw length and total face height were calculated using available Cartesian coordinates. The data was then combined with the measurements of upper and lower jaw length and total face height available directly from the AAOF Craniofacial Growth Legacy Collection (see Figure D-1).

Measurements of the upper jaw length, lower jaw length and total face height for 1,229 lateral cephalograms of 129 individuals (65 males and 64 females) are directly available. Measurements were carried out at the University of the Pacific as part of another project and later added to the collection[31].

Cartesian coordinates are available on 2,018 lateral cephalograms of 236 individuals (110 males and 126 females). Landmark identification and coordinate recording were performed with replication at the CRIL[32]. Based on anatomic landmarks and Cartesian coordinates (Figure D-2), the upper jaw length, lower jaw length and total face height were calculated using the principles of frame of reference and paired Cartesian coordinates (Barry, 2015; Baumrind & Miller, 1980). All calculations were performed using Stata software (StataCorp, 2015b).

The upper jaw length (Figure D-3) is denoted by the condyle–point-A distance (COPAD) which is measured as a linear distance (in millimetres) between anatomic landmarks Condyle and Point-A. The lower jaw length (Figure D-4) is recorded as condyle–pogonion distance (COPOD) measured in millimetres between anatomic landmarks Condyle and Pogonion. The total face height (Figure D-5) is measured as a distance (in millimetres) between anatomic landmarks Nasion and Menton along a perpendicular dropped on the Frankfurt horizontal plane from the Nasion. The measurement is recorded as the total face height nasion perpendicular (TFHNP).

[31] Personal communication, Sean Curry, 7/10/2016.
[32] Personal communication, Dr Sheldon Baumrind, 25/10/2016.

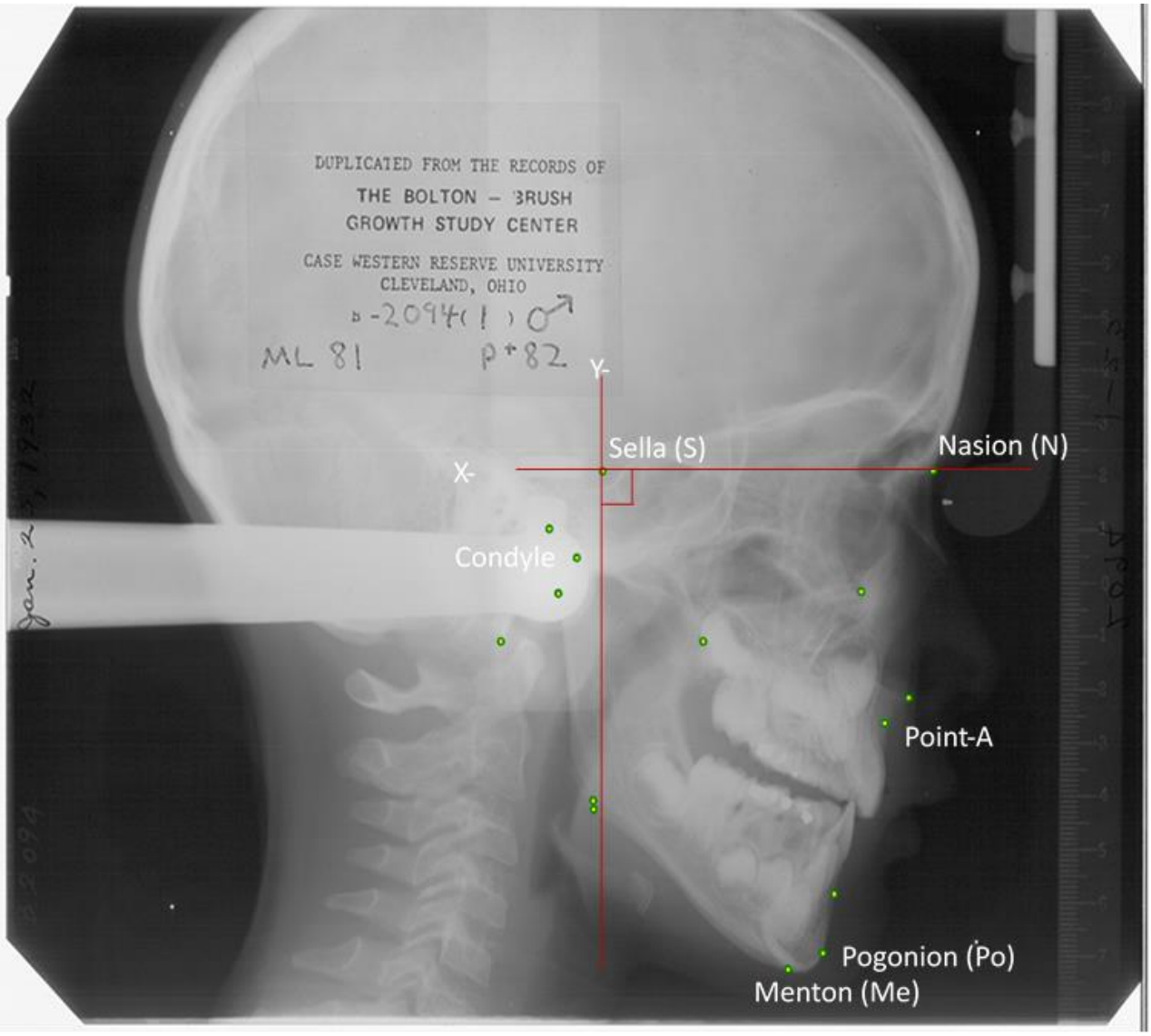


Figure D-2. Cephalometric landmarks and reference planes used to calculate upper jaw length, lower jaw length and total face height. The lateral cephalogram image used for illustration purposes is of a male participant (ID B2094) from the Bolton-Brush growth study. The image is the first lateral cephalogram (indicated by 1 in parentheses), taken at seven years and six months of age.[33]

[33] Image available (to orthodontists) for downloading from the AAOF Craniofacial Growth Legacy Collection. https://www.aaoflegacycollection.org/aaof_multiview.html?collectionID=CASEBolton&subjectID=B2094&imageType=1&imageNumber=1

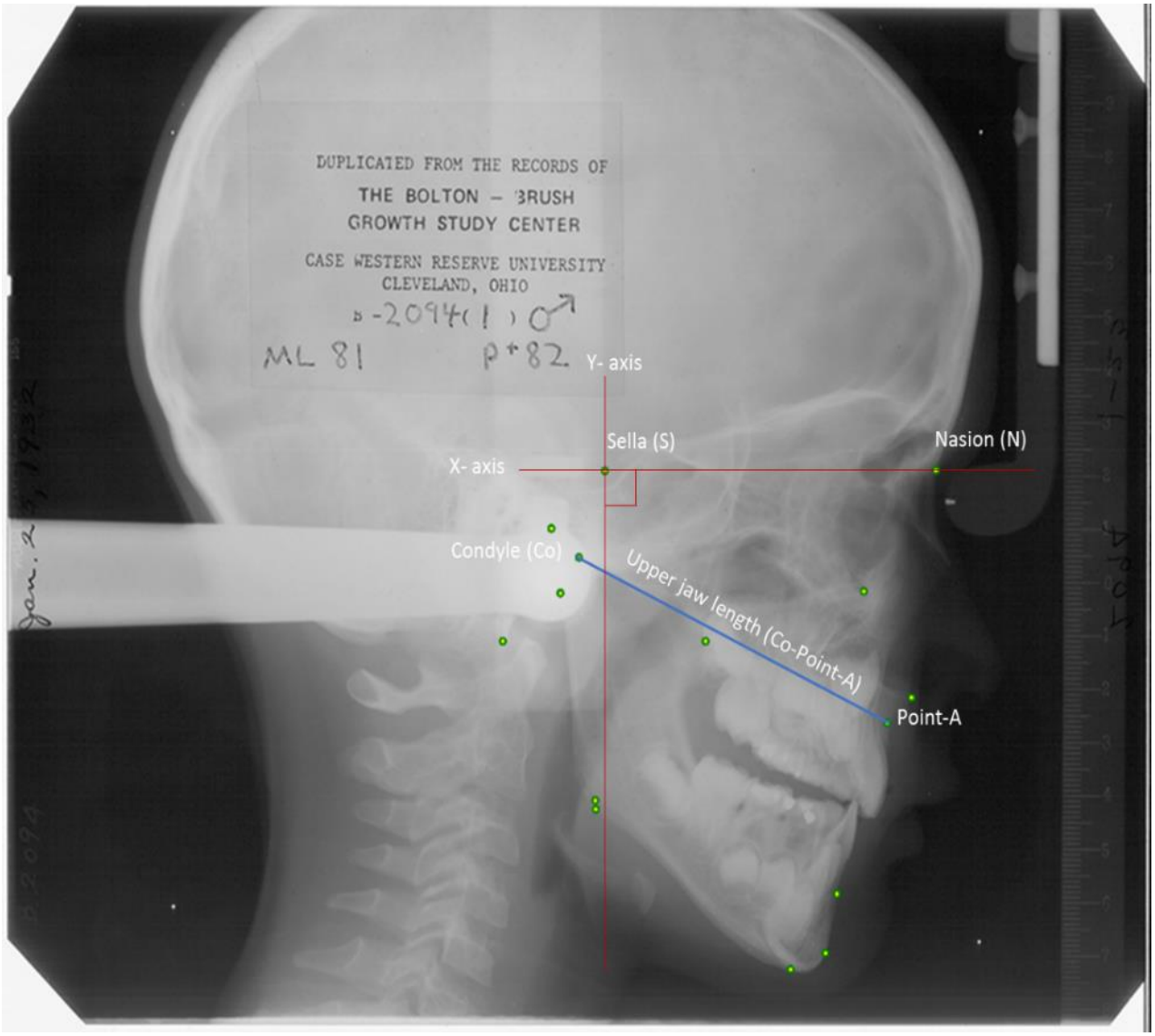


Figure D-3. Upper jaw length measured as linear distance (in millimetres) between condyle and point-A on lateral cephalogram. The upper jaw length (COPAD) shown as blue line.

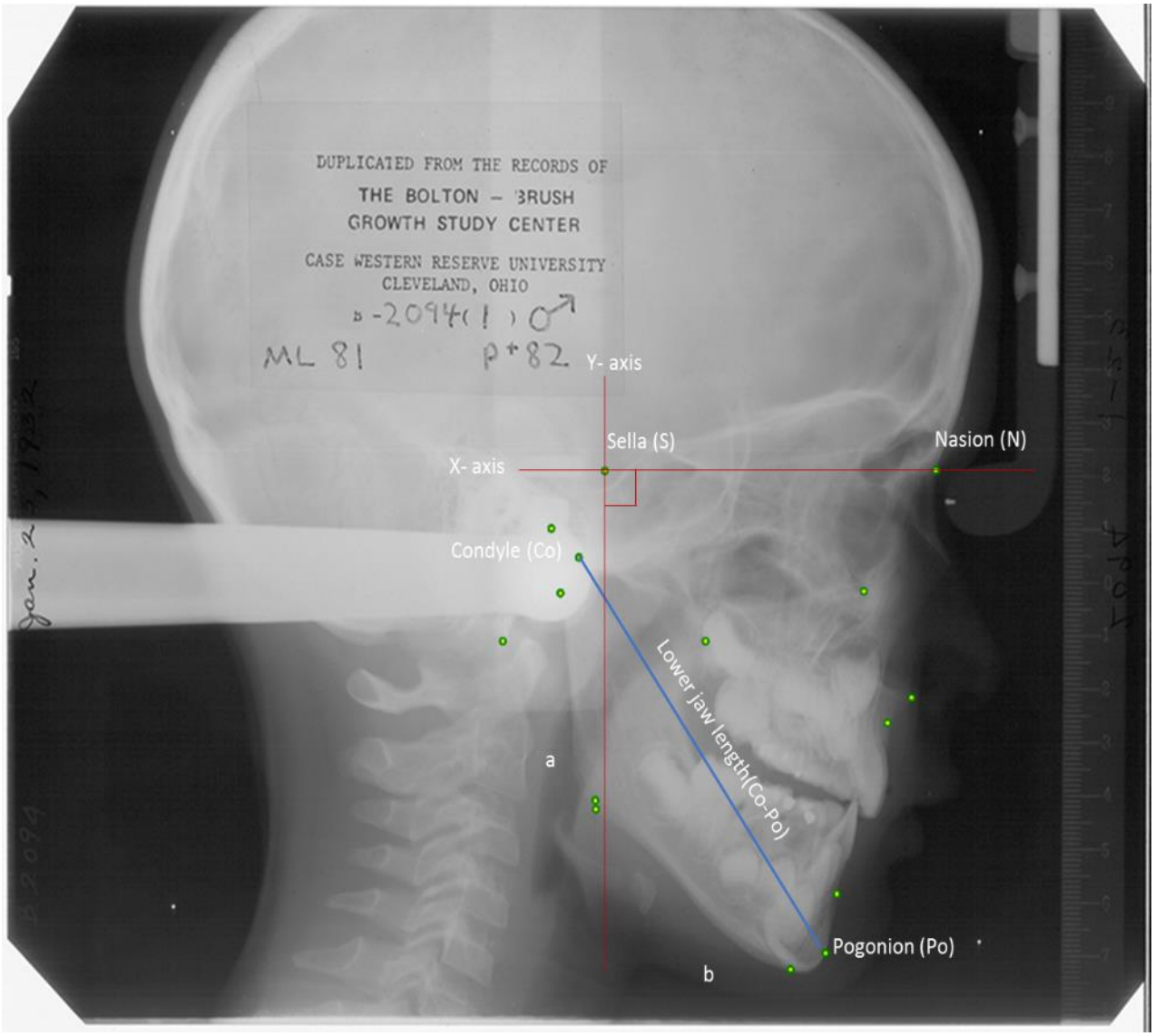


Figure D-4. Lower jaw length measured as a linear distance (in millimetres) between condyle and pogonion on lateral cephalogram. The lower jaw length (COPOD) shown as blue line.

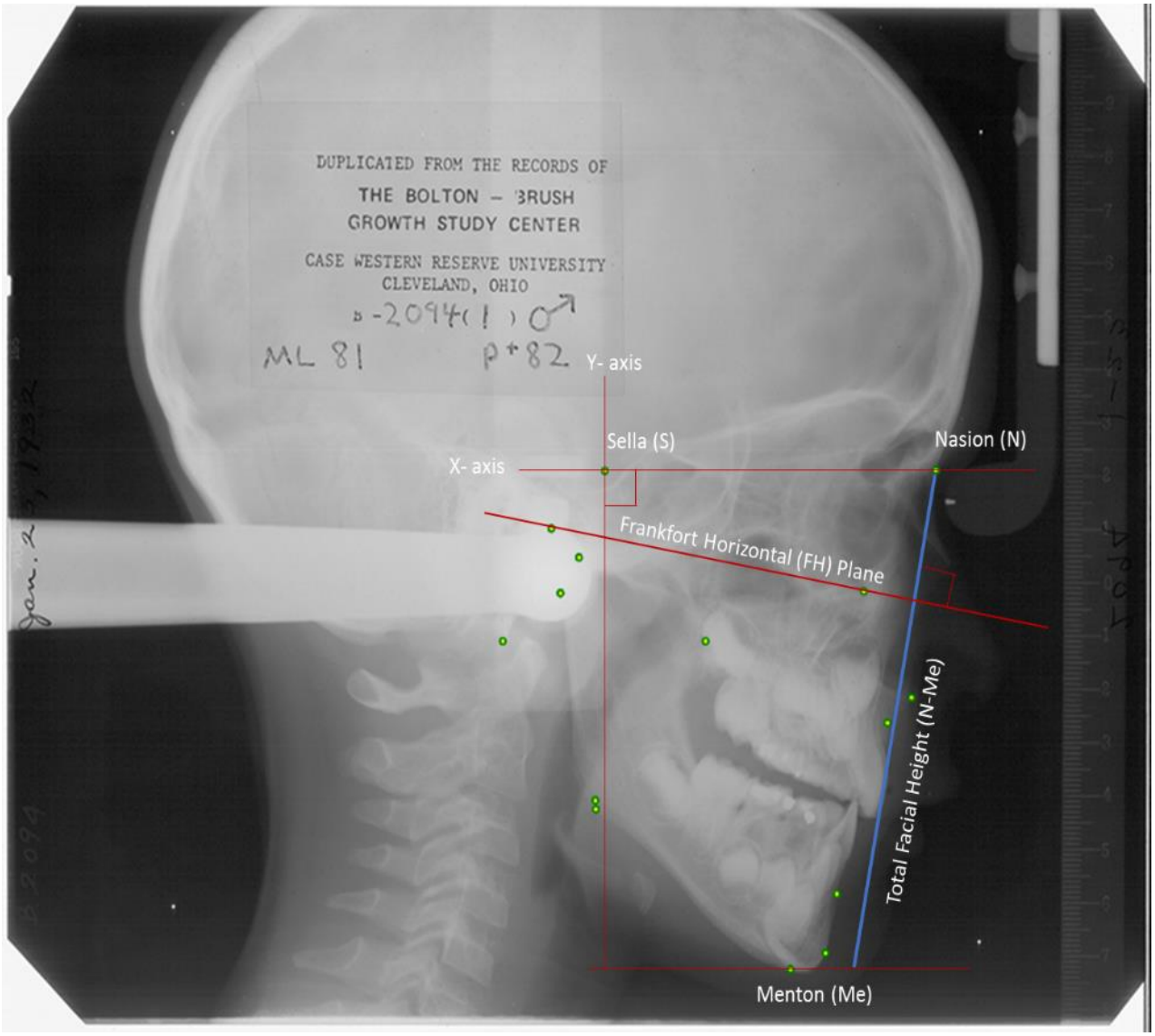


Figure D-5. Total face height measured as linear distance (in millimetres) between menton and nasion on lateral cephalogram. The total face height (shown as the blue line) is measured along the nasion perpendicular on the Frankfort horizontal plane (TFHNP).

Data calculated using Cartesian coordinates (236 individuals) and measurements available directly from the AAOF Craniofacial Growth Legacy Collection (129 individuals) were merged one-to-one for each individual and the age of growth assessment. This revealed duplicate records of 95 individuals.

The availability of duplicate records (i.e., same measurements available from both the source) allowed me to test the precision of my calculations. Table D-2 provides summary statistics for the comparison of the upper length, lower jaw length and total face height calculated using Cartesian coordinates and measurements available directly from the collection. The results show no difference except for a rounding error.

Table D-2. Accuracy evaluation of measurements calculated using cartesian coordinates. Results show comparison between measurements calculated using cartesian coordinates and measurements directly available from the AAOF Craniofacial Growth Legacy Collection database.

	Available from the AAOF Craniofacial Growth Legacy Collection				Calculated using Cartesian coordinates			
	Mean	SD [a]	Min [b]	Max [c]	Mean	SD [a]	Min [b]	Max [c]
Upper jaw length (COPAD)	87.87	7.65	67.50	114.90	87.87	7.65	67.45	114.90
Lower jaw length (COPOD)	108.22	10.58	82.30	151.80	108.22	10.59	82.30	151.80
Total face height (TFHNP)	111.94	11.10	84.80	157.40	111.94	11.10	84.80	157.40

COPAD: condyle–point A measurement in millimetres; COPOD: condyle–pogonion measurement in millimetres; TFHNP: total face height measurement in millimetres

[a] Standard deviation [b] Minimum [c] Maximum

After removing duplicate records (95 individuals), data were available for 270 individuals (130 males and 140 females). Of these, data for one male and one female were excluded as they did not provide any growth measurement between seven and 18 years of age. Also, as the Fels growth study contributed growth data for only one individual (a male), it was excluded from the data. Thus, the final data comprised measurements of the upper and lower jaw length and total face height from 2,060 lateral cephalograms of 267 individuals (128 males and 139 females; mean age 11.60 years, SD 2.90, range 7–18 years).

APPENDIX E. SEARCH STRATEGIES TO IDENTIFY AND RETREIVE STUDIES INCLUDED IN THE SYSTEMATIC REVIEW (CHAPTER 2)

Table E-1. MEDLINE search strategy (via Ovid) to identify and retrieve studies included in the systematic review.

#	Searches	Results
1	exp Epidemiologic Study/	2112667
2	exp observational Study/	43502
3	exp Cohort Study/	1707164
4	exp prospective Study/	462882
5	(epidem$ adj (study or studies)).ab,ti.	76507
6	(observ$ adj (study or studies)).ab,ti.	77269
7	(cohort adj (study or studies)).ab,ti.	143540
8	(prospective adj (study or studies)).ab,ti.	148079
9	(comparative adj (study or studies)).ab,ti.	90275
10	(follow up adj (study or studies)).ab,ti.	43826
11	cohort analy$.ab,ti.	5761
12	control analy$.ab,ti.	4869
13	longitudinal.ab,ti.	191950
14	follow up.ab,ti.	795935
15	prospective$.ab,ti.	576742
16	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15	2924163
17	Craniofacial growth Mandible/or Maxilla/or cranial base/or Maxillofacial Development/or Facial Bones/or Skull/or Face/	107168
18	Craniofacial development Mandible/or Maxilla/or cranial base/or Maxillofacial Development/or Facial Bones/or Skull/or Face/	107168
19	(cranio$ adj (growth or development)).ab,ti.	2097
20	(facial$ adj (growth or development)).ab,ti.	1427
21	(Craniofacial$ or facia$ or face$ or jaw$).tw.	366171
22	(mandible$ or mandiblular$ or lower jaw$).tw.	37077
23	(maxilla$ or maxillary$ or upper jaw$).tw.	61555
24	(maxillo-mandibular$ or jaw relation$).tw.	981
25	(face and growth).ab,ti,mp,af.	8869
26	(facial and growth).ab,ti,mp,af.	9718
27	(craniofacial and growth).ab,ti,mp,af.	7430
28	(jaw$ and growth).ab,ti,mp,af.	4185
29	(mandible and growth).ab,ti,mp,af.	5296
30	(maxilla$ and growth).ab,ti,mp,af.	5168
31	17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30	490845
32	exp Cephalometry/	25211
33	Cephalomet$.ab,ti,mp,af.	26833
34	Cephalogram$.ab,ti,mp,af.	3365
35	Cephalometric, Face/or Skull/or Facial Bones/or Mandible/or Maxilla/or Cephalometry/	112344
36	Cephalometry/or cephalometric.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	26549

37	cephalo$.tw.	40370
38	32 or 33 or 34 or 35 or 36 or 37	143540
39	Tooth Movement/or Orthodontics, Corrective/or Orthodontics, Preventive/or Orthodontics, Interceptive/or orthodontic treatment.mp.	22022
40	orthodontic appliance.mp. or Orthodontic Appliances/or appliance.mp. or appliance.tw. or "functional appliance".tw. or muscle.tw. or flour$.tw. or headgear$.tw. or therapy.tw. or implant$.tw. or tooth$.tw. or molar$.tw. or incisor$.tw. or amelob$$.tw. or impacted $$.tw. or transpose $.tw. or nasophary$.tw. or treatment.ti. or treatment.ab. or crowd$.ab.	5336566
41	Temporomandibular Joint Disorders.mp. or Temporomandibular Joint Disorders/or Disorder$.tw.	933691
42	Obstructive Sleep Apnea.mp. or Sleep Apnea, Obstructive/or Apnea.tw.	37633
43	Cleft Palate/or Cleft Lip/or cleft.mp. or cleft.tw. or anomol$.tw. or palsy$.tw. or dysplasia.tw. or disea$.tw.	3254927
44	surgery.mp. or Surgery/or surgery.tw. or fracture$.tw. or trauma$.tw. or reconstruction.tw. or restorat$.tw. or tumor$.tw.	2821904
45	Syndrome.mp. or Syndrome/or syndrome.tw.	1047646
46	Surgical Flaps/or Young Adult/or Craniosynostoses/or Skin Neoplasms/or Middle Aged/or Anthropometry/or Adult/or Aged/or Photography/or Photographic.mp.	6543078
47	39 or 40 or 41 or 42 or 43 or 44 or 45 or 46	12923160
48	Cross-Sectional Studies/or "cross sectional".mp. or "cross sectional".tw.	363609
49	("cone-beam" or "cone beam" or CBCT).tw.	9571
50	47 or 48 or 49	12997769
51	(16 and 31 and 38) not 50	1016
52	limit 51 to humans	803
53	exp Puberty/or exp Adolescent/or exp Growth/or spurt.mp. or exp Human Growth Hormone/	2521106
54	52 and 53	**286**

Table E-2. EMBASE search strategy (via Ovid) to identify and retrieve studies included in the systematic review.

#	Searches	Results
1	Epidemiology/or Epidemiologic.mp.	110490
2	exp observational Study/	43502
3	exp Cohort Study/	1707164
4	exp prospective Study/	462882
5	(epidem$ adj (study or studies)).ab,ti.	76507
6	(observ$ adj (study or studies)).ab,ti.	77269
7	(cohort adj (study or studies)).ab,ti.	143540
8	(prospective adj (study or studies)).ab,ti.	148079
9	(comparative adj (study or studies)).ab,ti.	90275
10	(follow up adj (study or studies)).ab,ti.	43826
11	cohort analy$.ab,ti.	5761
12	control analy$.ab,ti.	4869
13	longitudinal.ab,ti.	191950
14	follow up.ab,ti.	795935
15	prospective$.ab,ti.	576742
16	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15	2623521
17	Craniofacial growth Mandible/or Maxilla/or cranial base/or Maxillofacial Development/or Facial Bones/or Skull/or Face/	107168
18	Craniofacial development Mandible/or Maxilla/or cranial base/or Maxillofacial Development/or Facial Bones/or Skull/or Face/	107168
19	(cranio$ adj (growth or development)).ab,ti.	2097
20	(facial$ adj (growth or development)).ab,ti.	1427
21	(Craniofacial$ or facia$ or face$ or jaw$).tw.	366171
22	(mandible$ or mandiblular$ or lower jaw$).tw.	37077
23	(maxilla$ or maxillary$ or upper jaw$).tw.	61555
24	(maxillo-mandibular$ or jaw relation$).tw.	981
25	(face and growth).ab,ti,mp,af.	8869
26	(facial and growth).ab,ti,mp,af.	9718
27	(craniofacial and growth).ab,ti,mp,af.	7430
28	(jaw$ and growth).ab,ti,mp,af.	4185
29	(mandible and growth).ab,ti,mp,af.	5296
30	(maxilla$ and growth).ab,ti,mp,af.	5168
31	17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30	490845
32	exp Cephalometry/	25211
33	Cephalomet$.ab,ti,mp,af.	26833
34	Cephalogram$.ab,ti,mp,af.	3365
35	Cephalometric, Face/or Skull/or Facial Bones/or Mandible/or Maxilla/or Cephalometry/	112344
36	Cephalometry/or cephalometric.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	26549

37	cephalo$.tw.	40370
38	32 or 33 or 34 or 35 or 36 or 37	143540
39	Tooth Movement/or Orthodontics, Corrective/or Orthodontics, Preventive/or Orthodontics, Interceptive/or orthodontic treatment.mp.	22022
40	orthodontic appliance.mp. or Orthodontic Appliances/or appliance.mp. or appliance.tw. or "functional appliance".tw. or muscle.tw. or flour$.tw. or headgear$.tw. or therapy.tw. or implant$.tw. or tooth$.tw. or molar$.tw. or incisor$.tw. or amelob$$.tw. or impacted $$.tw. or transpose $.tw. or nasophary$.tw. or treatment.ti. or treatment.ab. or crowd$.ab.	5336566
41	Temporomandibular Joint Disorders.mp. or Temporomandibular Joint Disorders/or Disorder$.tw.	933691
42	Obstructive Sleep Apnea.mp. or Sleep Apnea, Obstructive/or Apnea.tw.	37633
43	Cleft Palate/or Cleft Lip/or cleft.mp. or cleft.tw. or anomol$.tw. or palsy$.tw. or dysplasia.tw. or disea$.tw.	3254927
44	surgery.mp. or Surgery/or surgery.tw. or fracture$.tw. or trauma$.tw. or reconstruction.tw. or restorat$.tw. or tumor$.tw.	2821904
45	Syndrome.mp. or Syndrome/or syndrome.tw.	1047646
46	Surgical Flaps/or Young Adult/or Craniosynostoses/or Skin Neoplasms/or Middle Aged/or Anthropometry/or Adult/or Aged/or Photography/or Photographic.mp.	6543078
47	39 or 40 or 41 or 42 or 43 or 44 or 45 or 46	12923160
48	Cross-Sectional Studies/or "cross sectional".mp. or "cross sectional".tw.	363609
49	("cone-beam" or "cone beam" or CBCT).tw.	9571
50	47 or 48 or 49	12997769
51	(16 and 31 and 38) not 50	988
52	limit 51 to humans	776
53	exp Puberty/or exp Adolescent/or exp Growth/or spurt.mp. or exp Human Growth Hormone/	2521106
54	52 and 53	**274**