Issues in modelling growth data within a life course framework

Richard Silverwood

2008

A thesis submitted for the degree of Doctor of Philosophy

University of London

Abstract

This thesis explores, develops and implements modelling strategies for studying relationships between childhood growth and later health, focusing primarily on the relationship between the development of body mass index (BMI) in childhood and later obesity. Existing growth models are explored, though found to be inflexible and potentially inadequate. Alternative approaches using parametric and nonparametric modelling are investigated.

A distinction between balanced and unbalanced data structure is made because of the ways in which missing data can be addressed. A dataset of each type is used for illustration: the Stockholm Weight Development Study (SWEDES) and the Uppsala Family Study (UFS). The focus in each application is obesity, with the first examining how the adiposity rebound (AR), and the second how the adiposity peak (AP) in infancy, relate to later adiposity. In each case a two-stage approach is used.

Subject-specific cubic smoothing splines are used in SWEDES to model childhood BMI and estimate the AR for each subject. As childhood BMI data are balanced, missingness can be dealt with via multiple imputation. The relationship between the AR and late-adolescent adiposity is then explored via linear and logistic regression, with both the age and BMI at AR found to be strongly and independently associated with late-adolescent adiposity.

In the UFS, where childhood BMI data are unbalanced, penalised regression splines are used within a mixed model framework to model childhood BMI and estimate the AP for each subject. The data correlations induced by the family structure of the observations are addressed by fitting multilevel models in the second stage. Both age and BMI at AP are found to be positively associated with later adiposity.

The two nonparametric modelling approaches are found to be effective and flexible. Whilst the thesis concentrates on BMI development in childhood and later adiposity, the techniques employed, both in terms the modelling of growth and the relating of the derived features to the outcomes, are far more widely applicable.

All models are wrong, some are useful.

G. E. P. Box [1]

Contents

List of Figures

List of Tables

Acknowledgements

I would first and foremost like to acknowledge the tremendous amount of help, guidance and encouragement provided by my supervisor, Bianca De Stavola. This, combined with the insightful input of my Advisory Committee of Tim Cole, Dave Leon and Chris Frost, and additional direction from Mike Kenward, has provided the inspiration for the development of this thesis. Equally important, however, has been the love and support of my partner, Nia Williams. Heartfelt thanks to you all.

The data used in this thesis have been made available by a variety of sources, to whom I am extremely grateful. The Stockholm Weight Development Study data were provided by Martin Neovius (Karolinska Institute, Stockholm), the Uppsala Family Study data by Dave Leon (LSHTM) and Ilona Koupil (Stockholm University and Karolinska Institute, Stockholm), and the National Survey of Health and Development data by Mike Wadsworth and Di Kuh (MRC Unit for Lifelong Health and Ageing, London). Data for both the National Child Development Study and the British Cohort Study were obtained from the UK Data Archive, funded by the Economic and Social Research Council. Additionally, although it does not feature within the thesis itself, many thanks to Vivienne Moore and Michael Davies (University of Adelaide) for allowing access to the Generation 1 Study.

The thesis is typeset using LaTeX 2_ε[2].

Acronyms

The following acronyms are used in the thesis:

AP	Adiposity peak	MI	Multiple imputation
AR	Adiposity rebound	ML	Maximum likelihood
BCS	British Cohort Study	MLE	Maximum likelihood estimation
BLUP	Best linear unbiased predictor	NCDS	National Child Development Study
BMI	Body mass index	NMAR	Not missing at random
CHB	Concordant high birthweight	NSHD	National Survey of Health and
CHD	Coronary heart disease		Development
CI	Confidence interval	OR	Odds ratio
CLB	Concordant low birthweight	OLS	Ordinary least squares
CV	Cross-validation	REML	Restricted maximum likelihood
DB	Discordant birthweight	SD	Standard deviation
EDF	Equivalent degrees of freedom	SE	Standard error
EM	Expectation maximisation	SES	Socioeconomic status
FP	Fractional polynomial	SPAWN	Stockholm Pregnancy and Women's
GCV	Generalised cross-validation		Nutrition
LRT	Likelihood ratio test	SWEDES	Stockholm Weight Development Study
MAR	Missing at random	UFS	Uppsala Family Study
MCAR	Missing completely at random	VIF	Variance inflation factor
MCMC	Markov chain Monte Carlo	WHO	World Health Organisation
MI	Multiple imputation	%BF	Percentage body fat

Chapter 1

Introduction

There is growing recognition that the risks of many adverse health outcomes in later life are affected not only by concurrent factors but also by early life variables. The study of the effects of exposures arising at different points throughout the life course is referred to as life course epidemiology.

The life course approach often incorporates longitudinal data — repeated measurement of the same variable at different occasions within the same individual. Analysis of longitudinal data is complicated by, often complex, correlations between the observed measurements at different occasions. This thesis focusses on life course analyses in which the repeated measures are of anthropological variables in childhood, and thus describe the growth of individuals.

In addition to issues surrounding the analysis of longitudinal data, observational studies are often subject to missingness. There is an expanding repertoire of techniques to handle missing data, but these approaches are not always pursed, jeopardising the validity of any conclusions.

There may also be more general issues surrounding data structure. If individuals can be considered as members of groups in which they are likely to be more similar to each other than to others outside of the group, then this must be taken into account in any analysis.

The fitting of models to the growth data for individuals can be seen as a tool to estimate values of the growth dimension at ages at which it was not observed, or to identify features of the growth curve, such as turning points or points of maximum velocity. Over the course of the last few decades many models have been developed to describe the growth of certain anthropometric dimensions, although more general statistical modelling approaches are also often used.

Due to its increased prevalence over recent years, obesity has become a major health concern worldwide. Because early life factors may prefigure later obesity, this has rightly become the focus of much life course research. In particular, attention has often focussed on the identification of periods of life which can be considered as 'critical periods' for later obesity. Although not necessarily ideal for this purpose, body mass index (BMI, defined as an individual's weight in kg divided by the square of their height in metres) has become the most frequently used measure in the assessment of obesity.

This thesis explores, develops and implements modelling strategies for studying relationships

between childhood growth and later health, focusing primarily on the relationship between the development of BMI in childhood and later obesity. In particular, interest lies in modelling individual childhood BMI growth curves in order to derive growth features of interest. These include the adiposity peak (AP, the maximum BMI obtained in infancy before BMI starts to decrease) and the adiposity rebound (AR, the period around 6 years of age when BMI begins to increase again following a nadir). Exploration of the relationships between the timings of, and levels of adiposity at, these growth features and later obesity can provide important insights into the development of obesity through the life course.

These relationships are investigated in two different datasets: the Stockholm Weight Development Study (SWEDES), a prospective longitudinal study of weight development in 481 children, and the Uppsala Family Study (UFS), a study of 602 families, each including two full siblings, where growth data is obtained via linkage to health records.

The thesis is divided into four Parts which correspond to 'Background', 'Approaches for balanced growth data only', 'General approaches' and 'Discussion'.

In Part I, Chapter 2 provides some background to the subject matter covered in the thesis. The general concepts of the life course approach and the modelling of growth are described, and the existing literature relating to obesity, which is central to the later applications, discussed. The aims of the thesis are presented in Chapter 3, and in Chapter 4 the datasets which are later used are introduced. The main statistical issues which are encountered in this type of life course analysis, such as data structure and missing data, are discussed in Chapter 5, along with statistical approaches which can be used to handle them. In Chapter 6, more specific subject-matter issues are discussed. The modelling of childhood growth is reviewed and the potential for using standardised measurements in life course studies is explored.

The distinction between balanced and unbalanced childhood growth data is an important one as it affects the approaches which can be used in an analysis. Thus in Part II (Chapters 7 and 8) methods are pursued which relate only to situations where the growth data are balanced, whilst Part III (Chapter 9) includes approaches which may be used for unbalanced growth data. However, as balanced data are effectively only a special case of unbalanced, the approaches of Part III are also appropriate in the balanced growth data setting.

Chapter 7 discusses the use of a naive multivariable regression analysis to relate BMI development to late-adolescent obesity (in terms of both BMI and percentage body fat (%BF)) in SWEDES. Here, the growth data at some or all of the measurement occasions are directly related to the later health outcome via multivariable regression, and a complete-case approach to missing data is employed.

This analysis is extended in Chapter 8 where the relationship between the AR and late-adolescent obesity is more explicitly explored. Childhood BMI trajectories are modelled using subject-specific cubic smoothing splines, from which the AR is estimated for each individual and related to measures of later adiposity. An essentially separate analysis is conducted in which multi-

ple imputation (MI) is used to handle missing data prior to the spline-fitting. A comparison of the two analyses provides both epidemiological and methodological insights. Further work considers whether the AR can be considered as a critical period for later obesity.

In Chapter 9 the relationship between the AP during infancy and later obesity is examined in the UFS, which includes unbalanced childhood growth data. Penalised regression splines are used within a mixed model framework to model childhood BMI growth and identify the AP for each subject, then the association is explored using mixed models to account for the structure in the dataset.

In Part IV (Chapter 10) the preceding epidemiological findings are brought together and the methodological considerations arising from the different applications discussed. Areas for future work are identified.

The Appendix reproduces a paper by Silverwood and Cole [3] regarding statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size.

Part I

Background

Chapter 2

Background

2.1 The life course approach

Kuh and Ben-Shlomo [4] define life course epidemiology as, "The study of long-term biological, behavioural, and psychosocial processes that link adult health and disease risk to physical or social exposures during gestation, childhood, adolescence, earlier in adult life, or across generations". This approach emerged to counteract the increasing polarisation of research in chronic disease etiology into either biological programming in utero or adult lifestyle factors. Life course epidemiology is built on the premise that various biological and social factors throughout life can independently, cumulatively and interactively influence health and disease in adult life [5]. Whilst the formal combination of these factors into a life course model provided a new way of thinking, the idea that childhood is important for adult health was not new in epidemiology or public health, being the prevailing model of health for the first half of the 20th century [6].

The life course approach has found many applications, in particular in the study of how patterns of early growth and other factors acting across the life course influence the onset and development of a wide array of common chronic diseases [4]. For example, women who grow faster in childhood and reach an adult height above the average for their menarche category have been found to be at particularly increased breast cancer risk [7].

Perhaps the best known example of a life course association is the developmental origins of health and disease (DOHaD) hypothesis. This expands upon the fetal origins of adult disease (FOAD) hypothesis, developed mainly by a group at the University of Southampton, led by Professor David Barker. Barker and colleagues have shown small size at birth or in infancy to be associated with an increased likelihood of adverse health outcomes in adulthood, including cardiovascular disease, coronary heart disease, stroke, hypertension, non-insulin dependent diabetes and impaired glucose tolerance [8]. This has led to the hypothesis that poor fetal nutrition, observed as small size at birth or subsequently, results in fetal adaptations that 'programme' the future propensity to chronic diseases in adulthood.

Whilst many of the studies of Barker and colleagues have shown a direct association between

small size in early life and adult health outcomes, in others the relationship has only emerged after body size at a later juncture (for example adult body mass index) is adjusted for. In the latter case it has been argued [9] that it is probably the change in size between points ('postnatal centile crossing') rather than fetal biology that is implicated.

Although much life course analysis has focussed on chronic disease epidemiology, this approach is also applicable within the context of infectious diseases and wider notions of health and wellbeing [6], for example in the investigation of prenatal and early life influences on the timing of menarche [10].

Some important conceptual models in the life course approach are *critical* and *sensitive periods*, and *accumulation of risk*. Critical and sensitive periods are both limited time windows in which a given exposure can have an effect on development and subsequent disease outcome. The difference between the two concepts is that outside of this window there is no excess disease risk associated with the exposure for critical periods, whilst for sensitive periods the excess risk is merely weaker [5]. Accumulation of risk occurs when the effect of an exposure accumulates gradually over the life course [6]. The ability to distinguish between these conceptual models is, however, often hampered by limited data being available at the relevant periods in the life course.

In general, even when data are available, life course epidemiology raises analytical challenges as both temporal and causal hierarchies among the exposures need to be taken into account [11]. If properly dealt with, such an approach allows the examination of dynamic processes and the identification of any critical or sensitive periods which may be present [12].

As noted above, there are also specific data quality issues. As different time periods and types of variables are usually examined, data from multiple sources are often merged, meaning that completeness, quality and coverage may vary. As a result, measurement errors and missing values affect life course studies to a greater extent than standard observational studies [11].

2.2 Growth

Human growth is the process of change in size and shape which occurs between conception and full maturity [13], generally defined in quantitative terms as the increase or decrease of some measurable quantity of tissue [14].

The formal study of growth has over 300 years of history [15], during which time many models to describe the changes in different anthropometric dimensions have been developed.

2.2.1 Dimensions of growth

Growth can be defined as the change in any one of many anthropometric variables. However, only height and weight are generally routinely measured in the clinical setting and included in medical records [16], thus it is propitious to focus attention on on these two dimensions, as well as composite measures obtained by their combination.

Further anthropometric variables for which measurements may be taken include head, waist and

hip circumferences, skinfold thicknesses (often measured at tricep and subscapular sites) and body composition. Body composition is usually considered as either a two-component (fat mass and fat-free mass) or four-component (fat mass, protein mass, water mass and mineral mass) model, measured using, for example, deuterium dilution, air-displacement plethysmography or dual-energy X-ray absorptionmetry [17].

Height

Height is a useful indicator of nutritional status [18]. Growth in height is generally a very regular process [19], which is often considered in three phases: infancy, childhood, and adolescence or puberty. *Infancy* is characterised by a high growth velocity immediately after birth and rapid deceleration until about 3 years of age [20]. This is followed by *childhood*, a period of lower, slowly decelerating velocity which lasts until the onset of puberty, although a slight increase in velocity, referred to as the *mid-growth spurt*, occurs between age 6 and 8 years in many children [19]. During *puberty* the *adolescent growth spurt* provides a marked acceleration of growth, then after the period of peak velocity there is deceleration until growth ceases [20]. Height changes little once final adult height is achieved.

Females are generally slightly shorter than males until adolescence, then at around age 11 years they become taller by virtue of their adolescent growth spurt occurring on average two years earlier than the males'. By approximately age 14 years, however, males are once again taller as their adolescent growth spurt has begun, whilst that of the females is nearly finished [19].

Secular increases in adult height (marked increases in the growth of successive generations of a population) over the last century or so have been seen globally, although decreasingly so over recent decades [21]. This is mirrored in children, though with an additional secular trend towards increased developmental tempo meaning that that the adolescent growth spurt is occurring at progressively younger ages [22].

Weight

Growth in *weight* is a somewhat less regular process than that in height, in that greater fluctuations, including decreases, are possible, though it still typically follows the phases of growth outlined above.

Females generally weigh a little less than males at birth, though they catch up and become heavier by age 9 or 10 years. Males become heavier again once females near the end of puberty at age 14 or 15 years [19].

Secular increases in weight have been reported in many parts of the world, both in adults and children [21]. Clearly this is in part due to the secular increases in height, but increasing adiposity has also been shown to contribute. Whilst increases in height have slowed over the last few decades, weight has continued to increase as part of the worldwide obesity epidemic [22].

Weight for height indices

Whilst weight and height are both, in some sense, reflections of the 'size' of an individual, often we are more interested in 'shape'. Information on shape can be obtained from height and weight measurements by removing the information on size. This is what *weight for height indices* try to achieve [23].

As weight is more variable than height, and thus more informative, height needs to be scaled up in some way when calculating weight for height indices. Two common forms of weight for height indices are

$$W - bH$$

and

$$\frac{W}{H^p}$$

where W and H are weight and height, and b and p specify how height should be scaled.

Relative weight A *relative weight* is obtained by expressing a subject's weight as a fraction of a reference weight, which is usually dependent on their height and sex. In adults weight and height are linearly related, so a regression line of weight on height in a reference population can be used to provide the reference weight for a given height [23]. Assuming the fitted regression line is

$$W = bH + c, \tag{2.1}$$

where b is the regression coefficient and c is the intercept, the reference weight, W_{ref}, for subject i with height H_i can be calculated as

$$W_{ref} = bH_i + c,$$

which corresponds to the average weight for all subjects of height H_i. Then if the weight of the subject is W_i their relative weight, W_{rel}, is

$$W_{rel} = \frac{W_i}{W_{ref}} = \frac{W_i}{bH_i + c}.$$

In children, however, weight and height are not linearly related so an approach using the regression line of weight on height is not appropriate. A reference weight could instead be obtained for a child's height indirectly using existing weight and height for age standards. The age at which the child's height, H_i, matches the median height in the height for age standard can be found, then the median weight at this age in the weight for age standard used as the child's reference weight, W_{ref} [23].

Indices of the form $\frac{W}{H^n}$ Commonly used weight for height indices such as the weight-height ratio ($\frac{W}{H}$), body mass index (BMI) or Quetelet's index ($\frac{W}{H^2}$) and ponderal index ($\frac{W}{H^3}$) are of the general form $\frac{W}{H^n}$, where n is a whole number.

Many studies have looked at the correlation between these different indices and height, preferring the index with the smallest correlation on the grounds that it best approximates relative weight. In men, BMI has been consistently found to be preferable, whilst in women BMI and the weight-height ratio are often seen to be equally useful. In adults, BMI has also been found to be the most highly correlated with various measures of body fat [23].

The weight-height ratio, BMI and ponderal index all change appreciably during childhood, so the indices need to be adjusted for age in children. This can be achieved by comparing the index calculated for a child with the same index calculated for a reference child of the same age and sex, creating a relative index. Relative BMI is particularly popular for this purpose as it has been found to be virtually uncorrelated with height for much of childhood, as well as having high correlation with measures of body fat [23].

A related concept to relative BMI is the *standardardisation* of BMI to create a *z-score* or *SD score*. Here, mean BMI in a reference dataset corresponding to the age and sex of a given child is subtracted from the calculated BMI for that child. This is then divided by the standard deviation (SD) of the BMI values for that age and sex in the reference dataset. If the distribution of the variable is skewed, as is often the case, then an additional parameter may be included to 'normalise' the data, as is employed in the LMS approach of Cole [24]. The z-score then indicates how many SDs above or below the mean BMI in the reference dataset the BMI for the child lies.

As BMI is the most widely used surrogate measure of adiposity, its involvement in the assessment of obesity is examined more closely in Section 2.3.2.

Benn index A more generalised form of the above index is the *Benn index* [25], $\frac{W}{H^p}$, where the exponent for height, p, is now allowed to take a non-integer value which is estimated from the population being studied. The value of p should be chosen so as to minimise the correlation between the index and height. Although there are several approaches to achieving this, Benn [25] advocates calculating the regression coefficient, b, of weight on height as in (2.1), then obtaining p as

$$p = \frac{b\bar{H}}{\bar{W}},$$

where $\bar{H}$ and $\bar{W}$ are the population means. Benn [25] also showed that, provided the correlation between height and relative adiposity does not differ too much from zero, this index will have a correlation with relative adiposity very near the maximum that can be achieved using this type of index.

Many studies have calculated p to be near to 2 in men and between 1 and 2 in women [23], though in children the optimal value of p changes with age. One way to use the Benn index with children is to analyse the data in narrow bands, calculating a potentially different value of p for

each, and thus adjusting simultaneously for both height and age. For much of childhood, including infancy, the optimal value of p has been found to be slightly greater than 2, although during puberty this increases to 3 [26]. It has, however, been suggested that Benn's assumption of low correlation between height and adiposity, which holds in adults, is not satisfied in children [27].

2.2.2 Modelling growth

Why model growth?

Many attempts have been made to find mathematical curves which fit, and thus summarise, the growth data of individuals. Indeed, Tanner [19] advises that fitting a curve to the individual values is the only way of extracting the maximum information about an individual's growth from the data.

The problem is effectively one of data smoothing. Given a number of points representing measurements taken on an individual at different ages, a smooth curve must be found which is believed to represent the underlying growth process more closely than the measurements themselves [28].

Growth models can successfully reduce large amounts of growth data to a small number of parameters. This is possible even when there is great variability in the number and spacing of measurements between individuals. It is then possible to compare growth between individuals, or even populations, using the parameters derived from the fitted models [29].

Whilst a major aim of any model must be to provide a satisfactory fit to the data, further features of growth models which are desirable include simplicity of the fitting procedure, biological interpretability of the model parameters, and model parsimony [30]. However, even the most rudimentary model should allow values of the variable to be estimated at ages between those where observations were made. Where differentiation of the growth model is possible, growth velocities and accelerations can also be examined. This allows identification of turning points (maxima and minima), as well as ages at maximum velocity and acceleration, for example peak height velocity during the adolescent growth spurt.

The fitting of growth curves can be considered as related to the area of statistics known as *functional data analysis* [31]. This is an approach for analysing data consisting of serial measurements, where each data series is first summarised as a smooth curve. Each curve is then treated as a single entity in the analysis.

Growth models

Many models have been used to describe human growth. Some relate to specific anthropometric variables over specific ranges of ages, whilst others are more general statistical modelling techniques.

There are several models which can describe the growth of either height or weight during the first few years of life. They achieve this by incorporating either an exponential function, as in the case of Jenss and Bayley [32], or a logarithmic function, as suggested by Count [33] and extended by Berkey and Reed [34].

Other models attempt to model height from birth or infancy right through to final adult height. The full model proposed by Count [33] and those of Bock and Thissen [35] and Karlberg [36] achieve this by modelling the growth curve as three separate phases, whilst Preece and Baines [29] derived a new family of mathematical functions with which to describe the height growth curve.

More general statistical modelling techniques such as polynomials, fractional polynomials and nonparametric modelling have also been employed to model growth, with varying degrees of success.

2.3 Obesity

Obesity, a condition of abnormal or excessive fat accumulation in adipose tissue, to the extent that health may be impaired, has become a major health concern worldwide in recent decades, with prevalence rising steeply [37]. In England, for example, the proportion of men classed as obese increased from 13.2% in 1993 to 23.1% in 2005, and from 16.4% to 24.8% in women, although there was no significant change in the proportion who were overweight [38].

Recently, the largest ever UK study into obesity concluded that dramatic and comprehensive action was required to stop the majority of the population becoming obese by 2050 [39], leading the Health Secretary to describe obesity as a "Potential crisis on the scale of climate change" [40].

Obesity is associated with increased risk of many adverse health conditions, including cardiovascular disease [41], type 2 diabetes [42], hypertension [42], and some cancers [43, 44]. Additionally, obesity may interact with other established risk factors. For example, the increased risks associated with small birth size for diseases in adulthood such as type 2 diabetes and hypertension are exacerbated in subjects who become obese in adulthood [45].

Obesity is a particularly intriguing aspect of life course research. Whilst patients with many adverse health outcomes are more likely to be obese, current obesity is dependent on the pattern of previous growth. This dual role as both an exposure and an outcome poses many interesting questions of interpretation [46].

2.3.1 Childhood obesity

Obesity in children and adolescents is a serious issue with many health and social consequences that often continue into adulthood. The prevalence of childhood obesity is increasing rapidly worldwide [37]. For example, obesity among boys aged 2 to 15 years in England rose from 10.9% in 1995 to 18.0% in 2005, and from 12.0% to 18.1% in girls [38]. Overweight among children and adolescents in the United States increased from 13.9% in 1999 to 17.1% in 2004 [47].

Reilly *et al* [48] found eight factors to be independently positively associated with obesity at age 7 in a UK cohort. They were: parental obesity, very early adiposity rebound, greater than 8 hours per week spent watching television at age 3 years, catch up growth, high weight z-score at ages 8 and 18 months, large weight gain in the first year, high birth weight and short sleep duration at age 3 years.

Obesity in childhood is associated with some immediate effects, including psychosocial out-

comes, with social isolation and peer problems more common in fatter children [49]. A population-based survey of 15 and 17 year olds in Sweden [50] found a significant association between adolescent obesity and depression and also experiences of shame, such as being degraded or ridiculed by others.

There is also evidence of childhood obesity being associated with the timing of puberty. Sandhu *et al* [51] found for men in the Christ's Hospital cohort that each SD increase in prepubertal BMI was associated with a 0.31 year decrease in the age at peak height velocity, which is used as an indicator of the timing of puberty.

However, most consequences of childhood obesity are deferred until adulthood. In a systematic review [52], childhood obesity was consistently found to be associated with most of the major cardiovascular risk factors, leading to the conclusion that obesity-mediated cardiovascular morbidity in adulthood can have its origins in childhood obesity. Adolescent obesity was also seen to be associated with adverse effects on social and economic outcomes in young adulthood.

Recently, Baker, Olsen and Sørensen [53] have found high BMI in childhood to be associated with increased risk of coronary heart disease (CHD) in adulthood in a large cohort of Danish schoolchildren. The risk of having a CHD event was seen to increase linearly with BMI z-score at each age in childhood, and also, for a given increase in BMI z-score, to increase as the age of the child increases.

There is a well established pattern of tracking of obesity from childhood to adulthood, meaning that even if overweight children avoid health problems in their youth, they have an increased likelihood of becoming overweight, and thus encountering the associated adverse health outcomes, as adults.

In a review of obesity tracking [54] it was found that about a third of obese preschool children and approximately half of obese school-age children were obese as adults. Generally, the risk of adult obesity was found to be at least twice as high for obese children as for nonobese children, and greater for children at higher levels of obesity and for children obese at older ages. However, most obese adults were not obese children. It was also suggested that the risk of obesity-related chronic diseases may be higher among obese adults who were *not* obese as children.

Eriksson *et al* [55] examined the relationship of adult obesity to childhood size in a Finnish cohort born in the 1920s and 30s. They found a 3-fold increase in obesity in men and women associated with having BMI greater than 16 kg/m^2 at age 7 years compared to BMI less than 14.5 kg/m^2.

Though the adverse health outcomes associated with obesity usually occur in adulthood [56], the ineffectiveness of the treatment of established obesity at this age is widely acknowledged [49]. It is thus often suggested that the problems of adult obesity may potentially be avoided by preventative measures taken in the more malleable climate of early childhood [57]. As Dr Ian Campbell of the National Obesity Forum says, "Clearly we are in the middle of an epidemic that is wreaking havoc on our children. The optimal time to intervene is in childhood, before irreversible damage has been done and while lifelong good habits can be learnt" [58].

2.3.2 Assessment of obesity

Whilst obesity, be it in childhood or adulthood, is thought to lead to adverse health outcomes, difficulties remain in assessing whether or not, or to what extent, an individual should be considered 'overweight' or 'obese'.

Body mass index

Accurate evaluation of obesity requires that both lean mass and fatness be taken into account [59], but the ideal definition for obesity, based on percentage body fat, is impractical for epidemiological use [60]. As a result of these conflicting requirements, BMI has become the preferred measure of adiposity for routine clinical and public health purposes [16].

BMI has been claimed to be a reliable and valid measure of adiposity in adults [16], giving an index that is broadly independent of height and equally applicable to men and women, which has proved exceptionally useful for large scale epidemiological work [61]. It has been found to be highly correlated with fat mass, to have a similar level of correlation with waist girth as fat mass does, and to have a similar level of correlation with abdominal visceral fat as both fat mass and waist girth do [62]. It is thus argued that BMI is perfectly adequate for clinical practice and population research.

However, BMI has been accused of having limited accuracy as it acts as a proxy for both lean mass and fat mass but can distinguish neither [59]. Maynard *et al* [63] found in a longitudinal study that despite BMI being highly correlated with both total body fat and percent body fat, it was also correlated with fat-free mass. Consequently individuals who are exceptionally muscular may be misclassified as overweight or obese. There is also much individual variability in the relationship between BMI and cardiovascular risk factors and long term health outcomes [64].

The use of BMI to investigate adiposity in children is complicated further by the manner in which BMI changes from birth through to early adulthood [61], with relationships between the fat and fat-free components of the body being affected by varying growth rates and maturity levels [63]. For most individuals, BMI increases from birth until about age 9 months, then decreases until around age 6 years, before increasing once more. This pattern is evident in the BMI curve for a typical child shown in Fig. 2.1. Despite these inherent complexities, BMI has been widely used in pediatrics owing to the ease with which measurements can be made on infants and children, and the often routine manner in which serial height and weight measurements are recorded.

Pietrobelli *et al* [65] found BMI to be strongly associated with both total body fat and percentage body fat measured by dual X-ray absorptiometry in a sample of Italian children and adolescents. Whilst this supports the use of BMI as a fatness measure in groups of children and adolescents, caution is recommended when comparing BMI across groups that differ in age.

BMI measurements may be standardised into age- and sex-specific z-scores using reference data, which is suggested as a useful approach for assessing adiposity cross-sectionally. As a measure for *change* in adiposity, however, the z-score may be less than ideal [66].

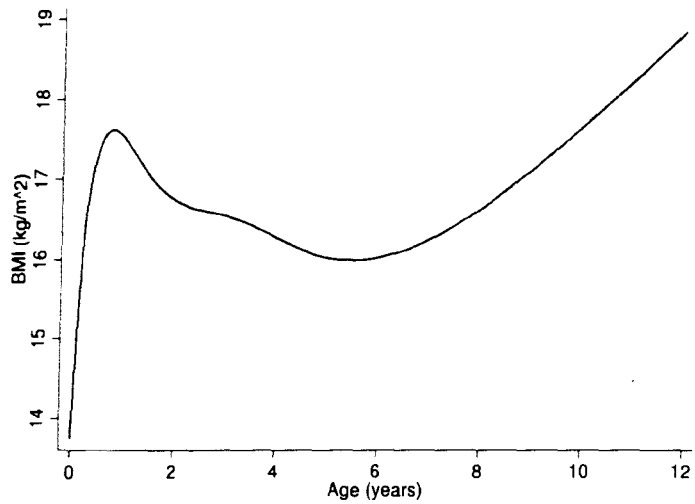

Fig. 2.1: A typical body mass index (BMI) growth curve through childhood.

Defining overweight and obesity in terms of body mass index

Whilst a large BMI value is likely to be indicative of an individual with greater adiposity, 'cut off' values have been sought with which to categorise individuals into different levels of obesity and overweight. A cut off point of 30 kg/m^2 is recognised internationally as a definition of adult obesity, but the World Health Organization (WHO) have gone further in defining a pragmatic adult classification system based on associations between BMI and all cause mortality [37], as given in Table 2.1.

BMI category	BMI (kg/m^2)
Underweight	<18.5
Ideal	18.5–24.9
Pre-obese ('overweight')	25.0–29.9
Obese class I	30.0–34.9
Obese class II	35.0–39.9
Obese class III	>40.0

Table 2.1: World Health Organisation body mass index (BMI) categories [37].

Whereas an approach based, albeit crudely, on known risk ratios for different levels of BMI is possible for an adult classification system, the fact that BMI in childhood changes substantially with age (as seen for the individual in Fig. 2.1), and a scarcity of equivalent data for children, makes it difficult to identify health based cut offs for children.

The most common method to overcome this is through the use of reference data to calculate age- and sex-specific BMI centiles or z-scores for individuals relative to the reference population. Subjects above a certain centile or z-score may then be defined as 'overweight' or 'obese', which has been found to be a reasonable approach for screening those at risk from excess adiposity [63]

with acceptable diagnostic accuracy for high body fat [67]. Commonly used BMI reference data include those for the UK dating from 1990 [68] and for the USA from 2000 [69].

A slightly different approach has been used to create explicit international BMI cut off values for children [60]. Here, data were pooled from several studies around the world, then the centiles corresponding to the adult BMI cut offs for overweight and obesity in Table 2.1 identified and extrapolated back through childhood, so that the same proportion of children are classified as overweight and obese at each age.

Childhood BMI growth references of this nature, however, represent only a snapshot of the reference population at one point in time, so do not reflect secular trends [16]. As secular increases in childhood BMI are well established [37], BMI growth references would need to be regularly updated if the aim was simply to describe the current BMI distribution. As this would then mask the secular trends, BMI growth references, for example the UK 1990 reference [70], may be intentionally 'frozen' at a certain point in time so that trends can be related to that fixed baseline [64].

Recently, the WHO have developed new international growth *standards* [71]. These differ from the growth *references* in that they summarise how children *ought* to grow, rather than merely how the children in the reference sample *do* grow. This is achieved by focussing on children who are growing 'optimally', and can thus be viewed as a model for other children to follow [72].

However, the use of cut off points to define obesity in terms of BMI is, whilst convenient, not ideal, as the use of arbitrary dichotomous (or categorical) classifications will inevitably result in a substantial number of individuals entering or leaving the 'obese' group over time [49].

2.3.3 The role of growth in obesity

As obesity is known to track from childhood through to adulthood [54, 73], obesity at any age in childhood leads to increased likelihood of obesity in adulthood. However, certain more specific patterns of growth in childhood have been found to be associated with the development of obesity. Dietz [74] posits three critical periods in childhood: gestation and early infancy, the adiposity rebound, and adolescence. This provides a convenient framework for exploring the life course approach to investigating obesity [75].

Infancy Whilst some studies have found there to be little evidence of size in infancy predicting later obesity [76], most find that infants who are at the highest end of the distribution for weight or BMI are at increased risk of later obesity [77].

Similarly, infants who grow more rapidly have been seen to be more likely to become obese [77]. For example, Ekelund *et al* [78] found increasing weight gain in infancy (from birth to age 6 months) (as well as in early childhood (age 3 to 6 years)) to be associated with greater BMI, fat mass, relative fat mass and waist circumference, but also fat-free mass, at age 17 years in a Swedish cohort study.

Rapid weight gain is closely related to the concept of 'catch-up growth'. Catch-up growth is

a property of human growth whereby children return to their genetic trajectory after a period of reduced growth, for example because of illness [79]. Whilst catch-up growth can thus occur at any age, it is most commonly observed during the first year or two of life in those of small size at birth, which is often taken as an indicator of intrauterine growth restriction. There is strong evidence that postnatal catch-up growth is positively associated with later obesity [80].

Ong *et al* [45] found that children who gained more than 0.67 z-scores of weight in the first two years of life had lower weight, length, and ponderal index at birth than other children, and were more likely to have been exposured to maternal smoking during pregnancy, indicating potential intrauterine growth restriction. These children generally became heavier, taller and fatter (in terms of BMI, percentage body fat and waist circumference) at five years than other children.

In a small study of low and normal birthweight children, Ibáñez *et al* [81] found the low birthweight children to have similar weight to the normal birthweight children (i.e. to exhibit catch-up growth) by age 2 years. By age 4 years, however, the low birthweight children gained more abdominal fat and body adiposity and less lean mass than the normal birthweight children.

Adiposity rebound The term '*adiposity rebound*' (AR) was introduced by Rolland-Cachera *et al* [82] to describe the period around 6 years of age when BMI begins to increase following a nadir. This feature of the BMI curve can be clearly seen for the individual in Fig. 2.1. Their initial work showed a relationship between age at AR and adiposity at age 16 years, with early AR (before age 5.5 years) being followed by a significantly higher adiposity level than a later AR (after age 7 years) [82]. Rolland-Cachera *et al* [83] went on to confirm that the predictive value of AR lasts until young adulthood. This is important as after AR, increasing adiposity with age might stop earlier among those with advanced AR than among those with delayed AR, removing the influence of age at AR on adult adiposity.

More recent studies investigating AR have drawn similar conclusions. Siervogel *et al* [84] found a negative correlation between age at AR and BMI at age 18 in a US longitudinal study. Whitaker *et al* [85] introduced a further explanatory variable, BMI at AR, and found adult obesity rates to be higher in those who were 'heavy' (BMI z-score ≥ 0.05) versus 'lean' (BMI z-score, < -0.54) at AR (24% versus 4%) and in those with early (age < 4.8 years) versus late (age ≥ 6.2 years) AR (25% versus 5%). Williams *et al* [86] found in a longitudinal study of a New Zealand cohort that BMI in early adulthood (ages 18 and 21 years) was associated with both age and BMI at AR. Guo *et al* [87] found an early AR to be associated with adult BMI overweight status in females, though not in males, in the Fels Longitudinal Study. In a US cohort Freedman *et al* [88] found subjects with an early AR (age less than 5 years) to be on average 4–5 kg/m^2 heavier in earlier adulthood than subjects with a late AR (age 7 years or later), although this association was not independent of childhood BMI levels.

There have been fewer studies concerning the AR in developing countries, but that of Corvalán *et al* [89] found increases in BMI between age 3 and 7 years to be strongly associated with adult fat mass and abdominal fat, though also associated with fat-free mass, in a study in Guatemala.

Whilst not focusing explicitly on the AR, the association between upward BMI centile crossing over the period of AR and later adiposity is supportive of previous findings regarding the AR.

Parental obesity, an established risk factor for obesity, has been found to be associated with earlier AR [90], suggesting that it may operate, at least in part, through influence on the timing of the AR.

The ability to predict adult fatness in early childhood has led to the suggestion that the AR is a 'critical period' of growth [74]. However Cole [91] disagrees, arguing that age at AR reflects level and rate of change of BMI centile at that age, with upward BMI centile crossing at the AR and other ages in childhood predicting later obesity. Cole goes on to suggest that, instead, the period leading up to the AR is in fact a critical period when children 'choose' a trajectory of static, rising or falling centile which predicts both their age and BMI at AR.

Whilst referred to as the 'adiposity' rebound, this feature is generally defined in terms of the BMI growth curve. Although BMI is relatively well correlated with measures of adiposity in childhood, there is only limited evidence that the AR is truly a feature of adiposity and not just of BMI. Thus the AR could perhaps be more accurately described as the 'BMI rebound' as it is elsewhere [87].

Adolescence Fewer studies concentrate on adolescence as a critical period for obesity, perhaps due to it being more temporally proximal to adulthood and thus less suitable for the application of interventions.

In the Fels Longitudinal Study, Guo *et al* [87] found maximum BMI velocity during pubescence to be associated with adult overweight status, with a 1 kg/m^2 per year increase in maximum BMI velocity being associated with almost three times the risk of being overweight in males and nearly double the risk in females. BMI level at maximum BMI velocity in pubescence was also associated with adult overweight status, with a 1 kg/m^2 increase in BMI leading to double the risk of being overweight in males and over three times the risk in females. Thus, in contrast to other studies reporting the importance of the AR for subsequent obesity, Guo *et al* provide evidence of the greater importance of adolescence.

In an Indian population-based cohort, whilst higher BMI and BMI gain in infancy and early childhood were found to predict adult lean mass more strongly than adult adiposity, greater BMI and BMI gain in late childhood and adolescence were found to predict increased adult adiposity [92], again illustrating the importance of this period for the development of obesity.

2.4 Summary

The life course approach is a useful framework in which to study a variety of health outcomes. One such outcome, which is encountering an increasing amount of interest due to its rising prevalence worldwide in recent years, is obesity. Whilst the life course approach has been seen to be a fruitful method by which to examine the development of obesity, many previous studies have had limitations.

Issues surrounding data availability mean that often only limited periods of the life course can be examined, or that the data for a given period are insufficient for an extensive analysis. Related to this, many studies also face problems of missing data, which are not always well handled. These are issues which further studies must address more thoroughly.

Many previous studies which utilise longitudinal growth data would benefit from more explicit modelling of growth. In particular, when the AR is being estimated so that its timing can be assessed for its influence on later obesity, growth modelling is often inadequate [87] or non-existent [88].

The timing of the AR has been consistently found to be inversely related to later obesity, leading to it being suggested as a critical period in the development of obesity [74]. However, others argue that the observed relationship is more statistical than physiological [91]. Further research regarding the effect of BMI centile crossing around the period of the AR may be illuminating. Of specific interest is whether the timing of the AR has any real predictive ability for later obesity beyond that of BMI and BMI velocity (or, equivalently, BMI centile crossing) at a similar age.

Whilst the relationship between the AR and later obesity has been widely examined, other features of the BMI growth curve have been less well studied in this context. The AR, as a turning point, is a readily identifiable feature of the typical BMI growth curve. So, however, is the point at which BMI reaches a maximum in infancy, which is seen clearly in the individual in Fig. 2.1, although this feature has received little interest. Examination of the association between the timing of this feature and later obesity may also prove fruitful. In particular, the combination of this with existing knowledge regarding the development of obesity around the period of AR and adolescence would make the study of obesity more truly a 'life course' discipline.

Chapter 3

Aims

The overall aim of this thesis is:

To explore, develop and implement modelling strategies for studying relationships between childhood growth and later health.

This aim is split into two sub-aims:

1. **To investigate and utilise existing methods for describing and modelling features of childhood growth, expanding and developing them where necessary;**

   The modelling of childhood growth involves complex correlated data, often affected by missingness. Existing growth models for anthropometric variables including height, weight and BMI are explored, and alternatives using parametric and nonparametric modelling investigated. The roles of data structure and data missingness are considered and approaches under different scenarios developed. Features of individual growth trajectories such as maxima, minima and periods of greatest growth velocity can then be derived from these models.

2. **To examine and implement methods for relating features of childhood growth to later outcomes;**

   Once derived, features of childhood growth can then be related to later health outcomes. The role of data structure is again important here, so mixed model approaches are considered alongside regression analysis techniques.

These approaches are illustrated using several datasets: the Stockholm Weight Development Study (SWEDES), the Uppsala Family Study, and three of the British national birth cohorts (National Survey of Health and Development, National Child Development Study and British Cohort Study).

The main relationship of interest is that between childhood BMI trajectory and later obesity.

Chapter 4

Introducing the datasets

This thesis utilises data from a variety of different datasets, which are briefly described in this chapter.

Two of the datasets (the Stockholm Weight Development Study (SWEDES), described in Section 4.1, and the Uppsala Family Study (UFS), described in Section 4.2) include longitudinal measurements of childhood growth, as well as several measures of later health outcomes. These datasets thus correspond to the type of data structure on which the thesis concentrates and are used in the exploration, development and implementation of modelling strategies. More specifically, it is the relationship between childhood growth in BMI and later obesity that is examined in each instance. The key difference between these two datasets with regards to the present analytical framework is in the longitudinal childhood growth data. In SWEDES these data are measured at common ages across all subjects and are thus balanced, whilst in the UFS measurements are not restricted to common ages, resulting in unbalanced data. This has implications for the type of analytical model which can be used and the manner in which missing (or sparse) data can be handled.

The remaining three datasets described in this chapter — the National Survey of Health and Development (NSHD) (Section 4.3.1), the National Child Development Study (NCDS) (Section 4.3.2) and the British Cohort Study (BCS) (Section 4.3.3) — are British national birth cohorts. The data collected at the various follow-up ages in each cohort provide longitudinal measures of childhood growth which are semi-balanced in the sense that there is a pre-specified age at which they were *intended* to be observed, but there is some degree of variability in the *actual* ages at which measurements were taken. Although suitable measures of later health outcomes could potentially be derived in each cohort, this is not pursued as these datasets are not used for the same purpose as SWEDES and the UFS. Instead, the British birth cohorts are used to illustrate the standardardisation of childhood BMI data into age- and sex-specific z-scores. In this context it is the national representativeness of and the temporal differences between the cohorts which are the important features of the datasets.

4.1 Stockholm Weight Development Study

The Stockholm Weight Development Study (SWEDES) is a prospective longitudinal study of weight development in the offspring of mothers who participated in the Stockholm Pregnancy and Women's Nutrition (SPAWN) Study. All mothers attending 14 maternity clinics in southern Stockholm over a 12 month period between 1984 and 1985 were invited to participate in the SPAWN study. Two thousand three hundred and forty-two women agreed to participate and were studied retrospectively during pregnancy at maternity clinics, and monitored prospectively for up to 1 year after delivery [93]. One thousand four hundred and twenty-three of these women completed the SPAWN study at 1 year follow-up and from these, 481 mothers and their offspring were invited to participate in the follow-up study (SWEDES) when the offspring were approximately 17 years of age [94].

As part of the SPAWN study, weight and length at birth of the offspring were recorded from hospital records and gestational age estimated from date of the last menstrual period reported by the mother. During infancy, height and weight were measured as part of routine visits to a child welfare centre by standard clinical procedures. Measurements were taken three further times after birth during the first year (at 6, 9 and 12 months) and annually thereafter until age 6 years [78]. From age 7 years onwards annual measurements of height and weight were recorded in journals by the offsprings' schools.

The SWEDES follow-up, when the offspring were approximately 17 years old, involved measurement of a variety of anthropometric, metabolic, psychological and lifestyle variables for the mothers, their offspring, or both, of which only those relevant to the analyses in this thesis are detailed here.

Anthropometric variables such as standing height, weight, waist circumference and body composition were measured in the same manner for both mothers and offspring at physical examinations. Standing height was measured to the nearest 0.5 cm with subjects stood in bare feet against a wall-mounted stadiometer. Body weight was measured to the nearest 0.1 kg using a BodPod scale with subjects wearing light clothing [94]. Body volume was measured using the BodPod system, which utilises air-displacement plethysmography. Fat mass, percentage fat mass (or percentage body fat, %BF) and fat-free mass were then calculated according to the equation of Siri [95] using the software provided by the manufacturer. Measurements were taken in duplicate with subjects wearing tight underwear or swimwear and a swimcap [78]. A further element to the SWEDES follow-up was a questionnaire covering maternal education, occupation and monthly income.

The sample within the SWEDES dataset represents a mixed metropolitan population from both the inner city and suburban districts of Stockholm, with a distribution in social groups that has been established to correspond reasonably well to the population in the Stockholm area [96].

The SWEDES data have previously been used in several published analyses, including a description of the associations between physical activity and fat mass in adolescents [94] and an examination of the associations between rapid weight gain in infancy and early childhood in relation to body composition in young adults [78].

4.2 Uppsala Family Study

The Uppsalastudien 'Familij och Hälsa' or Uppsala Family Study (UFS) was designed to survey and examine families comprised of two full-siblings and their biological mother and father. All families with at least two consecutive singleton children delivered at term (38–41 weeks of gestation) and within 36 months of each other at the Uppsala Academic Hospital, Uppsala, Sweden, between 1987 and 1995 were considered potentially eligible for the study. Children also had to share the same biological father and families had to live within Uppsala county at the time of the study, with both parents of Nordic origin. If there were more than two children in the same family fulfilling these criteria then the eldest pair of siblings were chosen [97].

By linkage between the Swedish Medical Birth Registry (a complete population-based register of births in Sweden) and the current population register, 5226 women and their 10,452 offspring were identified as fulfilling the above criteria and hence comprised the sampling frame for the study. In March 2000 letters were sent to the families inviting them to take part. A small number of respondent families were excluded at this point either because the father was not living within Uppsala county or because one or both parents were born outside of the Nordic area [98].

Initially the focus for the linked dataset was to study early and maternal effects on blood pressure and cardiovascular disease [99]. To increase statistical efficiency it was decided to invite only families where the siblings were either both in the top or bottom quarter of the birthweight distribution ('concordant high' or 'concordant low' birthweight) or the sex-adjusted difference in birthweight between them was 0.4 kg or more ('discordant' birthweight).

Of the respondents to the initial letters, 1,967 families fell into one of these sampling groups and were invited to take part in the study. 71% of these families responded, with just under half agreeing to take part, leading to the eventual recruitment of 602 families (31% of those eligible). Of these, 328 sibling pairs had discordant birthweight, 137 sibling pairs had concordant low birthweight, and 137 sibling pairs had concordant high birthweight. Participation rates were very similar across all three sampling groups.

Children's birth data, including gestational age, birthweight, length and head circumference, and placental weight, were obtained from mothers' obstetric records through the Swedish Medical Birth Registry. Parental birth data were obtained from grandmothers' obstetric records. Children's postnatal growth data, including serial measurements of height and weight, were obtained from health records, kept by Child Health Centres (if the child was younger than 6 years) or at schools.

All children, all mothers and 569 (94.5% of) fathers had a physical examination between May 2000 and November 2001 when children were aged 5–13 years, at which the following measurements were recorded: blood pressure, height, sitting height, weight, tricep and subscapular skinfolds, waist and hip circumference, and children's Tanner stage [19]. All anthropometric measurements were taken three times and the mean value used. In particular, height was measured with a wall-fixed stadiometer to an accuracy of 0.1 cm with subjects walking around the room between measurements and weight was measured with the subject wearing underwear to an accuracy of 0.1 kg using electronic scales [98]. From the concurrent observed height and weight values, body mass

index (BMI) was obtained, from which BMI z-scores were calculated using the Swedish population reference values [100].

Parents were also asked to complete questionnaires (one for each of the four family members) concerning demographic and socioeconomic circumstances, lifestyle, health-related behaviour and medical history. These were returned by 581 (96.5% of) younger children, older children and mothers and 552 (91.7% of) fathers.

The main analyses for this study have been published and show an inverse association between childhood systolic blood pressure (SBP) and birthweight of -2.3 mmHg/kg (95% CI -4.4 to -0.3) within families and -1.5 mmHg/kg (95% CI -3.1 to 0.0) between families, after adjustment for gestational age, sex, and height and weight at examination [101]. The existence of an inverse association of birthweight with SBP within families demonstrates that factors that vary between pregnancies in the same woman can influence later blood pressure of offspring. Also, morning cortisol has been found to have no association with size at birth, and to not mediate the birthweight-blood pressure association [102].

4.3 National birth cohorts

Three prospective, longitudinal national birth cohorts, dating respectively from 1946, 1958 and 1970 are utilised in the thesis. These cohorts are by design nationally representative and all three remain important ongoing, multidisciplinary studies.

4.3.1 National Survey of Health and Development

The National Survey of Health and Development (NSHD) was the first of the British national birth cohort studies and remains one of the longest running large-scale studies of human development in the world. It began as a national maternity survey designed to investigate the cost of childbirth and the quality of associated health care following concern over falling birth rates [103].

The target sample for the first data collection was the 16,695 registered births in England, Scotland and Wales that occurred in the first week of March 1946, of which 13,687 were successfully surveyed. From this original population a sample totalling 5,362 children and consisting of all those whose fathers were non-manual or agricultural workers and a randomly selected one in four sample of children of other manual workers was selected from the population of married mothers having single births. A weighting can be applied in analyses in order to adjust for this sampling procedure [104].

This sample has now been studied 21 times, most recently at age 53 years [103]. At different follow-up ages data have been obtained to address questions regarding growth, development, morbidity, educational experience and attainment, delinquency, income, occupation, and physical and mental function using various methods of data collection, including via midwives, obstetricians, health visitors, school nurses and doctors, teachers, postal questionnaires, interviewers and research nurses [105]. Many findings and publications have resulted from the NSHD.

Analyses in this thesis use anthropometric data from the collections at ages 4, 6, 7, 11 and 15 years in the NSHD. At these ages children were measured and weighed in their underclothes by school doctors or nurses [106]. Electronic data were obtained directly from the Medical Research Council NSHD team at University College London.

4.3.2 National Child Development Study

The National Child Development Study (NCDS) takes as its subjects all the people born in England, Scotland and Wales in one week in March 1958. It has its origins in the Perinatal Mortality Survey, which initially included over 17,000 subjects and aimed to identify social and obstetric factors linked to stillbirth and neonatal death. From this original focus, the NCDS has broadened its scope to include many aspects of health, education, and social development [107].

Following the initial birth survey in 1958, there have to date been six attempts to trace all members of the birth cohort, at ages 7, 11, 16, 23, 33 and 42 years. At birth, information was obtained from the mother and from medical records by the midwife. At the first three surveys, information was obtained from parents, head teachers, class teachers, school health services and the subjects themselves via interviews, questionnaires and medical examinations. At the later surveys information was gathered using professional survey research interviewers. The birth cohort was augmented by including immigrants born in the relevant week in the target sample for the first three follow-ups [108]. There have been over 900 publications involving the NCDS to date, and the cohort has been extremely influential in its impact on policy and practice [107].

This thesis uses anthropometric data from the NCDS at follow-up at ages 7, 11 and 16 years, at which children were measured and weighed in their underclothes as part of their medical examination [106]. Electronic data for the NCDS were obtained from the UK Data Archive [109] and relevant variables identified with the help of the data dictionary provided by the Centre for Longitudinal Studies [110].

4.3.3 British Cohort Study

The British Cohort Study (BCS) follows a similar pattern to the NCDS, taking as its subjects all those living in England, Scotland and Wales who were born in one week in April 1970. Data were collected about the births and families of over 17,000 subjects, initially focussing on the medical management of pregnancy and birth. Since then, however, the scope has broadened to include physical, educational, social and economic development [111].

Since birth there have been six further attempts to gather information from the whole cohort, at ages 5, 10, 15, 26, 30 and 34 years. Information at birth was collected using a questionnaire completed by the midwife and supplemented by data from clinical records. Data at later surveys were collected using a variety of interviews, questionnaires, medical examinations, tests of ability and postal surveys. Additional people born in the same week who immigrated to the UK or were identified subsequently have been added to the cohort [112]. To date there have been over 300 publications based on analysis of data from the BCS [111]

Anthropometric variables measured by school medical staff with a standardised technique [113] at follow-up ages 10 and 16 years are used in this thesis. Weight was not measured at age 5 years so this follow-up age is not included. Electronic data for the BCS were obtained from the UK Data Archive [114], with relevant variables identified with the help of the data dictionary provided by the Centre for Longitudinal Studies [115].

Chapter 5

Statistical issues and methods

There are several statistical issues which are pertinent to the present aim of relating a later health outcome to longitudinal growth data collected earlier in life. Data structure, both in terms of the longitudinal data and the potential of further overall hierarchical structure, are discussed in Section 5.1. Missing data are an issue in almost all epidemiological studies. The nature of missing data and statistical methods to deal with this are outlined in Section 5.2.

Two methodological approaches relevant to the work presented in this thesis are then described. The first, a single-stage analysis, relates the later outcome directly to the earlier observed values. This is discussed in Section 5.3. An alternative two-stage approach, whereby the longitudinal growth data for each individual is first modelled, then the later outcome related to one or more features of the fitted growth curve, is introduced in Section 5.4. Commonly-used methods are only discussed briefly whilst for more novel approaches greater details are given.

Section 5.5 provides an overview of the statistical issues and methods discussed in this chapter.

5.1 Data structure

It is important to acknowledge the structure of a dataset as part of any analysis. General hierarchical structure is considered in Section 5.1.1, and the role of longitudinal data, which is central to the life course approach, is discussed in Section 5.1.2.

5.1.1 Hierarchical data

In many situations it is natural to consider individuals as belonging to groups, such as families, school classes or geographical areas. The members of these groups, or *clusters*, are likely to be more similar to each other than to other members of the population, for example the physical characteristics of siblings being more similar than those of unrelated individuals. Thus the cluster in this case is the family. Clusters may also be nested within one another. For example, families may be considered as belonging to towns. As a result, these type of data are often referred to a *hierarchical data*.

Many statistical models are based on the assumption that separate observations in a sample are independent of one another, meaning that the value of one observation is not influenced by the value of another [116]. But clustered data of this nature means that this assumption of independence is unlikely to hold, and to analyse such data as if they were independent would lead to bias [117]. One approach to handling hierarchical data is through the use of *mixed models*, as discussed in Section 5.3.3.

5.1.2 Longitudinal growth data

In *longitudinal* studies several measurements are taken on the same individual over time, in contrast to cross-sectional studies in which measurements are taken at a single time point. This enables direct study of the change in a variable over time. Longitudinal data can be collected either *prospectively*, following subjects forwards in time, or *retrospectively*, by extracting multiple measurements on each individual from historical records [118].

Longitudinal studies are a special case of hierarchical data. Here, the 'clusters' are the subjects, with repeated observations on the same subject likely to be more similar to each other than to observations on other subjects [116]. As a result, longitudinal data require special statistical methods which take into account this hierarchical structure in order to draw valid inferences [118].

In the present setting of relating a single later outcome to earlier longitudinal growth data, the aim is not explicitly to describe the pattern of growth observed. However, it may be advantageous to do this as the first stage in a two-stage analysis, as described in Section 5.4. If growth is to be modelled for more than one individual using data which have been collected longitudinally, then the structure of the data must be taken into account. This is again achievable using *mixed models*, as described in Section 5.4.1.3.

5.1.2.1 Balanced and unbalanced longitudinal growth data

The terms 'balanced' and 'unbalanced' are often used with slightly varying connotations. Here, they are taken be descriptive of *study design* rather than data missingness, and the concern is only with the collection of longitudinal growth data.

Balanced longitudinal growth data are defined as data resulting from studies where the anthropometric variable of interest is *intended* to be observed at the same set of common ages for each subject in the study. Whether the variable is *actually* observed for a given individual at a given age is immaterial. *Unbalanced* longitudinal growth data, on the other hand, occur when there is no intention to observe the anthropometric variable at a common set of ages for each subject.

More formally, let there be m subjects in a longitudinal study designed so that subject i, $i = 1, \ldots, m$, is observed n_i times at ages x_{ij}, $j = 1, \ldots, n_i$. If, for each value of j, $x_{ij} = x_{i'j}$ for all i and i', the the longitudinal dataset is *balanced*. Implicit in this is that, since $x_{in_i} = x_{i'n_{i'}}$ for all i and i', both the intended number of observations and the age at the final observation are the same for every subject. If, however, for *any* value of j there exists an age $x_{i'j}$ so that $x_{ij} \neq x_{i'j}$, then the longitudinal dataset is *unbalanced*. Under this scenario there is no necessity for either the

number of observations or the age at the final observation to agree between subjects.

Distinguishing between balanced and unbalanced longitudinal growth data is important as it has implications on the statistical approaches which can be utilised in any analysis. In particular, many methods of analysis can only cope with balanced data [118]. For example, a multivariable regression analysis of a later outcome (for example, overweight in adulthood) on a longitudinally observed anthropometric variable (for example, height) at each observation time would not be possible if the data are unbalanced.

5.2 Missing data

Missing data occur whenever a datum which was expected to be present in a dataset is unavailable. This could, for example, be because an individual has refused to answer a certain question in a survey, a sample was accidently destroyed in a laboratory, or a study ran out of funding so was unable to complete the data collection to the intended extent. Missing data are an almost unavoidable problem in many epidemiological studies, and the nature of life course research means that the problem may be particularly acute under this approach [119].

However, data can only be 'missing' if, in some sense, they are 'expected'. Thus, when considering longitudinal growth data, missing data can only be defined when the data are *balanced*, as defined in Section 5.1.2.1. In *unbalanced* longitudinal growth data there are no specific ages at which observations are expected, so the concept of 'missingness' cannot be considered in the same way. However, there may still be periods when an individual has few or no observations, which is of similar concern. This is referred to as *data sparseness*.

Missing data patterns and mechanisms are introduced in Section 5.2.1. Then several different approaches to the handling of missing data are outlined: complete-case analysis (Section 5.2.2), single imputation (Section 5.2.3), and multiple imputation (Section 5.2.4).

5.2.1 Missing data patterns and mechanisms

Little and Rubin [120] suggest that it is useful to distinguish the *missing data pattern*, which describes which values are observed and which are missing, and the *missing data mechanism*, which concerns the relationship between missingness and the values of the variables.

Consider a dataset including p variables, Y_j, $j = 1, \ldots, p$. Here no distinction is drawn between explanatory and outcome variables. Let y_{ij} be the value of variable Y_j for subject i, $i = 1, \ldots, n$. Then let $Y = (y_{ij})$ represent the $n \times p$ data matrix. Now define the *missing data indicator matrix* $M = (m_{ij})$, with $m_{ij} = 1$ if y_{ij} is missing and $m_{ij} = 0$ if y_{ij} is non-missing. The matrix M then defines the *missing data pattern* [120].

When considering longitudinal data, a distinction may wish to be made between data missing intermittently or due to dropout. Using this notation, missing values occur due to *dropout* if

whenever y_{ij} is missing, so are y_{ik} for all $k \geq j$. Otherwise we say that the missing values are *intermittent*. In general, dealing with intermittent missing values is more difficult than dealing with missing values due to dropout [118].

Missing data mechanisms describe the relationship between missingness and the values of the variables, and are crucial as the properties of missing data methods depend very strongly on the nature of these dependencies [120]. The following framework for discussing missing data mechanisms is based on the definitions of Little and Rubin [120].

Data are said to be *missing completely at random* (MCAR) if missingness does not depend on the values of the data Y, either missing or observed, such that

$$P(M|Y) = P(M).$$

Let Y_{obs} denote the observed components of Y, and Y_{mis} the missing components, so that $Y = (Y_{\text{obs}}, Y_{\text{mis}})$. Then data are said to be *missing at random* (MAR) if missingness depends only on the components of the data Y that are observed (Y_{obs}), and not on those that are missing (Y_{mis}), such that

$$P(M|Y) = P(M|Y_{\text{obs}}).$$

If missingness depends on the components of Y that are missing (Y_{mis}) then the missing data mechanism is said to be *not missing at random* (NMAR).

The missing data mechanism has implications on the level of bias affecting different analyses, as well as the methods which are needed to correct for such bias. MAR is the minimal condition under which explicit incorporation of the missing data mechanism is not required. Thus the distinction between MAR and NMAR is often an important one. However, the observed data in a given dataset cannot be used to distinguish between MAR and NMAR mechanisms without additional untestable assumptions [121].

5.2.2 Complete-case analysis

A *complete-case analysis* restricts attention to the subsample of subjects with complete cases, excluding all individuals who have missing values for any of the variables being considered, whether outcome or explanatory.

A complete-case analysis is generally easy to carry out since standard statistical analyses intended for use with fully complete datasets can be applied without modification. This approach may be satisfactory with small amounts of missing data, but when this is not the case the loss of information in discarding incomplete cases can be great. This results in not just a loss of precision due to the reduced sample size, but also bias if the missing data mechanism is not MCAR [120].

As a result, this strategy is generally inappropriate [120]. One possible exception is when there is a specific interest in the sub-population of completers [118], although this situation is rather unusual.

Complete-case analysis is currently the most often used approach to handling missing data in life course epidemiology [119], and remains widely used in many epidemiological analyses, though efforts are being made to move beyond this [122].

5.2.3 Single imputation

Imputation describes a collection of methods whereby missing values are 'filled in' or *imputed* with a value which is, in some sense, plausible. Standard statistical procedures for complete data analysis can then be used to analyse the imputed dataset, with the imputed values treated identically to the non-missing values [123]. Here it is necessary to distinguish between the *imputation model*, used to obtain the value to be imputed, and the *analyst's model*, which is then fitted to the set of observed and imputed data [119].

Imputation is a general and flexible method for handling missing data problems [120] which incorporates many different specific approaches. *Single imputation*, whereby each missing value is imputed only once, contrasts with *multiple imputation*, described in Section 5.2.4, in which each value is imputed several times.

There are many simple approaches by which values for imputation can be obtained. *Mean* and *regression imputation* are described briefly here.

Mean imputation involves replacing the missing value (say y_{ij}, the value of variable Y_j for subject i) with the mean value of that variable ($\bar{Y}_j$) over the non-missing values within the sample.

Whilst this imputation method is incredibly simple, it is not recommended as its imputation model is never likely to be realistic, meaning that, even if the data are MCAR, the resulting estimates of the analyst's model will almost always be biased [119].

Regression imputation involves replacing each missing value by a prediction of its expected value given the other values that are observed for that subject. For example, if Y_k is a continuous variable which is missing for subject i then

$$\hat{y}_{ik} = \hat{\beta}_0 + \hat{\beta}_1 y_{i1} + \ldots + \hat{\beta}_{k-1} y_{ik-1} + \hat{\beta}_{k+1} y_{ik+1} + \cdots + \hat{\beta}_p y_{ip}$$

could be used as the imputed value. Here, $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{k-1}, \hat{\beta}_{k+1}, \ldots, \hat{\beta}_p$ are first estimated by fitting the linear regression of Y_k on $Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_p$ for all subjects with complete data.

Single imputation may, similarly to a complete-case analysis, be quite reasonable if the proportion of missing values is small [123]. However, the imputed values are effectively treated as known and thus, without special adjustments, single imputation cannot reflect the uncertainty surrounding the prediction of the missing values, meaning that inference will overstate precision. Although single imputation remains in wide use for handling missing data in many studies, it is becoming increasingly discouraged [122].

5.2.4 Multiple imputation

Multiple imputation (MI) is one of several *proper* methods for dealing with missing data. Here, as elsewhere [119], *proper* is used to refer to missing data methods which can be used for data which are MAR, regardless of the missing value pattern, and which provide unbiased estimates of the parameters and their standard errors in the analyst's model.

MI takes the idea of single imputation a step further by replacing each missing value in the dataset with a set of plausible values which are drawn from the predictive distribution of the missing data given the observed data. The inclusion of a random component reflects that imputed values are estimated rather than known with certainty. The MI procedure results in multiple datasets, each completed with independently imputed values, which are individually analysed using standard complete data procedures. The results from these analyses are then combined, using essentially the same process regardless of the complete data analysis used. The variability among the results of the analyses provides a measure of the uncertainty due to missing data, which, when combined with measures of ordinary sample variation, lead to a single inferential statement about the parameters of interest [124].

MI was originally developed for handling missing data in complex surveys used to create public-use datasets [125]. Consequently it is a powerful tool for more general large datasets with missing values across many variables.

Statistical assumptions

The key assumption underlying MI is that of *ignorability* (or *ignorable missingness*). Ignorability is made up of two parts—the assumption of data being MAR and the distinctness of parameters—which must both be satisfied.

The MAR assumption is as defined in Section 5.2.1. Although the MAR assumption cannot be verified with the data and may be questionable in some situations, the assumption becomes more plausible as more variables are included in the imputation model [126].

For ignorable missing data, the parameters θ of the data model and the parameters ξ of the model for the missing data indicators in M must also be *distinct*, meaning that any joint prior distribution applied to (θ, ξ) must factor into independent marginal priors for θ and ξ [124]. That is, knowing the values of θ does not provide any additional information about ξ, and vice versa.

Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a collection of methods for simulating random draws from nonstandard distributions via *Markov chains* [123]. Markov chains are sequences of random variables in which the distribution of each element depends on the value of the previous one [124]. Markov chains can be constructed so that they stabilise, or *converge*, to a distribution of interest. By repeatedly simulating steps from such a chain, draws are simulated from the distribution.

In a MI setting, MCMC is used to create independent imputations for the missing values, which are then used for repeated imputation inference. MCMC is one of the primary methods for

generating imputations in non-trivial missing data problems [124].

The aim is to impute independent realisations of $P(Y_{\text{mis}}|Y_{\text{obs}})$, the *posterior predictive distribution* of the missing data given the observed data. Suppose that $Y = (Y_{\text{obs}}, Y_{\text{mis}})$ follows a parametric model $P(Y|\theta)$ where θ has a prior distribution and Y_{mis} is ignorably missing. Now $P(Y_{\text{mis}}|Y_{\text{obs}})$ may be rewritten as

$$P(Y_{\text{mis}}|Y_{\text{obs}}) = \int P(Y_{\text{mis}}|Y_{\text{obs}}, \theta)P(\theta|Y_{\text{obs}})\mathrm{d}\theta,$$

where $P(Y_{\text{mis}}|Y_{\text{obs}}, \theta)$ is the *conditional predictive distribution* of Y_{mis} given θ and $P(\theta|Y_{\text{obs}})$ is the *observed-data posterior* of θ [124]. An imputation for Y_{mis} can thus be created by first simulating a random draw of the unknown parameters from their observed-data posterior

$$\theta^* \sim P(\theta|Y_{\text{obs}}) \tag{5.1}$$

followed by a random draw of the missing values from their conditional predictive distribution [123]

$$Y^*_{\text{mis}} \sim P(Y_{\text{mis}}|Y_{\text{obs}}, \theta^*). \tag{5.2}$$

Often, however, (5.1) cannot be easily summarised or simulated. Augmentation of Y_{obs} by an assumed value of Y_{mis} to give a complete-data posterior of

$$P(\theta|Y_{\text{obs}}, Y_{\text{mis}})$$

gives a more easily handled alternative [124]. Thus, consider an iterative, two-step process in which, given a current guess $\theta^{(t)}$ of the parameter, a value for the missing data is first drawn from the conditional predictive distribution for Y_{mis}

$$Y^{(t+1)}_{\text{mis}} \sim P(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)}). \tag{5.3}$$

Then, conditioning on the value obtained in (5.3), a new value of θ is drawn from a simulated complete-data posterior

$$\theta^{(t+1)} \sim P(\theta|Y_{\text{obs}}, Y^{(t+1)}_{\text{mis}}). \tag{5.4}$$

Repeating (5.3) and (5.4) from a starting value $\theta^{(0)}$ yields a stochastic sequence

$$\{(\theta^{(t)}, Y^{(t)}_{\text{mis}}) : \ t = 1, 2, \ldots\}$$

whose stationary distribution is

$$P(\theta, Y_{\text{mis}}|Y_{\text{obs}}).$$

Hence the sequences

$$\{(\theta^{(t)}), t = 1, 2, \ldots\}$$

and

$$\{(Y_{\text{mis}}^{(t)}), t = 1, 2, \ldots\}$$

have stationary distributions of

$$P(\theta|Y_{\text{obs}})$$

and

$$P(Y_{\text{mis}}|Y_{\text{obs}})$$

respectively [124]. Thus for a suitably large t, $\theta^{(t)}$ can be considered an approximate draw from $P(\theta|Y_{\text{obs}})$ and $Y_{\text{mis}}^{(t)}$ an approximate draw from $P(Y_{\text{mis}}|Y_{\text{obs}})$. The imputation of the missing value in (5.1) is often referred to as the *Imputation* (or *I*-) step, while the drawing of θ from the complete-data posterior in (5.2) is the *Posterior* (or *P*-) step.

However, in general it is not advisable to use successive iterates of $Y_{\text{mis}}^{(t)}$ as they tend to be correlated [124]. Thus *subsampling* may be utilised, whereby every kth iterate (i..e $Y_{\text{mis}}^{(k)}, Y_{\text{mis}}^{(2k)}, \ldots$) is instead used, where k is large enough so that the draws are approximately independent.

Assessing convergence

Investigation of the convergence of the MCMC process is essential to confirm that sufficient iterations have passed for the results to be reliable.

Time-series plots Convergence may be assessed by examining the iterates of θ from the simulation run. When θ is multidimensional, the behaviour of various components of θ, for example variable means and variances, can be investigated separately. Plotting successive estimates of a given component, say $\zeta = \zeta(\theta)$, at each iteration against the iteration number t forms a *time-series plot.* This provides a quick and easy way to assess convergence for that component [124], with long-term increasing or decreasing trends indicating that successive iterations are highly correlated and that the series of iterations has not yet converged [127].

Autocorrelation plots Autocorrelation plots also provide a more explicit means to examine the relationships between successive component estimates. The lag-k autocorrelation for a stationary series $\{\zeta^{(t)} : t = 1, 2, \ldots, m\}$ is defined to be [124]

$$\rho_k = \frac{\text{Cov}(\zeta^{(t)}, \zeta^{(t+k)})}{\text{Var}(\zeta^{(t)})}.$$

A sample estimate of ρ_k is given by [124]

$$r_k = \frac{\sum_{t=1}^{m-k}(\zeta^{(t)} - \bar{\zeta})(\zeta^{(t+k)} - \bar{\zeta})}{\sum_{t=1}^{m}(\zeta^{(t)} - \bar{\zeta})^2}$$

where $\bar{\zeta}$ is the mean of the series. A plot of r_k against k provides a useful summary of linear serial dependence, with long-term trends in ζ indicating slow convergence to stationarity [124].

Inference

The following approach for multiple imputation inference is based on that presented by Rubin [128].

Suppose again that $Y = (Y_{\text{obs}}, Y_{\text{mis}})$ and let Q denote a generic scalar quantity which is to be estimated, for example a mean, correlation or regression coefficient. Let $\hat{Q} = \hat{Q}(Y_{\text{obs}}, Y_{\text{mis}})$ be the estimate of Q that would be used if no data were missing. Let $U = U(Y_{\text{obs}}, Y_{\text{mis}})$ be the estimated variance associated with $\hat{Q}$, so that $\sqrt{U}$ is the complete data standard error. With m imputations there are m independent simulated versions of Y_{mis}: $Y_{\text{mis}}^{(1)}, \ldots, Y_{\text{mis}}^{(m)}$. Thus there can be calculated m different versions of $\hat{Q}$ and U. Let

$$\hat{Q}^{(t)} = \hat{Q}(Y_{\text{obs}}, Y_{\text{mis}}^{(t)})$$

and

$$U^{(t)} = U(Y_{\text{obs}}, Y_{\text{mis}}^{(t)})$$

be the point and variance estimates for the tth set of imputed data, $t = 1, \ldots, m$. Then the multiple imputation point estimate for Q is simply the arithmetic mean of the m point estimates,

$$\bar{Q} = \frac{1}{m} \sum_{t=1}^{m} \hat{Q}^{(t)}.$$

To obtain a variance estimate for $\bar{Q}$, both the *within-imputation variance* and the *between-imputation variance* must be considered. The within-imputation variance is the mean of the m variance estimates,

$$\bar{U} = \frac{1}{m} \sum_{t=1}^{m} U^{(t)}.$$

However, this assumes that all the observations are actually observed, so use of this alone would provide an underestimate of the variance. It is thus necessary to include a measure of the between-imputation variance (the variance of the m point estimates),

$$B = \frac{1}{m-1} \sum_{t=1}^{m} (\hat{Q}^{(t)} - \bar{Q})^2.$$

So the *total variance* is defined as

$$T = \bar{U} + (1 + m^{-1})B,$$

and inferences are based on Student's t-approximation

$$(Q - \bar{Q})\sqrt{T} \sim t_{\nu},$$

with degrees of freedom

$$\nu = (m-1)\left(1 + \frac{\bar{U}}{(1+m^{-1})B}\right)^2.$$

A measure of the *relative increase in variance* due to missing data is provided by

$$r = \frac{(1+m^{-1})B}{\bar{U}}$$

and the *rate of missing data* is approximately [123]

$$\lambda = \frac{r}{1+r}. \tag{5.5}$$

These results generalise to situations involving more than a single parameter, although some complexities are introduced.

How many imputations are required?

The *relative efficiency* (RE) of an estimate based on m imputations to one based on an infinite number of imputations is approximately [123]

$$\text{RE} = \left(1 + \frac{\lambda}{m}\right)^{-1}, \tag{5.6}$$

where λ is the rate of missing information as defined in (5.5). It can be seen from (5.6) that even with 50% missing information, an estimate based on $m = 10$ imputations has over of 95% the efficiency of one based on an infinite number of imputations. This has lead to the suggestion that unless rates of missing information are unusually high, there tends to be little or no practical benefit to using more than 5 to 10 imputations [123]. However, it has more recently been suggested [121] that far greater values of m may be more appropriate, with 100–200 required in some instances. With recent increases in available computing power meaning that it is practicable to use relatively large numbers of imputations, there would appear little reason not to do so.

5.3 Single-stage analysis

In a single-stage analysis the raw longitudinal anthropometric measurements in childhood are related directly to the distal outcome. Methods to achieve this include linear regression (discussed in Section 5.3.1), logistic regression (Section 5.3.2) and mixed modelling (Section 5.3.3).

However, these modelling approaches all require the anthropometric measurements to occur at the same ages in each individual. In other words, these techniques are restricted to datsets where the longitudinal data are *balanced*, as defined in Section 5.1.2.1.

5.3.1 Linear regression

Linear regression is a statistical approach which can be used to examine the dependency of a continuous outcome on one or more explanatory variables. When only a single explanatory variable

is being considered it is referred to as *simple* linear regression, and when there are more than one it is *multivariable* or *multiple* linear regression.

The general formulation of the linear regression model includes explanatory variables of arbitrary nature, but in the context of a single-stage analysis of the relationship between childhood growth and a later outcome, some or all of the explanatory variables may be repeated observations of the same anthropometric variable.

Consider a continuous outcome variable y (for example, BMI in adulthood) and p explanatory variables, x_j, $j = 1, \ldots, p$, which may be continuous (for example, BMI at a given age in childhood), dichotomous (for example, overweight vs. normal at a given age in childhood) or categorical (for example, obese vs. overweight vs. normal at a given age in childhood). Dichotomous explanatory variables should be coded 0 and 1, whilst each category of a categorical explanatory variable should be represented relative to a baseline category using dummy indicator variables, also coded 0 and 1. Let y_i and x_{ij} be, respectively, the observed values of y and x_j for subject i, $i = 1, \ldots, m$. Then the multivariable linear regression model for y on x_j, $j = 1, \ldots, p$, is given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i, \tag{5.7}$$

where the ε_i are independent and identically distributed with $\varepsilon_i \sim N(0, \sigma^2)$, for $i = 1, \ldots, m$.

The interpretation of the parameter β_j, $j = 1, \ldots, p$, in (5.7) differs depending on the type of variable that x_j is. If x_j is continuous then β_j is the estimated increase in the outcome y associated with a unit increase in x_j with all other explanatory variables (x_k, $k \neq j$) held constant (i.e. adjusting for all other explanatory variables). If x_j is dichotomous then β_j is the estimated increase in the outcome y associated with the exposure x_j, adjusting for all other explanatory variables. If x_j is a dummy indicator variable corresponding to a categorical variable then β_j is the estimated increase in the outcome y associated with the relevant category of the categorical variable relative to the baseline category, again adjusting for all other explanatory variables.

As all the estimated regression coefficients are mutually adjusted, any potentially confounding factors can be easily handled by including them in the multivariable linear regression model. Thus, for example, in a linear regression model of adult BMI on BMI measured at various ages through childhood, if the actual age at which adult BMI is observed differs between subjects, this may want to be taken into account. By including the age at measurement as a variable in the model, the estimated relationship between childhood BMI and adult BMI will be adjusted accordingly. Effect modification can also be assessed through the introduction of interaction terms to the regression model.

The regression coefficients $\beta_1, \ldots, \beta p$ can be estimated using the method of ordinary least squares (OLS), which minimises the sum of the squared residuals between the observed data points and the fitted regression function [116].

Multivariable linear regression is well suited to life course analysis involving serial anthropometric observations, although this approach does have several limitations. Firstly, each subject requires measurements to have been taken at *every* time point which appears in the model in order

to be included in the analysis. Secondly, the longitudinal data needs to be *balanced*, as defined in Section 5.1.2.1. Also, the regression coefficients are not constrained to vary smoothly across age, which would seem a more biologically plausible description of the relationship being studied [46].

Matrix notation

Let y be a continuous outcome variable and x_j, $j = 1, \ldots, p$, be p explanatory variables. Let y_i and x_{ij} be, respectively, the observed values of y and x_j for subject i, $i = 1, \ldots, m$. Let

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & \mathbf{x}_1^T \\ \vdots & \vdots \\ 1 & \mathbf{x}_m^T \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_0 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$

where

$$\mathbf{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}.$$

Then (5.7) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5.8}$$

referred to as the *general linear model* representation. The OLS estimator of $\boldsymbol{\beta}$ is then given by [129]

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{5.9}$$

Life course plots

In the case when the explanatory variables x_j, $j = 1, \ldots, p$, are observations of the same variable at different ages (possibly with additional explanatory variables for adjustment), *life course plots* can prove a useful aid in the interpretation of multivariable linear regression coefficients.

Initially, consider a continuous outcome variable y and an anthropometric variable x which is measured at two different ages to provide the explanatory variables x_1 and x_2. These are then converted to age- and sex-specific z-scores (z_1 and z_2) with a mean of 0, a SD of 1 and a normal distribution, so that the regression coefficients will be comparable [130]. From (5.7) the multivariable linear regression model for y on z_1 and z_2 can be given by

$$E(y) = \beta_0 + \beta_1 z_1 + \beta_2 z_2. \tag{5.10}$$

This can be rewritten as [130]

$$E(y) = \beta_0 - \beta_1(z_2 - z_1) + (\beta_1 + \beta_2)z_2 \tag{5.11}$$

or [46]

$$E(y) = \beta_0 + (\beta_1 + \beta_2)z_1 + \beta_2(z_2 - z_1), \tag{5.12}$$

where $(z_2 - z_1)$ is the change in z-score (or *growth*) of the anthropometric variable between the two measurements.

These three different parameterisations of the same model illustrate the duality of size and growth. However, the effect of a 1 SD increase in growth between the two measurements is seen to differ between (5.11) and (5.12). In (5.11) the result is an increase of $-\beta_1$ (or a decrease of β_1) in y, whilst in (5.12) it is a β_2 increase in y. This is because they are conditioned differently. In (5.11) adjustment is for z_2, whilst in (5.12) it is for z_1 [46].

Life course plots, introduced by Cole [130], are a graphical device which can help disentangle the effects of both the size and growth components of the anthropometric variable through time. The coefficients from the multivariable linear regression (5.10) (z_1 and z_2) are plotted against the corresponding ages at measurement, with connecting lines between the coefficients. The life course plot can be easily extended to include more than two occasions of measurement [46], with the regression coefficients plotted and connected in the same manner.

Life course plots show the effect of size in terms of the mutually adjusted regression coefficients at different ages. In addition, the difference between pairs of coefficients is proportional to the size of the regression coefficient for growth between the two corresponding ages [46]. The most important function of the life course plot is to emphasise the dual nature of size and growth, so that both appear on the same graph [130].

5.3.2 Logistic regression

It is often the case that the outcome variable in an analysis is measured on a dichotomous scale. An example of this would be an assessment of whether an individual is overweight or not. The logistic regression model has become, in many fields, the standard method of analysis in this situation [131].

Consider a dichotomous outcome variable y and p explanatory variables, $x_1, \ldots, x_p$, which may again be continuous, dichotomous or categorical. Let x represent the set of explanatory variables,

$$x = \{x_1, \ldots, x_p\}.$$

Define $\pi(x)$ to be the expected value of y given the observed values of x or, equivalently, the probability of y being equal to 1 given the observed values of x,

$$\pi(x) = E(y|x) = P(y = 1|x). \tag{5.13}$$

Allow $\pi(x)$ to be represented by the *logistic regression model*

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p}} \tag{5.14}$$

and define $g(x)$ to be the *logit function*

$$g(x) = \log\left(\frac{\pi(x)}{1-\pi(x)}\right). \tag{5.15}$$

The function $g(x)$ is thus the logarithm of the *odds* of y taking value 1 given the observed values of x, where

$$\text{odds}(y=1|x) = \frac{P(y=1|x)}{P(y=0|x)}. \tag{5.16}$$

Substituting (5.14) into (5.15) it can be seen that

$$g(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p.$$

Thus $g(x)$ is similar to the multivariable linear regression model (5.7), sharing many of its desirable properties such as being linear in its parameters and being able to take any value between $-\infty$ and ∞ [131].

The *odds ratio* (OR) is the ratio of the odds of the event of interest happening in an 'exposed' group to the odds of the event of interest happening in an 'unexposed' group [116]. When the explanatory variable (say x_j) is dichotomous then the OR compares the odds in the two levels of the variable,

$$\text{OR}(x_j) = \frac{\text{odds}(y=1|x_j=1)}{\text{odds}(y=1|x_j=0)}. \tag{5.17}$$

Similarly, if x_j is a dummy indicator variable corresponding to a categorical variable then the OR compares the odds in that category to the odds in the baseline category. If, however, x_j is continuous then the OR relates to the change in odds due to a unit increase in x_j,

$$\text{OR}(x_j) = \frac{\text{odds}(y=1|x_j=a+1)}{\text{odds}(y=1|x_j=a)}. \tag{5.18}$$

It can be shown [131] using (5.17) or (5.18), (5.16), (5.13) and (5.14) that

$$\text{OR}(x_j) = e^{\beta_j}, \tag{5.19}$$

when all the other explanatory variables (x_k, $k \neq j$) are kept constant. Due to their ease of interpretation, ORs are usually the parameters of interest in a logistic regression analysis rather than the regression coefficients themselves. The simple relationship (5.19) is the fundamental reason why logistic regression has proven to be such a powerful analytic research tool [131].

As with multivariable linear regression, the estimated logistic regression coefficients (and hence the ORs) are mutually adjusted, meaning that potential confounding factors can be accommodated through inclusion in the logistic regression model. Effect modification can also be assessed through the introduction of interaction terms to the regression model.

The logistic regression model is generally fitted using *maximum likelihood estimation* (MLE).

Matrix notation

Let y be a dichotomous outcome variable and x_j, $j = 1, \ldots, p$, be p explanatory variables. Let y_i and x_{ij} be, respectively, the observed values of y and x_j for subject i, $i = 1, \ldots, m$. Let

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & \mathbf{x}_1^T \\ \vdots & \vdots \\ 1 & \mathbf{x}_m^T \end{pmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}$$

where

$$\mathbf{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}.$$

Then the *general logistic regression model* can be written as

$$\text{logit}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, \tag{5.20}$$

which is the same form as the general linear model (5.8) but with logit($\mathbf{y}$) as the outcome rather than $\mathbf{y}$. The MLE estimator of $\boldsymbol{\beta}$ is then given by (5.9).

5.3.3 Mixed models

Hierarchical data, whereby members of *clusters* are likely to be more similar to each other than to other members of the population, were introduced in Section 5.1.1. The statistical methods for relating longitudinal data in childhood to a later outcome discussed so far (linear and logistic regression) rely on subjects being independent of one another, so their use with hierarchical data could lead to bias. *Mixed models* (also known as *random effect*, *multilevel* and *hierarchical models*), however, provide an extensive and flexible class of models suitable for handling such data [132].

Mixed models allow data to be viewed as a series of levels nested within one another to form a hierarchy. Explicitly defining the structure in this way as part of the modelling process enables the influences of variables at different levels to be examined and the induced clustering effects to be correctly accounted for.

Random intercepts model

Consider a study of school children who belong to different classes in a school, where a continuous outcome variable y (for example, BMI at age 11 years) and a single explanatory variable x (for example, BMI at age 5 years) are observed for each child. Let y_{ij} and x_{ij} be, respectively, the observed values of y and x for subject j, $j = 1, \ldots, n_i$, in class i, $i = 1, \ldots, m$. Then the *random intercepts* linear model, representing the simplest mixed model approach, is given by

$$y_{ij} = (\beta_0 + u_i) + \beta_1 x_{ij} + \varepsilon_{ij} \tag{5.21}$$

where $u_i \sim N(0, \sigma_u^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ are both independent and identically distributed, and all u_i are independent of all ε_{ij}.

Now $\beta_0 + \beta_1 x_{ij}$ gives the 'average' relationship between the outcome y and the explanatory variable x, with β_0 and β_1 referred to as the *fixed effects* or *fixed parameters*. The u_i are the class-specific *random effects* (or *level-2 residuals*) and the ε_{ij} are the *level-1 residuals*, both modelled as random draws from normally distributed random variables. As (5.21) contains both fixed and random effects, it is known as a *mixed model*.

Fitting an ordinary linear regression line to the data, ignoring their hierarchical nature, would give a biased estimate of the true relationship. The fitting of a mixed model allows the structure of the data to be explicitly accounted for.

The class-specific regression lines estimated by the mixed model draw strength from the mean regression line, with classes with fewer observations drawing greater strength [132]. In this way, mixed models can be used to handle missing (or sparse) data.

Intra-class correlation The random intercepts model allows for within-class correlation. The covariance between the observed outcome y for two subjects, j and j', $j \neq j'$, in class i is

$$\text{cov}(y_{ij}, y_{ij'}) = \text{var}(u_i) = \sigma_u^2$$

and the variance for an observed outcome y_{ij} for subject j in class i is

$$\text{var}(y_{ij}) = \text{var}(u_i) + \text{var}(\varepsilon_{ij}) = \sigma_u^2 + \sigma_\varepsilon^2,$$

resulting in a correlation coefficient of

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}. \tag{5.22}$$

This is more generally referred to in mixed modelling as the *intra-class correlation* and is a measure of how much more similar a subject is to others in their cluster than to individuals outside their cluster.

Random intercepts and slopes model

The random intercepts model, with the relationship in each class being restricted to linearity and to taking the same slope as in every other class, is often insufficient to study accurately the relationships inherent in the data. One natural extension is to allow each class to have their own slope in addition to their own intercept, creating a *random intercepts and slopes* linear model, given by

$$y_{ij} = (\beta_0 + u_{0i}) + (\beta_1 + u_{1i})x_{ij} + \varepsilon_{ij}, \tag{5.23}$$

where the terms u_{0i}, u_{1i} and ε_{ij} are considered as random variables with $u_{0i} \sim N(0, \sigma_{u_0}^2)$, $u_{1i} \sim N(0, \sigma_{u_1}^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. Now u_{0i} and $u_{0i'}$ are independent of each other for $i \neq i'$, u_{1i} and

$u_{1i'}$ are independent of each other for $i \neq i'$, ε_{ij} and $\varepsilon_{i'j'}$ are independent of each other unless $i = i'$ *and* $j = j'$, u_{0i} and $\varepsilon_{i'j}$ are independent of each other for all i, i' and j, and u_{1i} and $\varepsilon_{i'j}$ are independent of each other for all i, i' and j. However, u_{0i} and u_{1i} may be correlated, with $\text{cov}(u_{0i}, u_{1i}) = \sigma_{u_0 u_1}$.

Further extensions

Further extensions to the mixed model can include allowing additional explanatory variables to have random effects (giving estimates of class-specific effects for the variable), adding further levels to the hierarchy, incorporating nonlinear relationships, and including multivariate responses.

Matrix notation

Let y be a continuous outcome variable and x be a single explanatory variable. Let y_{ij} and x_{ij} be, respectively, the observed values of y and x for subject j, $j = 1, \ldots, n_i$, in 'cluster' i, $i = 1, \ldots, m$. Let

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_i \\ \vdots \\ \mathbf{y}_m \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},$$

$$\mathbf{Z} = \begin{pmatrix} \mathbf{1}_{n_1 \times 1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2 \times 1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_m \times 1} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{pmatrix}$$

where

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in_i} \end{pmatrix}$$

with

$$E\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \quad \text{and} \quad \text{Cov}\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{pmatrix} = \begin{pmatrix} \sigma_u^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I} \end{pmatrix}.$$

Then the random intercepts model (5.21) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}. \tag{5.24}$$

Here, $\boldsymbol{\beta}$ are the fixed effects and $\mathbf{u}$ and $\boldsymbol{\varepsilon}$ are random effects, both are assumed to be Normally distributed. The matrices $\mathbf{X}$ and $\mathbf{Z}$ are design matrices. $\text{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ summarises the fixed component of the model, $\mathbf{Zu}$ describes the between-subject random effects and $\boldsymbol{\varepsilon}$ the within-subjects random effects.

General linear mixed model

Indeed, *any* mixed model for Normal responses can be expressed in the form (5.24) with

$$E\begin{pmatrix}\mathbf{u}\\ \boldsymbol{\varepsilon}\end{pmatrix} = \begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix} \quad \text{and} \quad \text{Cov}\begin{pmatrix}\mathbf{u}\\ \boldsymbol{\varepsilon}\end{pmatrix} = \begin{pmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{pmatrix}.$$

This is referred to as the *general linear mixed model* representation [129]. $\mathbf{G}$ and $\mathbf{R}$ denote the variance-covariance matrices for $\mathbf{u}$ and $\boldsymbol{\varepsilon}$ respectively.

It can now be seen that the general linear model (5.8) is just a special case of the general linear mixed model (5.24) with $\mathbf{Z} = \mathbf{0}$.

Best linear unbiased prediction (BLUP) Estimation of $\boldsymbol{\beta}$, prediction of $\mathbf{u}$, and estimation of the parameters in $\mathbf{G}$ and $\mathbf{R}$ in the general linear mixed model (5.24) can be obtained via the notion of *best linear unbiased prediction* (BLUP). Estimates are *linear* in the sense that they are linear functions of the data, $\mathbf{y}$, *unbiased* in the sense that the average value of the estimate is equal to the average value of the quantity being estimated, *best* in the sense that they have minimum mean squared error within the class of linear unbiased estimators, and *predictors* to distinguish them from estimators of fixed effects [133].

The BLUP solutions for $\boldsymbol{\beta}$ and $\mathbf{u}$ can be shown [133] to be

$$\text{BLUP}(\boldsymbol{\beta}) = \tilde{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \tag{5.25}$$

and

$$\text{BLUP}(\mathbf{u}) = \tilde{\mathbf{u}} = \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \tag{5.26}$$

where $\mathbf{V} = \text{cov}(\mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$.

One derivation of the BLUP solutions [133] additionally assumes that $\mathbf{u}$ and $\boldsymbol{\varepsilon}$ are normally distributed and leads to the BLUP criterion

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \mathbf{u}^T\mathbf{G}^{-1}\mathbf{u}. \tag{5.27}$$

From this the BLUP of $(\boldsymbol{\beta}, \mathbf{u})$ can also be written as [134]

$$\begin{pmatrix}\tilde{\boldsymbol{\beta}}\\ \tilde{\mathbf{u}}\end{pmatrix} = (\boldsymbol{\gamma}^T\mathbf{R}^{-1}\boldsymbol{\gamma} + \mathbf{B})^{-1}\boldsymbol{\gamma}^T\mathbf{R}^{-1}\mathbf{y}$$

with fitted values

$$\text{BLUP}(\mathbf{y}) = \boldsymbol{\gamma}(\boldsymbol{\gamma}^T\mathbf{R}^{-1}\boldsymbol{\gamma} + \mathbf{B})^{-1}\boldsymbol{\gamma}^T\mathbf{R}^{-1}\mathbf{y} \tag{5.28}$$

where $\boldsymbol{\gamma} = (\mathbf{X}\mathbf{Z})$ and $\mathbf{B} = \begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}^{-1}\end{pmatrix}$.

Estimation of covariance matrices The BLUPs of $\boldsymbol{\beta}$ and $\mathbf{u}$ given in (5.25) and (5.26) depend on $\mathbf{G} = \text{cov}(\mathbf{u})$ and $\mathbf{R} = \text{cov}(\boldsymbol{\varepsilon})$, either directly, indirection through $\mathbf{V} = \text{cov}(\mathbf{Zu}+\boldsymbol{\varepsilon}) = \mathbf{ZGZ}^T+\mathbf{R}$, or both.

The parameters in these covariance matrices are typically estimated via maximum likelihood (ML) or *restricted maximum likelihood* (REML). The main advantage of REML over ML is that REML takes into account the degrees of freedom for the fixed effects in the model. For small sample sizes REML is expected to be more accurate than ML, but for large samples there is little difference between the two approaches [134].

In practice, the BLUPs of $\boldsymbol{\beta}$ and $\mathbf{u}$ given in (5.25) and (5.26) are usually replaced by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y}$$

and

$$\hat{\mathbf{u}} = \hat{\mathbf{G}}\mathbf{Z}^T\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

where $\hat{\mathbf{G}}$ and $\hat{\mathbf{V}}$ are obtained by plugging in the estimates of their parameters.

5.4 Two-stage analysis

In the single-stage analysis described in Section 5.3 the raw longitudinal anthropometric measurements in childhood are related directly to the distal outcome. However, this approach is confined to datasets in which the longitudinal data are *balanced*, as defined in Section 5.1.2.1. As many datasets from observational studies are in fact *unbalanced*, it is important to consider alternative modelling approaches.

One obvious approach would be to create balanced data out of the unbalanced data by deriving values for the anthropometric variables at common time points for each subject, which could be achieved via linear interpolation between the observed data points. Once values are defined at common time points then the single-stage approaches described in Section 5.3 can be utilised in exactly the same way as previously.

However, linear interpolation is effectively just the simplest example of a fitted *growth model* for each individual, which could take a variety of forms. Indeed, modelling each subject's growth in this manner need not just be for the purpose of deriving estimates of the anthropometric variable at common time points. Alternative features of the growth curve, such as turning points and ages at maximum velocities and accelerations, can also be derived given a suitable fitted model. These can then be used as exposures and related to later outcomes.

Clearly this type of analysis may also be of interest when dealing with balanced longitudinal data. Indeed, issues such as collinearity between the anthropometric measurements at different ages and missing data, which may affect balanced longitudinal data, may also be addressed via the fitting of growth curves.

Thus the general two-stage analysis approach becomes one whereby in the first stage 'growth' through childhood is modelled for each individual using the longitudinal anthropometric measurements. 'Growth features' are then derived from the growth curves and related to the later outcome in the second stage.

Alternative approaches for the modelling of longitudinal growth data are described in Section 5.4.1, then in Section 5.4.2 methods for relating the derived growth features to the later outcome are discussed.

5.4.1 Modelling growth

There are many existing models for human growth which are commonly used, differing in which anthropometric variables they can describe and over what range of ages. As alternatives to these, more general statistical modelling approaches can also be employed.

Models which have been developed specifically to describe growth are not addressed here as they are discussed in detail in Section 6.1.1. However, the more general statistical models are introduced. The parametric approaches of polynomial (Section 5.4.1.1) and fractional polynomial (Section 5.4.1.2) modelling are briefly discussed, as well as the use of mixed models (Section 5.4.1.3) in the context of growth modelling. Two nonparametric methods are also introduced: smoothing splines (Section 5.4.1.4) and regression splines (Section 5.4.1.5).

5.4.1.1 Polynomials

Polynomials can represent a wide variety of curve shapes, so have often been used in the modelling of growth. Polynomial growth models can be fitted for an individual in a similar manner to the linear regression model in Section 5.3.1. Whilst in the linear regression model a later outcome is modelled as a function of an anthropometric variable observed at a set of common time points *across* individuals, in subject-specific polynomial growth curves the anthropometric variable is modelled as a function of the ages at which it is observed (raised to a set of exponents) *within* an individual.

More specifically, let y be a continuous anthropometric variable. For subject i, $i = 1, \ldots, m$, consider the n_i observations of y, y_{ij}, $j = 1, \ldots, n_i$, made at age x_{ij}. Then the *degree p polynomial model* for subject i is given by

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \ldots + \beta_p x_{ij}^p + \varepsilon_{ij}, \tag{5.29}$$

where the ε_{ij} are are independent and identically distributed with $\varepsilon_{ij} \sim N(0, \sigma^2)$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$.

Matrix notation

Let

$$\mathbf{y} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{i1} & x_{i1}^2 & \dots & x_{i1}^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{in_i} & x_{in_i}^2 & \dots & x_{in_i}^p \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{i0} \\ \vdots \\ \varepsilon_{in_i} \end{pmatrix}$$

where y_{ij}, x_{ij} and ε_j are defined as above. Then (5.29) can again be written in the general linear model form (5.8),

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

5.4.1.2 Fractional polynomials

Although conventional polynomials, as described in Section 5.4.1.1, are a popular modelling tool, low degree polynomials are severely limited in their range of curve shapes and higher degree polynomials often produce undesirable artifacts, such as 'edge effects' and 'waves'. *Fractional polynomials* (FPs), introduced by Royston and Altman [135], extend the range of models afforded by conventional polynomials by allowing parameters to also take fractional powers.

Let y be a continuous anthropometric variable. For subject i, $i = 1, \dots, m$, consider the n_i observations of y, y_{ij}, $j = 1, \dots, n_i$, made at age x_{ij}. Then a FP of degree m with powers $p_1, \dots, p_m$ for subject i is defined as

$$y_{ij} = \beta_1 x_{ij}^{p_1} + \beta_2 x_{ij}^{p_2} + \dots + \beta_m x_{ij}^{p_m}$$

where, by convention, x_{ij}^0 is defined to be $\log(x_{ij})$. As a result, all values of x_{ij} must be greater than zero.

If one or more power in the model is duplicated then the model will include 'repeated powers'. A FP of degree m with m powers of p is defined as

$$y_{ij} = \beta_1 x_{ij}^p + \beta_2 x_{ij}^p \log(x_{ij}) + \dots + \beta_m x_{ij}^p (\log(x_{ij}))^{m-1},$$

though a general FP may include some unique and some repeated powers.

The powers are chosen from a predetermined set, usually taken to be $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$. Whilst entirely feasible, there has been found to be little advantage in adding intermediate fractional powers to this set [136].

Estimation of the best fitting FP for a given dataset involves both a systematic search for the best power or combination of powers from the permitted set and estimation of the associated parameter coefficients. This selection process includes fitting a model for each combination of powers in the permitted set. This means, for example, that fitting a FP of degree 2 using the standard set detailed above would involve fitting a different model for each of the 36 permissable combinations of powers. From these models the one with lowest deviance is chosen to be optimal.

FPs include many useful curves and can include features such as asymptotes and single points of inflection. They give at least as good a fit to data as a conventional polynomial of corresponding degree and often offer a better fit than conventional polynomials of higher degree.

However, polynomials, even when extended to include FPs, have limitations as a method of producing smooth growth curves. Their fitting is global rather than local so that changes in coefficient values to improve the fit of the curve at one age may have unwanted effects at other ages [117].

5.4.1.3 Mixed models

Mixed models, introduced in Section 5.3.3 as a method to incorporate data structure into a single-stage analysis, are also a useful tool for modelling longitudinal growth data [137]. Instead of it being responses *within a group* of individuals which are likely to be more similar (more highly correlated), in the modelling of longitudinal growth data it is the anthropometric measurements *within an individual.*

Random intercepts model

Consider a longitudinal study of an anthropometric dimension y which is measured repeatedly in a sample of m children. Let y_{ij} be the observed measurement for subject i, $i = 1, \ldots, m$, at age x_{ij}, $j = 1, \ldots, n_i$. Now time is the level-1 variable and subjects are the level-2 variable in the hierarchy, whereas previously (in Section 5.3.3) subjects were the level-1 variable.

The random intercepts linear model is again given by (5.21), but now $\beta_0 + \beta_1 x_{ij}$ gives the 'population average' growth trajectory. The parameters β_0 and β_1 are again fixed effects and the u_i are now *subject*-specific random effects, which model the deviation of the growth curve of subject i from the population average growth curve.

Intra-class correlation The intra-class correlation (5.22) now provides a measure of the degree to which a measurement for an individual is more similar to their own other measurements than to those for other people [138].

Random intercepts and slopes model

Allowing each subject to have their own slope in addition to their own intercept gives the random intercepts and slopes linear model (5.23). The parameters are as defined in Section 5.3.3 but u_{1i} is now the *subject*-specific slope.

Further extensions

Mixed models for longitudinal growth data can be extended in the same way as detailed in Section 5.3.3. One particularly fruitful advance has been the incorporation of smoothing methods into the mixed model framework. This is examined in Section 5.4.1.5.

When measurements for an individual are taken sufficiently close together in time, then the assumption of independence among the ε_{ij} may not hold. This can be dealt with via the explicit modelling of the autocorrelation structure [139].

Matrix notation

When used to model individual growth curves, the mixed model can again be written in the general linear mixed model form (5.24), with matrices as defined in Section 5.3.3.

5.4.1.4 Smoothing splines

An alternative to the parametric modelling approaches discussed thus far is provided by the related nonparametric approaches of smoothing splines, discussed in this section, and regression splines, discussed in Section 5.4.1.5.

Let y be a continuous anthropometric variable. For subject i, $i = 1, \ldots, m$, consider the n_i observations of y, y_{ij}, $j = 1, \ldots, n_i$, made at age x_{ij}. Suppose that a growth curve g is fitted to the longitudinal growth data of subject i. Then the goodness of fit of g can be assessed via the residual sum of squares

$$\sum_{j=1}^{n_i} \{y_{ij} - g(x_{ij})\}^2. \tag{5.30}$$

Smoothing splines use a *roughness penalty* approach to quantify the 'roughness' of a fitted curve and examine the trade-off between this and the goodness of fit of the curve. One widely used method of quantifying the roughness of a twice-differentiable curve g, a function of x defined on the interval $[a, b]$, is to calculate its integrated squared second derivative,

$$\int_a^b \{g''(x)\}^2 dt. \tag{5.31}$$

Now suppose that $x_{i1}, \ldots, x_{in_i}$ lie in the interval $[a, b]$ and satisfy $a < x_{i1} < \ldots < x_{in_i} < b$. Given a smoothing parameter $\alpha > 0$, the goodness of fit (5.30) and the roughness penalty (5.31) can be combined to give the *penalised sum of squares* [140]

$$\sum_{j=1}^{n_i} \{y_{ij} - g(x_{ij})\}^2 + \alpha \int_a^b \{g''(x)\}^2 dx, \tag{5.32}$$

with the *penalised least squares estimator* $\hat{g}$ defined to be the minimiser of (5.32) over all twice-differentiable functions g.

The smoothing parameter α represents the rate of exchange between residual error and local variation [141]. For a given α, $\hat{g}$ will be the 'best' compromise between smoothness and goodness of fit. Large α emphasises the roughness penalty term in (5.32), leading to little curvature in $\hat{g}$. As α tends to infinity the roughness penalty term dominates (5.32), so $\hat{g}$ will approach the linear regression fit. Small α emphasises the residual sum of squares term in (5.32), leading to a $\hat{g}$ which follows the meanders of the data closely. Thus as α tends to zero the roughness penalty disappears from (5.32) and $\hat{g}$ will approach an interpolating curve.

It can be shown [140] that $\hat{g}$ is necessarily a *natural cubic spline* with knots at ages $x_{i1}, \ldots, x_{in_i}$, meaning that [140]

1. on each interval $(a, x_{i1}), (x_{i1}, x_{i2}), \ldots, (x_{in}, b)$, $\hat{g}$ is a cubic polynomial,

2. these polynomials fit together at the interval boundaries $x_{i1}, \ldots, x_{in_i}$ in such a way that $\hat{g}$ itself and its first and second derivatives are continuous at each knot x_{ij}, and hence on the whole of $[a, b]$, and

3. $\hat{g}$ is linear on the two extreme intervals $[a, x_{i1}]$ and $[x_{in_i}, b]$.

Cross-validation

Thus in order to obtain a fitted cubic smoothing spline growth curve for the data of subject i, the only parameter which needs to be specified is the smoothing parameter α. There have been a number of 'automatic' procedures proposed for choosing α, probably the best well known being *cross-validation* (CV). The basic principal is to leave the data points out one at a time, choosing the value of α for which the remaining data points best predict the missing data point. More formally, let g_α^{-ij} be the smoothing spline calculated from all the data pairs except (x_{ij}, y_{ij}), under a smoothing parameter value of α. The CV choice of α is then the value of α minimising the *cross-validation score* [141]

$$CV(\alpha) = \frac{1}{n_i} \sum_{j=1}^{n_i} \{y_{ij} - g_\alpha^{-ij}(x_{ij})\}^2.$$

Generalised cross-validation (GCV) is a modified form of cross-validation which has some computational advantages [140].

Equivalent degrees of freedom

Although when fitting nonparametric curves parameters do not arise in the same way as in the parametric equivalents, it is often desirable to obtain an indication of the effective number of parameters for a fitted spline. In parametric regression the number of fitted parameters, and thus the number of degrees of freedom, can be calculated as

$$\text{trace}(A)$$

where A is the hat matrix for the fitted curve. The nonparametric analogy of this is the *equivalent degrees of freedom* (EDF), defined as

$$\text{trace}(A(\alpha))$$

where $A(\alpha)$ is the hat matrix associated with spline smoothing with smoothing parameter α, often referred to as the *smoother matrix*. EDF allows direct comparison with polynomials fits as a spline with ν EDF summarises the data to about the same extent as a $(\nu - 1)$-degree polynomial [134]. Using EDF, as opposed to the smoothing paramenter α itself, may well provide a more intuitive way of specifying the 'complexity' of the fitted cure [140].

5.4.1.5 Regression splines

Regression splines are another nonparametric approach which can be used for the modelling of individual growth curves. Although related to the smoothing splines described in Section 5.4.1.5, they have practical advantages in certain circumstances.

Models, bases and knots

Again, let y be a continuous anthropometric variable. For subject i, $i = 1, \ldots, m$, consider the n_i observations of y, y_{ij}, $j = 1, \ldots, n_i$, made at age x_{ij}. Suppose that an individual growth curve for subject i is to be fitted.

Linear regression models If the relationship between y and x appears to be linear for subject i then the *simple linear regression model*,

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}, \tag{5.33}$$

may be thought suitable, where ε_{ij} are the residuals associated with the jth fitted value which are assumed to be independent realisations of a random variable with mean zero. The right hand side of the simple linear regression model can be obtained as a linear combination of the functions

$$1 \text{ and } x. \tag{5.34}$$

These functions are referred to as the *basis* for the simple linear regression model. Similarly, the basis for the *quadratic regression model*,

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \varepsilon_{ij},$$

is

$$1, \ x \text{ and } x^2.$$

Linear regression spline models In situations where parametric models are not sufficiently flexible to capture the shape of a curve, further functions can be added to the basis. One extension to the simple linear regression model of (5.33) would be to allow the model to have two differently sloped sections which meet at, say, κ. This model is a *linear regression spline model* with 1 *knot*. The basis for this model would be formed by adding an additional function to (5.34) which is 0 to the left of κ and positively sloped from κ onwards. Define

$$x_+ = \max(0, x). \tag{5.35}$$

Then this additional function can be written as $(x - \kappa)_+$, the basis for the linear regression spline model with 1 knot as

$$1,\ x \text{ and } (x-\kappa)_+,$$

and the model itself as

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_1(x_{ij} - \kappa)_+ + \varepsilon_{ij}.$$

The number of linear sections in the linear regression spline model, and hence the amount of detail it can represent, can be increased by increasing the number of knots. More generally, a *linear regression spline model* with K knots at $\kappa_1, \ldots, \kappa_K$ has basis

$$1,\ x,\ (x-\kappa_1)_+, \ldots,\ (x-\kappa_K)_+$$

and model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \sum_{k=1}^{K} u_k(x_{ij} - \kappa_k)_+ + \varepsilon_{ij}. \tag{5.36}$$

Higher degree regression spline models The fitting of linear regression spline models as given in (5.36) results in continuous piecewise linear functions, which is unlikely to be appropriate for the modelling of growth. *Quadratic regression spline models* include an additional x^2 term in the basis as well as replacing each $(x-\kappa_k)_+$ by $(x-\kappa_k)_+^2$. As the resulting function is piecewise quadratic it will have a continuous first derivative meaning a much smoother appearance than the linear spline model.

Clearly the degree of the regression spline model can be increased further, leading to the generalisation of a *regression spline model of degree p*, with basis

$$1, x, \ldots, x^p, (x-\kappa_1)_+^p, \ldots, (x-\kappa_K)_+^p,$$

referred to as the *truncated power basis of degree p*, and model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \ldots + \beta_p x_{ij}^p + \sum_{k=1}^{K} u_k(x_{ij} - \kappa_k)_+^p + \varepsilon_{ij}. \tag{5.37}$$

A regression spline of degree p will be continuous on $p-1$ derivatives, meaning that higher degree regression spline models become increasingly more smooth.

Penalised regression spline modelling

Penalised linear regression spline models Consider the linear spline model with K knots as given in (5.36) and let

$$\mathbf{y} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} (x_{i1} - \kappa_1)_+ & \dots & (x_{i1} - \kappa_K)_+ \\ \vdots & \ddots & \vdots \\ (x_{in_i} - \kappa_1)_+ & \dots & (x_{in_i} - \kappa_K)_+ \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \quad \text{and} \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_K \end{pmatrix}. \tag{5.38}$$

Now let

$$\boldsymbol{\gamma} = (\mathbf{XZ}) = \begin{pmatrix} 1 & x_{i1} & (x_{i1} - \kappa_1)_+ & \dots & (x_{i1} - \kappa_K)_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{in_i} & (x_{n_i} - \kappa_1)_+ & \dots & (x_{in_i} - \kappa_K)_+ \end{pmatrix}$$

$$\text{and} \quad \boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ u_1 \\ \vdots \\ u_K \end{pmatrix}. \tag{5.39}$$

Also define the *norm* of a vector $\mathbf{v}$, denoted $||\mathbf{v}||$, to be

$$||\mathbf{v}|| = \sqrt{\mathbf{v}^T\mathbf{v}}.$$

Then the OLS fit of the linear regression spline model for subject i can be written as

$$\hat{\mathbf{y}} = \boldsymbol{\gamma}\hat{\boldsymbol{\delta}}, \text{ where } \hat{\boldsymbol{\delta}} \text{ minimises } ||\mathbf{y} - \boldsymbol{\gamma}\boldsymbol{\delta}||^2. \tag{5.40}$$

As unconstrained fitting of $u_1, \dots, u_K$ will result in a 'wiggly' fit [134], a constraint such as $\sum_{k=1}^{K} u_k^2 < C$ for some constant C may be imposed. Letting

$$\mathbf{D} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times K} \\ \mathbf{0}_{K\times 2} & \mathbf{I}_{K\times K} \end{pmatrix}, \tag{5.41}$$

this minimisation problem can be written as

$$\text{minimise } ||\mathbf{y} - \boldsymbol{\gamma}\boldsymbol{\delta}||^2 \text{ subject to } \boldsymbol{\delta}^T\mathbf{D}\boldsymbol{\delta} \leq C.$$

Using Lagrange multipliers it can be shown [134] that this is equivalent to minimising

$$||\mathbf{y} - \boldsymbol{\gamma}\boldsymbol{\delta}||^2 + \lambda^2\boldsymbol{\delta}^T\mathbf{D}\boldsymbol{\delta} \tag{5.42}$$

for some $\lambda \geq 0$. This has the solution [134]

$$\hat{\boldsymbol{\delta}}_\lambda = (\boldsymbol{\gamma}^T\boldsymbol{\gamma} + \lambda^2\mathbf{D})^{-1}\boldsymbol{\gamma}^T\mathbf{y}$$

with fitted values given by

$$\hat{\mathbf{y}} = \boldsymbol{\gamma}(\boldsymbol{\gamma}^T\boldsymbol{\gamma} + \lambda^2\mathbf{D})^{-1}\boldsymbol{\gamma}^T\mathbf{y}. \tag{5.43}$$

The term $\lambda^2\boldsymbol{\delta}^T\mathbf{D}\boldsymbol{\delta}$ penalises fits that are not sufficiently smooth, so is referred to as the *roughness penalty*. The amount of smoothing is controlled by λ, the *smoothing parameter*. For $\lambda = 0$ the fitted model corresponds to the unconstrained case given in (5.40). As λ increases the fit becomes increasing less rough until, as λ approaches infinity, the least-squares linear regression line is approached.

Higher degree penalised regression spline models Consider now fitting the generalised regression spline model of degree p as given in (5.37) to the growth data of subject i. The vectors $\mathbf{y}$ and $\mathbf{u}$ remain the same as in (5.38) but now

$$\mathbf{X} = \begin{pmatrix} 1 & x_{i1} & \dots & x_{i1}^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{in_i} & \dots & x_{in_i}^p \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} (x_{i1}-\kappa_1)_+^p & \dots & (x_{i1}-\kappa_K)_+^p \\ \vdots & \ddots & \vdots \\ (x_{in_i}-\kappa_1)_+^p & \dots & (x_{in_i}-\kappa_K)_+^p \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix},$$

$$\boldsymbol{\gamma} = (\mathbf{XZ}) = \begin{pmatrix} 1 & x_{i1} & \dots & x_{i1}^p & (x_{i1}-\kappa_1)_+^p & \dots & (x_{i1}-\kappa_K)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{in_i} & \dots & x_{in_i}^p & x_{in_i}-\kappa_1)_+^p & \dots & (x_{in_i}-\kappa_K)_+^p \end{pmatrix}$$

$$\text{and} \quad \boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \\ u_1 \\ \vdots \\ u_K \end{pmatrix}.$$

Following a similar argument to above the fitted values can be shown to be [134]

$$\hat{\mathbf{y}} = \boldsymbol{\gamma}(\boldsymbol{\gamma}^T\boldsymbol{\gamma} + \lambda^{2p}\mathbf{D})^{-1}\boldsymbol{\gamma}^T\mathbf{y} \tag{5.44}$$

where now

$$\mathbf{D} = \begin{pmatrix} \mathbf{0}_{(p+1)\times(p+1)} & \mathbf{0}_{(p+1)\times K} \\ \mathbf{0}_{K\times(p+1)} & \mathbf{I}_{K\times K} \end{pmatrix}.$$

5.4.1.6 Regression splines as mixed models

Recall the *general linear mixed model* representation (5.24) and the *best linear unbiased predictor* (BLUP) criterion (5.27) given in Section 5.3.3. Considering again the linear regression spline model given in (5.36), suppose that $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}$. Because

$$\boldsymbol{\gamma}\boldsymbol{\delta} = (\mathbf{X}\mathbf{Z}) \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad \text{and} \quad \boldsymbol{\delta}^T \mathbf{D} \boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{pmatrix}^T \mathbf{D} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{pmatrix} = ||\mathbf{u}||^2,$$

(5.42) can be rewritten as

$$||\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}||^2 + \lambda^2 ||\mathbf{u}||^2.$$

Dividing this by σ_ε^2 gives

$$\frac{1}{\sigma_\varepsilon^2} ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}||^2 + \frac{\lambda^2}{\sigma_\varepsilon^2} ||\mathbf{u}||^2. \tag{5.45}$$

By treating $\mathbf{u}$ as a set of random coefficients with

$$\text{Cov}(\mathbf{u}) = \sigma_u^2 \mathbf{I} \quad \text{where} \quad \sigma_u^2 = \frac{\sigma_\varepsilon^2}{\lambda^2} \tag{5.46}$$

(5.45) becomes

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T (\sigma_\varepsilon^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \mathbf{u}^T (\sigma_u^2 \mathbf{I})^{-1} \mathbf{u}.$$

Setting $\mathbf{G} = \sigma_u^2 \mathbf{I}$ and $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}$ this becomes precisely the BLUP criterion (5.27). As a result, the penalised regression spline can be represented in the linear mixed model form (5.24), namely

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

with

$$\text{Cov} \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{pmatrix} = \begin{pmatrix} \sigma_u^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I} \end{pmatrix}. \tag{5.47}$$

This mixed model representation means that penalised regression spline models can be easily implemented using standard statistical software.

To illustrate the relationship between penalised regression splines and mixed models, consider the expression for the fitted values from the mixed model using the BLUP estimates of $\boldsymbol{\beta}$ and $\mathbf{u}$ as given in (5.28). Letting $\mathbf{G} = \sigma_u^2 \mathbf{I}$ and $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}$ as in (5.47), $\mathbf{D} = \begin{pmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times K} \\ \mathbf{0}_{K\times 2} & \mathbf{I}_{K\times K} \end{pmatrix}$ as in (5.41) and $\lambda^2 = \frac{\sigma_\varepsilon^2}{\sigma_u^2}$ as in (5.46), this becomes

$$\boldsymbol{\gamma} \left(\boldsymbol{\gamma}^T \boldsymbol{\gamma} + \lambda^2 \mathbf{D} \right)^{-1} \boldsymbol{\gamma}^T \mathbf{y},$$

which is precisely the expression for the fitted values from the penalised regression spline given in (5.43).

Subject-specific penalised regression splines

Penalised regression spline models have thus far been described as a method for fitting a growth curve to the longitudinal data for a single individual (the generic 'subject i'). Whilst curves could be fitted in this way for each subject in a dataset, penalised regression splines provide a far more succinct approach to obtaining subject-specific curves.

In Section 5.4.1.3 mixed models were described as a method for obtaining subject-specific growth curves. However, the type of curve shape available when using mixed models is restricted when only parametric modelling approaches are considered.

As penalised regression splines can be handled within the mixed model framework they can also be easily extended in this manner. This fusion between parametric mixed modelling and smoothing is referred to as *semiparametric mixed modelling* [134].

Consider the linear regression spline model of (5.36) with K knots at $\kappa_1, \ldots, \kappa_K$,

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \sum_{k=1}^{K} u_k (x_{ij} - \kappa_k)_+ + \varepsilon_{ij}, \tag{5.48}$$

where y_{ij} denotes the observed response for subject i, $i = 1, \ldots, m$, at time x_{ij}, $j = 1, \ldots, n_i$. This model can be extended via the inclusion of subject-specific random parameters which model the deviation of a given individual's curve from the population average curve. Whilst, in the simplest cases, random intercept or random slope terms could be introduced, given that the underlying population average function is a spline, in many instances it will be necessary for the subject-specific deviations from this to also be modelled as splines. So, for example, the linear regression spline model given by (5.48) can be extended to give

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \sum_{k=1}^{K} u_k (x_{ij} - \kappa_k)_+ + a_{i0} + a_{i1} x_{ij} + \sum_{k=1}^{K} v_{ik} (x_{ij} - \kappa_k)_+ + \varepsilon_{ij} \tag{5.49}$$

where $u_k \sim N(0, \sigma_u^2)$, $(a_{i0}, a_{i1})^T \sim N(0, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is an unstructured 2×2 covariance matrix, $v_{ik} \sim N(0, \sigma_v^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. Now u_k and $u_{k'}$ are independent of each other for $k \neq k'$, a_{i0} and $a_{i'0}$ are independent of each other for $i \neq i'$, a_{i1} and $a_{i'1}$ are independent of each other for $i \neq i'$, v_{ik} and $v_{i'k'}$ are independent of each other unless $i = i'$ *and* $k = k'$, ε_{ij} and $\varepsilon_{i'j'}$ are independent of each other unless $i = i'$ *and* $j = j'$, u_k and a_{i0} are independent of each other for all i and k, u_k and a_{i1} are independent of each other for all i and k, u_k and $v_{ik'}$ are independent of each other for all i, k and k', u_k and ε_{ij} are independent of each other for all i, j and k, a_{i0} and $a_{i'1}$ are independent of each other for $i \neq i'$, a_{i0} and $v_{i'k}$ are independent of each other for all i, i' and k, a_{i0} and $\varepsilon_{i'j}$ are independent of each other for all i, i' and j, a_{i1} and $v_{i'k}$ are independent of each other for all i, i' and k, a_{i1} and $\varepsilon_{i'j}$ are independent of each other for all i, i' and j. However, a_{i0} and a_{i1} may be correlated

The fitted subject-specific curve for each subject is now the sum of the linear regression spline population average curve and a further subject-specific linear regression spline which models the

deviation from this. All the subject-specific parameters, a_{i0}, a_{i1} and $v_{i1}, \ldots, v_{iK}$, are modelled as random effects with mean 0.

Letting

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_i \\ \vdots \\ \mathbf{y}_m \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},$$

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{X}_1 & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{Z}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{Z}_2 & \mathbf{0} & \mathbf{X}_2 & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{Z}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_m & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{X}_m & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{Z}_m \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_K \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{pmatrix}$$

where

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{pmatrix}, \quad \mathbf{Z}_i = \begin{pmatrix} (x_{i1} - \kappa_1)_+ & \cdots & (x_{i1} - \kappa_K)_+ \\ \vdots & \ddots & \vdots \\ (x_{in_i} - \kappa_1)_+ & \cdots & (x_{in_i} - \kappa_K)_+ \end{pmatrix},$$

$$\mathbf{a}_i = \begin{pmatrix} a_{i0} \\ a_{i1} \end{pmatrix}, \quad \mathbf{v}_i = \begin{pmatrix} v_{i1} \\ \vdots \\ v_{iK} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in_i} \end{pmatrix}$$

with

$$\mathbf{G} = \text{Cov}(\mathbf{u}) = \begin{pmatrix} \sigma_u^2 \mathbf{I} & 0 & 0 \\ 0 & (\text{blockdiag } \boldsymbol{\Sigma})_{1 \le i \le m} & 0 \\ 0 & 0 & \sigma_v^2 \mathbf{I} \end{pmatrix},$$

this model can again be fitted using the general linear mixed model form of (5.24).

The model (5.49) can be easily extended so that either the population average curve, the subject-specific deviations from this, or both, are of degree greater than one.

5.4.2 Relating derived growth parameters to later outcomes

In the first stage of the two-stage analysis approach, individual growth curves are fitted for the anthropometric variable of interest. From these fitted curves, estimates of the anthropometric

variable at common time points or 'growth features', such as turning points and ages at maximum velocities and accelerations, can be derived.

In the second stage, these derived explanatory variables are related to the distal outcome. The statistical approaches for this second stage differ little from the methods detailed for the single-stage analysis in Section 5.3. If the outcome variable is continuous then linear regression (Section 5.3.1) can be used. Likewise, if the outcome is dichotomous then logistic regression (Section 5.3.2) may be suitable. Again, if any further hierarchical structure is present in the data then it it is important to take account of this through mixed modelling (Section 5.3.3).

A two-stage analysis approach is used in two different analyses in this thesis. In the first (Chapter 8), the second stage involves relating late-adolescent body mass index (BMI) and percentage body fat (%BF) to derived features of the childhood BMI growth curve, particularly the location of the adiposity rebound (AR, see Section 2.3.3). Continuous values of late-adolescent BMI and %BF are related to the AR location using linear regression. The measurements of late-adolescent BMI and %BF are also used to define 'overweight' and 'overfat' status for each individual, which is related to the AR location using logistic regression.

In Chapter 9 a two-stage analysis approach is also used, in which BMI z-score observed later in childhood is related to the derived location of the adiposity peak (AP) seen in infancy. The dataset used in this instance includes sibling pairs — this is taken into account through the use of a mixed modelling approach.

5.5 Methodological overview

When studying relationships between childhood growth and later health, there are several key issues which must be considered:

- Data structure, in particular hierarchical structure, can create correlations between individuals. This should be taken into account when relating later health to childhood growth, for example through the use of mixed modelling.
- When the childhood growth data are balanced they can be subject to missing data, and when they are unbalanced they can be subject to the related issue of data sparsity. Either case can be addressed via the fitting of individual growth models as the first stage of a two-stage analysis approach. If the data are balanced then an alternative is to use MI.
- Repeated measures of childhood growth within an individual are likely to be correlated. This can lead to problems of collinearity if repeated measures are used in a distal outcome model. The fitting of individual growth models can again be used as a tool to overcome this.
- If individual growth models are to be used, there are a variety of different approaches.

It is the aim of this thesis to explore, develop and implement statistical methods to address these issues.

Fig. 5.1 and Fig. 5.2 provide schematic overviews of the statistical methods used in this thesis for, respectively, balanced and unbalanced childhood growth data. It should be emphasised that these diagrams do not try to include all potential analysis approaches which could be considered for a given scenario. Clearly there could be many further viable alternatives.

Whilst unbalanced data are inherently more difficult to deal with, it is, somewhat paradoxically, the balanced data diagram (Fig. 5.1) which is the more complex. This is because all the approaches which are available for unbalanced data can be used with balanced data, but there are also additional balanced data-specific approaches.

The following comments relate to the labels in Fig. 5.1, the diagram concerning balanced childhood growth data:

1. If the data are incomplete then MI, as described in Sect 5.2.4, may be used. If MI is used, the result is several multiply imputed datasets, which can be partitioned into childhood growth data and outcome variables. If MI is not used, the original raw data can be similarly partitioned into childhood growth data and outcome variables (hence the two paths labelled 'No' emanating from the 'Use multiple imputation?' decision node).

2. If derived features of growth, such as estimated values, velocities or ages at maxima or minima, are required then either the raw childhood growth data (if MI is not used) or the multiply imputed childhood growth data (if MI is used) are used in the 'Growth modelling' section of the diagram. This results in a two-stage analysis approach, as described in Section 5.4. If derived features of growth are not required, then the raw childhood growth data or the multiply imputed childhood growth data are used directly in the 'Distal outcome modelling' section. This is a single-stage analysis approach, as described in Section 5.3.

3. In the 'Growth modelling' section, individual growth curves are fitted to either the raw childhood growth data (if MI is not used) or to the multiply imputed childhood growth data (if MI is used). If there is an existing growth model which is adequate for the purpose, then this may be utilised. Otherwise a more general statistical approach, as described in Section 5.4.1, may be employed. Models may be developed within a mixed model framework, as described for linear models in Section 5.4.1.3 and for regression splines in Section 5.4.1.6, or fitted as entirely subject-specific curves. From the fitted growth models, the required growth features may be derived. These are then used in the 'Distal outcome modelling' section.

4. In the 'Distal outcome modelling' section, the outcome variables, which may be either raw or multiply imputed, are related to the explanatory variables of interest. In a single-stage analysis these will be either the raw or multiply imputed childhood growth data, and in a two-stage analysis these will be the derived growth features, which potentially also result from multiply imputed data. If no further data structure, for example of a hierarchical nature, needs to be taken into account, then simple regression models such as those described in Section 5.3.1 and Section 5.3.2 and referred to in Section 5.4.2 can be used. If further data

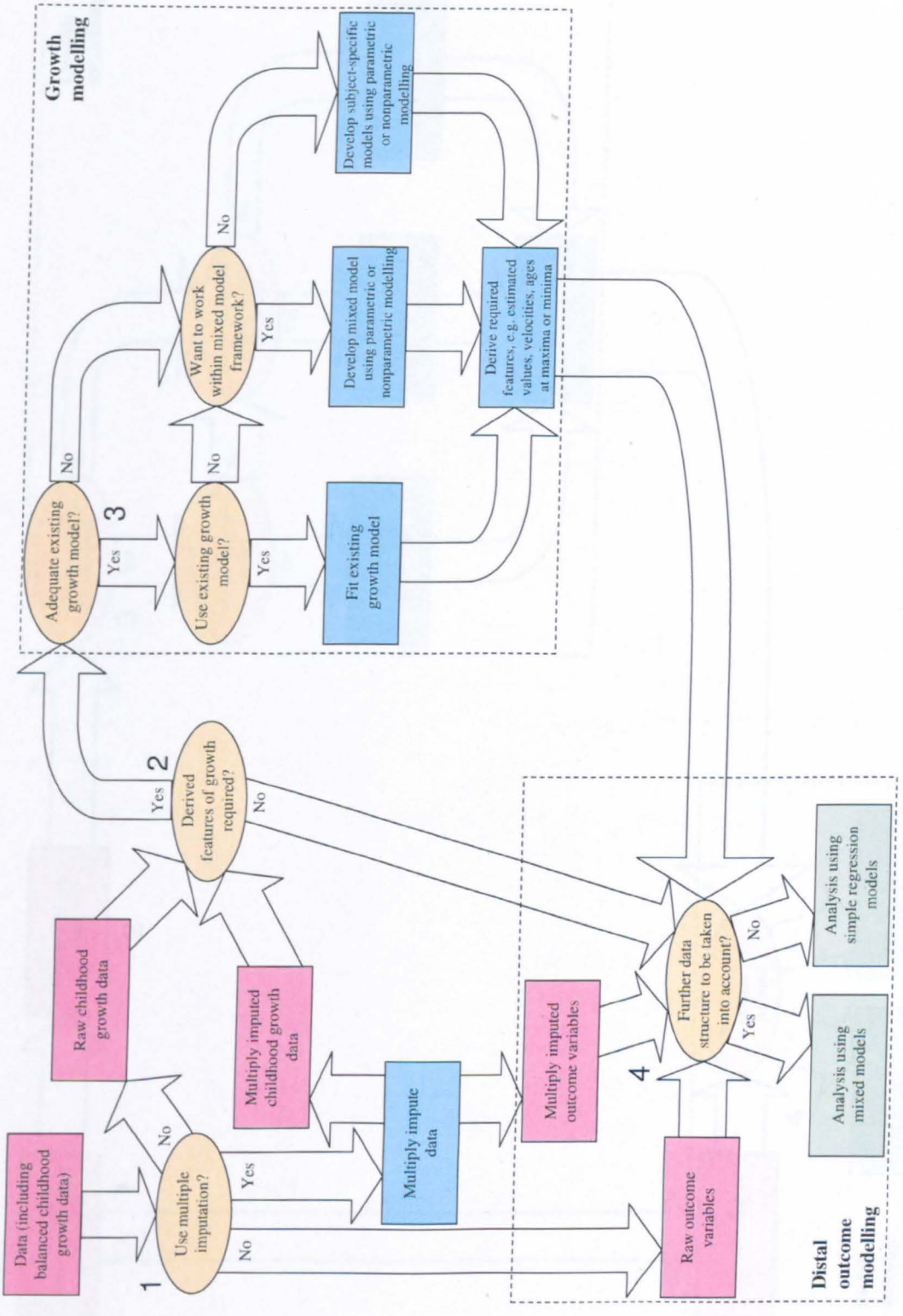

Fig. 5.1: Diagrammatic overview of statistical methods for balanced data.

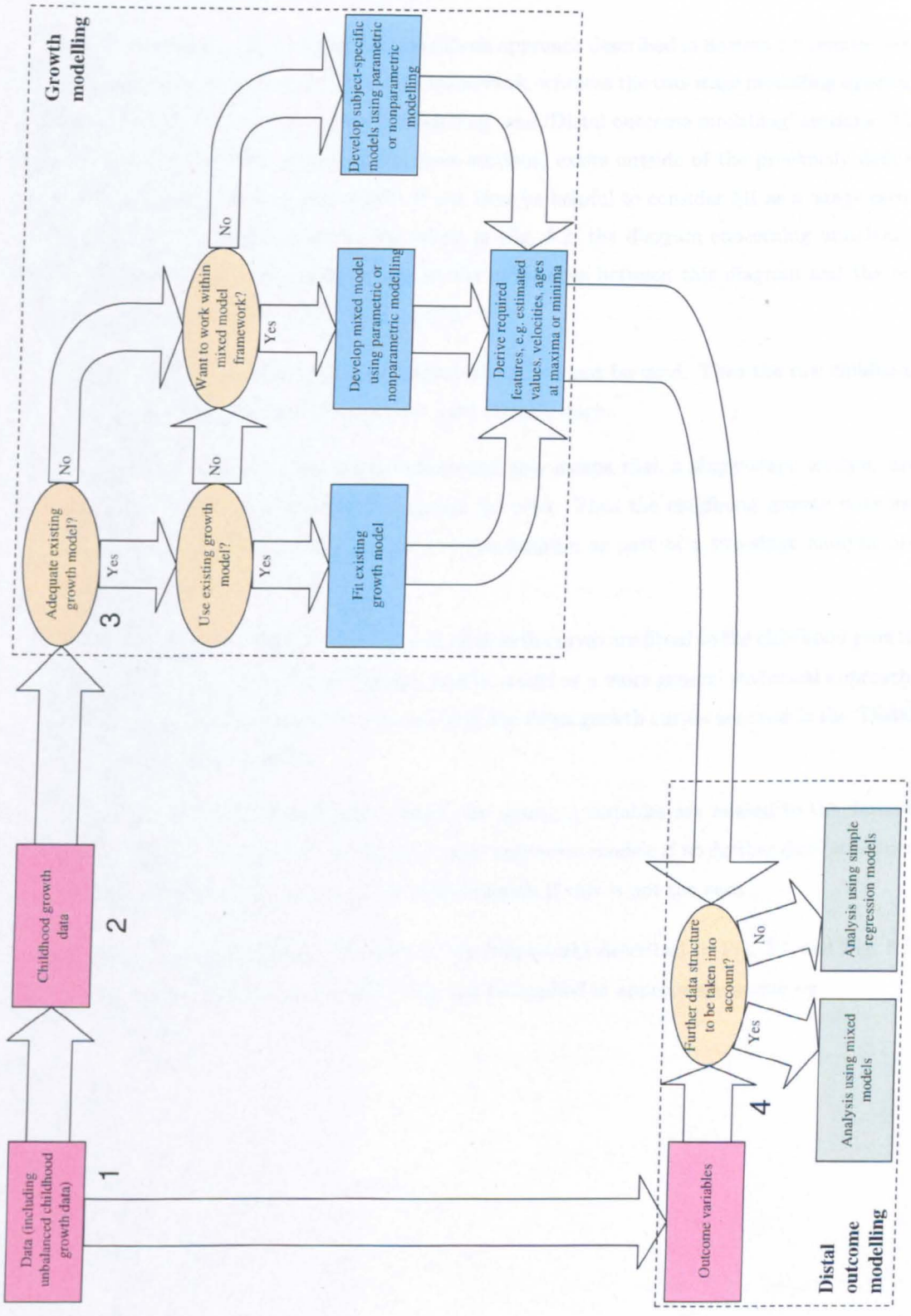

Fig. 5.2: Diagrammatic overview of statistical methods for unbalanced data.

structure does need to be taken into account then a mixed model approach, as described in Section 5.3.3 and referred to in Section 5.4.2, should be used.

Thus it can be seen that the single-stage analysis approach described in Section 5.3 requires only the 'Distal outcome modelling' section of the framework, whereas the two-stage modelling approach of Section 5.4 includes both the 'Growth modelling' and 'Distal outcome modelling' sections. The potential use of MI, which precedes both these sections, exists outside of the previously defined single- or two-stage modelling framework. It can thus be helpful to consider MI as a 'stage zero'.

The following comments relate to the labels in Fig. 5.2, the diagram concerning unbalanced childhood growth data, and describe some of the differences between this diagram and the one concerning balanced childhood growth data (Fig. 5.1):

1. As the childhood growth data are unbalanced, MI cannot be used. Thus the raw childhood growth data and outcome variables are used at each stage.

2. The childhood growth data being unbalanced also means that a single-stage analysis approach, as described in Section 5.3, cannot be used. Thus the childhood growth data are used in the 'Growth modelling' section of the diagram as part of a two-stage analysis approach, as described in Section 5.4.

3. In the 'Growth modelling' section, individual growth curves are fitted to the childhood growth data. This can again be via an existing growth model or a more general statistical approach. The required growth features are derived from the fitted growth curves are used in the 'Distal outcome modelling' section.

4. In the 'Distal outcome modelling' section, the outcome variables are related to the derived growth features. Again, this can be via simple regression models if no further data structure needs to be taken into account, or by mixed models if this is not the case.

In the remainder of this thesis, elements of the frameworks described in Fig. 5.1 and Fig. 5.2 are developed in more detail in order that they can be applied in appropriate scenarios.

Chapter 6

Subject-matter issues

This Chapter addresses two 'subject-matter issues', namely the modelling of growth (Section 6.1) and the standardisation of anthropometric variables into z-scores (Section 6.2).

Many different models for describing human growth have been developed over the last few decades, with varying degrees of success. These models often differ in the anthropometric variables and range of ages for which they can be used. Additionally to these very specific models are more general statistical modelling approaches, both parametric and nonparametric, which have also sometimes been used for modelling growth. In Section 6.1 these various models and modelling approaches are reviewed and illustrated.

Observations of anthropometric variables are often standardised to create *z-scores* or *SD scores*, as briefly introduced in Section 2.2.1. Calculated z-scores provide a measure of how many standard deviations (SDs) above or below the mean of some distribution the observed measurement lies. When considering a given anthropometric variable observed at two different ages, either within the same individual or across different individuals, a comparison of the measurements is difficult to interpret. This is because the distribution of the variable, and hence its expectation, is age-dependent. However, if both measurements are transformed onto the z-score scale using distributions which correspond to the age at which the measurements were taken, then the z-scores no longer have an age-dependent expectation. This makes a direct comparison much more meaningful. Issues surrounding the standardisation of anthropometric variables into z-scores are explored in Section 6.2.

6.1 Modelling growth

Models are often sought to reduce large amounts of growth data for an individual to a small number of parameters. Many different models have been suggested over the course of the last few decades for this purpose, differing in which anthropometric variables they describe and over what range of ages. Some have been developed specifically for modelling growth (Section 6.1.1), whilst others are more general statistical modelling approaches (Section 6.1.2).

The aim of this chapter is to provide a brief review of the most influential of these individual

growth models, along with examples of their fits to real data which are presented simply for illustrative purposes. The chapter is not intended to be a comprehensive review and no formal comparisons between the models are made. In particular, providing examples where the fit of a given model to a given set of data appears to be unsatisfactory is in no way intended to 'prove' the model to be inadequate.

Only individual growth models, as opposed to those intended to be fitted on populations (for example in the development of growth references), are discussed. Some of the latter are covered in relation to fetal growth by Silverwood and Cole [3] in the Appendix.

Besides obtaining a satisfactory fit to the data, desirable features of a growth model include simplicity of the fitting procedure, biological interpretability of the model parameters, and model parsimony [30], so these will be considered in what follows.

A further feature of interest is whether there is any subjectivity involved in the model fitting, for example by having to examine the data for an individual to determine over what range of ages a certain part of the model needs to be fitted. If input from the user is required in this manner for each subject then it impacts on the 'automatability' of the model, which is of obvious concern with larger datasets.

The data used in this chapter concern a selected group of subjects partaking in the Uppsala Family Study (UFS, see Section 4.2).

Because several of the models are intended to model height between birth and age 6 years, participants with relevant profiles over this period, referred to as 'Subject A' and 'Subject B', were selected. Subject A has more 'typical' observed height values, whilst Subject B displays slightly more unusual growth. In particular, Subject B has an observed height value at approximate age 3 years which is somewhat greater than may be expected. Whilst it is possibly the case that this data point is erroneous, perhaps as a result of measurement error, and thus should not be taken into account when modelling the height of the individual, it does remain within the bounds of biological plausibility, so its incorporation into the height model may be deemed important.

Where the models are suggested for use with weight as well as height, they are applied to 'Subject C' and 'Subject D'. Subject C is again a more 'typical' pattern of weight development from birth to age 6 years, whilst Subject D deviates from this somewhat. Given the consistency of this deviation seen in the observed weights, as opposed to the single anomalous height measurement seen in Subject B, this would appear to be the 'true' growth pattern, meaning that models should ideally be able to handle it.

Several of the models describe growth in height from birth or infancy right through to adult height. To illustrate these models an individual, 'Subject E', is selected who has a good coverage of data points throughout this period and a final observed height at a relatively late age.

To evaluate models that have been proposed for studying growth in BMI, particularly around the adiposity rebound (AR, see Section 2.3.3), two additional UFS participants, identified as 'Subject F' and 'Subject G', have been selected. Again, the former is a more 'typical' pattern of BMI

development around this period, whilst the latter is somewhat atypical.

6.1.1 Models developed specifically for growth

Many of the models in common use have been developed specifically for the purpose of modelling human growth, as opposed to being more general statistical modelling techniques. The most influential of these are discussed here in chronological order. The Jenss-Bayley and Berkey-Reed models cover only the first few years of life but have been suggested for use with multiple anthropometric measures. The Count, Bock-Thissen, Preece-Baines, Karlberg and JPPS models, on the other hand, describe height from birth or infancy right through to final adult height. The models also differ in their complexity and the number and interpretability of their parameters.

6.1.1.1 Jenss-Bayley

Because of the complexity of modelling the growth curve in its entirety, many early modelling attempts concentrated on shorter periods of the growth curve [142]. Jenss and Bayley [32] presented the first widely used model in 1937, describing either height or weight during the first 6 years of life. The Jenss-Bayley model is given by

$$y = \alpha_1 + \alpha_2 t - e^{\alpha_3 + \alpha_4 t} \tag{6.1}$$

where y is height or weight at time t, and α_1, α_2, α_3 and α_4 are the parameters to be estimated.

The exponential component in (6.1) accommodates the rapidly decelerating growth usually seen during infancy, then approaches the linear asymptote. After infancy the exponential component makes negligible contribution to the model so growth is effectively linear with growth velocity α_2.

One feature of the Jenss-Bayley curve is that the value of e^{α_4} gives a measure of the acceleration of growth at any point relative to the acceleration one unit of time prior to that. This is referred to as the 'growth constant' and it is independent of the scale used. Jenss and Bayley [32] suggest using this to compare the growth of different characteristics within the same child or across different children.

Berkey [30] fitted the Jenss-Bayley model to height and weight data for a sample of children from Boston and found it to be robust to variability in either the number or location of ages at which measurements are available. Mean residuals from both fitted models were found to be small, except perhaps at age 6 months when considering height.

Other applications of the Jenss-Bayley curve include modelling height and weight between birth and age 8 years in a study of US children [143], modelling height, weight and head circumference in the first year of life in a sample of Indian children [144], and within a mixed model framework to provide parameters describing growth in height for use in screening for Turner syndrome [145].

Although there are now many alternatives to the Jenss-Bayley model, it remains popular for modelling both height and weight in early life [146].

Fig. 6.1 shows fitted Jenss-Bayley height curves for Subject A and Subject B and Fig. 6.2 shows fitted Jenss-Bayley weight curves for Subject C and Subject D. Model fitting is carried out using the `nl` procedure in Stata [147], which allows the fitting of nonlinear functions using least squares regression. There is no subjectivity involved in the model fitting, making it an extremely simple procedure.

The fitted Jenss-Bayley height curve for Subject A can be seen to fit the data very well. For Subject B the fit is not quite so good, with the model appearing to overestimate height around age 1 to 2 years and the curve remaining virtually linear between age 2 and 4 years even though this results in a poor fit to the point at age 3 years. This illustrates the inflexible nature of the Jenss-Bayley curve.

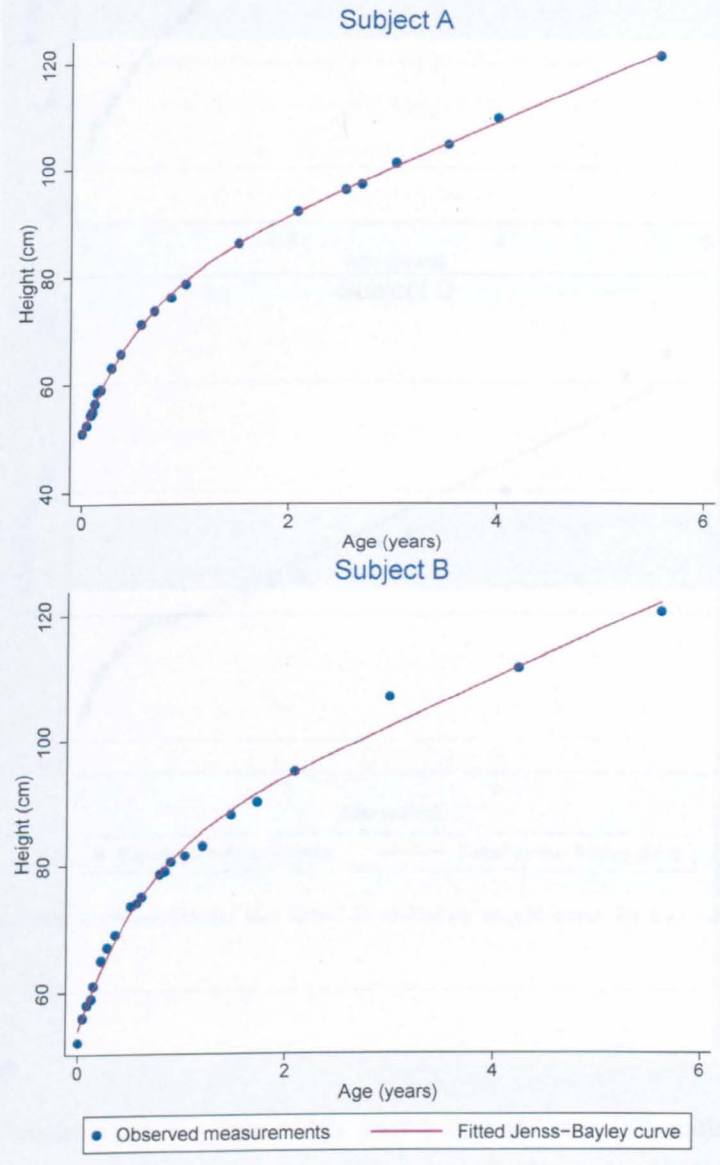

Fig. 6.1: Observed height measurements and fitted Jenss-Bayley height curve for two subjects in the Uppsala Family Study.

The fitted weight curve for Subject C in Fig. 6.2 again provides an excellent fit to the data.

For Subject D, however, the model systematically underestimates weight between age 6 months and 1 year, then overestimates it up to around age 5 years. Whilst this pattern of weight growth may be somewhat extreme, this again shows that the rigid form of the Jenss-Bayley model given by the combination of exponential and linear components may not always be appropriate.

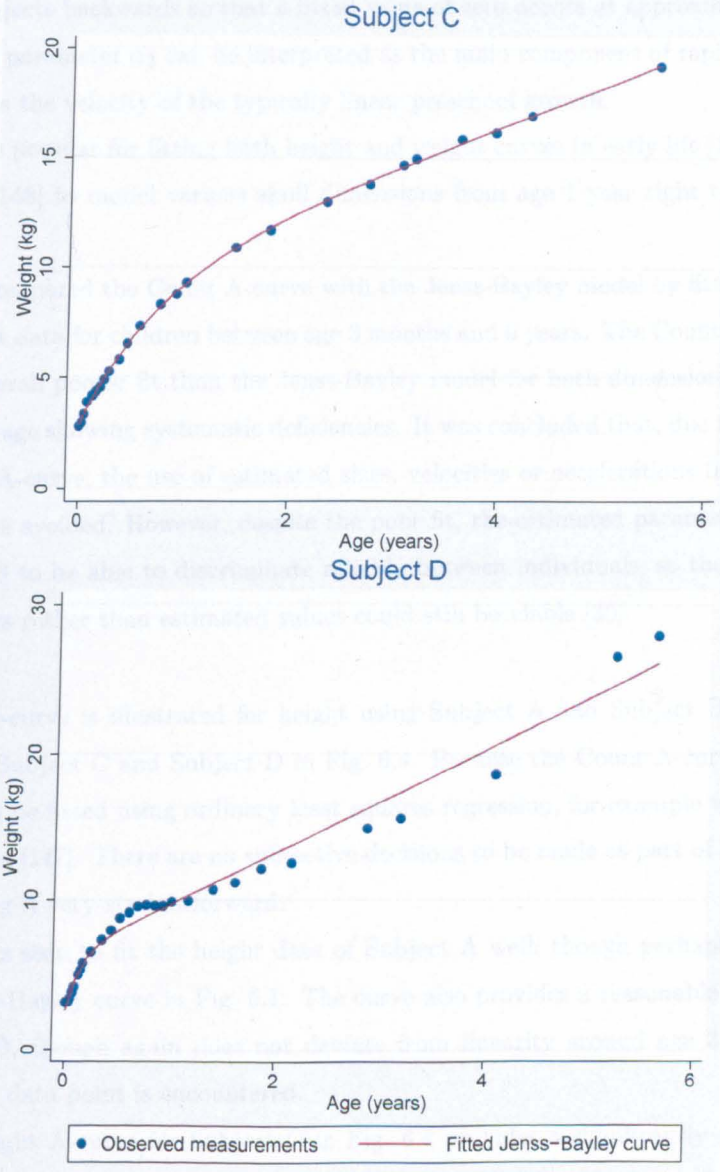

Fig. 6.2: Observed weight measurements and fitted Jenss-Bayley weight curve for two subjects in the Uppsala Family Study.

6.1.1.2 Count

Another growth model which has been widely used is that of Count [33], dating from 1943. Although presented as a model for height in three sections, often only the first, more generally applicable, of these (the 'A-curve') is used, with modelling restricted to the first 6 years of life after birth.

The A-curve (no relation to 'Subject A' in the present context) is the logarithmic curve given

by

$$y = \alpha_1 + \alpha_2 t + \alpha_3 \log t, \tag{6.2}$$

where y is the fitted dimension at time t, and α_1, α_2 and α_3 are the parameters to be estimated. The A-curve projects backwards so that a fitted value of zero occurs at approximately the time of conception. The parameter α_3 can be interpreted as the main component of rapid early childhood growth and α_2 as the velocity of the typically linear preschool growth.

This model is popular for fitting both height and weight curves in early life [146] and has been used by Count [148] to model various skull dimensions from age 1 year right through to age 16 years.

Berkey [30] compared the Count A-curve with the Jenss-Bayley model by fitting them to both height and weight data for children between age 3 months and 6 years. The Count model was found to provide an overall poorer fit than the Jenss-Bayley model for both dimensions, with the mean residuals at each age showing systematic deficiencies. It was concluded that, due to the inadequate fit of the Count A-curve, the use of estimated sizes, velocities or accelerations from the model at any age should be avoided. However, despite the poor fit, the estimated parameters of the Count model were found to be able to discriminate reliably between individuals, so that analyses based on the parameters rather than estimated values could still be viable [30].

The Count A-curve is illustrated for height using Subject A and Subject B in Fig. 6.3 and for weight using Subject C and Subject D in Fig. 6.4. Because the Count A-curve is linear in its parameters it can be fitted using ordinary least squares regression, for example with the `regress` procedure in Stata [147]. There are no subjective decisions to be made as part of the model fitting procedure, making it very straightforward.

The A-curve is seen to fit the height data of Subject A well, though perhaps marginally less so than the Jenss-Bayley curve in Fig. 6.1. The curve also provides a reasonable fit to the height data of Subject B, though again does not deviate from linearity around age 3 years when the anomalously high data point is encountered.

The fitted weight A-curve for Subject C in Fig. 6.4 provides a significantly poorer fit to the data than the equivalent Jenss-Bayley curve, with systematic underestimation of weight up to age 1 year, then overestimation up to age 4 years. The fitted curve for Subject D is also a poor fit, though similarly so to the Jenss-Bayley curve in Fig. 6.2.

When considering height throughout the period of growth, Count [33] advises the addition of a further two sections to the model. From approximately age 6 years to age 11 years (between 'first-molar time' and 'second-molar time', as Count describes it) there is a simple step-up of velocity which is accounted for by the 'B-curve',

$$y = \beta_1(\text{A-curve value}) + \beta_2,$$

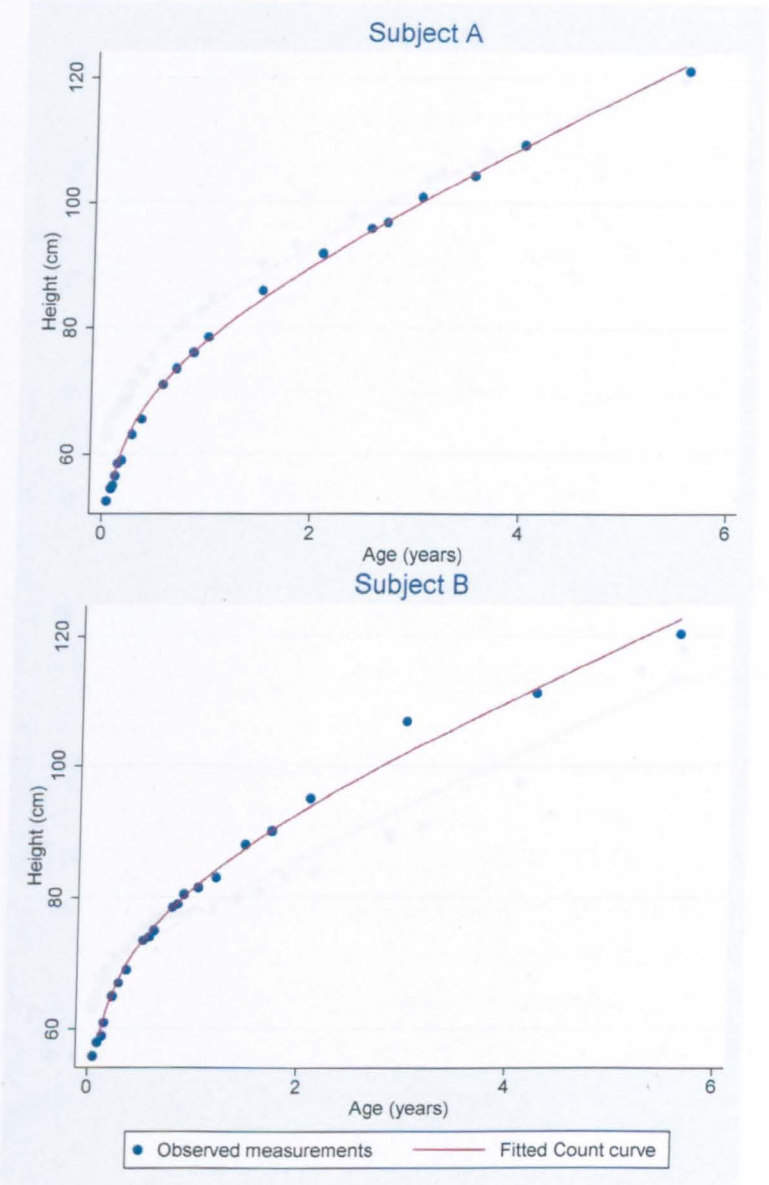

Fig. 6.3: Observed height measurements and fitted Count height A-curve for two subjects in the Uppsala Family Study.

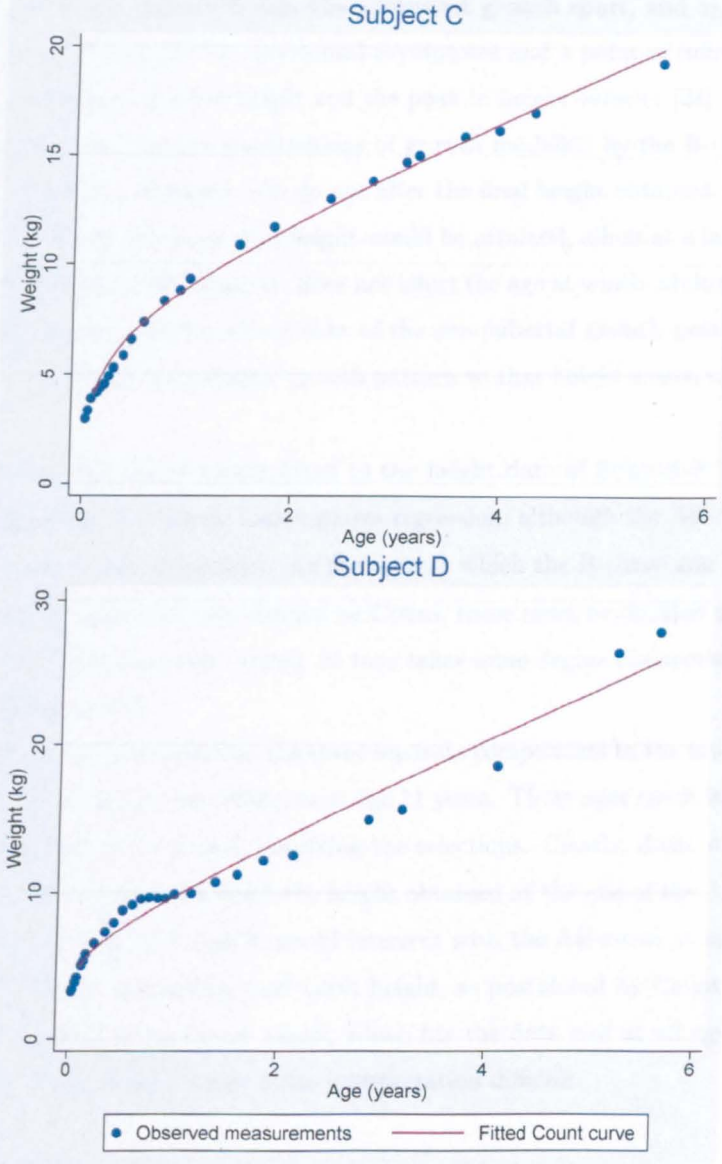

Fig. 6.4: Observed weight measurements and fitted Count weight A-curve for two subjects in the Uppsala Family Study.

where β_1 and β_2 are the parameters to be estimated.

After this age there is then a further increase in growth velocity during adolescence before height flattens out. This adolescent growth spurt is modelled using a logistic function, referred to as the 'AH-curve',

$$y = \gamma_1 + \frac{\gamma_2}{1 + e^{\gamma_3 + \gamma_4 t}},$$

where γ_1 is the value of y attained before the adolescent growth spurt, and γ_2, γ_3 and γ_4 are to be estimated. The AH-curve has two horizontal asymptotes and a point of inflection so is able to model both the attainment of adult height and the peak in height velocity [28].

Count [33] argues that the two accelerations of growth modelled by the B-curve and the AH-curve speed up the process of growth but do not alter the final height obtained. Without the first acceleration (the B-curve), the same final height would be attained, albeit at a later date. The second acceleration (the AH-curve), however, does not affect the age at which adult height is achieved, it merely increases growth velocity above that of the pre-pubertal growth pattern initially, then reduces it below that of the pre-pubertal growth pattern so that height ceases to increase.

Fig. 6.5 shows the full Count model fitted to the height data of Subject E. Both the A- and B-curves can be fitted using ordinary least squares regression, although the AH-curve requires the use of nonlinear least squares regression. As the ages at which the B-curve and AH-curve should be introduced are only approximately defined by Count, these must be decided upon, introducing a level of subjectivity into the model fitting. It thus takes some degree of experience to be able to fit the Count model optimally.

The upper plot in Fig. 6.5 illustrates the three separate components to the model. The B-curve is introduced at age 6 years and the AH-curve at age 11 years. These ages result in a good fit to the data for each component of the model, justifying the selections. Clearly, if the A-curve continued until a later age it would eventually reach the height obtained at the end of the AH-curve. Also, if the B-curve continued it appears that it would intersect with the AH-curve at approximately the same age as the AH-curve is reaching final adult height, as postulated by Count. The lower plot in Fig. 6.5 shows the final fitted Count model, which fits the data well at all ages. However, the large number of model parameters may make interpretation difficult.

6.1.1.3 Berkey-Reed

Berkey and Reed [34] tried to improve upon the Count A-curve by adding an additional term to give

$$y = \alpha_1 + \alpha_2 t + \alpha_3 \log t + \alpha_4 \left(\frac{1}{t}\right), \tag{6.3}$$

where y is the fitted dimension at time t and α_1, α_2, α_3 and α_4 are the parameters to be estimated. The additional term behaves not dissimilarly to the exponential term in the Jenss-Bayley model, enhancing the flexibility of the Count model. In particular, this means that growth can be described

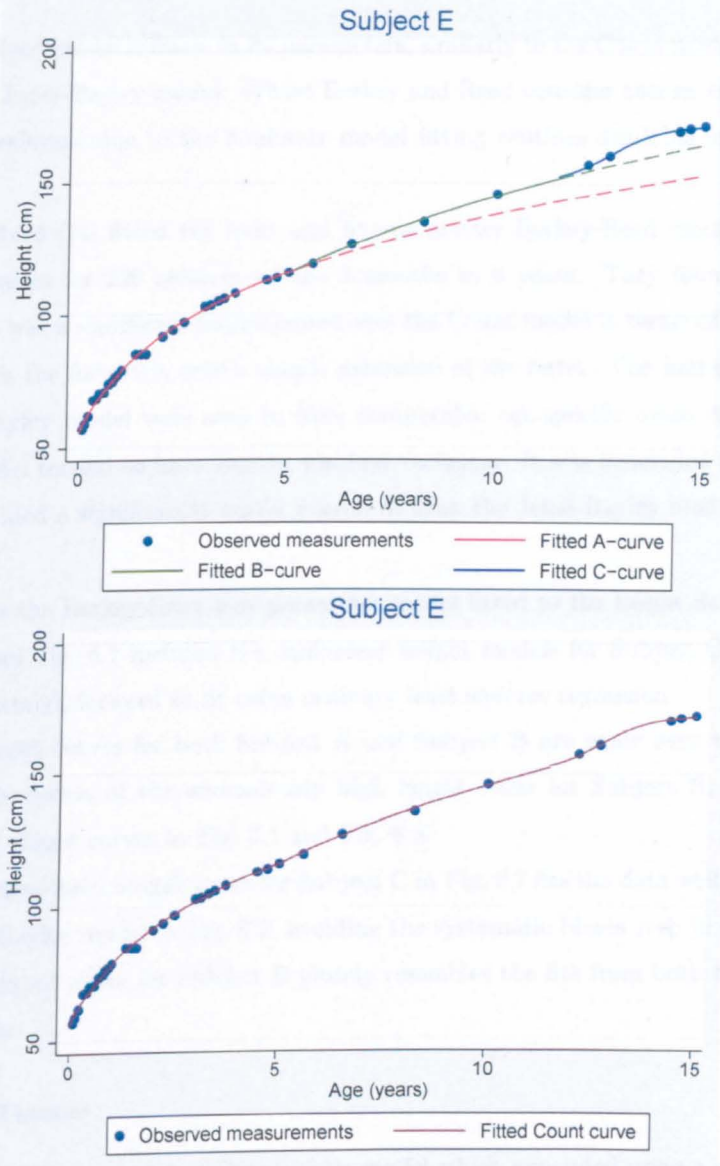

Fig. 6.5: Observed height measurements and fitted Count curve for a subject in the Uppsala Family Study. Upper plot shows the separate components of the model. Lower plot shows the final fitted model.

in which the velocity does not simply decelerate smoothly but fluctuates, causing inflection points in the growth curve. Whilst being mainly intended for the modelling of growth in height, Berkey and Reed [34] suggest the model may also be appropriate for weight and head circumference.

As well as the four-parameter model given in (6.3), a five-parameter model can be obtained by the inclusion of '$\alpha_5 \left(\frac{1}{t}\right)^2$', allowing an additional inflection point in the growth curve. Further inflection points could also be allowed for, if deemed necessary, by inclusion of $\alpha_6 \left(\frac{1}{t}\right)^3$, $\alpha_7 \left(\frac{1}{t}\right)^4$, etc.

The Berkey-Reed model is linear in its parameters, similarly to the Count model but as opposed to the nonlinear Jenss-Bayley model. Whilst Berkey and Reed consider this as an advantage, this is now largely irrelevant due to the nonlinear model fitting routines available in most statistical software.

Berkey and Reed [34] fitted the four- and five-parameter Berkey-Reed models to recumbent length measurements for 229 children of age 3 months to 6 years. They found that the four-parameter model was a significant improvement over the Count model in terms of fit to the growth data, even though the former is only a simple extension of the latter. The four-parameter model and the Jenss-Bayley model were seen to have comparable age-specific mean residuals, but the Berkey-Reed model tended to have smaller residual variances. It was concluded that the Berkey-Reed model provided a significantly better overall fit than the Jenss-Bayley model.

Fig. 6.6 shows the Berkey-Reed four-parameter model fitted to the height data for Subject A and Subject B and Fig. 6.7 includes the equivalent weight models for Subject C and Subject D. The models are straightforward to fit using ordinary least squares regression.

The fitted height curves for both Subject A and Subject B are again seen to fit well to the data, with the exception of the anomalously high height value for Subject B, similarly to the Jenss-Bayley and Count curves in Fig. 6.1 and Fig. 6.3.

The fitted Berkey-Reed weight curve for Subject C in Fig. 6.7 fits the data well, similarly to the equivalent Jenss-Bayley model in Fig. 6.2, avoiding the systematic biases seen in the Count curve in Fig. 6.4. The fitted curve for Subject D closely resembles the fits from both the Jenss-Bayley and Count models.

6.1.1.4 Bock-Thissen

Bock and Thissen [35] developed a triple-logistic model which expanded upon a previous double-logistic version [149]. The model describes growth in height from age 1 year to adulthood using separate components for 'early childhood', 'middle childhood' and 'adolescence' given by

$$y = \frac{\delta\epsilon}{1 + e^{-\alpha_1(t-\alpha_2)}} + \frac{\delta\zeta}{1 + e^{-\beta_1(t-\beta_2)}} + \frac{y_A - \delta}{1 + e^{-\gamma_1(t-\gamma_2)}}.$$

Here, y is fitted height at time t, y_A is final adult height and δ is the contribution of pre-pubertal growth to adult height. α_1, β_1 and γ_1 are the maximum growth velocities within, respectively, the early childhood, middle childhood and adolescent components, and α_2, β_2 and γ_2 are the ages at

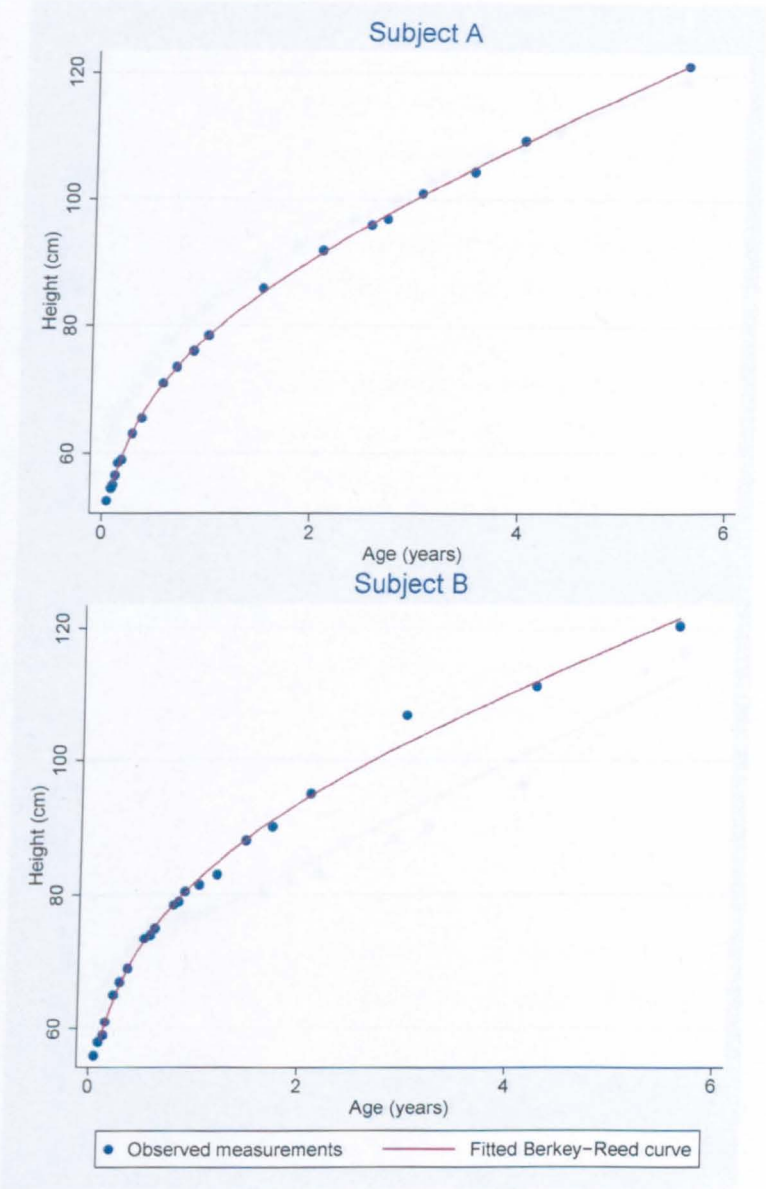

Fig. 6.6: Observed height measurements and fitted Berkey-Reed height curve for two subjects in the Uppsala Family Study.

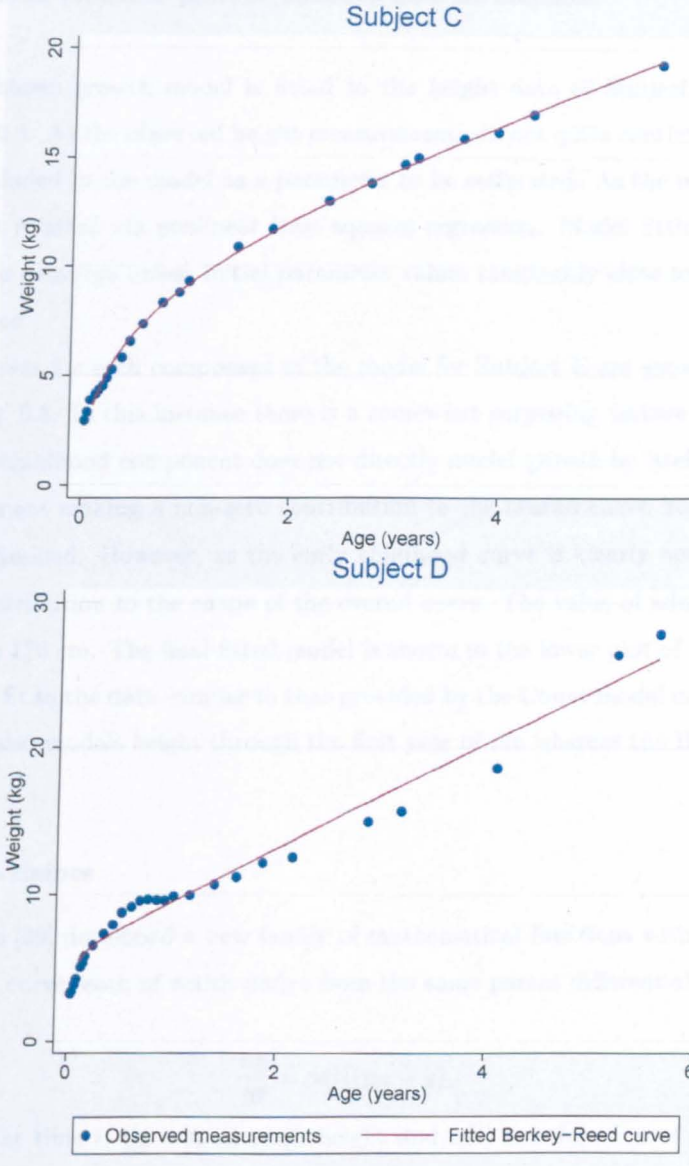

Fig. 6.7: Observed weight measurements and fitted Berkey-Reed weight curve for two subjects in the Uppsala Family Study.

which these velocities occur. ϵ and $\zeta = 1-\epsilon$ are the proportion of pre-adolescent growth attributable to the early and middle childhood components respectively, and $y_A - \delta$ is the contribution of the adolescent component to adult height [150].

The parameter y_A is generally suggested as being *observed* final adult height, the inclusion of which results in a fitted curve which is constrained to pass through this value [142]. However, if adult height is not known there appears to be little reason why it cannot be included in the model as a parameter to be estimated provided sufficient data are available.

The Bock-Thissen growth model is fitted to the height data of Subject E from age 1 year onwards in Fig. 6.8. As the observed height measurements do not quite continue until adult height is reached, is included in the model as a parameter to be estimated. As the model is nonlinear in its parameters it is fitted via nonlinear least squares regression. Model fitting is simple but the model may fail to converge unless initial parameter values reasonably close to the final estimated values are supplied.

The fitted curves for each component of the model for Subject E are shown separately in the upper plot of Fig. 6.8. In this instance there is a somewhat surprising feature of the fitted curves in that the early childhood component does not directly model growth by itself due to the middle childhood component making a non-zero contribution to the overall curve from the very start of the age range examined. However, as the early childhood curve is clearly non-zero itself, it still makes a large contribution to the shape of the overall curve. The value of adult height estimated from the model is 176 cm. The final fitted model is shown in the lower plot of Fig. 6.8 and is seen to provide a good fit to the data, similar to that provided by the Count model in Fig. 6.5. However, the Count curve also models height through the first year of life whereas the Bock-Thissen model does not.

6.1.1.5 Preece-Baines

Preece and Baines [29] developed a new family of mathematical functions with which to describe the height growth curve, each of which derive from the same parent differential equation,

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \alpha(t)(y_A - y), \tag{6.4}$$

where y is height at time t, y_A is final adult height and $\alpha(t)$ is a function of time which differs between the models.

Three models derived from (6.4) are described by Preece and Baines and found to be superior to previous models. Their 'Model 1', which describes height from age 2 years to maturity, was found to be especially accurate and robust [29]. It is given by

$$y = y_A - \frac{2(y_A - y_\beta)}{e^{\alpha_1(t-\beta)} + e^{\alpha_2(t-\beta)}}, \tag{6.5}$$

where α_1 and α_2 are rate constants, β is a time constant and y_β is height at $t = \beta$. In the Preece-Baines model adult height is included as a parameter to be estimated, allowing use of this model

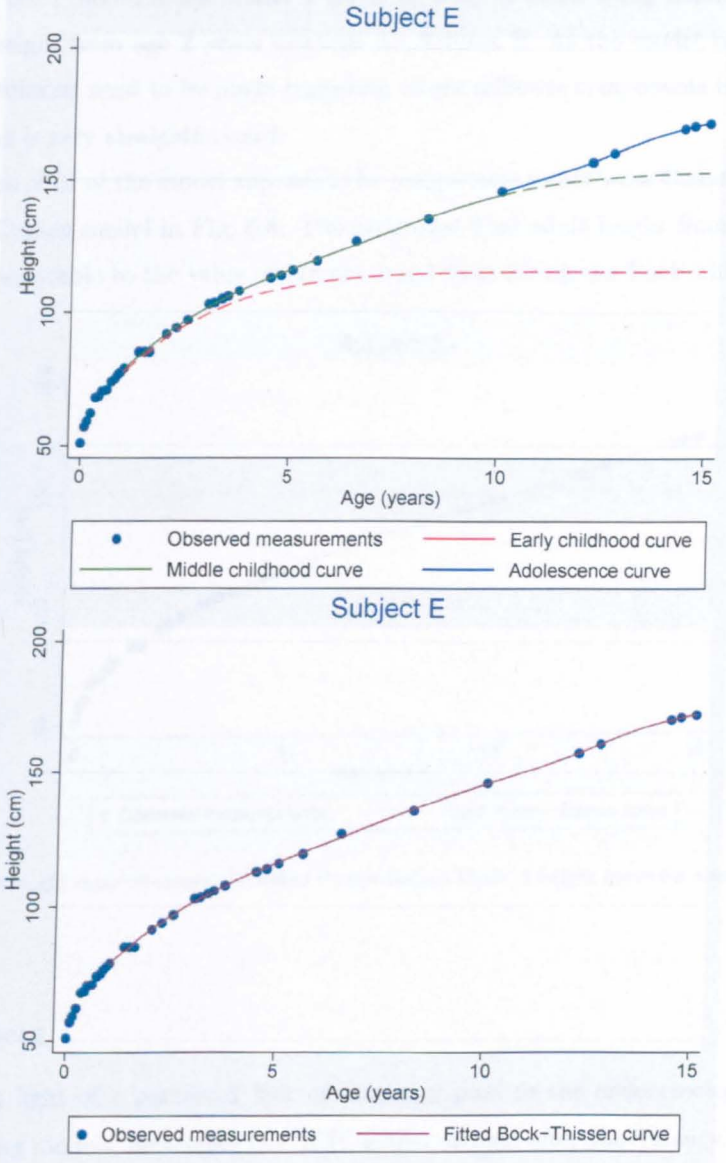

Fig. 6.8: Observed height measurements and fitted Bock-Thissen height curve for a subject in the Uppsala Family Study. Upper plot shows the separate components of the model. Lower plot shows the final fitted model.

for data where final size is not known [142].

This model has been found to be useful for summarising the dynamics of the pubertal growth spurt [146] and has been used extensively for fitting longitudinal data on height [28]. However, the inadequacy of the Preece-Baines model for fitting data concerning infants if often seen as a disadvantage [151].

In Fig. 6.9 the Preece-Baines Model 1 given in (6.5) is fitted using nonlinear least squares estimation to height from age 2 years onwards for Subject E. As the model is fitted in a single stage and no decisions need to be made regarding where different components begin and end, the modelling fitting is very straightforward.

The goodness of fit of the model appears to be comparable to both the Count model in Fig. 6.5 and the Bock-Thissen model in Fig. 6.8. The estimated final adult height from the model is 178 cm, which is comparable to the value of 176 cm found from fitting the Bock-Thissen model.

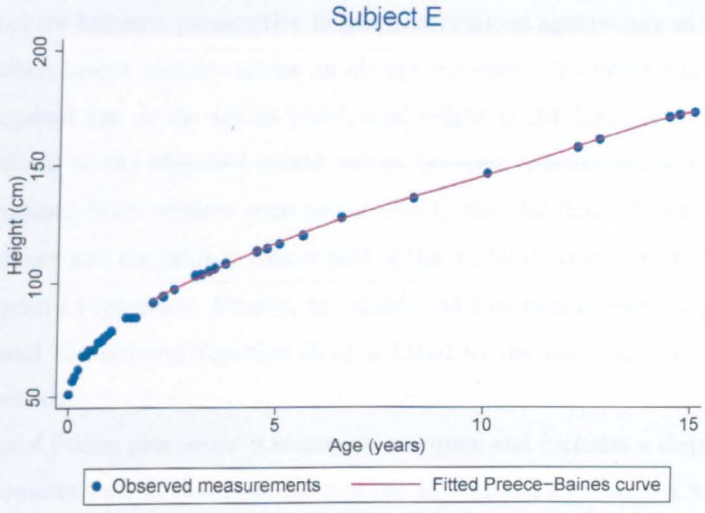

Fig. 6.9: Observed height measurements and fitted Preece-Baines Model 1 height curve for a subject in the Uppsala Family Study.

6.1.1.6 Karlberg

Karlberg [36], in light of a perceived lack of attention paid to the endocrinology of the growth process by existing models, developed the 'ICP' model, named after the 'infancy', 'childhood' and 'puberty' components into which it is split. The components are additive and partly superimposed, with each phase describing growth using a different function.

The infancy component consists of a constantly decelerating function which effectively starts before birth then continues through infancy before tailing off by age 3–4 years. This is represented by the exponential function

$$y = \alpha_1 + \alpha_2(1 - e^{-\alpha_3 t}). \tag{6.6}$$

The childhood phase, starting during the first year of life at age t_C, slowly decelerates until final height is obtained at age t_E and is modelled using a simple quadratic function,

$$y = \beta_1 + \beta_2 t + \beta_3 t^2. \tag{6.7}$$

The final component, puberty, accounts for the additional growth experienced during the adolescent growth spurt. Height accelerates until age at peak velocity (t_V), then decelerates until growth ceases at t_E. This phase is modelled by the logistic function

$$y = \gamma_1/(1 + e^{-\gamma_2(t-t_V)}). \tag{6.8}$$

In each of these functions y denotes height at time t and α_i, β_i and γ_i are the parameters to be estimated.

Karlberg [36] recommends the fitting of the ICP model to be done in a sequential manner. Firstly, the ages at t_C and t_E should be identified. The former can be determined from a plot of calculated velocities between consecutive height observations against age as the age during the first year of life when height velocity shows an abrupt increase. The latter can be identified from a plot of height against age as the age at which final height is obtained. Secondly, the childhood function (6.7) is fitted to the observed height values between approximately age 3 years and age 11 years used ordinary least squares regression. Next, the childhood function is extrapolated backwards into infancy and the infancy component of the model (6.6) is fitted to the residuals using nonlinear least squares regression. Finally, the childhood function is also extrapolated forwards into adolescence and the puberty function (6.8) is fitted to the residuals, again using nonlinear least squares regression.

Clearly the model fitting procedure is somewhat complex and includes a degree of subjectivity. Indeed, Karlberg admits that to make the most of the ICP model a researcher would need considerable experience of this sequential approach [36].

Fig. 6.10 shows the Karlberg ICP curve fitted for Subject E. The model is quite difficult and time-consuming to fit, with the required identification of t_C and t_E, in particular, meaning that an element of subjectivity is introduced.

The upper plot illustrates how the separate components contribute to the final model, with each fitting the observed data points well. Whilst the adolescent growth spurt, as modelled by the puberty component, results in relatively little deviation from the fitted childhood component, the curve still provides a good fit to the data during this period. The overall curve shown in the lower plot of Fig. 6.10 illustrates a similarly good fit to the Count (Fig. 6.5), Bock-Thissen (Fig. 6.8) and Preece-Baines (Fig. 6.9) models. Of these alternatives, however, only the Count curve includes growth within the first year of life as the Karlberg ICP model does.

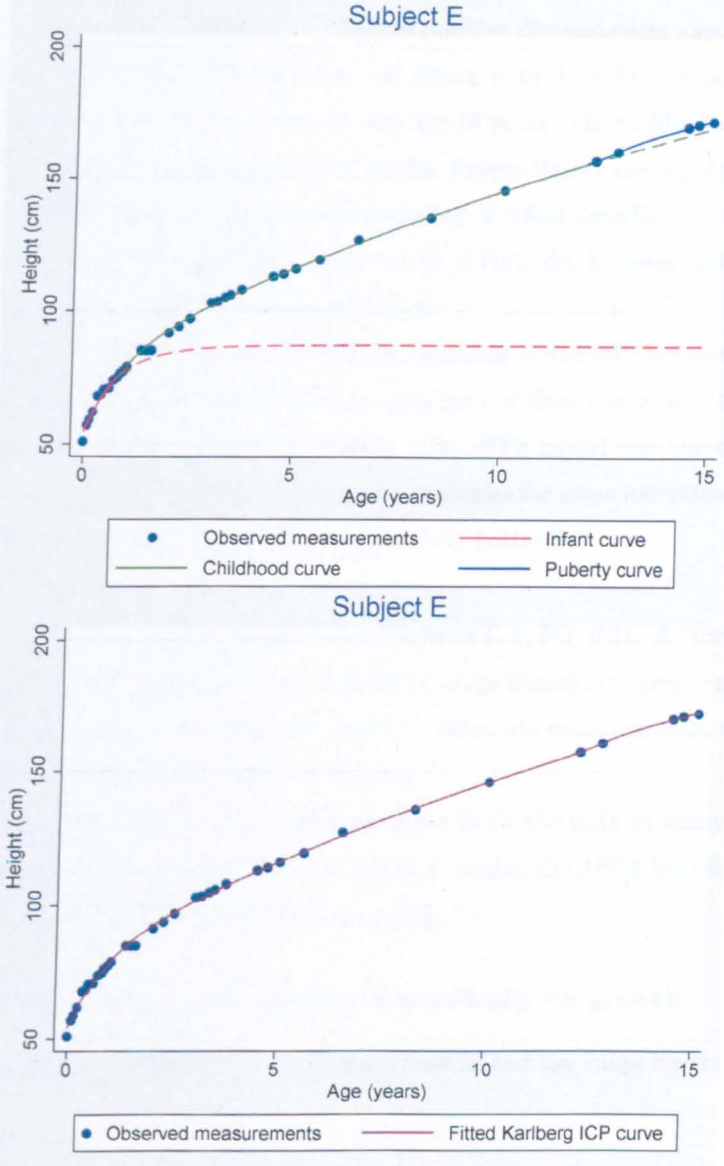

Fig. 6.10: Observed height measurements and fitted Karlberg ICP height curve for a subject in the Uppsala Family Study.

6.1.1.7 Jolicoeur-Pontier-Pernin-Sempé

The seven-parameter model introduced by Jolicoeur, Pontier, Pernin and Sempé [151] (the 'JPPS' model) describes growth in height from birth until maturity. The model is given by

$$y = A\left(1 - \frac{1}{1 + (t_t/\alpha_1)^{\beta_1} + (t_t/\alpha_2)^{\beta_2} + (t_t/\alpha_3)^{\beta_3}}\right),$$

where t_t is 'total age', which takes as its origin the point of conception. The parameters $\alpha_1, \ldots, \alpha_3$ are positive time-scale factors, while $\beta_1, \ldots, \beta_3$ are positive dimensionless exponents.

Jolicoeur *et al* [151] illustrated the model by fitting it to data from a sample of individuals observed longitudinally between age 1 month and age 19 years. The residual sum of squares were found to be 7.5 times greater on average than for the Preece-Baines model. As the Preece-Baines model was never proffered as a solution to the modelling of infant data [29], much of this difference is understandably seen at younger ages. Jolicoeur *et al* [151] do, however, acknowledge that the JPPS model is unable to model to mid-growth spurt.

Ledford and Cole [152] also compared the performance of the JPPS model with that of the Preece-Baines model, though using only data for ages greater than 1 year due to the acknowledged deficiencies of the Preece-Baines model in infancy. The JPPS model was found to be less easy to fit than the Preece-Baines model, with convergence problems for some individuals. In spite of this, the JPPS model was observed to provide as consistently better fit.

The JPPS model is fitted to the height data of Subject E in Fig. 6.11. As time since conception is not known explicitly, t_t is taken to be age plus the average duration of pregnancy (0.75 years), as has been practiced elsewhere [151, 152]. The model is fitted via nonlinear least squares regression, and no problems were experienced with convergence.

The model can be seen to provide a similarly good fit to the data as many of the previously described models for height. As with the Karlberg ICP model, the JPPS model also benefits from being able to model growth within the first year of life.

6.1.1.8 Summary of the models developed specifically for growth

Table 6.1 summarises the dimensions it is possible to model and age range covered for each specific growth model.

6.1.2 General statistical modelling approaches

As opposed to the models developed specifically for growth discussed in Section 6.1.1, many more generally statistical modelling approaches have been suggested to describe individual growth trajectories. Polynomials, fractional polynomials and nonparametric modelling techniques are discussed here. As these approaches are more general they can potentially be used to model growth in any anthropometric variable over any age range, although limitations inherent in the techniques may restrict their usefulness.

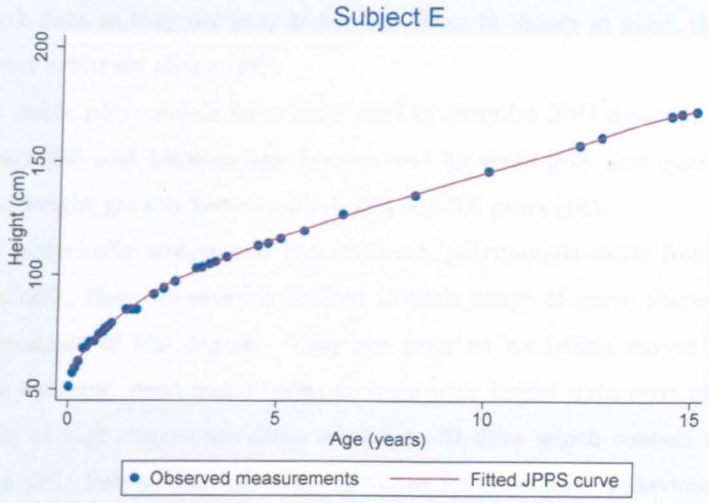

Fig. 6.11: Observed height measurements and fitted JPPS height curve for a subject in the Uppsala Family Study. Upper plot shows the separate components of the model. Lower plot shows the final fitted model.

Model	Dimension(s) modelled	Age range covered
Jenss-Bayley [32]	Height Weight	Birth–6 years
Count A-curve [33]	Height Weight Skull dimensions	Birth–6 years
Count (full model) [33]	Height	Birth–final adult height
Berkey-Reed [34]	Height Weight Head circumference	Birth–6 years
Bock-Thissen [35]	Height	1 year–final adult height
Preece-Baines [29]	Height	2 years–final adult height
Karlberg [36]	Height	Birth–final adult height
Jolicoeur-Pontier-Pernin-Sempé [151]	Height	Birth–final adult height

Table 6.1: Summary of the models developed specifically for growth.

6.1.2.1 Polynomials

Polynomial models, as described in Section 5.4.1.1, were relied upon in much of the early statistical analyses of growth data as they are easy to fit [30]. Also, in theory at least, they can be made to fit curves of almost arbitrary shape [28].

For example, cubic polynomials have been used to describe BMI development between age 2 years and 18 years [84] and between age 2 years and 25 years [87], and quartic polynomials to model height and weight growth between birth and age 2.5 years [145].

Despite their historically widespread use, however, polynomials suffer from a number of disadvantages. Generally, they are severely limited in their range of curve shapes, especially when considering polynomials of low degree. They are poor at modelling curves that approach an asymptote so, for example, need many terms to cope with height data near maturity [28]. Similarly, polynomials of high degree are often needed to fit data which contain observations prior to one year of age [30]. Polynomials are also notorious for their poor behaviour near the ends of the age range covered by the data [34], so-called 'edge effects'. A further concern may be that estimated parameters corresponding to a given polynomial have no real biological meaning.

Whilst these issues have led some to conclude that the use of polynomials to describe human growth should be avoided [153], others argue that polynomials can have their uses, especially when growth is studied over relatively short periods [28].

Polynomials of different degrees are fitted to height data for Subject A and Subject B in Fig. 6.12 and to BMI data for Subject F and Subject G in Fig. 6.13. As polynomials of any degree are linear in their parameters they can be fitted via ordinary least squares regression.

For both Subject A and Subject B, the degree 2 (quadratic) polynomial can be seen to provide a poor fit to the height data. It underestimates height at younger ages whilst overestimating it at older ages. For Subject B it even shows height to be decreasing at around age 5 years, which is clearly implausible. The degree 3 (cubic) polynomial provides a better fit to the data, but because of the predetermined shape it is forced to take the resulting trajectory shows increasing height velocity for both individuals from approximately age 4 years onwards. Again, this contradicts what would be anticipated. The degree 4 (quartic) polynomial is a further improvement on the fit to the data points, but once more the inflexibility of the curve shape makes it far from ideal, with height appearing to reach an asymptote around the age of the last measurement in both cases.

As the pattern of BMI development differs greatly between Subject F and Subject G, the quality of fit of the polynomials in Fig. 6.13 does also. For Subject F, all the polynomial curves fit the data relatively well, though this clearly improves as the degree of the polynomial increases. The BMI development of Subject G, on the other hand, is not well described by any of the polynomial curves. The main features of the data are high and relatively constant BMI for the first few years, with similarly high BMI at the end of the age range and a minimum in between. The quadratic curve does not describe any of these features at all. The cubic curve fares a little better, but shows a maximum at age 3 years which is clearly not present in the data. The quartic curve fits these

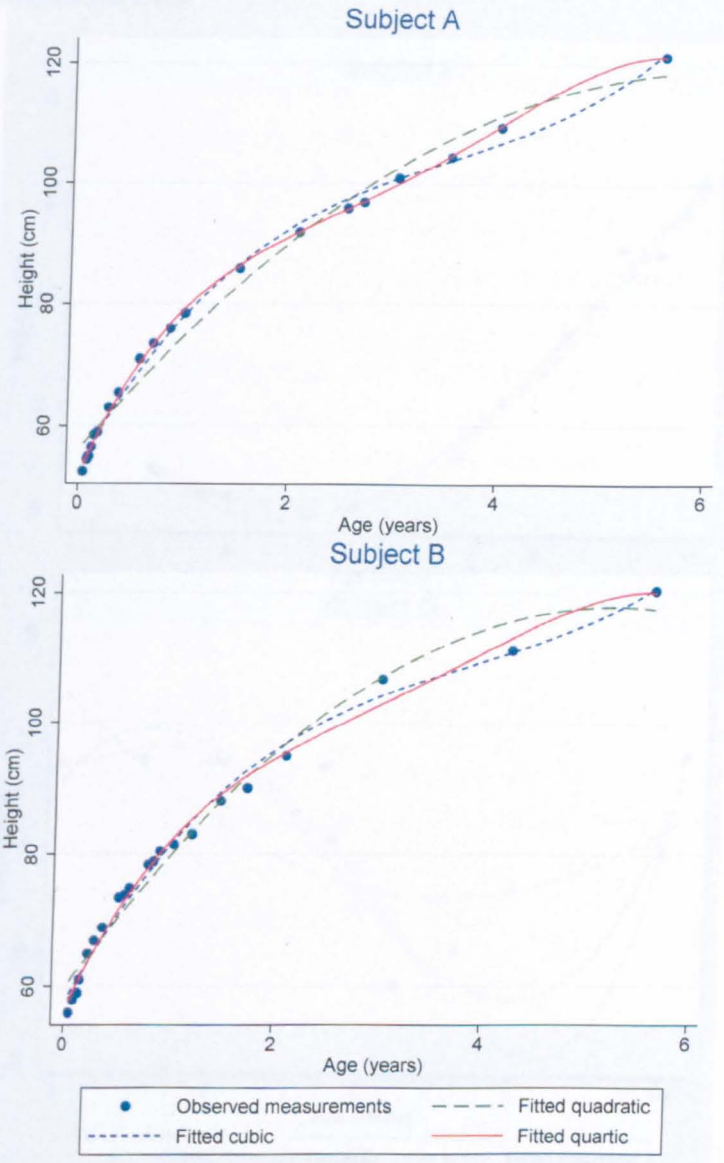

Fig. 6.12: Observed height measurements and fitted polynomial curves for two subjects in the Uppsala Family Study.

early data somewhat better, but if there was interest in modelling the sharp fall in BMI at around age 5.5 years, then none of these polynomials would be appropriate. Similarly, identification of the age at which the minimum BMI occurs using these models would give unreliable results as the fit around this age is poor.

Whilst increasing the degree of the polynomial further would possibly improve the fit of the curve to the data points, it is unlikely to remove the unwanted 'edge effects' as these are inherent aspects of the polynomial form.

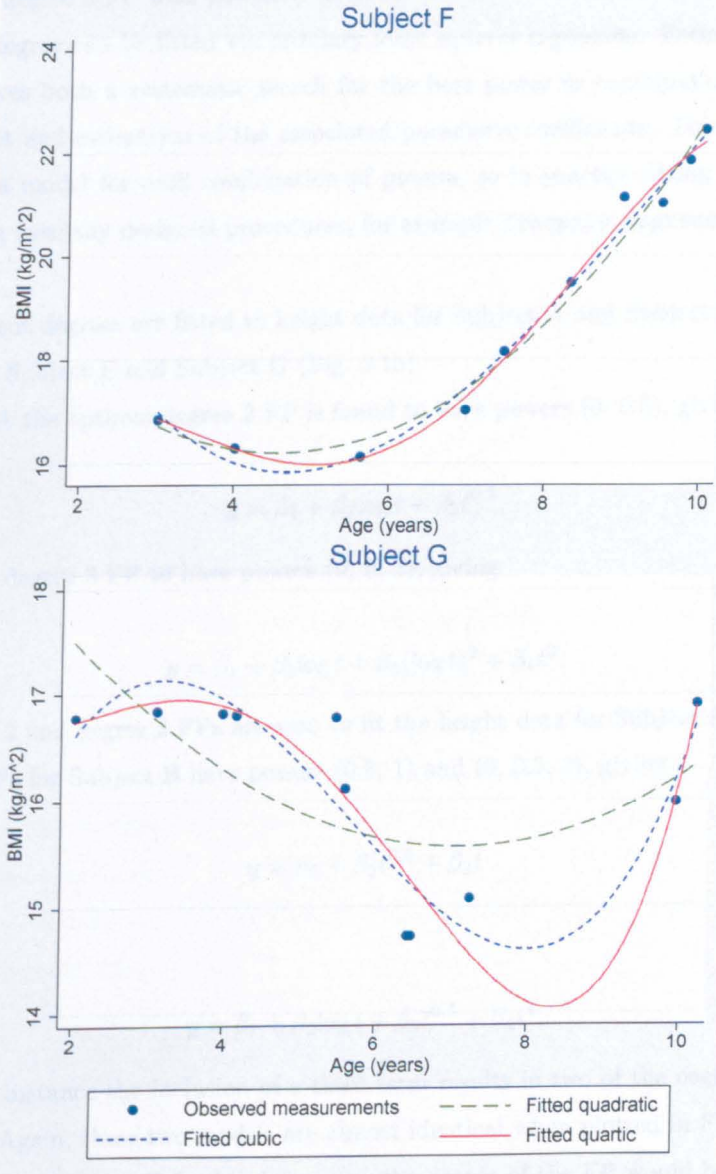

Fig. 6.13: Observed body mass index (BMI) and fitted polynomial curves for two subjects in the Uppsala Family Study.

6.1.2.2 Fractional polynomials

Fractional polynomials (FPs), described in Section 5.4.1.2, extend the range of models afforded by conventional polynomials by allowing parameters to also take fractional powers. This means that

FPs provide many useful curves and can include features such as asymptotes and single points of inflection. They give at least as good a fit to data as a conventional polynomial of corresponding degree and often offer a better fit than a conventional polynomial of higher degree. Similarly to conventional polynomials, however, estimated parameter values often have no obvious biological meaning.

FPs are frequently used in growth modelling, although often not explicitly referred to as such. For example, the Count A-curve (6.2) is a degree 2 FP with powers (0, 1) and the Berkey-Reed model (6.3) is a degree 3 FP with powers (−1, 0, 1).

FPs of any degree can be fitted via ordinary least squares regression. Estimation of the best fitting FP involves both a systematic search for the best power or combination of powers from the permitted set and estimation of the associated parameter coefficients. This selection process includes fitting a model for each combination of powers, so in practice fitting of FPs is usually carried out using specially designed procedures, for example `fracpoly regress` in Stata [147].

FPs of different degrees are fitted to height data for Subject A and Subject B (Fig. 6.14) and to BMI data for Subject F and Subject G (Fig. 6.15).

For Subject A the optimal degree 2 FP is found to have powers (0, 0.5), giving the model

$$y = \beta_1 + \beta_2 \log t + \beta_3 t^{0.5}, \tag{6.9}$$

and the optimal degree 3 FP to have powers (0, 0, 2), giving

$$y = \beta_1 + \beta_2 \log t + \beta_3 (\log t)^2 + \beta_4 t^2. \tag{6.10}$$

Both the degree 2 and degree 3 FPs are seen to fit the height data for Subject A similarly well.

The fitted FPs for Subject B have powers (0.5, 1) and (0, 0.5, 3), giving

$$y = \beta_1 + \beta_2 t^{0.5} + \beta_3 t \tag{6.11}$$

and

$$y = \beta_1 + \beta_2 \log t + \beta_3 t^{0.5} + \beta_4 t^4, \tag{6.12}$$

although in this instance the inclusion of a third term results in two of the coefficients becoming non-significant. Again, these two models are almost identical when plotted in Fig. 6.14. Thus for both Subject A and Subject B further increasing the degree of the FP would be unlikely to lead to significant improvements in the fitted curves.

The degree 2 FP models for Subject A and Subject B ((6.9) and (6.11)) both differ from the Count A-curve by only one term, with Subject A retaining the logarithmic term but replacing the linear term and Subject B retaining the linear term but replacing the logarithmic term. Although this may seem somewhat counter-intuitive for Subject B, these changes manifest themselves in Fig. 6.14 as allowing the curvature seen in the first few years of life to continue throughout the

range of ages observed. This is as opposed to the clearly linear curves seen for older ages when using the Count model in Fig. 6.3. Thus, as the degree 2 FP powers required for the Count A-curve are available when fitting the height curves for Subject A and Subject B yet are not chosen as optimal, it can be concluded that (6.9) and (6.11) provide better fits to the data than the Count A-curve. Indeed, this can be more formally assessed via comparison of the deviance for each model. For Subject A this is 37.4 (degree 2 FP) vs. 56.0 (Count A-curve) and for Subject B this is 62.4 vs. 78.3, so in both cases the degree 2 FP fit is a clear improvement.

The fitted degree 3 FPs ((6.10) and (6.12)) differ from the Berkey-Reed model, although both retain the same underlying logarithmic function. This again shows that the fitted FP models provide a better fit to the data than the Berkey-Reed model. However, comparison of the plots in Fig. 6.14 with those in Fig. 6.6 illustrates the lack of any real difference in the curve shapes. This is reinforced by an examination of the model deviances, which are smaller in the FPs, though not markedly so: 18.0 (degree 3 FP) vs. 20.5 (Berkey-Reed) for Subject A and 61.5 vs. 65.7 for Subject B.

The fitted degree 2 and 3 FPs for the BMI data of Subject F in Fig. 6.15 have powers (0.5, 1) and (3, 3, 3) respectively, giving models

$$y = \beta_1 + \beta_2 t^{0.5} + \beta_3 t.$$

and

$$y = \beta_1 + \beta_2 t^3 + \beta_3 t^3 \log t + \beta_4 t^3 (\log t)^2.$$

Whilst the parameters in these models are completely different, the fitted curves they produce are again very similar, suggesting there may be little to gain by fitting FPs of higher degree. This can be tested more formally, again using the deviance. The degree 3 FP has a deviance of 0.88, compared to 0.03 for the degree 4 FP. Although this does show some improvement in fit, it is far from being statistically significant (P = 0.96).

The fitted degree 2, 3 and 4 FPs for Subject G have optimal powers (3, 3), (3, 3, 3) and (1, 1, 1, 1) respectively, giving models

$$y = \beta_1 + \beta_2 t^3 + \beta_3 t^3 \log t,$$

$$y = \beta_1 + \beta_2 t^3 + \beta_3 t^3 \log t + \beta_4 t^3 (\log t)^2$$

and

$$y = \beta_1 + \beta_2 t + \beta_3 t \log t + \beta_4 t (\log t)^2 + \beta_5 t (\log t)^3.$$

Interestingly, the degree 3 FP takes the same powers as the degree 3 FP for Subject F. Additionally, whilst the degree 3 FP for Subject G can be seen to be a simple extension of the degree 2 model, the degree 4 model is completely different.

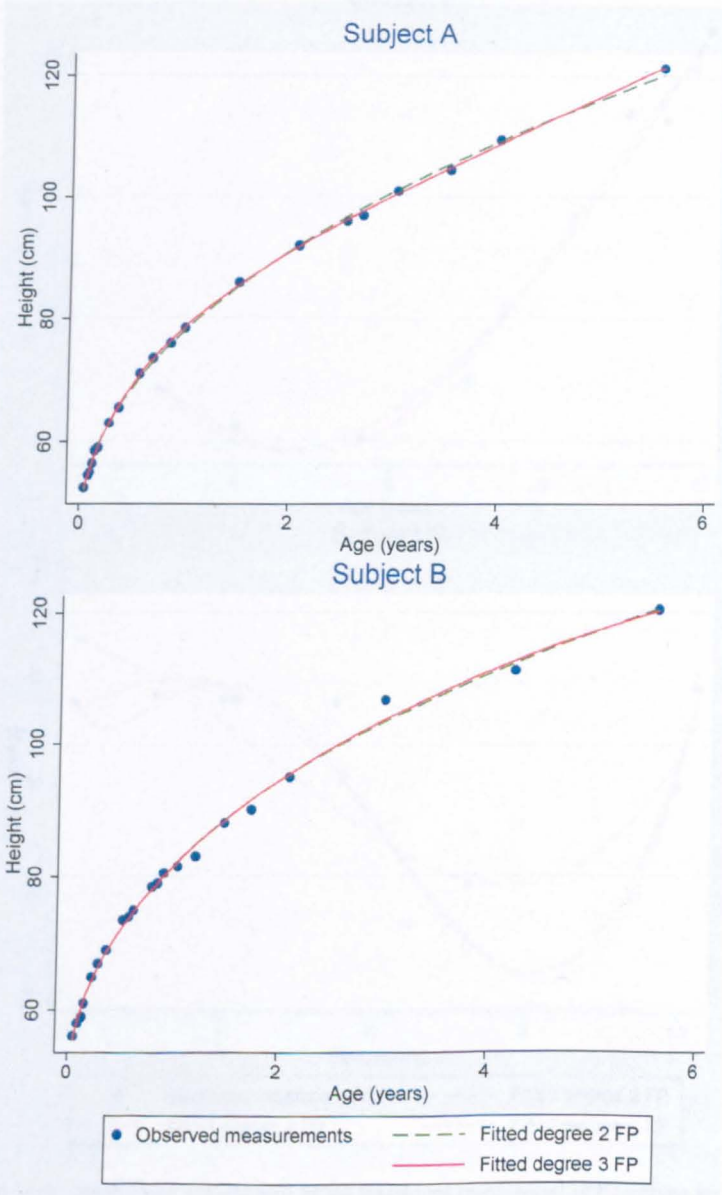

Fig. 6.14: Observed height measurements and fitted fractional polynomial (FP) curves for two subjects in the Uppsala Family Study.

The resulting curves, however, are seen to be poor fits to the data. The degree 2 FP does not describe even the major features of the data, whilst the degree 3 and 4 FPs struggle to model the sharp decline in BMI and both feature unwanted edge effects at young ages. Indeed, the fitted FP curves are not dissimilar from the quadratic, cubic and quartic polynomial curves in Fig. 6.13, showing that FPs are also inadequate to model the delicate features of this growth pattern.

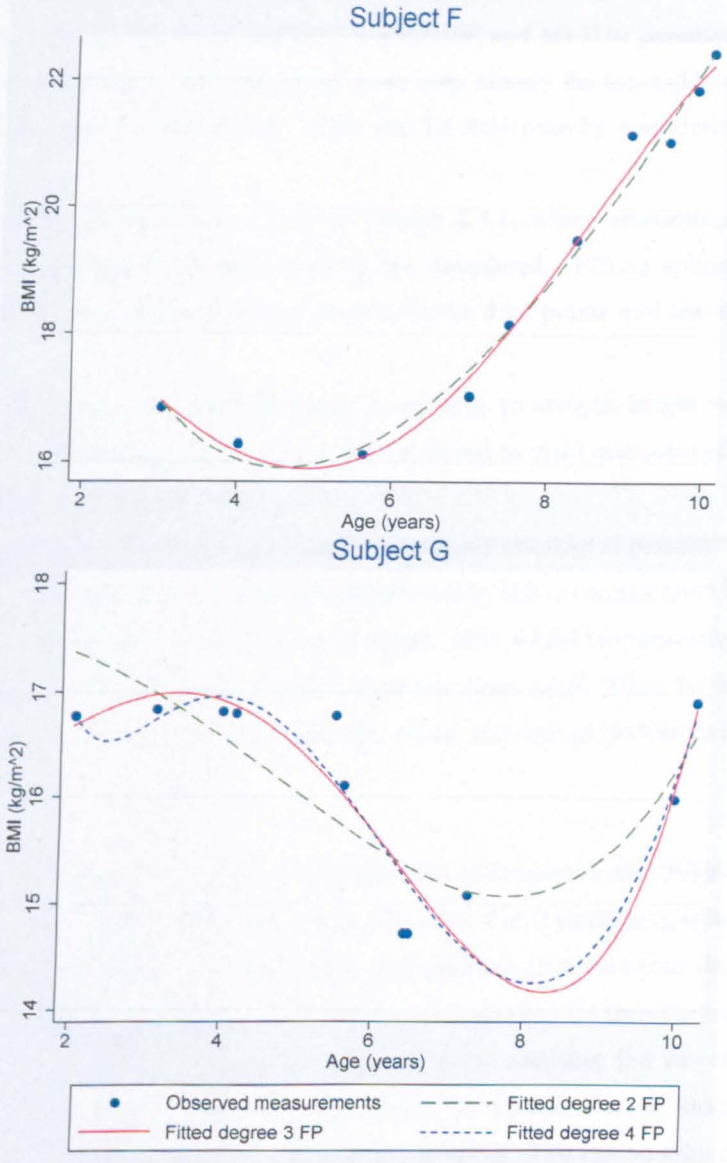

Fig. 6.15: Observed body mass index (BMI) and fitted fractional polynomial (FP) curves for two subjects in the Uppsala Family Study.

A further concern regarding FPs may be that, even for a given degree, the optimal FPs for two individuals may differ in the powers that they take, making it impossible to compare the estimated parameters between them. This is the case in three of the four examples used here. If this comparability of model parameters is a desirable feature of the model fitting, then all subjects should be forced to have FPs with the same combination of powers. This could be decided upon by, for example, initially allowing a FP to be optimally fitted for each individual, analysing the

distribution of the resultant power combinations, then refitting the FPs with each forced to take the most commonly observed combination of powers. However, this would result in sub-optimally fitted FP models for some individuals.

6.1.2.3 Nonparametric modelling

All the methods discussed thus far impose a set algebraic form upon the fitted growth curve, although for FPs a greater degree of flexibility is available, and are thus parametric. One problem with this type of approach is that the given form may simply be too rigid to model the true complexities of the growth process [28]. This can be overcome by considering nonparametric modelling.

Nonparametric modelling is introduced in Section 5.4.1, where smoothing splines (Section 5.4.1.4) and regression splines (Section 5.4.1.5) are considered. Fitting splines can be considered as a compromise between achieving a close fit to the data points and the smoothness of the curve [28].

Cubic spline functions were used by Largo *et al* [154] to smooth height velocity data in a longitudinal study. The resultant curves were then analysed to yield estimates of certain points of interest, for example peak height velocity.

Whilst a potential disadvantage of some of the previously considered parametric models is that the estimated parameters lack any biological interpretability, this is compounded in nonparametric approaches by the number of parameters often involved. Also, whilst the derivatives of a spline can be generally obtained, they will be less smooth than the curve itself. Thus, for example, the first derivative of a cubic spline is piecewise quadratic, whilst the second derivative consists of linear sections.

Cubic smoothing splines are fitted to the height data of Subject A and Subject B in Fig. 6.16 and the BMI data of Subject F and Subject G in Fig. 6.17. Cubic smoothing splines are piecewise cubic polynomials which employ a roughness penalty approach to ensure that the fit of a curve is determined not only by its goodness of fit to the data but also but its smoothness [140].

Although there exist several 'automatic' procedures for selecting the smoothing parameter used in the model fitting (for example cross-validation, see Section 5.4.1.4), manually specifying the degree of smoothing may be preferable in growth modelling. This can be achieved by specifying the smoothing parameter itself or by using equivalent degrees of freedom (EDF, see Section 5.4.1.4).

For each individual, splines are fitted using several different EDF values. These have been selected to try to illustrate cases of 'underfitting' (where the curve is too smooth and thus provides a poor fit to the data), 'overfitting' (where the curve is not sufficiently smooth, resulting in an implausibly 'wiggly' growth curve), and a reasonable compromise between the two. However, it should be noted that a 'reasonable compromise' is both subjective and dependent on the aims of the curve fitting.

The splines are fitted using the `smooth.spline` function in R [155]. Basic curve fitting with

user-defined EDF values is easily implemented.

The fitted cubic smoothing spline for the height data of Subject A in Fig. 6.16 with an EDF of 3 can be seen to provide a poor fit to the data as too much emphasis is placed on the roughness penalty. With an EDF of 20 (the number of data points to which the spline is fitted) the curve interpolates the data points, although in this instance the points lie almost on a smooth trajectory so this is perhaps not wholly unreasonable. The final fitted cubic smoothing spline, with an EDF of 7, is a compromise between these two fits. Whilst this curve is by no means constrained to pass through the data points, it is very similar to the interpolating curve in this instance.

The plot for Subject B again shows that with a low EDF value the model is underfitted. Increasing the EDF so that it is equal to the number of data points again provides an interpolation, but with this individual this results in a less biologically plausible curve. It does, however, mean that the fitted curve acknowledges the unexpectedly high height value at approximately age three years, which other models have failed to do in any meaningful way. If it is not believed that this observed height value is the true value of height at this age (for example if the measurement could be expected to be subject to measurement error), then a compromise can be reached whereby the curve describes a 'growth spurt' at this age without being constrained to interpolate the data points. This is the case with an EDF of 6, which is also shown to provide a good fit to the data at other ages. This exemplifies how smoothing splines have scope for 'fine-tuning' which allows the user to try to identify specific features of the growth curve to greater or lesser extents.

In Fig. 6.17 the fitted smoothing splines for the BMI data of Subject F follow a similar pattern, with an EDF of 4 providing a smooth curve which is a poor fit to the data and an EDF of 10 resulting in a curve which passes through all the data points but is insufficiently smooth. Using an EDF of 6 appears to be a reasonable compromise.

As the pattern of BMI development for Subject G is somewhat more complex, there is greater variability in the shape of the fitted splines. Whilst a low EDF again underfits, the interpolation is in this case clearly implausible. Using an EDF of 6 provides a curve which describes the underlying trajectory of the growth reasonably well, but for this individual it is somewhat less obvious exactly what this is, meaning that a greater level of subjectivity is involved.

This ability to 'fine-tune' the fitted spline means that if identification of a particular feature is deemed a priority, for example the point of minimum BMI, this can be achieved, although potentially to the detriment of the fit of the curve at other ages. This aspect of spline modelling also has implications for the automatability of the process, as the 'optimal' degree of smoothing may differ between individuals and may only be assessable manually (as opposed to using an automatic procedure such as CV). Similarly to the approach suggested for FPs, the spline for each individual could be forced to take the same predetermined degree of smoothing, though again this may result in sub-optimal fits for some subjects.

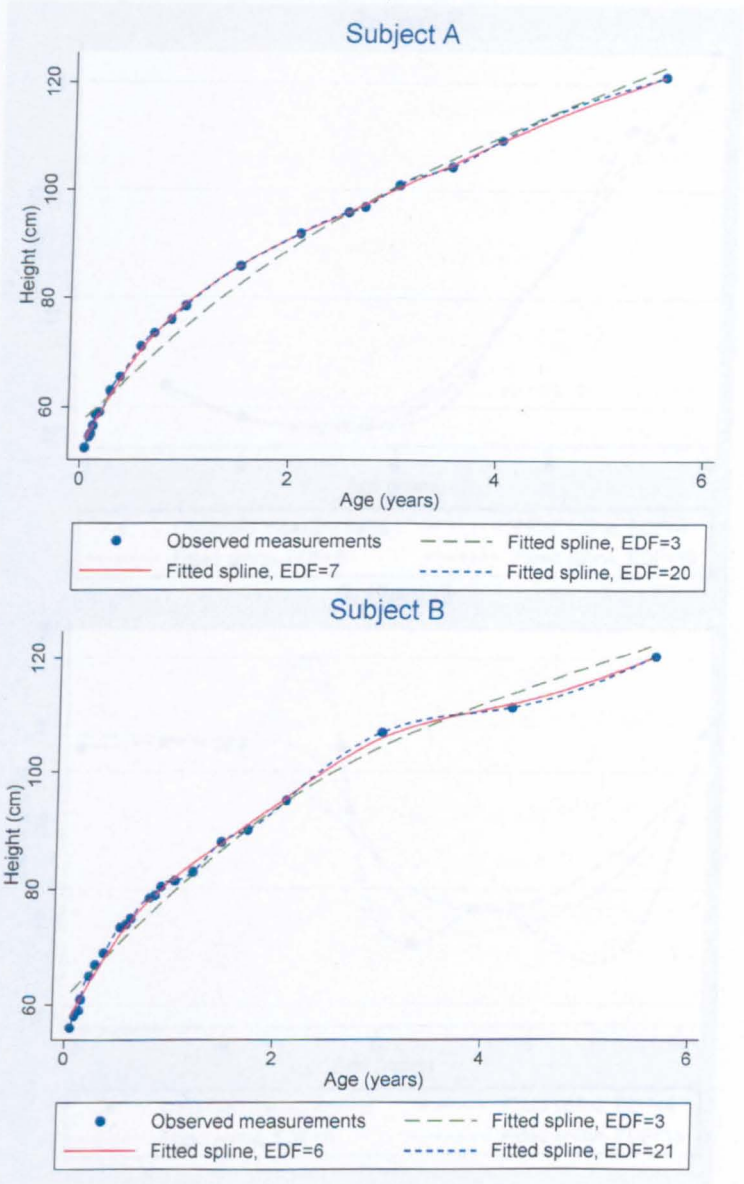

Fig. 6.16: Observed height measurements and fitted spline curves for two subjects in the Uppsala Family Study.

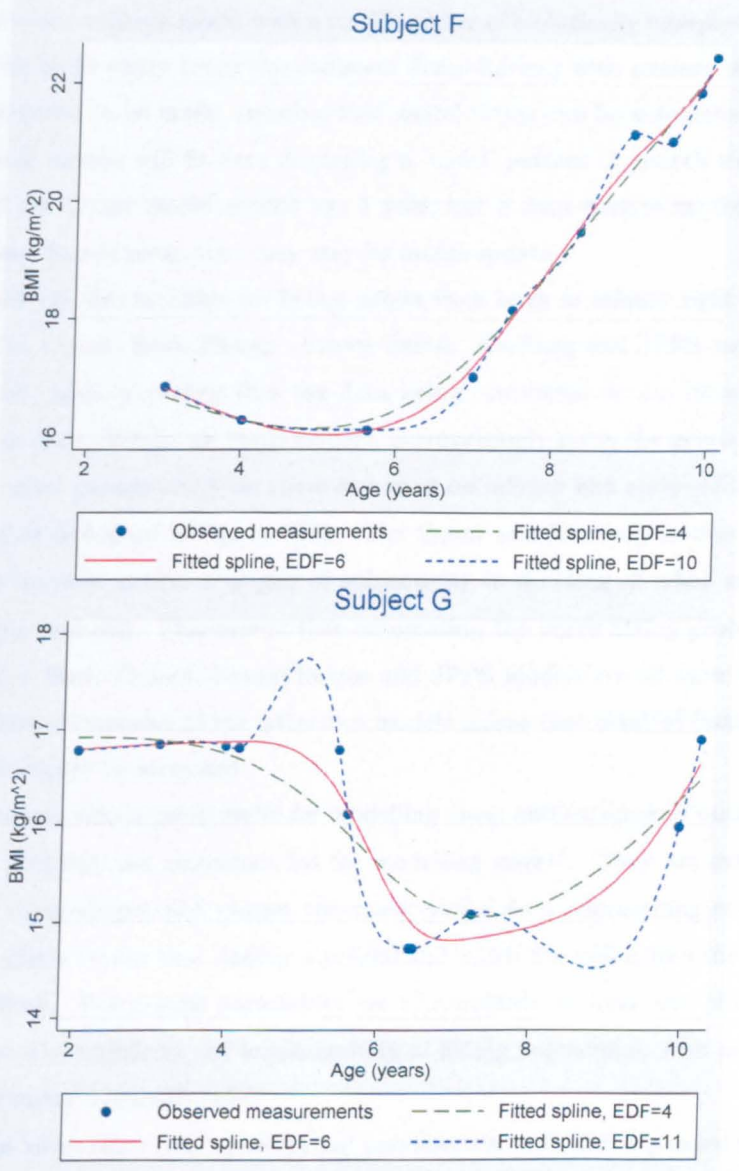

Fig. 6.17: Observed body mass index (BMI) measurements and fitted spline curve for a subject in the Uppsala Family Study.

6.1.3 Discussion

There is a vast array of both specifically developed growth models and more general modelling approaches available, each with its own advantages and disadvantages. The user must therefore consider carefully what they require from their model before selecting one.

If the aim is to fit models for height, weight, or possibly other anthropometric variables, through infancy and early childhood then the Jenss-Bayley, Count A-curve or Berkey-Reed models may suffice. Each provides a simple model with a small number of biologically interpretable parameters. These models can be fit easily (even the nonlinear Jenss-Bayley) with modern software and with no subjective decisions to be made, meaning that model fitting can be automated across multiple individuals. These models will fit data displaying a 'usual' pattern of growth well, perhaps with the exception of the Count model around age 1 year, but if data deviate far from this then the rigid form of these models mean that they may be inappropriate.

Several models are also available for fitting height from birth or infancy right through to final adult height. The Count, Bock-Thissen, Preece-Baines, Karlberg and JPPS models all achieve this objective well, again providing that the data being considered do not deviate too far from the expected trajectory. Whilst all these models, unsurprisingly given the greater ranges of ages covered, include more parameters than those focussing on infancy and early childhood, they still retain some level of biological interpretability. The Count and Karlberg models are more time-consuming to fit as they involve a degree of subjectivity in deciding at what ages components of the model begin and end. This means that automating the curve fitting process becomes far more difficult. The Bock-Thissen, Preece-Baines and JPPS models are all more straightforward to fit, but the relative simplicity of the latter two models means that detailed features such as the mid-growth spurt cannot be identified.

Polynomial models, whilst being useful for modelling many anthropometric variables over short time-frames, are generally not recommended for modelling growth. They are extremely limited in their range of curve shapes and cannot effectively model data approaching asymptotes. The presence of edge effects means that finding a polynomial which fits well across the entirety of the data is often difficult. Polynomial parameters are also unlikely to have any obvious biological meaning. However, the simplicity and automatabilty of fitting polynomials with modern software means that they remain frequently used.

FPs offer some advantages over conventional polynomials, with the expanded range of curve shapes meaning that asymptotes and points of inflection can be handled more easily. Again, with modern statistical software the fitting of FPs is simple and can be fully automated. However, the presence of edge effects can still be troublesome in some applications and the possibility of having differing power combinations across individuals may also lead to problems with interpretation.

The flexibility of nonparametric approaches is particularly appealing. The lack of a pre-defined algebraic form means that they can provide models for arbitrary anthropometric measures. The ability to 'fine-tune' the amount of smoothing of spline curves means that different curves can be fitted depending on the aims of the analysis, although this does also have implications for

automatability. The lack of a concise functional form may be seen as a disadvantage, but given that estimated fitted values and derivatives can be simply obtained for any given age using most modern statistical software this is not necessarily a problem.

To conclude, many of the models which have been developed specifically for growth are perfectly adequate for applications which involve the anthropometric variables and cover the range of ages for which they were designed. However, when wishing to model a variable for which an explicit growth model has not been specifically developed, then alternatives must be sought. In these situations, the use of a nonparametric modelling approach would appear preferable to the parametric approaches examined. An example of an application where nonparametric modelling could prove fruitful is in describing individual trajectories of BMI in order to identify ages at which turning points in the growth curves occur. This method is pursued in later chapters.

6.2 Standardisation of measurements

In this section the standardisation of anthropometric variables into z-scores is examined. The general issues are introduced in Section 6.2.1, then in Section 6.2.2 the use of contemporary references datasets to calculated BMI z-scores in historical datasets is investigated.

6.2.1 Issues

Standardisation of anthropometric variables to create *z-scores* or *SD scores* is introduced in Section 2.2.1. Generally, z-scores are a way of comparing an observation of a variable to some relevant distribution. The observed value is transformed into a z-score by subtracting the mean value of the distribution, then dividing by the standard deviation (SD) of the distribution. The z-score of the observation then expresses in terms of SDs how far the observation lies from the centre of the distribution.

There is thus flexibility in the choice of distribution to which the observed value is compared. Often a reference dataset is chosen which is nationally representative. The calculated z-score then provides a measure of how many SDs above or below the national average the observed value lies. An alternative approach, if the observation being considered comes from a larger sample of data, is to use the sample mean and SD of the observed values themselves. The calculated z-score then indicates the position of each observation relative to the other observed values. When a separate reference dataset is used, the z-scores are said to be *externally* standardised, and when they are related to their own distribution they are *internally* standardised.

The choice of reference dataset, or of the subset of data to standardise within if standardisation is internal, depends on whether observations can be considered to come from the same distribution. For example, if height in adulthood is being considered, there are acknowledged differences between males and females. Thus, for externally standardised data, only reference data pertaining to individuals of the same sex as the individual under consideration should be used. Similarly, internally standardised z-scores should only be standardised within individuals of the same sex.

This approach produces sex-adjusted z-scores, as the calculated value for an individual indicates the proximity of their observation to the average value for their sex.

When the distribution of the variable being considered is age-dependent, for example height in childhood, standardisation should only involve individuals of the same sex and age as the individual whose observed value is being considered. This produces sex- and age-adjusted z-scores. Age-dependent variables such as height in childhood are generally difficult to compare between different ages, and z-scores calculated in this manner have emerged as a useful tool to facilitate comparison. *Tracking* of an anthropometric variables is defined as the maintenance of a relative position within a distribution of values in a population over time [156]. The calculation of z-scores thus provides a means of identifying and monitoring tracking in individuals.

Subtraction of the mean and division by the SD of a distribution will only produce reliable z-score values if the distribution is approximately normally distributed. If the distribution is skewed then a transformation may first be used to normalise the distribution. A generalistion of this approach to age-dependent variables is provided by the LMS method of Cole [24], in which the skewness of the distribution, as well as the median and variability, is allowed to vary with age.

6.2.2 Standardisation of historical data using contemporary reference datasets

In this section the use of contemporary references datasets to calculated BMI z-scores in historical datasets is investigated.

6.2.2.1 Introduction

BMI has become the most widely used surrogate measure of adiposity. Although BMI has shortcomings, not least the inability to differentiate between lean mass and fat mass, it is widely used in pediatrics owing to the ease with which measurements can be made on infants and children, and the often routine manner in which serial anthropometric measurements are recorded.

The use of BMI to investigate adiposity in children is complicated further by the manner in which BMI shows profound changes from birth through to early adulthood [61], with relationships between the fat and fat-free components of the body being affected by varying growth rates and maturity levels [63]. However, one tool which is often used to facilitating comparisons across ages is the calculation of BMI z-scores.

There exist contemporary BMI growth reference data, notably the 1990 reference data for the United Kingdom and the 2000 Centers for Disease Control and Prevention (CDC) reference data for the United States, which are frequently used to standardise BMI values. Standardisation of a measurement using an external reference dataset allows an assessment of the position of the measurement within the reference distribution. However, it is unclear whether these growth references are useful as comparisons to less contemporary data. Specifically, given the widely acknowledged increases in childhood BMI over recent years, it may be expected that, on average, childhood BMI in historical datasets would be lower than in the contemporary growth references,

leading to a preponderance of negative BMI z-scores. If standardisation does lead to z-scores which do not follow a standard normal distribution, then there are implications for any analysis using these standardised values.

The aim of the present analysis is to assess how useful contemporary BMI growth references are when looking at historical British datasets. This is achieved by the calculation and analysis of standardised BMI measurements (z-scores) using both the 1990 UK and 2000 CDC growth references for three different British birth cohorts. These cohorts (National Survey of Health and Development (NSHD), National Child Development Study (NCDS) and British Cohort Study (BCS)) are chosen for their national representativeness, range of years of birth (1946–1970), range of ages for which BMI data are available (4–16 years) and longitudinal nature, meaning that the same children can be examined at several follow-up ages in each cohort.

6.2.2.2 Subjects

Three prospective, longitudinal national birth cohorts, dating respectively from 1946, 1958 and 1970 are examined. These cohorts are by design nationally representative. The BMI values in each of these cohorts of children are standardised using the 1990 UK and 2000 CDC BMI references. The cohorts and reference datasets are detailed below.

As all three cohorts analysed are made up of children resident in the UK, the 1990 UK growth references would be the more appropriate choice for standardisation of the data and thus are presented first. However, as the 2000 CDC growth references would also often be used, their application remains of great interest. Since this reference dataset is both temporally and geographically less similar to the historical datasets, it may be expected that the BMI z-scores calculated would lie further away from zero.

National birth cohorts

Data from the National Survey of Health and Development (NSHD), National Child Development Study (NCDS) and British Cohort Study (BCS) are used. These datasets are described in more detail in Section 4.3.1. Briefly, the three datasets are prospective, longitudinal, nationally representative birth cohorts, dating respectively from 1946 (NSHD), 1958 (NCDS) and 1970 (BCS). The present analysis includes data from follow-up at ages 4, 6, 7, 11 and 15 years from the NSHD, 7, 11 and 16 years from the NCDS, and 10 and 16 years from the BCS.

BMI growth references

1990 BMI reference curves for the United Kingdom (1990 UK) BMI reference curves for UK children were developed for the first time in the mid-1990s [68] based on data collected between 1978 and 1990. Data from 11 distinct surveys were combined, between them recording BMI from birth to age 23, with most being representative of England, Scotland and Wales and all but one being cross-sectional. Summary centile curves were fitted using the LMS method and penalised likelihood [24].

2000 Centers for Disease Control and Prevention growth charts for the United States (2000 CDC) The 2000 Centers for Disease Control and Prevention (CDC) growth charts for the United States represent a revised version of the 1977 National Center for Health Statistics (NCHS) growth charts and include BMI-for-age charts [157]. Most of the data came from the National Health and Nutrition Examination Survey (NHANES) cross-sectional studies conducted from 1963 to 1994, though some supplementary data sources were also utilised. Initial curve smoothing for selected major percentiles was accomplished with various parametric and nonparametric procedures, then a normalisation procedure was used to generate z-scores that closely match the smoothed percentile curves [69].

6.2.2.3 Methods

For each child at each follow-up age in each cohort BMI z-scores are calculated using both the 1990 UK and 2000 CDC BMI reference datasets. For a BMI z-score to be calculated for a given child, and thus for the child to be included in the analysis, data for age, sex and BMI are required. Although each follow-up in each cohort was planned at a specific age, the actual measurements occur over a range of ages. Thus a further stipulation imposed is that all children included at a given follow-up age must have had their measurement within 6 months of the median age at measurement within that follow-up age group. This ensures some degree of homogeneity of age within each age group.

The calculation of BMI z-scores using the 1990 UK and 2000 CDC growth references uses the LMS method developed by Cole and Green [24]. The LMS method summarises the changes in BMI distribution through childhood in a reference dataset by three curves representing the median (M), coefficient of variation (S) and a measure of skewness (L) based on the Box-Cox power required to transform the data to normality. The three parameters are constrained to change smoothly with age, and estimated using penalised maximum likelihood. Once the L, M and S parameters are defined for a reference dataset they can then be used to calculate the BMI value corresponding to any given percentile or z-score, enabling in the construction of growth charts. Conversely, given a BMI measurement, the L, M and S parameters can be used to calculate where, in terms of percentile or z-score, said measurement would occur relative to the distribution of the reference dataset.

The 1990 UK BMI-for-age LMS parameters are extracted from the Microsoft Excel add-in lmsGrowth [158], with equivalent parameters for the 2000 CDC growth reference obtained via the CDC website [159].

The z-score (z) for a given BMI measurement (X) is calculated as

$$z = \frac{(X/M)^L - 1}{LS} \qquad \text{if } L \neq 0$$

or

$$z = \frac{\log(X/M)}{S} \qquad \text{if } L = 0$$

where L, M and S are the growth reference LMS parameters corresponding to the age of the child.

If the BMI values for a study agree closely with the growth reference then the z-scores calculated should be normally distributed with mean and standard deviation 0 and 1, respectively. Once calculated, BMI z-scores in each cohort may then be assessed at each follow-up age for any systematic deviation from this (i.e. any systematic difference from the growth reference).

BMI measurements are deemed implausible if they correspond to an absolute z-score (using the 1990 UK growth reference) greater than six and are thus excluded, as has been practiced elsewhere [160].

Representativeness of the data

The extent to which any results can be extrapolated is dependent of the representativeness of the data. In the present analysis this is affected by both the proportion of individuals for whom data were successfully collected and, within those for whom data are available, the proportion who are included in the present analysis.

Table 6.2 details these characteristics for each follow-up age in each birth cohort. 'Target sample' in each instance is the maximum possible number of individuals for whom data could potentially be collected after the exclusion of the dead, those living abroad and permanent refusals. 'Achieved sample' is the number of individuals for whom at least one response was recorded. 'Sex, age or BMI missing' for an individual means that their BMI z-score cannot be calculated so they are excluded from the analysis. 'Age > 6 months from follow-up median' for an individual means that the age at which their BMI was observed is not sufficiently similar to the other ages within the age group to allow their inclusion in the analysis.

In the NSHD the achieved sample at each follow-up age was between 90 and 96% of the target sample. There are between 9 and 17% individuals excluded from the analysis due to missing sex, age or BMI at each follow-up age, though virtually all measurements occur within the required 12 month interval. As a result, of the achieved sample between 83 and 90% are included in the analysis.

The NCDS includes similarly high levels of achieved sample at each follow-up age (87–92%), though a greater degree of missing sex, age or BMI data, particularly at age 16 years (25%). Thus between 73 and 82% of the achieved sample at each follow-up age is included in the analysis.

Whilst the proportion of the target sample achieved in the BCS was of a similar magnitude to the other cohorts at follow-up age 10 years, at age 16 years the data collection was noticeably handicapped by a teachers' strike [111]. Additionally, over 50% of children have either sex, age or BMI values missing so cannot be included in the analysis, meaning that at age 16 years the BCS cannot be considered as nationally representative as the other cohorts.

From Table 6.2 it can be seen that the requirement for data to have been recorded within 6 months of the median age at each follow-up age rarely results in the exclusion of a significant amount of data and never more than 4% of the achieved sample.

Cohort	Year of birth	Initial cohort size	Follow-up age	Target sample at follow-up[A]	Achieved sample at follow-up[A] (%[B])	Excluded from analysis		Included in analysis (%[C])
						Sex, age or BMI missing (%[C])	Age > 6 months from follow-up median (%[C])	
NSHD	1946	5,362	4	4,900	4,700 (95.9%)	520 (11.1%)	23 (0.5%)	4,157 (88.4%)
			6	4,858	4,603 (94.8%)	758 (16.5%)	13 (0.3%)	3,832 (83.0%)
			7	4,838	4,480 (92.6%)	542 (12.1%)	5 (0.1%)	3,933 (87.8%)
			11	4,799	4,281 (89.2%)	402 (9.4%)	9 (0.2%)	3,870 (90.4%)
			15	4,790	4,274 (89.2%)	698 (16.3%)	10 (0.2%)	3,566 (83.4%)
NCDS	1958	17,634	7	16,727	15,425 (92.2%)	2,168 (14.1%)	589 (3.8%)	12,668 (82.1%)
			11	16,754	15,337 (91.5%)	2,848 (18.6%)	0 (0.0%)	12,489 (81.4%)
			16	16,901	14,647 (86.7%)	3,609 (24.8%)	299 (2.0%)	10,739 (73.3%)
BCS	1970	17,287	10	17,275	14,874 (84.9%)	2,901 (19.5%)	419 (2.8%)	11,554 (77.7%)
			16	17,529	11,621 (66.3%)	5,905 (50.8%)	262 (2.3%)	5,454 (46.9%)

Table 6.2: Representativeness of the data. NSHD is the National Survey of Health and Development, NCDS is the National Child Development Study and BCS is the British Cohort Study. [A] Information taken from [161] (NSHD), [107] (NCDS) and [111] (BCS). [B] Percentage of target sample at follow-up. [C] Percentage of achieved sample at follow-up.

6.2.2.4 Results

The split between males and females and summaries of the age and BMI distributions for the subset of cohort members who are included in the analysis are shown in Table 6.3. In each cohort at each follow-up age, except the less-representative age 16 years follow-up in the BCS, there are slightly more males than females. Due to the skewed nature of the age and BMI distributions, medians and inter-quartile ranges (IQRs) are presented. Both the magnitude and the variability of BMI can be seen to increase after about age 7 years.

The distributions of the calculated BMI z-scores for each birth cohort using the 1990 UK and 2000 CDC growth references are shown in Table 6.4 and Table 6.5, respectively. Once more, medians and IQRs are presented due to the skewed nature of the distributions.

There is clearly a great deal of variation in the median values of BMI z-score in the cohorts at different follow-up ages. Median z-scores are generally positive in early childhood before decreasing, often becoming negative, then increasing once more. These results are more easily interpretable when plotted graphically.

Fig. 6.18 and Fig. 6.19 show the median BMI z-score plotted against the median age at each follow-up age in the three cohorts. Fig. 6.18 displays the BMI z-scores calculated using the 1990 UK (upper plot) and 2000 CDC (lower plot) growth references for males, and Fig. 6.19 shows the equivalent plots for females. Whilst the four plots show all three cohorts to exhibit similar patterns of BMI z-score throughout childhood, there are some cohort-, sex- and growth reference-specific features.

For the males of all three cohorts, using the 1990 UK growth references (Fig. 6.18, upper plot) results in a median BMI z-score that is positive but decreasing through early childhood, reaching a minimum around age 11 years before increasing once more. In the NSHD (the oldest birth cohort) this minimum value corresponds to a BMI z-score of approximately zero, whereas in the other cohorts the minima are clearly negative. Use of the 2000 CDC growth references (Fig. 6.18, lower plot) results in a similar pattern of median BMI z-score through early childhood. In this case, however, all three cohorts cross into negativity, with more extreme minimum values exhibited, then, rather than returning to positivity, merely level off and remain negative.

Over the age range for which data are available for more than one cohort a cross-cohort comparison can be made. It can be seen that at age 6–7 years the median BMI z-scores for the NSHD and NCDS are very similar whereas at later ages it is the NCDS and the BCS that take similar values with those for the NSHD clearly greater, especially around age 11 years.

The pattern of BMI z-score over age in the females (Fig. 6.19) is not dissimilar to that in the males, though the growth reference-specific differences are less marked. Under both growth references the median BMI z-score is positive though decreasing through early childhood before crossing into negativity, with all three cohorts reaching a minimum of about -0.2 around age 11 years. Median BMI z-scores then increase once more to exhibit positive values in adolescence.

Cohort	Follow-up age	Male/female split	Age (years)				BMI (kg/m^2)							
							Males				Females			
			Min.	Median	Max.	IQR	Min.	Median	Max.	IQR	Min.	Median	Max.	IQR
NSHD	4	52.5/47.5	4.2	4.3	4.8	0.0	11.2	16.2	22.9	1.9	10.7	15.9	22.6	2.2
	6	52.7/47.3	5.9	6.0	6.5	0.1	11.8	15.9	22.6	1.7	11.9	15.6	23.2	1.7
	7	51.8/48.2	6.9	7.0	7.5	0.1	11.0	15.8	24.8	1.7	11.8	15.5	26.2	1.8
	11	52.0/48.0	10.7	10.8	11.3	0.1	12.5	16.9	29.8	2.3	11.4	17.0	32.9	2.9
	15	52.5/47.5	14.3	14.5	15.0	0.3	13.0	19.3	33.8	2.8	12.1	20.3	39.8	3.5
NCDS	7	51.7/48.3	7.1	7.3	7.8	0.2	10.7	15.8	29.0	1.7	10.0	15.6	28.2	2.0
	11	51.1/48.9	10.9	11.4	11.8	0.1	11.7	16.8	32.9	2.4	10.9	17.1	37.7	3.1
	16	51.7/48.3	15.4	15.8	16.3	0.2	13.0	19.8	43.9	2.9	12.5	20.6	41.1	3.5
BCS	10	51.6/48.4	10.1	10.5	11.0	0.3	10.9	16.4	29.4	2.2	10.2	16.6	30.9	2.8
	16	48.7/51.3	16.3	16.7	17.2	0.3	13.0	20.5	67.6	3.4	13.0	21.0	48.1	3.8

Table 6.3: Distributions of key variables for subjects included in the analysis, by sex. NSHD is the National Survey of Health and Development, NCDS is the National Child Development Study and BCS is the British Cohort Study. BMI is body mass index. IQR is the inter-quartile range.

Cohort	Follow-up age	BMI z-score							
		Males				Females			
		Min.	Median	Max.	IQR	Min.	Median	Max.	IQR
NSHD[D]	4	−5.59	0.41	3.85	1.45	−5.44	0.21	3.40	1.41
	6	−4.14	0.29	3.23	1.29	−3.07	0.08	3.14	1.06
	7	−5.33	0.16	3.44	1.17	−3.13	−0.07	3.32	1.10
	11	−3.55	0.02	3.16	1.13	−4.43	−0.19	3.37	1.35
	15	−4.70	0.08	3.15	1.20	−5.30	0.23	3.63	1.30
NCDS	7	−5.85	0.14	4.09	1.15	−5.57	−0.08	3.65	1.18
	11	−4.88	−0.14	3.37	1.29	−5.28	−0.28	3.73	1.45
	16	−5.15	0.01	3.84	1.24	−5.10	0.08	3.69	1.28
BCS	10	−5.75	−0.16	3.17	1.19	−5.97	−0.27	3.18	1.35
	16	−5.50	0.03	4.66	1.37	−4.89	0.09	4.15	1.37

Table 6.4: Distributions of calculated body mass index (BMI) z-scores using the 1990 United Kingdom (UK) growth reference, by sex. NSHD is the National Survey of Health and Development, NCDS is the National Child Development Study and BCS is the British Cohort Study. [D] Results weighted to adjust for the one in four sampling of children from manual and self-employed workers.

Cohort	Follow-up age	BMI z-score							
		Males				Females			
		Min.	Median	Max.	IQR	Min.	Median	Max.	IQR
NSHD[D]	4	−6.46	0.53	3.70	1.49	−8.35	0.50	2.84	1.38
	6	−5.07	0.36	2.50	1.23	−3.89	0.25	2.46	1.05
	7	−7.04	0.18	2.52	1.11	−3.83	0.06	2.49	1.10
	11	−3.90	−0.12	2.33	1.05	−4.68	−0.17	2.48	1.22
	15	−5.03	−0.13	2.37	1.10	−5.54	0.21	2.49	1.08
NCDS	7	−7.90	0.15	2.79	1.10	−8.05	0.04	2.64	1.17
	11	−5.42	−0.29	2.46	1.20	−5.61	−0.24	2.66	1.30
	16	−5.51	−0.21	2.87	1.15	−5.47	0.08	2.47	1.08
BCS	10	−6.75	−0.28	2.33	1.13	−6.73	−0.24	2.38	1.24
	16	−5.98	−0.23	3.43	1.27	−5.45	0.06	2.59	1.16

Table 6.5: Distributions of calculated body mass index (BMI) z-scores using the 2000 Centers for Disease Control and Prevention (CDC) growth reference, by sex. NSHD is the National Survey of Health and Development, NCDS is the National Child Development Study and BCS is the British Cohort Study. [D] Results weighted to adjust for the one in four sampling of children from manual and self-employed workers.

The main difference between the two plots in Fig. 6.19 is that when using the 2000 CDC growth reference the median BMI z-score is noticeably greater through early childhood, resulting in it becoming negative slightly later.

In terms of the differences between the cohorts within each plot the pattern is somewhat similar to that seen for the males, with median BMI z-score in the NSHD and the NCDS similar at age 6–7 years then median BMI z-score in the NSHD becoming increasingly greater than in the other two cohorts at older ages.

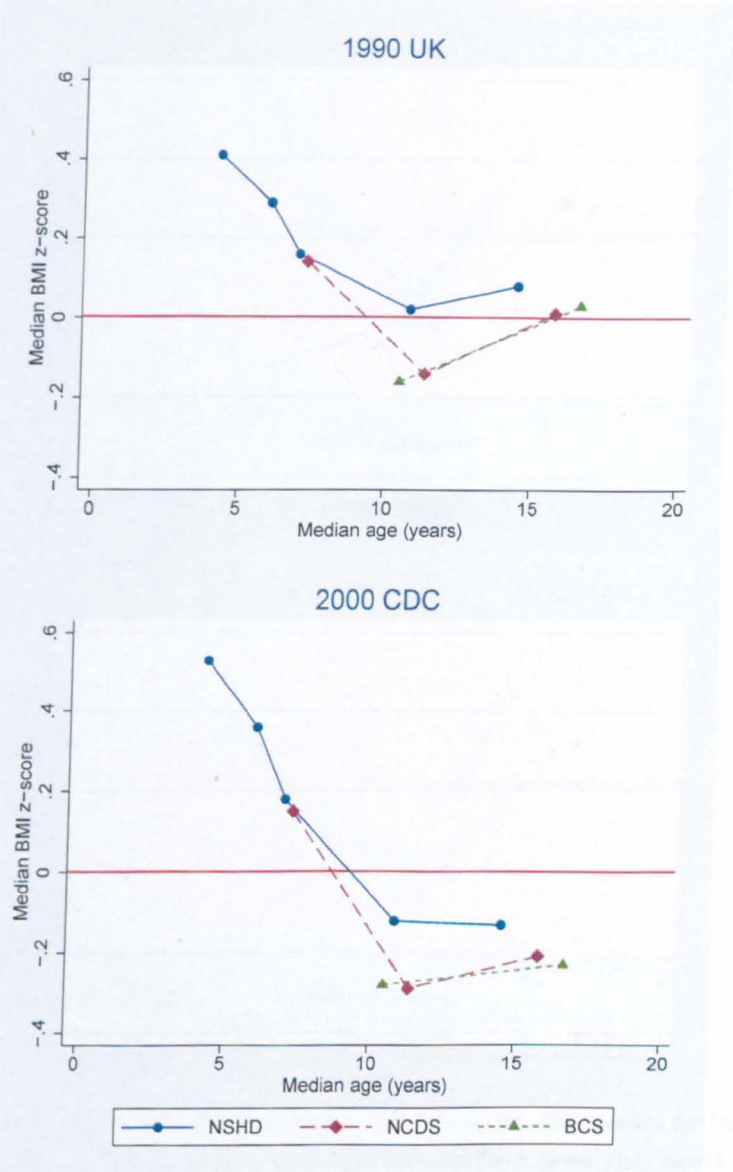

Fig. 6.18: Plots of body mass index (BMI) z-score calculated using the 2000 Centers for Disease Control and Prevention (CDC) (upper plot) and 1990 United Kingdom (UK) (lower plot) growth references against age for males. NSHD is the National Survey of Health and Development, NCDS is the National Child Development Study and BCS is the British Cohort Study.

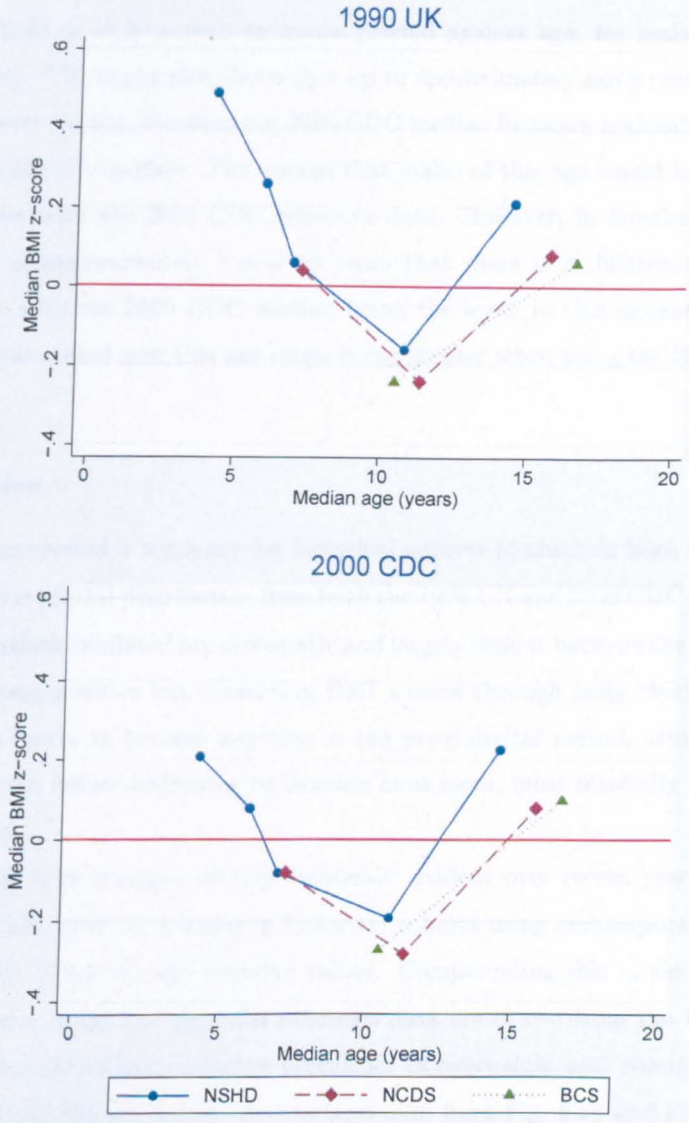

Fig. 6.19: Plots of body mass index (BMI) z-score calculated using the 2000 Centers for Disease Control and Prevention (CDC) reference (upper plot) and 1990 United Kingdom (UK) (lower plot) growth references against age for females. NSHD is the National Survey of Health and Development, NCDS is the National Child Development Study and BCS is the British Cohort Study.

6.2.2.5 Differences between the growth references

Whilst the overall trends in median BMI z-score profile are clearly similar under the two growth references, there are some differences. The 2000 CDC reference data appear to decrease the z-score value relative to the 1990 UK data somewhat in males at follow-up ages of 10 years and older, whereas females of follow-up age 7 years and younger see an increased BMI z-score. These observations correspond to the differences between the growth references evident in Fig. 6.20, showing median BMI in each growth reference plotted against age, for males (upper plot) and females (lower plot). The upper plot shows that up to approximately age 8 years, the two medians for the males are very similar, but then the 2000 CDC median becomes noticeably and increasingly greater than the 1990 UK median. This means that males of this age would have a reduced BMI z-score if calculated with the 2000 CDC reference data. However, in females (lower plot), it is between the ages of approximately 3 and 10 years that there is a difference between the two reference medians, with the 2000 CDC median being the lower in this instance. This results in any BMI z-scores calculated over this age range being greater when using the 2000 CDC reference data.

6.2.2.6 Discussion

This analysis has uncovered a tendency for historical cohorts of children born between 1946 and 1970 to differ in terms of BMI distribution from both the 1990 UK and 2000 CDC growth references. Moreover, the deviations exhibited are systematic and largely similar between the historical cohorts. All three cohorts have positive but decreasing BMI z-score through early childhood. There is a general trend for z-scores to become negative in the pre-pubertal period, attaining a minimum value in early puberty, before beginning to increase once more, most markedly in females, in the late-pubertal period.

Given the widely acknowledged obesity 'epidemic' evident over recent years [37], it may be expected that the calculation of z-scores in historical cohorts using contemporary reference data would lead to largely, if not wholly, negative values. Compounding this in the case of the 2000 CDC growth reference is the fact that the reference data are drawn from the US population, a country generally thought to have a higher prevalence of overweight and obesity than that from which the historical cohorts are drawn. As has been seen from Fig. 6.18 and Fig. 6.19, however, median BMI z-score in the historical cohorts is frequently not negative.

Aside from the growth reference-specific effects of the standardisation, it may be expected that the historical cohorts show a temporal ordering, with those born more recently having higher BMI z-scores. In addition to its less temporal proximity, one may also expect childhood BMI in the NSHD to be lower than in the other cohorts due to cohort members' nutrition being influenced by food rationing which continued after the war until 1954 [162]. However, from Fig. 6.18 and Fig. 6.19 it can be seen that for both males and females BMI z-score in the NSHD is at least a high as, and generally higher than, that seen in the other cohorts.

To expect patterns in BMI z-score through childhood in the historical cohorts to be merely

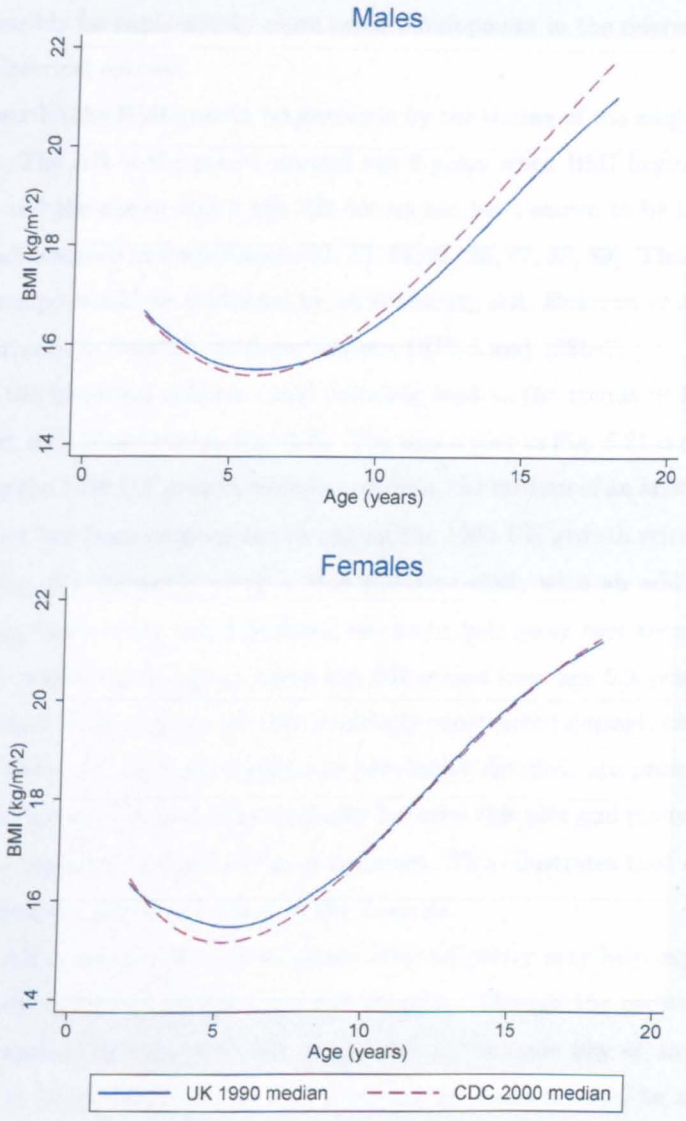

Fig. 6.20: Plots of median body mass index (BMI) in the 1990 United Kingdom (UK) and 2000 Centers for Disease Control and Prevention (CDC) growth references against age for males (upper plot) and females (lower plot).

'negative' may be something of an over-simplification. For median BMI z-score in a given cohort to take a constant value of, say, -0.2 across the entire range of ages would mean the median BMI within the cohort being equal to the median BMI in the reference dataset minus 0.2 of a standard deviation at each follow-up age. Implicit in this is that the median BMI growth trajectories in the historical cohort and the reference dataset follow the same shape. However, acknowledged secular changes in growth patterns over the last century, particularly a trend towards a faster developmental tempo [22], mean this may not be true. Indeed, the results observed in the present analysis could plausibly be explained by more rapid development in the reference data relative to that seen in the historical cohorts.

One way to describe the BMI growth trajectory is by the timing of the adiposity rebound (AR, see Section 2.3.3). The AR is the period around age 6 years when BMI begins to increase again following a nadir, and the age at which the AR occurs has been shown to be inversely associated with adiposity in adolescence and adulthood [82, 83, 84, 85, 86, 87, 88, 89]. Thus a secular increase in developmental tempo would be evidenced by an advancing AR. Eriksson *et al* [163] have shown precisely this occurrence in Swedish children between 1973–5 and 1985–7.

A later AR in the historical cohorts could certainly lead to the trends in BMI z-score which have been identified, as is illustrated in Fig. 6.21. The upper plot in Fig. 6.21 is of BMI against age including, alongside the 1990 UK growth reference median, the median of an artificially constructed dataset. This dataset has been constructed by taking the 1990 UK growth reference median BMI values and stretching the timescale using a multiplicative shift, with an additional component allowing this slowing down of the developmental tempo to fade away over time. The outcome of this manipulation is a shift in the age at which the AR occurs from age 5.9 years in the 1990 UK median to age 6.7 years. BMI z-scores for this artificially constructed dataset, calculated using the 1990 UK reference data and the same method as previously detailed, are presented in the lower part of Fig. 6.21, plotted against age. The similarity between this plot and many of the equivalent plots for the cohorts included in this analysis is apparent. This illustrates that delayed AR alone could plausibly explain the patterns evident in the analysis.

That an earlier AR is associated with increased later adiposity may help explain the positive BMI z-scores in early childhood in the historical cohorts. Though the growth references and historical cohorts examined here are not truly comparable in the same way as, say, two individuals in the same cohort, it is not inconceivable that a similar mechanism could be at work, with the earlier adiposity rebound of the reference data leading to increased adiposity at a later date. In this way, the positive z-scores evident in early childhood could be attributed solely to the earlier adiposity rebound in the reference data, with the possibility of greater adiposity in adulthood in the more contemporary reference dataset bringing the findings more in line with recent trends. This does, however, conflict somewhat with the positive z-scores around age 15 years, particularly in the NSHD, which remain more difficult to explain.

If the children in the historical datasets and the children in the growth references are indeed following slightly different growth trajectories then comparison of the two groups of children at

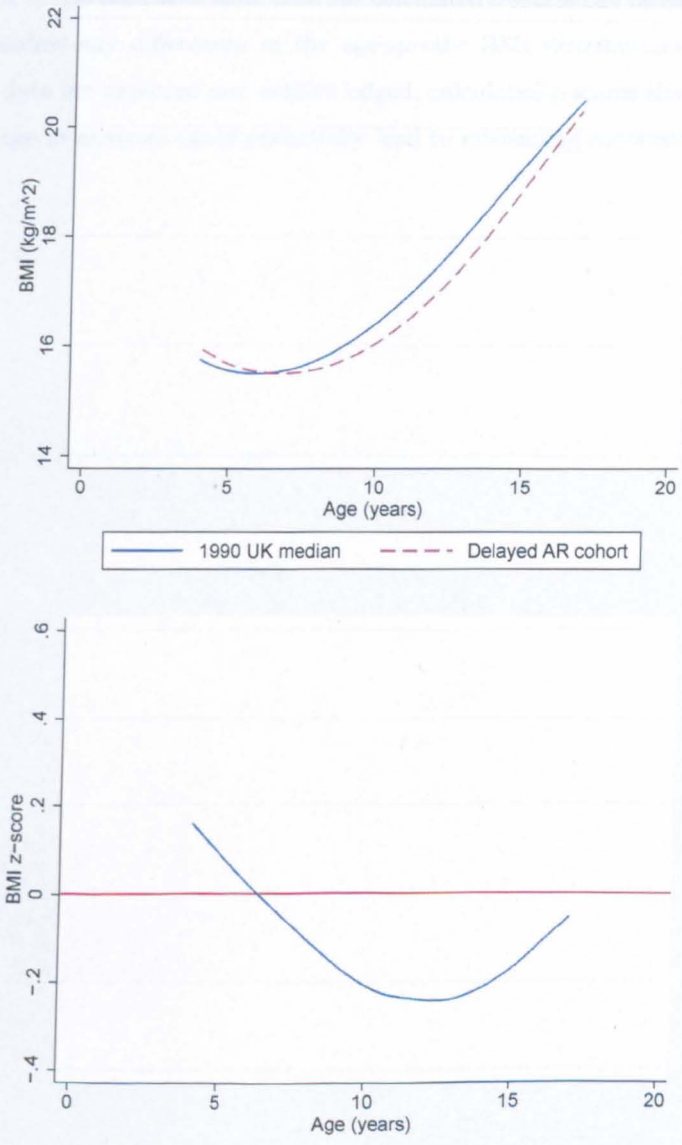

Fig. 6.21: Plots of body mass index (BMI) (upper plot) and body mass index z-score calculated using the the 1990 United Kingdom (UK) growth reference (lower plot) against age for the artificially constructed delayed adiposity rebound (AR) cohort.

specific ages is not comparing like-for-like, thus may be considered inappropriate.

That the BMI z-score trends are seen even in samples of individuals as temporally similar as the BCS subjects (born in 1970) and the subjects who contributed to the 1990 UK growth references (data collected between 1978 and 1990) suggests that it is not necessarily the 'historical' aspect of the data which is the cause. Indeed, regardless of the relative points in time at which the data and the reference data were collected, if the age-specific BMI distributions differ between the two in a systematic manner, as has been seen here, then the calculated z-scores can be easily misinterpreted.

To conclude, unless any differences in the age-specific BMI distributions between the data and the reference data are explored and acknowledged, calculated z-scores should be viewed with caution and their use in analyses could potentially lead to misleading conclusions.

Part II

Approaches for balanced growth data only

This thesis focuses on relating childhood growth, in the form of repeated observations of an anthropometric variable for each child, to a later health outcome. An important distinction with regards to the analytical approaches which may be utilised in this scenario is between *balanced* and *unbalanced* childhood growth data. In Chapters 7 and 8 modelling strategies for use with balanced childhood growth data are explored, developed and implemented.

Balanced growth data, as defined in Section 5.1.2.1, are data resulting from studies where the anthropometric variable of interest is *intended* to be observed at the same set of common ages for each subject in the study. Whether the variable is *actually* observed for a given individual at a given age is immaterial.

One important consequence of balanced growth data is that a single-stage analysis, outlined in Section 5.3, may be used to related the later health outcome directly to the observed growth variable. This could involve, for example, a linear regression of the later health outcome on the growth variable observed at each age. This type of analysis approach is explored in Chapter 7, where measures of late-adolescent adiposity (body mass index (BMI) and percentage body fat (%BF)) are related to observed values of BMI between age 1 and 10 years in the Stockholm Weight Development Study (SWEDES).

Balanced growth data also mean that there are specific ages at which data are 'expected'. Thus missing data can be considered in the traditional sense, as discussed in Section 5.2. There are many approaches for dealing with missing data in balanced datsets. In Chapter 8 multiple imputation (MI), as described in Section 5.2.4, is used to handle the issue of missing data in SWEDES when relating the location of the adiposity rebound (AR, see Section 2.3.3) to late-adolescent adiposity (BMI and %BF).

Chapter 7

Naive multivariable regression analysis

A situation often encountered in life course analysis is that of balanced longitudinal data observed in earlier life being related to a single outcome in later life. Examples of this include childhood height being related to breast cancer risk [7], childhood length/height and weight to adolescent blood pressure [164], and childhood BMI and dietary intake to BMI at age 8 years [165].

One common approach to this situation is to regress the later outcome on some or all of the earlier longitudinal data. This is referred to as *multivariable regression.* However, results obtained when including many childhood measurements in a regression model may be difficult to interpret, especially if observations are close together in time, due to their respective conditioning [11]. This is referred to as *multiplicity.* Further to this, as the longitudinal data may be highly correlated within individuals, *collinearity* can affect the analysis. An additional concern is *data missingness*, which will often occur in this type of application.

The main issues surrounding multivariable regression are discussed in Section 7.1 and an illustrative application, regarding the relationship between childhood BMI development and late-adolescent adiposity in the Stockholm Weight Development Study, provided in Section 7.2.

7.1 Issues

The main issues to consider when using multivariable regression are problems with model interpretation due to multiplicity and collinearity, as well as the potential effects of missing data.

7.1.1 Multiplicity and collinearity

When using a multivariable model to relate a later outcome to several measurements of the same variable taken through childhood, the estimated papameters may be difficult to interpret, particularly if the if measurements are taken close together in time, due to their respective conditioning [11]. This is referred to as *multiplicity.*

Additionally, *collinearity* may further obscure interpretation. Collinearity occurs when there is an almost linear relationship between two or more explanatory variables in a multivariable regression analysis. Collinearity means that changes in one explanatory variable can effectively be compensated for by changes in other variables, so that very different sets of regression coefficients provide similarly good fits to the data [117]. This results in large standard errors corresponding to the estimated regression coefficients when near-collinear covariates are included in the same model, making interpretation difficult. Including collinear explanatory variables in a regression model can lead to the erroneous conclusion that the collinear variables are not associated with the outcome [116].

In longitudinal data, where data are merely earlier or later observations of the same variable for a given subject, both multiplicity and collinearity can be especially prevalent, with the likelihood of encountering problems increasing as the length of interval between successive measurements is reduced.

7.1.2 Missing data

Missing data can affect both the longitudinal growth data observed earlier in life and the outcome in later life. In the datasets utilised in this thesis (see Chapter 4), the childhood growth data are collated retrospectively from multiple sources, mainly through linkage to existing datasets, making them particularly liable to missingness. However, even if the childhood growth data were collected prospectively, missing data may still occur through attrition or for logistical reasons. The outcome variables in these studies are measured prospectively and specifically for the study, making data missingness less likely. Thus attention is focused more on missing values within the longitudinal growth data.

Missing values arise in longitudinal data whenever the sequences of measurements from one or more subjects are incomplete, in the sense that intended measurements are not available for some reason [118]. When faced with a longitudinal dataset with missing values the problem is then whether or not the values on an individual that are available can be used in an analysis and, if so, how [117]. One simple way to deal with this issue is to discard all incomplete sequences, an approach known as *complete-case analysis*, as described in Section 5.2.2.

The most appealing feature of complete-case analysis is its simplicity since it allows standard statistical analysis to be applied without modifications. However, discarding incomplete cases also results in a loss of information, which manifests itself as a loss of precision, and the potential introduction of bias. Precision is lost by virtue of the reduced sample size and bias may be introduced when the missing data mechanism is not one of 'missing completely at random' [120] (MCAR, see Section 5.2.1), meaning a systematic difference between those included in and those excluded from the analysis.

7.2 Investigating the relationship between childhood body mass index development and late-adolescent adiposity in the Stockholm Weight Development Study

In this section, a single-stage analysis approach is used to investigate the relationship between childhood BMI development and late-adolescent adiposity (in terms of both BMI and percentage body fat (%BF)) in the Stockholm Weight Development Study (SWEDES). A multivariable linear regression model is used which includes all available explanatory variables with no concern for the collinearity between them. Missing data are dealt with via a complete-case analysis approach. Whilst this type of naive analysis is often used, it is clearly far from ideal.

7.2.1 Introduction

A critical period for overweight or obesity is defined as a time associated with an increased risk of onset, complications or persistence of overweight or obesity [166]. Such periods are important to identify in order to target interventions to prevent children at high risk from becoming overweight or obese, especially given the widely-reported increases in prevalence of obesity over recent years.

Several periods during childhood have been suggested as critical for adverse adiposity development [74]. However, few studies have been able to examine the effects of BMI development throughout the entirety childhood. The aim of this study is to investigate critical periods of childhood BMI development for adiposity in late-adolescence.

The Stockholm Weight Development Study (SWEDES) is a prospective longitudinal study which provides a healthy contemporary birth cohort in which to investigate the relationships between childhood BMI development and late-adolescent adiposity. Annual weight and height measurements are available throughout childhood, allowing BMI development to be examined in detail. Measurement of many anthropometric variables at follow-up when the SWEDES participants were approximately 17 years old, in particular BMI and percentage body fat (%BF), provide measures of late-adolescent adiposity.

The SWEDES dataset brings with it the issue of missing data, particularly among the childhood BMI values. This is dealt with via a complete-case analysis approach, whereby all relevant childhood BMI values as well as the outcome variable are required for a subject to be included. Analysis is by standard multivariable regression.

This analysis formed part of the work presented at the 4th World Congress on Developmental Origins of Health and Disease (DOHaD), held 13-16 September 2006 at the University of Utrecht in The Netherlands [167].

7.2.2 Subjects

A general introduction to the Stockholm Weight Development Study (SWEDES) can be found in Section 4.1. As the present analysis focuses on the offspring as opposed to the mothers, the terms

'subject' and 'individual' will henceforth refer to the offspring in the study. Similarly, 'examination' should be taken to mean the occasion of the measurement of the anthropometric variables as part of the SWEDES follow-up when the offspring are approximately 17 years old.

For the purposes of the present analysis the data are reduced to the subset of annual observations between age 1 and 10 years inclusive. This is to avoid using as predictors variables that are temporally too close to the outcomes.

7.2.3 Methods

Multivariable regression models

BMI and %BF at examination are related to childhood BMI development using standard multivariable regression models, as introduced in Section 5.3.1. Childhood BMI development is defined as either each of the 10 childhood BMI observations or the 9 childhood BMI velocities calculated by taking the differences of consecutive BMI observations and dividing by the time between them (one year).

Each model includes either all 10 childhood BMI observations or BMI at age 1 year plus all 9 childhood BMI velocities, as well as age at examination. It is necessary to adjust for age at examination in this manner as both outcomes are age dependent meaning that the relationships could potentially be confounded by the age at which the measurements are taken. Models are fitted separately for males and females due to acknowledged differences in the BMI growth curves [68].

The model relating childhood BMI to BMI at examination in either males or females is thus

$$\begin{aligned} E(\mathrm{BMI_{exam}}) &= \beta_0 + \beta_1\mathrm{BMI_1} + \beta_2\mathrm{BMI_2} + \beta_3\mathrm{BMI_3} + \beta_4\mathrm{BMI_4} + \beta_5\mathrm{BMI_5} + \beta_6\mathrm{BMI_6} + \beta_7\mathrm{BMI_7} \\ &\quad + \beta_8\mathrm{BMI_8} + \beta_9\mathrm{BMI_9} + \beta_{10}\mathrm{BMI_{10}} + \delta\,\mathrm{age_{exam}} \\ &= \beta_0 + \sum_{i=1}^{10}\beta_i\mathrm{BMI}_i + \delta\,\mathrm{age_{exam}} \end{aligned} \tag{7.1}$$

where $\mathrm{BMI_{exam}}$ is BMI at examination, BMI_i is BMI at age i years, $i = 1, \ldots, 10$, and $\mathrm{age_{exam}}$ is age at examination. Here, β_i represents the conditional effect attributed to a 1 $\mathrm{kg/m^2}$ increase in BMI at age i years when all of the other variables (i.e. BMI at all other ages and age at examination) are held constant.

Similarly, the model relating childhood BMI velocity to BMI at examination in either males or females is

$$\begin{aligned} E(\mathrm{BMI_{exam}}) &= \gamma_0 + \gamma_1\mathrm{BMI_1} + \gamma_2\mathrm{BMIvel_2} + \gamma_3\mathrm{BMIvel_3} + \gamma_4\mathrm{BMIvel_4} + \gamma_5\mathrm{BMIvel_5} + \gamma_6\mathrm{BMIvel_6} \\ &\quad + \gamma_7\mathrm{BMIvel_7} + \gamma_8\mathrm{BMIvel_8} + \gamma_9\mathrm{BMIvel_9} + \gamma_{10}\mathrm{BMIvel_{10}} + \delta\,\mathrm{age_{exam}} \\ &= \gamma_0 + \gamma_1\mathrm{BMI_1} + \sum_{i=2}^{10}\gamma_i\mathrm{BMIvel}_i + \delta\,\mathrm{age_{exam}} \end{aligned} \tag{7.2}$$

where BMIvel_i is BMI velocity between age $i-1$ and i years, $i = 2, \ldots, 10$. Now γ_1 represents the effect of a 1 kg/m^2 increase in BMI at age 1 years controlled for all childhood BMI velocities and age at examination, whilst γ_i, $i = 2, \ldots, 10$, is the effect of a 1 kg/m^{2}year increase in BMI velocity between age $i-1$ and i years controlled for BMI age 1 year, all other BMI velocities and age at examination.

The models for %BF at examination are identical to (7.1) and (7.2) with the outcome changed to $\%\text{BF}_{\text{exam}}$.

As

$$\text{BMIvel}_i = \frac{\text{BMI}_i - \text{BMI}_{i-1}}{1} = \text{BMI}_i - \text{BMI}_{i-1},$$

(7.2) can be rewritten as

$$\begin{aligned} E(\text{BMI}_{\text{exam}}) &= \gamma_0 + \gamma_1 \text{BMI}_1 + \sum_{i=2}^{10} \gamma_i (\text{BMI}_i - \text{BMI}_{i-1}) + \delta\, \text{age}_{\text{exam}} \\ &= \gamma_0 + \sum_{i=1}^{9} (\gamma_i - \gamma_{i+1}) \text{BMI}_i + \gamma_{10} \text{BMI}_{10} + \delta\, \text{age}_{\text{exam}} \end{aligned} \tag{7.3}$$

Comparing (7.3) to (7.1) it can be seen that

$$\gamma_i = \beta_i \quad \text{for } i = 0 \text{ and } 10 \tag{7.4}$$

and

$$\gamma_i - \gamma_{i+1} = \beta_i \quad \text{for } i = 1, \ldots, 9.$$

Thus for $i = 1, \ldots, 9$,

$$\gamma_i = \beta_i + \gamma_{i+1} = \sum_{j=i}^{10} \beta_j. \tag{7.5}$$

So the coefficient for BMI velocity over each interval, γ_i, being the sum of all the conditional effects associated with BMI between age i and 10 years $(\beta_i, \ldots, \beta_{10})$, is the cumulative effect of increasing each BMI measurement between age i and 10 years by 1 kg/m^2. This is equivalent to an upwards shift of 1 kg/m^2 across the *entire* BMI trajectory from age i onwards.

Thus it can be seen that (7.1) and (7.2) are merely reparameterisations of the same model. This has been shown elsewhere for similar, though often less complex, models [130, 11, 46].

Since, additionally, $\gamma_{10} = \beta_{10} = \sum_{j=10}^{10} \beta_j$ from (7.4), this equivalence means that the reparameterisation in (7.2) can be rewritten as

$$E(\text{BMI}_{\text{exam}}) = \beta_0 + \beta_1 \text{BMI}_1 + \sum_{i=2}^{10} \left(\left(\sum_{j=i}^{10} \beta_j \right) \text{BMIvel}_i \right) + \delta\, \text{age}_{\text{exam}}.$$

Collinearity

Collinearity occurs when there is a near-linear relationship between two or more explanatory variables. If only two variables are involved then this is merely correlation so can be identified from the correlation matrix of the explanatory variables.

If collinearity occurs between more than two explanatory variables then this may not be obvious from their pairwise correlation coefficients. Collinearity can often be identified by a comparison of the standard errors of the regression coefficients for an explanatory variable in univariable and multivariable models. For example, the univariable model relating BMI at a given age in childhood to BMI at examination may be compared with the multivariable model relating BMI though childhood to BMI at examination. If collinearity is high then there will be a dramatic increase in the standard error of the regression coefficient [116].

Additionally, a more formal statistic which can be used to measure possible collinearity is the *variance inflation factor* (*VIF*). For a given explanatory variable, say BMI_i in (7.1), the VIF is defined as

$$VIF = \frac{1}{1 - R^2_{\text{BMI}_i}} \tag{7.6}$$

where $R^2_{\text{BMI}_i}$ is the proportion of the variability in BMI_i that is explained by the other variables when BMI_i is the dependent variable in a regression on all the remaining explanatory variables (BMI_j, $j \neq i$, and age_{exam}).

A VIF of 1 (which occurs when $R^2_{\text{BMI}_i} = 0$) indicates orthogonality of the explanatory variables, whilst a high VIF may imply a problem with collinearity. Suggested values of the VIF above which it is appropriate to be concerned with collinearity differ somewhat, with 'rule of thumb' cut off values of both 5 [129] and 10 [117] suggested.

Missing data

Using complete-case multivariable regression means that for an individual to be included in the analysis they must have data present for each variable in the model. Thus in the present analysis a subject must have all 10 childhood BMI values present as well as BMI (or %BF) and age at examination. When the number of explanatory variables in the model is large, as in this case, even if the proportion of missing data on each variable is small, this can result in a large proportion of individuals being excluded from the analysis.

In the SWEDES dataset all 481 subjects have BMI at examination recorded along with their age at this measurement and only seven individuals have missing %BF at examination. It is thus missing childhood BMI values which are the most troublesome in this instance. Table 7.1 shows the number of recorded childhood BMI observations for each subject. It can be seen that approximately half of individuals have recorded BMI values at all 10 ages ('complete childhood BMI data') and half have observed BMI at fewer that 10 ages ('incomplete childhood BMI data'). The percentage of subjects with complete childhood BMI data is virtually identical in males and females

(47.5 vs. 47.7). Around 20% of subjects have no childhood BMI data whatsoever. The next most frequent number of observed values is 4, occurring in 13.5% of subjects, which is quite anomalous given the infrequency with which similar numbers of observed values occur. Of those subjects with 4 observations virtually all are at ages 7, 8, 9 and 10 years (results not shown), indicating that data from their school journals (covering age 7 years onwards) are present, whilst data from their health care centre journals (covering ages up to 7 years) are not.

Number of observations	Frequency	Percentage
0	95	19.8
1	1	0.2
2	16	3.3
3	7	1.5
4	65	13.5
5	4	0.8
6	6	1.3
7	6	1.3
8	23	4.8
9	29	6.0
10	229	47.6
Total	481	100

Table 7.1: Number of recorded childhood body mass index observations per subject.

Table 7.2 compares the mean BMI values through childhood as well as the mean BMI and %BF at examination between subjects with complete childhood BMI data and those with incomplete childhood BMI data. The differences are seen to be relatively small for the childhood BMI values but more significant for both BMI and %BF at examination. Conducting t-tests between the two subgroups reinforces this observation: of the twenty t-tests for childhood BMI, only one is significant at the 5% level (females age 5 years, $P = 0.03$), which is what would be expected by chance alone, whereas all four t-tests for the measurements taken at examination are significant (BMI in males, $P = 0.02$; BMI in females, $P = 0.02$; %BF in males, $P < 0.001$; and %BF in females, $P = 0.05$). However, these comparisons between the complete and incomplete subgroups, particularly those for childhood BMI, must be treated with some caution as the number of subjects with incomplete data which contribute to them are often only a small proportion of those who should contribute (for example, only 14 of the 106 incomplete males contribute at age 6 years). This means that those individuals who do contribute may not be representative of the larger subgroup, rendering the comparison somewhat questionable.

In the present analysis, both BMI and %BF at examination are used as outcomes. Fig. 7.1

Variable	Males ($n = 202$)				Females ($n = 279$)			
	Complete ($n = 96$)		Incomplete ($n = 106$)		Complete ($n = 133$)		Incomplete ($n = 146$)	
	Frequency	Mean	Frequency	Mean	Frequency	Mean	Frequency	Mean
BMI (kg/m^2)								
at age (years)								
1	96	17.6	27	17.4	133	17.2	43	17.2
2	96	16.9	23	16.7	133	16.5	38	16.8
3	96	16.3	24	16.1	133	16.1	36	15.9
4	96	15.8	27	15.6	133	15.6	40	15.9
5	96	15.6	23	15.8	133	15.5	29	16.1
6	96	15.5	14	15.5	133	15.5	22	15.9
7	96	15.5	51	15.9	133	16.0	80	16.3
8	96	16.2	49	16.5	133	16.5	84	16.6
9	96	16.6	50	17.0	133	16.9	76	16.9
10	96	17.2	46	17.7	133	17.4	75	17.7
At examination								
BMI (kg/m^2)	96	20.6	106	21.6	133	21.0	146	21.9
%BF	95	14.3	105	18.0	130	28.6	144	30.1

Table 7.2: Comparison of mean body mass index (BMI) through childhood and mean body mass index and percentage body fat (%BF) at examination between subjects with complete childhood body mass index data and those with incomplete childhood body mass index data.

illustrates the univariate and bivariate distributions of these variables separately for males and females. Each plot is restricted to the subset of subjects with complete childhood BMI data. From the histograms it can be seen that there is perhaps a slight positive skew to the distributions of BMI, though %BF appears closer to normality. Meanwhile, the scatterplots show a clear positive association between the two measures of adiposity. Indeed, the correlations between the two dimensions, calculated using log-transformed BMI due to the skew, are 0.51 and 0.66 in males and females respectively.

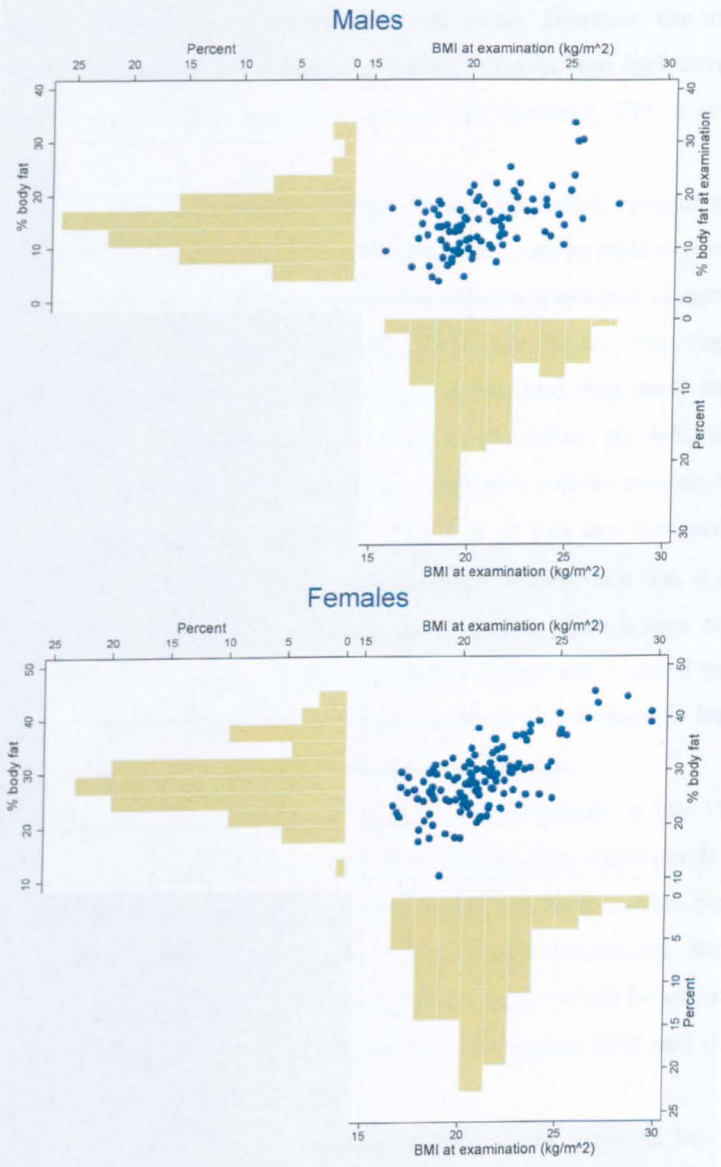

Fig. 7.1: Univariate and bivariate distributions of body mass index (BMI) and percentage body fat (%BF) at examination, by sex, for the 229 subjects with complete childhood body mass index data.

Software

The fitting of the multiple regression models is carried out using the `regress` procedure in Stata [147].

7.2.4 Results

Correlation between childhood BMI at different ages

As discussed in Section 7.2.3, collinear explanatory variables may potentially be identified from their correlation coefficients. Table 7.3 shows the correlation matrices for BMI through childhood, for males and females separately, restricted to those subjects with complete childhood BMI data. It can be seen for both sexes that BMI at a given ages is strongly and positively correlated with BMI at similar ages, meaning that subjects tend to retain a similar level of BMI relative to their contemporaries (BMI *tracking*) over a period of several years. However, the magnitude of these correlations diminishes as the interval between the measurements (the *lag*) increases. For a given lag, the strength of the correlation is seen to increase as age increases. This is evidence of stronger tracking at older ages.

Another way to view the correlation coefficients is via correlation contour plots, as shown in Fig. 7.2 and Fig. 7.3 for males and females, respectively. Here, age at BMI measurement is plotted on both axes with the correlation coefficient corresponding to a given pair of ages displayed by the appropriate colour according to the key on the right hand side. In this way, regions (i.e. pairs of ages) with similar levels of correlation will be the same colour and thus easily identifiable.

The patterns for males and females are seen to be largely similar. By definition, when the two ages being considered are the same the correlation coefficient will be one, so the points on the line $y = x$ will be the darkest shade of purple. Either side of this line the correlation is seen to decrease, albeit relatively slowly. The distance that a given contour line (i.e. a given correlation) lies away from the line $y = x$ tends to increase as age increases. This is seen most clearly in the females. This means that, for example, the correlation between age 2 and 3 years (a lag of one year) is approximately equal to the correlation between age 6 and 8 years (a lag of two years) in females, and is again evidence of BMI tracking increasing with age.

In both plots there is a region around 6–7 years where correlation is less than would be expected. There are two possible explanations for this. Firstly, this corresponds to the age when the measurement of height and weight are transferred from the child welfare centre to schools so this could imply some level of discontinuity in the measurement procedures. Secondly, this is the age around which the adiposity rebound (AR, see Section 2.3.3) would be expected to occur. At this age those individuals who are pre-AR will still have decreasing BMI and those post-AR will be increasing so correlation is likely to be reduced.

Whilst the observed features of the correlation contour plots could all be deduced from the values in the correlation matrices, the graphical display does aid interpretation.

The high levels of correlation, particularly between measurements only one year apart, means that in a multiple regression model including BMI at all ages collinearity may potentially be a problem.

Males ($n = 96$)										
BMI at	BMI at age (years)									
age (years)	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.75	1								
3	0.63	0.77	1							
4	0.58	0.76	0.80	1						
5	0.58	0.71	0.73	0.82	1					
6	0.51	0.58	0.65	0.69	0.89	1				
7	0.55	0.59	0.59	0.61	0.72	0.75	1			
8	0.42	0.52	0.51	0.59	0.73	0.71	0.90	1		
9	0.39	0.51	0.50	0.54	0.69	0.68	0.86	0.95	1	
10	0.36	0.48	0.43	0.46	0.60	0.60	0.81	0.90	0.94	1

Females ($n = 133$)										
BMI at	BMI at age (years)									
age (years)	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.64	1								
3	0.66	0.82	1							
4	0.64	0.73	0.84	1						
5	0.53	0.64	0.73	0.88	1					
6	0.47	0.59	0.66	0.80	0.92	1				
7	0.44	0.53	0.60	0.73	0.83	0.83	1			
8	0.34	0.46	0.52	0.65	0.78	0.80	0.92	1		
9	0.31	0.43	0.48	0.62	0.75	0.79	0.89	0.94	1	
10	0.30	0.37	0.41	0.56	0.67	0.72	0.82	0.85	0.93	1

Table 7.3: Estimated correlation coefficients between body mass index (BMI) at different ages through childhood, by sex.

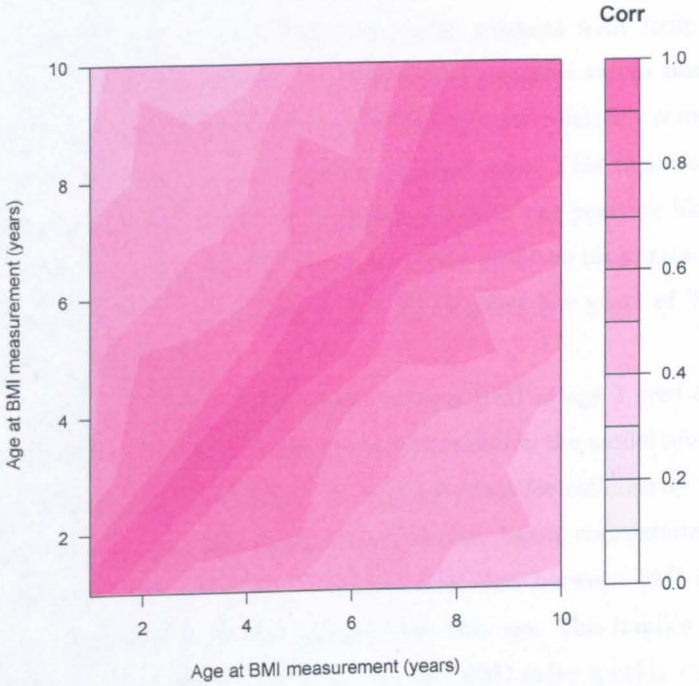

Fig. 7.2: Correlation contour plot for body mass index (BMI) through childhood in males.

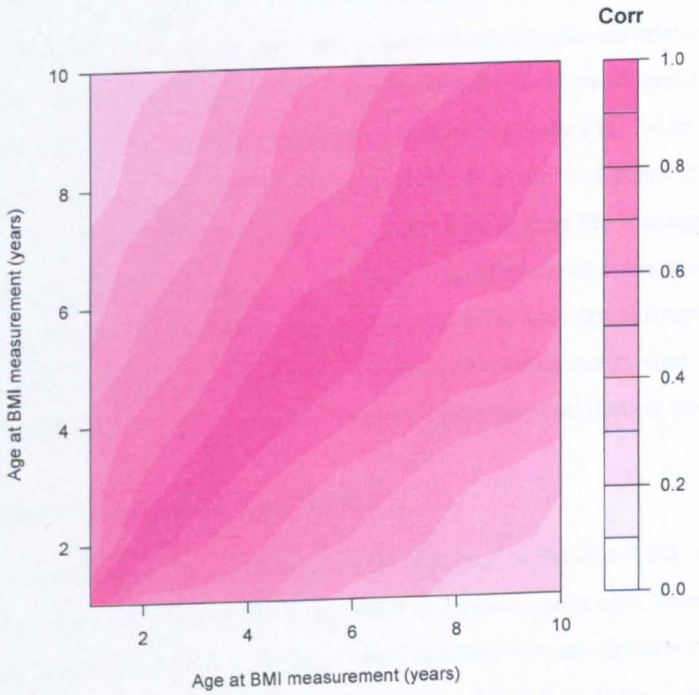

Fig. 7.3: Correlation contour plot for body mass index (BMI) through childhood in females.

Correlation between childhood BMI velocity at different ages

Table 7.4 displays the correlation matrices for BMI velocity at each age through childhood, separately for males and females, again restricted to those subjects with complete childhood BMI data. The correlation coefficients take both positive and negative values but are generally close to zero. There are few obvious features, though a tendency for relatively strong negative correlations corresponding to a 1 year lag can be seen at younger ages. This signifies that those subjects who increase BMI more quickly than their contemporaries in one year are likely to increase BMI relatively less quickly the next year, and vice versa. This could be interpreted as subjects having 'spurts' of increasing BMI at different times through the first few years of life but all obtaining similar levels of BMI eventually.

Also shown in Table 7.4 are the correlations between BMI at age 1 year and BMI velocity at each age through childhood. As BMI at age 1 year is included in the model relating childhood BMI velocity to late-adolescent BMI or %BF, (7.2), the potential for collinearity between BMI age 1 year and BMI velocity at any age should also be considered. Again, correlations are generally weak, with the only discernible pattern being a negative correlation between BMI age 1 year and BMI velocity in the first year (females) or two (males) after this age. This implies that the individuals who have low BMI at age 1 year are likely to increase BMI more quickly over the next year or two, and vice versa. Again, this can be considered in terms of spurts of growth at different ages in infancy against a backdrop of similar underlying growth patterns.

Fig. 7.4 and Fig. 7.5 show the correlation contour plots for BMI velocity through childhood for males and females, respectively. The colours used are the same as those in Fig. 7.2 and Fig. 7.3 so that the plots are directly comparable. Aside from the high correlation along the line $y = x$, which is equal to 1 by definition, there is a lack of any regions of meaningful correlation. The only obvious pattern in the correlation which is present is the previously mentioned negative correlations at 1 year lag through infancy (note that the seemingly positive region in Fig. 7.4 in infancy is an artifact of the positive correlation between BMI velocity at age 1–2 and 3–4 years rather than a positive correlation at 1 year lag). This trend appears to extend later into life amongst males.

The generally weak correlations seen between the variables mean there there is less likely to be problems with collinearity when BMI at age 1 year and BMI velocity through childhood are used as explanatory variables in (7.2). However, it should be remembered that collinearity between more than two variables may not be evident from the pairwise correlation coefficients.

Childhood BMI and late-adolescent BMI

Before applying the multivariable regression models detailed in Section 7.2.3, in which the relationship between late-adolescent BMI or %BF and BMI at a given age in childhood is adjusted for BMI at every other age in childhood, it is useful to first consider the *unadjusted* relationships between the late-adolescent outcomes and BMI at each age in childhood in turn. These relationships can be assessed using separate regression models, all of which should again include age at examination. The results obtained, when compared to their adjusted equivalents from the multivariable

Males ($n = 96$)									
BMI velocity	BMI at	BMI velocity at age (years)							
at age (years)	age 1 year	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9
1–2	−0.26	1							
2–3	−0.25	−0.32	1						
3–4	−0.17	0.06	−0.43	1					
4–5	0.03	−0.11	−0.05	−0.25	1				
5–6	0.04	−0.13	0.20	−0.17	0.18	1			
6–7	0.10	−0.02	−0.17	−0.00	−0.20	−0.30	1		
7–8	−0.13	0.18	−0.01	0.20	0.16	−0.14	−0.10	1	
8–9	0.05	0.14	−0.04	−0.16	0.16	0.11	0.06	0.01	1
9–10	0.02	0.07	−0.19	0.01	0.00	0.02	0.20	0.15	0.04

Females ($n = 133$)									
BMI velocity	BMI at	BMI velocity at age (years)							
at age (years)	age 1 year	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9
1–2	−0.51	1							
2–3	−0.06	−0.43	1						
3–4	0.07	−0.10	−0.17	1					
4–5	−0.22	0.08	−0.03	−0.06	1				
5–6	−0.06	0.06	−0.08	0.01	0.04	1			
6–7	0.12	−0.03	0.03	0.08	−0.03	−0.21	1		
7–8	−0.16	0.10	0.01	0.00	0.31	0.08	−0.12	1	
8–9	−0.01	0.05	−0.03	0.14	0.06	0.13	0.01	−0.09	1
9–10	0.06	−0.10	−0.01	0.09	−0.04	0.09	0.09	−0.10	0.21

Table 7.4: Estimated correlation coefficients between body mass index (BMI) at age 1 year and body mass index velocity at different ages through childhood, by sex.

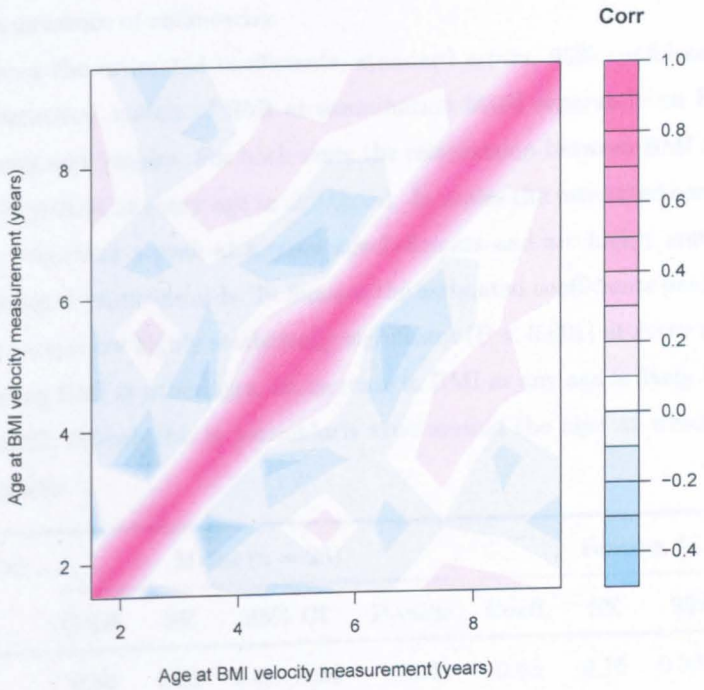

Fig. 7.4: Correlation contour plot for body mass index (BMI) velocity through childhood in males.

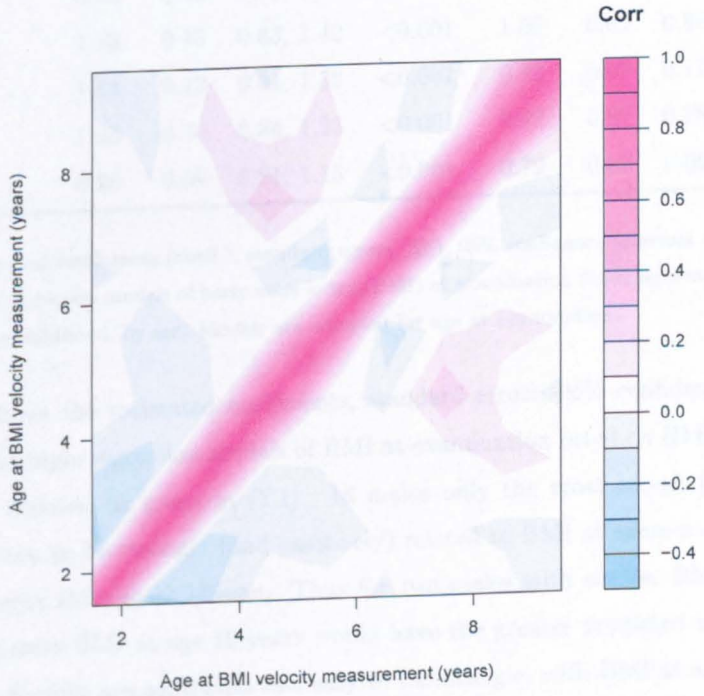

Fig. 7.5: Correlation contour plot for body mass index (BMI) velocity through childhood in females.

regression models, allow an assessment of the effect of adjustment for BMI at the other ages in childhood and a comparison of the standard errors for the equivalent estimated coefficients may help indicate the presence of collinearity.

Table 7.5 shows the estimated coefficients, standard errors, 95% confidence intervals and P-values for the regression models of BMI at examination fitted separately on BMI at each age in childhood for males and females. For both sexes the relationship between BMI at examination and childhood BMI is positive at every age in childhood. In males the estimated coefficients increase in magnitude before reaching a peak at around age 7–8 years and are highly statistically significant ($P < 0.001$) from age 5 years onwards. In females the estimated coefficients peak at little earlier, at around age 6 years, and are highly statistically significant ($P < 0.001$) at every age examined. This shows that, ignoring BMI at other ages, an increase in BMI at any age is likely to lead to increased late-adolescent BMI, though this is particularly true around the ages at which the magnitude of the coefficients peaks.

BMI (kg/m^2) at	Males ($n = 96$)				Females ($n = 133$)			
age (years)	Coeff.	SE	95% CI	P-value	Coeff.	SE	95% CI	P-value
1	0.50	0.22	0.06, 0.94	0.03	0.63	0.15	0.33, 0.92	<0.001
2	0.61	0.20	0.22, 1.01	0.003	0.97	0.15	0.67, 1.27	<0.001
3	0.50	0.22	0.07, 0.94	0.03	0.99	0.17	0.65, 1.33	<0.001
4	0.78	0.23	0.32, 1.25	0.001	0.99	0.15	0.69, 1.29	<0.001
5	0.93	0.22	0.49, 1.37	<0.001	1.15	0.14	0.87, 1.43	<0.001
6	0.85	0.18	0.48, 1.21	<0.001	1.19	0.12	0.95, 1.43	<0.001
7	1.13	0.15	0.83, 1.42	<0.001	1.06	0.09	0.88, 1.23	<0.001
8	1.14	0.12	0.91, 1.38	<0.001	0.93	0.09	0.77, 1.10	<0.001
9	1.05	0.10	0.84, 1.25	<0.001	0.92	0.07	0.78, 1.07	<0.001
10	0.98	0.09	0.81, 1.15	<0.001	0.79	0.07	0.66, 0.92	<0.001

Table 7.5: Estimated coefficients (coeff.), standard errors (SE), 95% confidence intervals (CI) and P-values from Wald tests for the regression models of body mass index (BMI) at examination fitted separately on body mass index at different ages in childhood, by sex. Models are adjusted for age at examination.

Table 7.6 shows the estimated coefficients, standard errors, 95% confidence intervals and P-values for the multiple regression models of BMI at examination fitted on BMI through childhood for males and females, as given in (7.1). In males only the most recent (age 10 years) BMI observation is seen to be strongly (and positively) related to BMI at examination, conditional on BMI at other ages through childhood. Thus for two males with similar BMI age 1 to 9 years, the one with greater BMI at age 10 years would have the greater predicted late-adolescent BMI. The results for females are somewhat less easy to disentangle, with BMI at ages 2, 7 and 9 years positively associated with BMI at examination but BMI at age 8 years negatively associated, meaning that having lower BMI at this age tends to lead to higher late-adolescent BMI. The lack

of a positive relationship between BMI at the last available age (10 years) and BMI at examination in females is likely to be due to confounding by pubertal stage. At this age, those females who are more developmentally advanced will experience the adolescent growth spurt in height [168] meaning that their BMI is reduced. As an early menarche is known to be associated with later obesity [75], these same females are likely to have a higher BMI at examination. Thus, for different reasons, females with both low and high BMI at age 10 years could be considered at risk for higher late-adolescent BMI, with the two effects cancelling each other out. This does not appear to be an issue for males, which may be expected given that the adolescent growth spurt occurs on average two years later in males than in females [168].

BMI (kg/m^2) at	Males ($n = 96$)				Females ($n = 133$)			
age (years)	Coeff.	SE	95% CI	P-value	Coeff.	SE	95% CI	P-value
1	0.06	0.24	−0.42, 0.55	0.80	0.09	0.14	−0.18, 0.37	0.51
2	0.02	0.29	−0.55, 0.59	0.95	0.56	0.19	0.18, 0.94	0.004
3	−0.44	0.28	−1.01, 0.12	0.12	0.00	0.27	−0.53, 0.52	0.99
4	0.45	0.36	−0.26, 1.16	0.21	−0.54	0.30	−1.14, 0.06	0.08
5	−0.71	0.48	−1.67, 0.25	0.14	−0.35	0.35	−1.05, 0.35	0.33
6	0.30	0.33	−0.35, 0.95	0.36	0.44	0.27	−0.09, 0.97	0.10
7	−0.14	0.34	−0.82, 0.55	0.69	0.61	0.24	0.13, 1.09	0.01
8	0.44	0.44	−0.43, 1.32	0.32	−0.61	0.28	−1.17, −0.06	0.03
9	0.09	0.43	−0.76, 0.94	0.84	0.79	0.29	0.21, 1.37	0.01
10	0.76	0.28	0.20, 1.33	0.01	0.16	0.16	−0.16, 0.49	0.32

Table 7.6: Estimated coefficients (coeff.), standard errors (SE), 95% confidence intervals (CI) and P-values from Wald tests for the multivariable regression models of body mass index (BMI) at examination fitted on body mass index through childhood, by sex. Models are adjusted for age at examination.

A comparison of the adjusted regression coefficients in Table 7.6 with their unadjusted equivalents in Table 7.5 illustrates the large impact that adjustment for BMI at other ages through childhood has. Whilst the unadjusted coefficients are all positive and largely highly statistically significant, their adjusted equivalents are markedly different, with some suggesting a negative relationship and few providing any strong support (i.e. a highly significant P-value) for a relationship in either direction. Additionally, although at younger ages the standard errors of the estimated coefficients are similar in the unadjusted and adjusted models, at older ages they are up to 3–4 times as great in the adjusted models.

Another way to present the results in Table 7.6 which may help in interpreting the conditional impact of each repeated measure [11] is in a life course plot [130], introduced in Section 5.3.1. Here, the regression coefficients are re-estimated using standardised childhood BMI values to provide comparable coefficients which are then plotted against age. The upper plot in Fig. 7.6 makes it clear that in males only the most recent BMI observation has any meaningful relationship with BMI

at examination. When the coefficients in a life course plot switch sign between two ages there is evidence that changes in the explanatory variable over this interval affect the outcome of interest [11]. Thus the corresponding plot for females, as well as reinforcing the previous observations, suggests that a relative reduction in BMI between age 2 and 4 years, a relative increase between age 4 and 7 years, a relative decrease between age 7 and 8 years and a relative increase between age 8 and 9 years are all associated with higher BMI at examination. It is also apparent from the life course plots that the confidence intervals (CIs) for the estimated coefficients tend to increase as age increases.

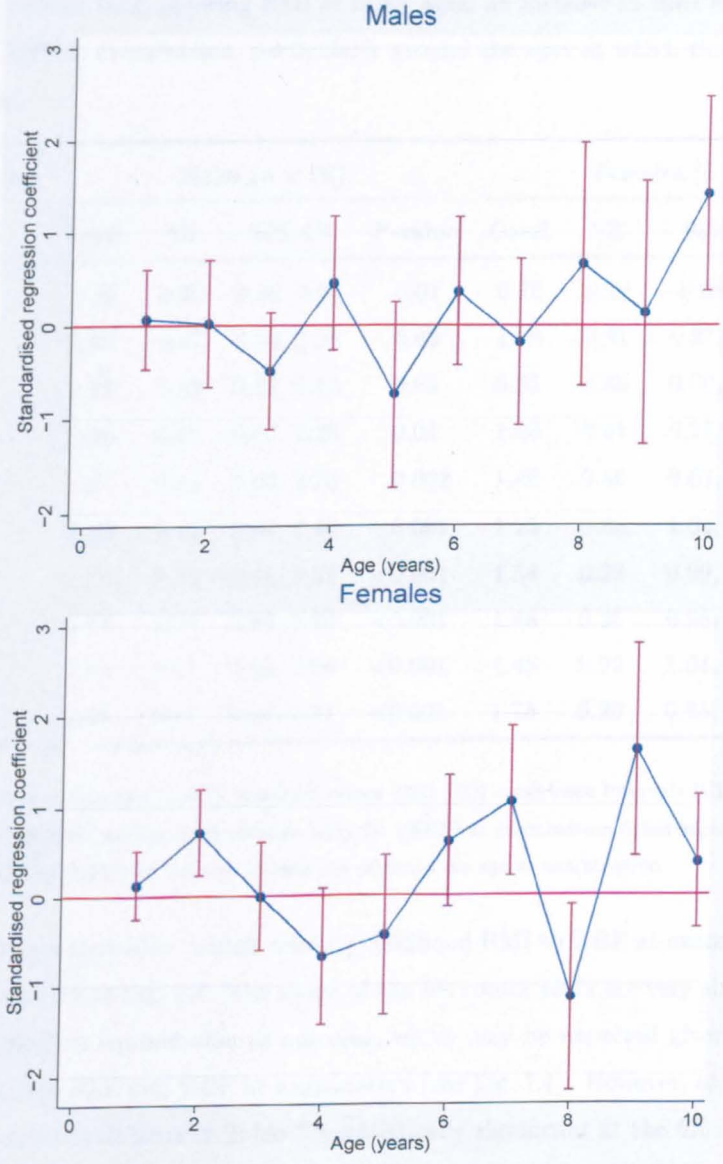

Fig. 7.6: Life course plots for models of body mass index (BMI) at examination on body mass index through childhood for males (upper plot) and females (lower plot).

Childhood BMI and late-adolescent %BF

Table 7.7 shows the estimated coefficients, standard errors, 95% confidence intervals and P-values for the regression models of %BF at examination fitted separately on BMI at each age in childhood for males and females. As with BMI at examination, the relationship between %BF at examination and childhood BMI is positive at every age in childhood for both sexes. Again, the estimated coefficients increase in magnitude through early childhood, reaching a peak at around age 7 years in males and age 6 years in females, before decreasing a little. In males the coefficients are highly statistically significant ($P < 0.001$) from age 7 years onwards and from age 5 years in females. This provides evidence that, ignoring BMI at other ages, an increase in BMI at any age tends to lead increased %BF at examination, particularly around the ages at which the magnitude of the coefficients peaks.

BMI (kg/m^2) at	Males ($n = 96$)				Females ($n = 133$)			
age (years)	Coeff.	SE	95% CI	P-value	Coeff.	SE	95% CI	P-value
1	1.30	0.50	0.30, 2.29	0.01	0.70	0.38	−0.06, 1.45	0.07
2	1.02	0.47	0.10, 1.95	0.03	1.08	0.41	0.27, 1.90	0.01
3	1.12	0.50	0.12, 2.12	0.03	0.91	0.46	0.00, 1.81	0.05
4	1.16	0.55	0.07, 2.25	0.04	1.03	0.41	0.21, 1.85	0.02
5	1.67	0.52	0.64, 2.70	0.002	1.46	0.40	0.68, 2.25	<0.001
6	1.53	0.44	0.66, 2.40	0.001	1.75	0.35	1.04, 2.45	<0.001
7	1.76	0.39	0.98, 2.54	<0.001	1.54	0.28	0.99, 2.09	<0.001
8	1.57	0.34	0.89, 2.26	<0.001	1.48	0.25	0.98, 1.98	<0.001
9	1.45	0.31	0.85, 2.06	<0.001	1.48	0.22	1.04, 1.93	<0.001
10	1.38	0.27	0.84, 1.91	<0.001	1.25	0.20	0.85, 1.66	<0.001

Table 7.7: Estimated coefficients (coeff.), standard errors (SE), 95% confidence intervals (CI) and P-values from Wald tests for the regression models of percentage body fat (%BF) at examination fitted separately on body mass index at different ages in childhood, by sex. Models are adjusted for age at examination.

Results for the multivariable models relating childhood BMI to %BF at examination are given in Table 7.8 and plotted in Fig. 7.7. The shape of the life course plots are very similar to those for the models with BMI at examination as outcome, which may be expected given the high degree of correlation between BMI and %BF at examination (see Fig. 7.1). However, at no ages in either sex are the estimated coefficients in Table 7.8 statistically significant at the 5% level, contrasting markedly with the estimated coefficients in Table 7.6. In the models with BMI at examination as outcome it may be expected that stronger relationships be found as the outcome is merely a later measurement of the explanatory variable. With %BF as outcome, however, this is not the case.

An alternative way to consider this is that, with %BF as outcome, we would ideally like to have longitudinal observations of childhood %BF as exposures. Similarly to the model with BMI at examination as outcome and childhood BMI observations as explanatory variables, both outcome

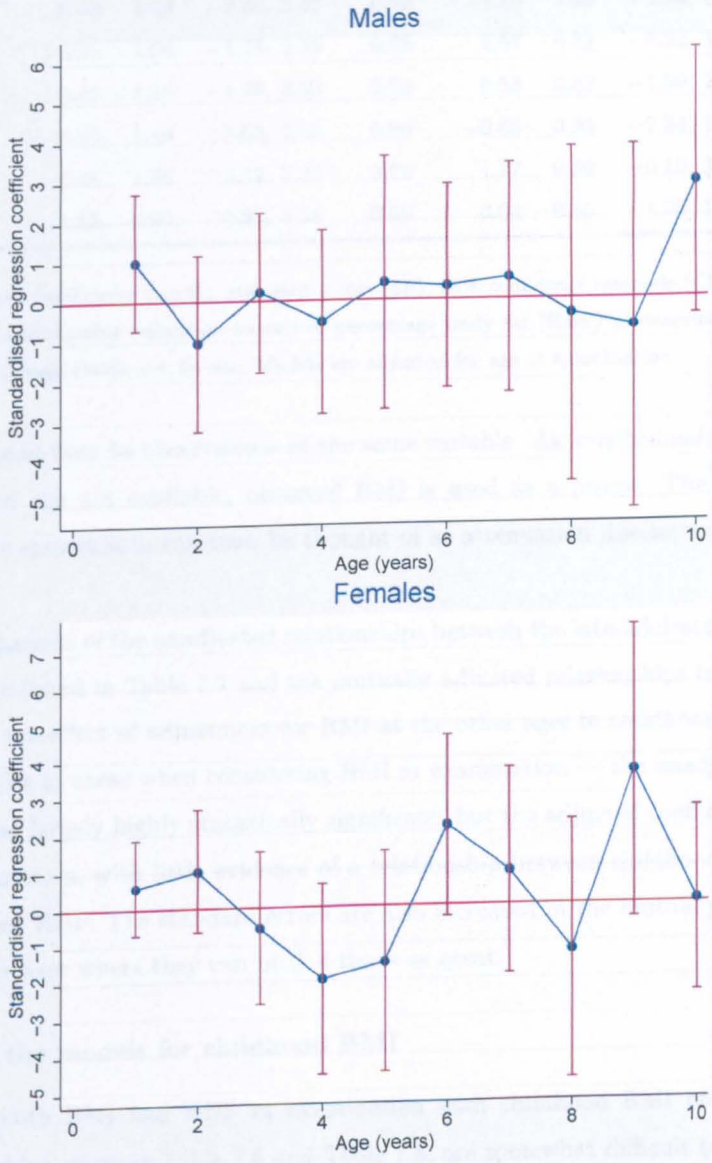

Fig. 7.7: Life course plots for models of percentage body fat (%BF) at examination fitted on body mass index (BMI) through childhood for males (upper plot) and females (lower plot).

BMI (kg/m^2) at	Males ($n = 95$)				Females ($n = 130$)			
age (years)	Coeff.	SE	95% CI	P-value	Coeff.	SE	95% CI	P-value
1	0.90	0.77	−0.64, 2.45	0.25	0.43	0.47	−0.51, 1.36	0.37
2	−0.85	0.91	−2.67, 0.96	0.35	0.80	0.66	−0.50, 2.10	0.23
3	0.20	0.90	−1.60, 2.00	0.83	−0.48	0.91	−2.27, 1.32	0.60
4	−0.51	1.13	−2.76, 1.74	0.65	−1.56	1.04	−3.61, 0.50	0.14
5	0.43	1.53	−2.60, 3.47	0.78	−1.19	1.19	−3.54, 1.16	0.32
6	0.29	1.04	−1.78, 2.35	0.78	1.57	0.91	−0.22, 3.37	0.09
7	0.42	1.10	−1.76, 2.60	0.70	0.53	0.82	−1.09, 2.16	0.52
8	−0.25	1.40	−3.03, 2.53	0.86	−0.66	0.95	−2.54, 1.21	0.49
9	−0.41	1.36	−3.12, 2.30	0.76	1.77	0.99	−0.19, 3.72	0.08
10	1.55	0.90	−0.23, 3.34	0.09	0.04	0.55	−1.05, 1.13	0.95

Table 7.8: Estimated coefficients (coeff.), standard errors (SE), 95% confidence intervals (CI) and P-values from Wald tests for the multivariable regression models of percentage body fat (%BF) at examination fitted on body mass index (BMI) through childhood, by sex. Models are adjusted for age at examination.

and exposures would then be observations of the same variable. As longitudinal measures of %BF through childhood are not available, observed BMI is used as a proxy. The reduction in the significance of the associations can then be thought of as attenuation due to 'measurement error' in the exposures.

Again, a comparison of the unadjusted relationships between the late-adolescent %BF and BMI at each age in childhood in Table 7.7 and the mutually adjusted relationships in Table 7.8 allows an assessment of the effect of adjustment for BMI at the other ages in childhood. The differences observed are similar to those when considering BMI at examination — the unadjusted coefficients are all positive and largely highly statistically significant, but the adjusted coefficients do not have such a coherent pattern, with little evidence of a relationship between childhood BMI at any age and late-adolescent %BF. The standard errors are also increased in the mutually adjusted model, especially at older ages where they can be 3–4 times as great.

Collinearity in the models for childhood BMI

The models for both BMI and %BF at examination with childhood BMI observations as the explanatory variables, given in Table 7.6 and Table 7.8, are somewhat difficult to interpret due to the estimated coefficients often changing sign and having large standard errors. Whilst difficulty of interpretation is always likely in multiple regression models including repeated measures taken close together in time due to multiplicity, collinearity may provide an alternative explanation for unexpected values of regression coefficients and large standard errors.

The high correlations already observed amongst the explanatory variables (see Table 7.3) may indicate that collinearity will be a problem in the multiple regression models. Additionally, comparisons of the multiple regression models with the univariate regression models (i.e. comparing

Table 7.6 with Table 7.5, and Table 7.8 with Table 7.7) shows the standard errors of the estimated regression coefficients to increase by up to 3–4 times. Again, this suggests that there may be collinearity present in the explanatory variables.

A more formal approach to the identification of potential collinearity is through use of the variance inflation factor (VIF), introduced in Section 7.2.3. Table 7.9 shows the calculated VIF for BMI at each age through childhood according to (7.6). It can be seen that for both sexes there are several ages when the VIF is greater than 10, and it is greater than 5 for the majority of ages. This again suggests that the results of the models with childhood BMI as explanatory variables may be affected by collinearity.

BMI at	VIF	
age (years)	Males ($n = 96$)	Females ($n = 133$)
1	2.7	2.0
2	4.4	3.2
3	3.7	5.0
4	4.9	8.0
5	9.4	11.0
6	6.0	7.4
7	7.6	9.5
8	16.2	14.9
9	19.1	19.4
10	10.4	7.7

Table 7.9: Variance inflation factor (VIF) for body mass index (BMI) at each age through childhood, by sex.

It should be noted that the correlations between the standardised childhood BMI values used in the life course plots (Fig. 7.6 and Fig. 7.7) will be identical to those between the unstandardised values in the original multivariable regression models (Table 7.6 and Table 7.8), and thus any collinearity between the explanatory variables will not be affected by the standardisation.

Childhood BMI velocity and late-adolescent BMI

Again, it is informative to look at the unadjusted as well as adjusted regression models. Table 7.10 shows the estimated coefficients, standard errors, 95% confidence intervals and P-values for the regression models of BMI at examination fitted separately on BMI velocity at each age in childhood for males and females. In males, there is little evidence of a relationship between BMI velocity and BMI at examination before age 5 years, though after this there is fairly strong evidence of a positive relationship, especially between age 7 and 8 years. In females, there is evidence of a positive relationship between BMI velocity and BMI at examination from age 3 years onwards. This relationship is seen to be particularly strong between age 5 and 9 years, apart from the rather anomalous result for age 7–8 years. The negative coefficient corresponding to BMI velocity

between age 2 and 3 years in both males and females should also be noted. Although not strong, there is some evidence that a high BMI velocity over this period is associated with a lower BMI at examination.

BMI velocity (kg/m^2year)	Males ($n = 96$)				Females ($n = 133$)			
at age (years)	Coeff.	SE	95% CI	P-value	Coeff.	SE	95% CI	P-value
1–2	0.43	0.31	−0.18, 1.04	0.16	0.28	0.19	−0.10, 0.66	0.15
2–3	−0.44	0.31	−1.06, 0.18	0.17	−0.49	0.29	−1.07, 0.09	0.10
3–4	0.42	0.37	−0.31, 1.15	0.25	0.61	0.32	−0.02, 1.23	0.06
4–5	0.54	0.40	−0.26, 1.34	0.19	0.75	0.36	0.05, 1.46	0.04
5–6	0.84	0.43	−0.02, 1.69	0.06	1.24	0.38	0.48, 2.00	0.002
6–7	0.80	0.26	0.28, 1.33	0.003	0.96	0.22	0.53, 1.39	<0.001
7–8	1.45	0.35	0.75, 2.15	<0.001	0.29	0.31	−0.33, 0.90	0.36
8–9	1.30	0.45	0.41, 2.19	0.01	1.31	0.31	0.70, 1.92	<0.001
9–10	1.29	0.40	0.53, 2.04	0.001	0.45	0.25	−0.05, 0.95	0.08

Table 7.10: Estimated coefficients (coeff.), standard errors (SE), 95% confidence intervals (CI) and P-values from Wald tests for the regression models of body mass index (BMI) at examination fitted separately on body mass index velocity at different ages in childhood, by sex. Models are adjusted for age at examination.

Table 7.11 shows the estimated coefficients, standard errors, 95% confidence intervals and P-values for the multiple regression models of BMI at examination fitted on BMI at age 1 year and BMI velocity through childhood for males and females, as given in (7.2). For both males and females BMI velocity at virtually every age is seen to be positively associated with BMI at examination (conditional on BMI at age 1 year and BMI velocity at every other age), though the strength of the relationship varies with age. In both sexes BMI at age 1 year is also strongly positively associated with BMI at examination conditional on BMI velocity through childhood showing that for a given childhood BMI trajectory those with higher BMI at age 1 year are likely to have higher BMI in late adolescence.

A comparison of the unadjusted regression models in Table 7.10 with the mutually adjusted regression models in Table 7.11 shows that adjustment for BMI velocity at other ages (as well at BMI age 1 year) generally leads to larger estimated coefficients and greater statistical significance at younger ages, but smaller coefficients and reduced statistical significance at older ages. This means that having a high BMI velocity at a younger age is not a very good predictor for late-adolescent BMI when taken on its own, but a high BMI velocity at a younger age will tend to lead higher BMI at examination for a fixed pattern of BMI velocity at older ages. Similarly, whilst a high BMI velocity at older ages is a relatively good predictor of high BMI at examination when taken on its own, perhaps as it is indicative of individuals who have a high BMI velocity throughout childhood, when it is considered in conjunction with a fixed pattern of BMI velocity up to that age

	Males (n = 96)				Females (n = 133)			
	Coeff.	SE	95% CI	P-value	Coeff.	SE	95% CI	P-value
BMI (kg/m^2) at age 1 year	0.83	0.18	0.47, 1.19	<0.001	1.16	0.13	0.91, 1.41	<0.001
BMI velocity (kg/m^{2}year) at age (years)								
1–2	0.77	0.26	0.26, 1.28	0.004	1.07	0.17	0.73, 1.40	<0.001
2–3	0.75	0.32	0.13, 1.38	0.02	0.50	0.23	0.06, 0.95	0.03
3–4	1.20	0.34	0.52, 1.88	0.001	0.51	0.21	0.09, 0.92	0.02
4–5	0.75	0.32	0.12, 1.38	0.02	1.05	0.24	0.57, 1.53	<0.001
5–6	1.46	0.32	0.82, 2.10	<0.001	1.40	0.26	0.89, 1.90	<0.001
6–7	1.16	0.20	0.75, 1.56	<0.001	0.95	0.15	0.66, 1.25	<0.001
7–8	1.29	0.29	0.72, 1.87	<0.001	0.35	0.21	−0.06, 0.75	0.10
8–9	0.85	0.32	0.21, 1.49	0.01	0.96	0.21	0.54, 1.38	<0.001
9–10	0.76	0.28	0.20, 1.33	0.01	0.16	0.16	−0.16, 0.49	0.32

Table 7.11: Estimated coefficients (coeff.), standard errors (SE), 95% confidence intervals (CI) and P-values from Wald tests for the multivariable regression models of body mass index (BMI) at examination fitted on body mass index velocity through childhood, by sex. Models are adjusted for age at examination.

the association is reduced. Additionally, the standard errors of the estimated regression coefficients in the multivariable models are generally very similar to those for the univariable models — indeed, the standard errors are often *reduced* in the mutually adjusted models.

In both sexes, though particularly in females, the period between age 1 and 2 years appears important for the development of late-adolescent BMI. In males a further key period is between 5 and 8 years of age, whilst females have a corresponding period between age 4 and 7 years. Both of these intervals cover ages when subjects will be experiencing the AR and this slight sex-specific difference corresponds to the earlier AR often observed in females [68]. Because BMI is at a minimum at AR, at ages either side of AR an increased BMI velocity is indicative of an earlier AR. This is borne out in Fig. 7.8 which shows artificially created BMI trajectories for the ages around AR (upper plot) and the corresponding BMI velocities (lower plot). The solid line in the upper plot represents a subject with an AR at age 6.5 years. The BMI velocity corresponding to this trajectory (in this case the derivative of the BMI function) is plotted, also with a solid line, in the lower plot. The age at which the BMI velocity crosses the x-axis corresponds to the age at AR. The dashed line in the lower plot corresponds to a subject with a BMI velocity that is consistently greater than that of the first subject. This results in the dashed line crossing the x-axis, and hence this individual having their AR, at a younger age (5.5 years). The dashed line in the upper plot shows a BMI trajectory that would correspond to this BMI velocity. Whilst this explanation is somewhat contrived, it serves to illustrate that increased BMI velocity around the AR is associated with an earlier AR. So, given that it is often suggested (see Section 2.3.3) that an earlier AR is predictive of higher later adiposity, it should be no surprise that increased BMI velocity around the AR is associated with higher BMI at examination.

It should also be noted that the estimated coefficients in Table 7.11, when compared to those in Table 7.6, conform to (7.4) and (7.5), as has previously been discussed by De Stavola *et al* [11]. For example, in males,

$$\gamma_1 = 0.83 = 0.06 + 0.02 - 0.44 + 0.45 - 0.71 + 0.30 - 0.14 + 0.44 + 0.09 + 0.76 = \sum_{j=1}^{10} \beta_j$$

and

$$\gamma_2 = 0.77 = 0.02 - 0.44 + 0.45 - 0.71 + 0.30 - 0.14 + 0.44 + 0.09 + 0.76 = \sum_{j=2}^{10} \beta_j.$$

Childhood BMI velocity and late-adolescent %BF

Table 7.12 shows the estimated coefficients, standard errors, 95% confidence intervals and P-values for the regression models of %BF at examination fitted separately on BMI velocity at each age in childhood for males and females. There is little evidence of an association between %BF and BMI velocity before age 4 years in either sex, though after this age there is a more obvious positive relationship. Evidence for the association is stronger in males age 4–5 and 8–10 years

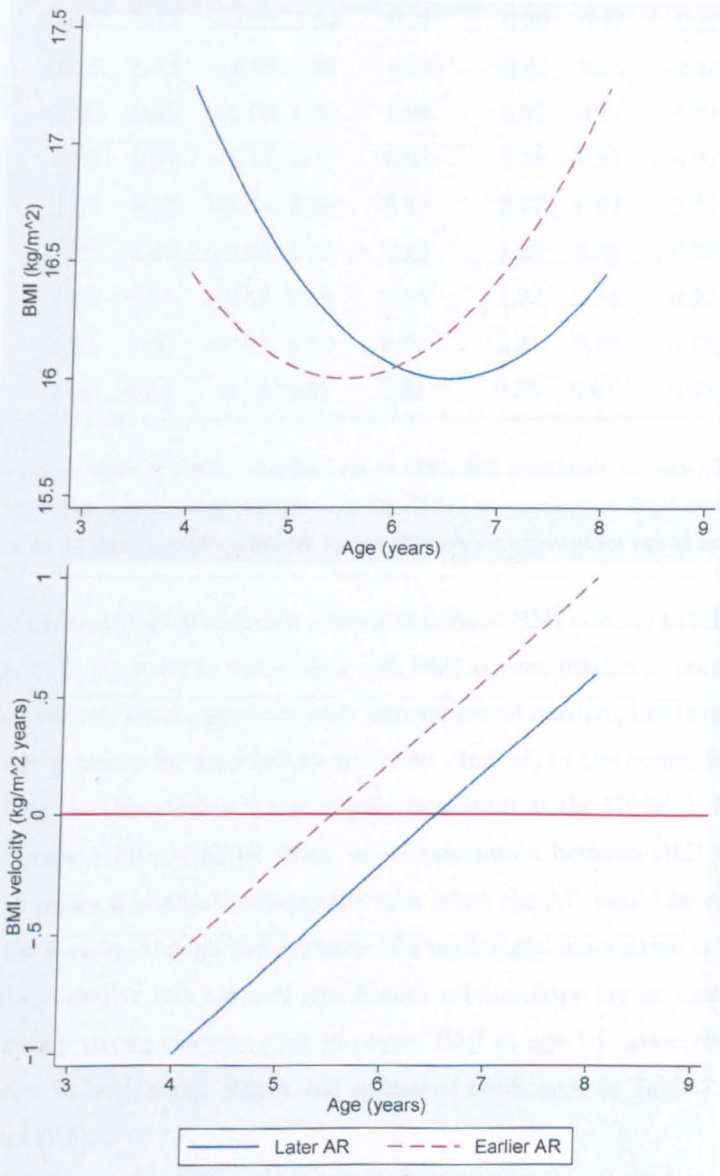

Fig. 7.8: Artificially created body mass index (BMI) trajectories around the adiposity rebound (AR) (upper plot) and their corresponding body mass index velocities (lower plot).

and is particularly strong in females age 5–6 and 8–9 years, where a 1 kg/m^{2}year increase in BMI velocity is estimated to lead to a 2–3% increase in %BF at examination..

BMI velocity (kg/m^{2}year)	Males (n = 96)				Females (n = 133)			
at age (years)	Coeff.	SE	95% CI	P-value	Coeff.	SE	95% CI	P-value
1–2	−0.17	0.71	−1.57, 1.24	0.81	0.30	0.47	−0.63, 1.23	0.53
2–3	−0.14	0.72	−1.58, 1.29	0.84	−0.99	0.73	−2.43, 0.44	0.17
3–4	−0.40	0.85	−2.08, 1.29	0.64	0.92	0.77	−0.61, 2.45	0.24
4–5	1.64	0.91	−0.17, 3.45	0.07	2.08	0.87	0.37, 3.80	0.02
5–6	1.53	0.99	−0.44, 3.50	0.13	3.07	0.92	1.24, 4.90	0.001
6–7	0.89	0.62	−0.35, 2.13	0.16	1.38	0.55	0.30, 2.47	0.01
7–8	1.16	0.87	−0.57, 2.88	0.19	1.23	0.74	−0.23, 2.70	0.10
8–9	1.94	1.06	−0.16, 4.04	0.07	2.31	0.77	0.78, 3.84	0.003
9–10	1.90	0.90	0.12, 3.68	0.04	0.59	0.61	−0.62, 1.80	0.34

Table 7.12: Estimated coefficients (coeff.), standard errors (SE), 95% confidence intervals (CI) and P-values from Wald tests for the regression models of percentage body fat (%BF) at examination fitted separately on body mass index (BMI) velocity at different ages in childhood, by sex. Models are adjusted for age at examination.

The results for the multivariable models relating childhood BMI velocity to %BF at examination are given in Table 7.13. Similarly to the models with BMI at examination as outcome (Table 7.11), the estimated BMI velocity coefficients over each interval are all positive, but here the CIs are much wider meaning that evidence for associations is weaker. Indeed, in the model for males there are no intervals in which the relationship is statistically significant at the 5% level. However, the ages at which there is some evidence, albeit weak, of an association between BMI velocity and late-adolescent %BF in males is similarly around the ages when the AR would be expected to occur. This is also true for females, though the evidence of a meaningful association is far stronger. The reasons behind these similar but reduced significance relationships are as discussed previously. There is also relatively strong evidence that increased BMI at age 1 is associated with increased %BF at examination in both sexes. Again, the estimated coefficients in Table 7.8 and Table 7.13 adhere to (7.4) and (7.5).

A comparison of the unadjusted models in Table 7.12 with their adjusted equivalents in Table 7.13 suggests that, similarly to when considering BMI at examination as the outcome, adjustment leads to increased estimated coefficients and greater statistical significance at younger ages, but decreased coefficients and reduced statistical significance at older ages. As the two outcomes are correlated, these similarities are hardly surprising. Again, the standard errors for each estimated regression coefficient are very similar in the unadjusted and adjusted models for both males and females.

	Males ($n = 96$)				Females ($n = 133$)			
	Coeff.	SE	95% CI	P-value	Coeff.	SE	95% CI	P-value
BMI (kg/m^2) at age 1 year	1.78	0.58	0.63, 2.92	0.003	1.25	0.43	0.40, 2.10	0.004
BMI velocity (kg/m^2year) at age (years)								
1–2	0.88	0.82	−0.74, 2.50	0.29	0.82	0.58	−0.32, 1.96	0.16
2–3	1.73	1.01	−0.28, 3.74	0.09	0.02	0.77	−1.50, 1.54	0.98
3–4	1.53	1.09	−0.63, 3.70	0.16	0.50	0.71	−0.89, 1.90	0.48
4–5	2.04	1.01	0.04, 4.05	0.05	2.06	0.82	0.43, 3.69	0.01
5–6	1.61	1.02	−0.43, 3.64	0.12	3.25	0.86	1.54, 4.95	<0.001
6–7	1.32	0.65	0.02, 2.62	0.05	1.67	0.51	0.67, 2.68	0.001
7–8	0.90	0.92	−0.94, 2.74	0.33	1.14	0.69	−0.24, 2.52	0.10
8–9	1.15	1.03	−0.90, 3.19	0.27	1.80	0.72	0.38, 3.23	0.01
9–10	1.55	0.90	−0.23, 3.34	0.09	0.04	0.55	−1.05, 1.13	0.95

Table 7.13: Estimated coefficients (coeff.), standard errors (SE), 95% confidence intervals (CI) and P-values from Wald tests for the multivariable regression models of percentage body fat (%BF) at examination fitted on body mass index (BMI) velocity through childhood, by sex. Models are adjusted for age at examination.

Collinearity in the models for childhood BMI velocity

The models for BMI and %BF at examination with childhood BMI velocities as the explanatory variables, given in Table 7.11 and Table 7.13, are far easier to interpret than those with childhood BMI observations as the explanatory variables due to all the estimated coefficients being positive and their standard errors being relatively small. In the same way in which collinearity is explored as an explanation for the results when childhood BMI observations are used as explanatory variables, it is insightful to investigate the extent of collinearity in the childhood BMI velocities.

As previously noted (see Table 7.4) the pairwise correlations between BMI velocities at different ages through childhood are generally very low. Although this does not allow assessment of possible collinearity between BMI velocity at three or more ages in childhood, it may be expected that if there is collinearity is most likely to be displayed BMI velocity values over adjacent time periods and there is little evidence of this.

Comparison of the unadjusted (Table 7.10 and Table 7.12) and mutually adjusted (Table 7.11 and Table 7.13) models shows little difference between the standard errors of the estimated regression coefficients. As a large increase in standard errors in the multivariable models relative to the univariable equivalents would be indicative of collinearity, this is again suggestive of a lack of collinearity.

Finally, the VIF for each BMI velocity, as well as BMI at age 1 year, is calculated and shown in Table 7.14. There are no variables with a VIF greater than approximately 2 in either sex, providing no evidence of collinearity among the explanatory variables.

	VIF	
	Males ($n = 96$)	Females ($n = 133$)
BMI at age 1 year	1.5	1.7
BMI velocity at age (years)		
1–2	1.7	2.0
2–3	2.3	1.5
3–4	2.0	1.1
4–5	1.4	1.2
5–6	1.3	1.1
6–7	1.3	1.1
7–8	1.4	1.2
8–9	1.1	1.1
9–10	1.2	1.1

Table 7.14: Variance inflation factor (VIF) for body mass index (BMI) velocity over each interval through childhood, by sex.

Thus, when considering childhood BMI velocities rather than BMI itself as the explanatory

variables there is evidence of reduced collinearity. This provides some explanation as to why the BMI velocity models are more easily interpretable.

7.2.5 Discussion

Conclusions

The multiple regression analyses using childhood BMI observations (as opposed to BMI velocities) as the explanatory variables are often difficult to interpret due to changing size and significance of the coefficients caused by the respective conditioning between BMI at different ages, although interpretation is aided somewhat by the use of life course plots. In males it can be seen that, conditional on BMI at other ages in childhood, only the most recent BMI measurement is associated with late-adolescent BMI. In females the relationships appear more complex, with BMI at different ages, and changes in BMI between different ages, through childhood being associated with later adiposity in different ways.

Although models including many repeated childhood measurements are often found to be difficult to interpret due to the respective conditioning [11], in this instance there appears to be evidence of collinearity between the childhood BMI measures which exacerbates the problem of interpretation. This is evidenced through high pairwise correlation coefficients, increased standard errors of estimated regression coefficients in multivariable models relative to the univariable equivalents, and high VIF values.

The reparameterisation of the model so that childhood BMI velocity is used as the explanatory variable makes interpretation somewhat simpler. Whilst high BMI velocity at any age is seen to tend to lead to higher late-adolescent BMI, it is between age 1 and 2 years and the period between age 4 and 7 years in females and between 5 and 8 years in males that this relationship is strongest. This latter observation suggests that an earlier AR is associated with higher late-adolescent BMI.

When using childhood BMI velocities there is little evidence of collinearity (low pairwise correlation, similar standard errors in the univariable and multivariable models, and low VIF values), which is likely to contribute to making the models more easily interpretable.

Whether considering BMI or BMI velocity, the relationship between childhood BMI development and later %BF is seen to be similar, though less strong, than that with later BMI. The similarities are likely due to the high correlation between BMI and %BF at examination, whilst the reduced significance of associations could be attributed to %BF not being merely a later observation of the exposure, as is the case when considering BMI as outcome.

Thus it can be concluded that the periods between age 1 and 2 years and around the AR appear to be critical periods of BMI development for late-adolescent adiposity.

Missing data For all the models examined and in both sexes the number of subjects contributing to the analysis is less than 50%. This high level of exclusion is as a result of relatively lower levels of missing data for childhood BMI at each age being compounded when using a complete-case analysis approach. This means that only if the missing individuals can be considered as 'missing completely

at random' [120] (MCAR, see Section 5.2.1) will the results remain unbiased. Whilst this is difficult to assess categorically, comparison of different variables between those with complete childhood BMI data and those with incomplete childhood BMI data can provide some indication. It has been seen that childhood BMI is similar between these two subgroups at each age, whilst both BMI and %BF at examination are differ somewhat. Whilst the number of individuals contributing to these comparisons are often small, this may suggest that those subjects with incomplete childhood BMI data who are excluded from the analysis may not be MCAR. If this is the case then the analysis may be biased and results for subjects with complete childhood BMI data may not be extrapolatable to the wider dataset, raising concerns about the conclusions drawn.

Previously published results Whilst, as discussed, the results obtained may be somewhat questionable due to the large proportion of excluded subjects, it is still of interest to compare them to previously published studies. However, relatively few have attempted to investigate critical periods of BMI development for adiposity in late adolescence in the same way as the present analysis. Many tend to focus on childhood weight, rather than BMI, development and often concentrate on younger ages ('catch-up growth'). Another novel aspect of the present analysis is the availability of annual BMI observations through childhood, a luxury which is afforded by few datasets. Nevertheless, where results can be compared with previous studies, they do largely agree.

Ong *et al* [45] found that 'catch-up growth' in the first 2 years of life is positively associated with obesity at age 5 in a recent British cohort. Whilst catch-up growth is here defined in terms of an increase in relative weight between birth and age 2, this is not completely removed from the positive association seen between high *BMI* velocity and later adiposity in the present study.

Corvalán *et al* [89] determined the associations between changes in BMI over several intervals covering childhood and adult BMI, %BF, abdominal circumference and fat-free mass in a Guatemalan cohort. Whilst they found change in BMI between age 3 and 7 years to be strongly and positively associated with all four adult body composition measures, change in BMI between age 1 and 3 years was not associated with any of them. They suggest, as in the present analysis, that their results support findings elsewhere that early AR predicts later fatness.

Alternative approaches

Collinearity is seen to be a potential issue in the present analysis when considering childhood BMI observations. The analysis also suffers from the large proportions of subjects who must be excluded from the multiple regression modelling. Alternative approaches should be considered which can deal with these problems. Other issues are also raised, such as the relationship between the timing of the AR and later adiposity, which could be better investigated using different methods.

Collinearity It has been seen that a reparameterisation of the original model, formulated in terms of childhood BMI, into one in terms of BMI velocity has reduced the collinearity between the explanatory variables and resulted in more sensible results.

An alternative approach to overcome the problems caused by collinearity, whilst still working within a multivariable regression framework, would be to remove some of the ages at which childhood BMI is observed from the model. A very simple way to achieve this would be by using, say, only BMI observations at even ages. The correlations at one year lag would then be removed, although this variable selection procedure is somewhat arbitrary. Variable removal could also be achieved via some form of stepwise selection procedure whereby variables are included or excluded from the model according to some predetermined criteria. Alternatively, ages for inclusion could be selected by studying the average BMI growth curve and picking ages which correspond to obvious features, such as the AR. This would then ensure that the effects of these features can still be estimated. One approach which is specific to the use of growth velocities is to initially fit the model on all available velocities then identify any consecutive intervals with similar estimated coefficients. The growth velocities can then by recalculated to cover the combined intervals and the model refitted [11]. Regardless of the approach used, the effect of the removal of a small proportion of the parameters should be negligible as the high collinearity means that fewer variables can still retain almost all the information present in the full model [117].

Alternative methods beyond multivariable regression which would also resolve the problems due to collinearity are discussed later.

Missing data This analysis has illustrated that use of a complete-case approach is potentially inappropriate, especially when many variables, each with missing data, are being used. Only when the proportion of missing data is small and uninformative, and thus the pay-off of exploiting the information in the incomplete cases minimal, may a complete-case analysis be justified [120]. Thus complete-case analysis is not generally a recommended approach, except perhaps in the rare cases when the question of interest is genuinely confined to the subpopulation of completers [118].

However, it should be noted that complete-case analyses remain in common use, often, seemingly, with little concern for the consequences of the exclusion of a relatively large proportion of the data, although efforts are being made to persuade researchers away from this approach [122]. Clearly alternatives need to be considered.

The removal of explanatory variables from the model, described above as a means to avoid the problems caused by collinearity, would also have the effect of increasing the proportion of subjects which could be included in the analysis. If, for example, 10 explanatory variables are used in the model and each individual has, independently, a probability of 0.9 of having each variable observed then the proportion of subjects with all 10 variables observed (i.e. the proportion used in a complete-case analysis) would be expected to be 0.35. However, if the number of explanatory variables is reduced to 5 then this proportion is increased to 0.59.

Because the SWEDES data are *balanced* (see Section 5.1.2.1), one viable approach which would handle the issue of missing data yet retain all the childhood BMI growth observations is *multiple*

imputation [120] (MI, see Section 5.2.4). Here, each missing value (i.e. each unobserved childhood BMI value) would be replaced by a set of plausible values which represent the uncertainly surrounding to value to be imputed. This would create multiple datasets, each completed with independently imputed values, which would then be analysed using multiple regression in an identical manner to the present analysis. As each subject in each dataset would then have complete childhood BMI data, none of them would be excluded from the analyses. The results from the separate analyses would then be combined, leading to a single inferential statement about the parameters of interest [124]. Use of MI would not, however, overcome the problems of collinearity. MI is used in the SWEDES dataset in Chapter 8, albeit in a slightly different application.

Beyond multivariable regression Structural equation models (SEMs) are an extension of standard regression models to include multiple outcomes, called 'endogenous variables', and unobservable 'latent' variables [169]. SEMs are made up of two components, the 'structural model' and the 'measurement model'. The measurement part specifies how proxy or manifest measures of unmeasured or unmeasurable latent variables are related to the latent variables. The structural part defines the relation between the latent variables and one or more outcomes.

An equivalent analysis would assume that a subject's BMI growth profile is determined by a latent process that influences their late-adolescent BMI and %BF. Childhood BMI development would then be parameterised in terms of 'true' BMI or BMI velocity at different ages, which would be latent variables manifested by the observed childhood BMI values, forming the measurement model. The structural part of the model then defines how these latent variables influence late-adolescent BMI or %BF [11].

Under the assumption that subjects without a given observed childhood BMI value are 'missing at random' (MAR, see Section 5.2.1), models can be fitted on all individuals with at least one observed childhood BMI value. Thus SEMs have the advantage of handling missingness directly.

The strong positive associations seen between BMI velocity around the AR and late-adolescent BMI and, to a lesser extent, %BF indicate a possible relationship between the timing of the AR and later adiposity. Estimation of the point at which the AR occurs in each individual would allow this relationship to be examined more explicitly. One crude approach would be to use the age at the minimum observed BMI value for an individual, perhaps restricted to a certain interval of ages, as an estimate of the age at AR. This would, however, restrict estimated ages at AR to being integer values, losing much information contained within the observed BMI values, and would also be very susceptible to measurement error.

As an extension to this, the fitting of subject-specific BMI growth curves to the observed longitudinal BMI data as an initial step of an analysis is an appealing and potentially fruitful approach. Indeed, going beyond repeated measures to understand trajectories is a theme that it has been suggested should be more often addressed in life course epidemiology [170]. From the fitted curves, minima can be derived to use as estimated locations for the AR. Childhood BMI could

be modelled using parametric (for example polynomial) or non-parameteric (for example spline) curves. The latter of these approaches is utilised in Chapter 8 to investigate the relationship between the AR and late-adolescent adiposity in SWEDES. Whilst clearly *some* childhood BMI values are required to be observed in order to fit the curve, the necessity for all 10 childhood BMI values to be present, as in the complete-case multivariable regression, could be relaxed. This would mean that a higher proportion of subjects could contribute to the analysis.

Chapter 8

Examining the relationship between the adiposity rebound and late-adolescent obesity in the Stockholm Weight Development Study

8.1 Introduction

The term 'adiposity rebound' (AR) is used to describe the period around 6 years of age when BMI begins to increase following a nadir. There is evidence that the age at AR is associated with later adiposity, with children displaying a earlier AR being at increased risk of obesity. Given the widely-reported increases in prevalence of obesity over recent years a more thorough understanding of the relationship between the AR and later adiposity is important.

The Stockholm Weight Development Study (SWEDES) is a prospective longitudinal study which provides a healthy contemporary birth cohort in which to investigate the relationships between the AR and late-adolescent adiposity. Annual weight and height measurements are available throughout childhood, allowing the BMI trajectory to be examined. Many anthropometric variables were also measured at follow-up when the SWEDES participants were approximately 17 years old. In particular, BMI and percentage body fat (%BF) provide measures of late-adolescent adiposity and are used as outcomes in the present analysis.

The epidemiological aims are to assess the extent to which the AR is associated with late-adolescent adiposity, and to investigate whether the period around the AR can be considered as a 'critical period' for later obesity. In order to achieve this, different methodological approaches

must be explored, developed and implemented.

As parametric techniques often fail to adequately model the BMI trajectory, subject-specific cubic smoothing splines are fitted to the childhood BMI values for each individual in the dataset. The fitted splines are then used to derive estimates of the age at adiposity rebound and the BMI at this age for each individual. These derived explanatory variables are then related to BMI and %BF in adolescence, through use of logistic and linear regression, to investigate the association between AR and adolescent adiposity.

The SWEDES dataset brings with it the issue of missing data, particularly among the childhood BMI values. With a balanced dataset such as SWEDES this can be dealt with via multiple imputation (MI). As the methodology being used here is of interest as well as the results of its application, in each instance both the results using the original data only and the results using the imputed datasets are presented and compared.

The fitted splines also allow the estimation of BMI and BMI velocity for any given age in childhood. These estimated values can then be used to try and investigate whether the AR can considered as a 'critical period' for adolescent adiposity.

The analysis using the original data only formed part of the work presented at the 4th World Congress on Developmental Origins of Health and Disease (DOHaD), held 13-16 September 2006 at the University of Utrecht in The Netherlands [167].

8.2 Subjects

A general introduction to the Stockholm Weight Development Study (SWEDES) can be found in Section 4.1. As in Chapter 7, the terms 'subject' and 'individual' continue to refer to the offspring in the study, and 'examination' to the occasion of the measurement of the anthropometric variables as part of the SWEDES follow-up when the offspring are approximately 17 years old.

The childhood growth data are again reduced to the subset of annual observations between age 1 and 10 years inclusive. These are referred to as the 'childhood BMI measurements'. The lower end of this range should be sufficiently low to capture any very early ARs yet late enough to avoid the additional curve-fitting complications caused by the BMI peak often observed within the first year of life. The upper end of this range should be sufficiently late to capture the entire range of plausible ages for AR without being so late as to confuse their identification by also including further undulations in the BMI trajectory associated with the pubertal period. The outcome variables are BMI and %BF measured at examination when the subjects are approximately 17 years old.

BMI through childhood and at examination is calculated as weight/height^2 (kg/m^2). Of the 481 individuals in the study, 95 (19.8%) have no BMI observations whatsoever (i.e. no concurrent height and weight observations) between age 1 and 10 years. Using the data in their initial form these individuals can contribute nothing to any analysis involving childhood BMI trajectory. Even under a MI approach they can only contribute if the *entire* childhood BMI trajectory is imputed,

which is rather unappealing. Thus these 95 individuals are excluded from the analysis at this stage leaving 386 eligible subjects.

This group of 386 individuals are referred to throughout as the 'original dataset'. Whilst clearly they are not the 'original' dataset in the sense that some of the initial 481 subjects have been excluded, this dataset remains 'original' in the context of the data themselves being unchanged. This is in opposition to the imputed datasets in which any values missing in the original dataset will be 'filled in'.

It is important to investigate whether those individuals with no observed BMI values who are excluded from the analysis differ from those with at least one observed BMI value who are included. Table 8.1 summarises by inclusion status the distributions of a variety of variables at birth and at examination.

The majority of the variables examined appear to have very similar distributions in those who are included in and those who are excluded from the analysis, although both included males and females perhaps seem to be a little heavier at birth. Several of the variables at examination have differing mean values but median values which differ markedly less. This is likely to be evidence of skew in the distribution or a small number of outlying values having a large effect on the mean, so should be of little concern. The standard deviations (SDs) in those who are excluded from the analysis are often seen to be greater, though because of the small sample sizes involved this may again be due to one or two outlying values.

That the distributions of these variables appear to be very similar in the two subsets is important as it suggests that the excluded individuals are little more than a random group of the SWEDES dataset — or, to use the language of Rubin (see Section 5.2.1), they are 'missing completely at random' (MCAR). The result of this is that their exclusion should not bias the results obtained using the remaining 386 individuals in the dataset.

8.3 Methods

In previous studies, several different approaches have been employed to estimate the location of the AR (meaning both the age and BMI at AR) in each individual. The most basic method is to take it to be the lowest observed BMI value [88], though more often the observed BMI values are plotted and the AR visually determined by identification of the point of lowest BMI [82, 83, 90, 165]. Alternatively, individual BMI curves may be fitted to the data, and from them the AR location derived. Cubic polynomials are often used for this purpose [84, 85, 86] and have also been extended to a random coefficients model [86]. A final approach is to obtain growth curves for the logarithm of height and the logarithm of weight for each individual using random coefficient cubic polynomial models. Velocity curves are then derived from the fitted curves and the age at AR found as the point when the velocity of the log-transformed weight curve exceeds twice the velocity of the log-transformed height curve [171, 172].

Although it has been suggested that estimating the AR location visually reflects the physiolog-

Males ($n = 202$)						
Variable	Included ($n = 159$))			Excluded ($n = 43$)		
	Mean	Median	SD	Mean	Median	SD
At birth						
Gestational age (weeks)	39.5	40	1.8	39.6	40	2.3
Weight (kg)	3.56	3.53	0.54	3.42	3.42	0.52
At examination						
Age (years)	16.9	16.9	0.4	17.0	17.0	0.4
Weight (kg)	68.2	66.2	11.0	70.5	67.1	14.9
Height (m)	1.80	1.80	0.06	1.79	1.79	0.06
BMI (kg/m^2)	20.9	20.2	2.8	22.0	20.5	4.5
Waist circumference (cm)	74.6	73	7.4	78.4	74	11.5
Hip circumference (cm)	92.5	91	7.3	94.5	92	9.2
%BF	15.5	14.2	6.7	19.0	15.9	9.2

Females ($n = 279$)						
Variable	Included ($n = 227$)			Excluded ($n = 52$)		
	Mean	Median	SD	Mean	Median	SD
At birth						
Gestational age (weeks)	39.5	40	1.6	39.8	40	1.4
Weight (kg)	3.43	3.48	0.48	3.39	3.40	0.47
At examination						
Age (years)	16.8	16.8	0.4	16.8	16.8	0.4
Weight (kg)	59.3	59.0	8.8	61.3	59.9	10.6
Height (m)	1.67	1.67	0.06	1.67	1.66	0.05
BMI (kg/m^2)	21.3	20.8	3.0	22.1	21.5	3.6
Waist circumference (cm)	71.0	70	6.7	73.1	71.5	8.6
Hip circumference (cm)	91.9	92	6.5	93.6	93	7.3
%BF	29.0	28.3	6.3	31.3	31.3	6.8

Table 8.1: Distributions of variables at birth and at examination for subjects in the Stockholm Weight Development Study, by inclusion in the analysis and sex. BMI is body mass index and %BF is percentage body fat.

ical basis of the AR better than estimation via polynomial fitting [173], visual inspection of each BMI curve may not be practicable for large datasets. Additionally, as BMI curve fitting has been restricted to cubic polynomials, the deficiencies seen in the model fitting approach may be due to the specific model used. In particular, the use of a parametric model may not provide sufficient flexibility of curve shape to model BMI around the period of AR, as discussed in Section 6.1. Thus in the present application non-parametric subject-specific cubic smoothing splines are used to model BMI growth and derive estimates of the AR location. This is described in Section 8.3.1.

An additional issue to be addressed in the present application is missing data, which particularly affects the childhood BMI values in SWEDES. As growth curves are to be fitted to the BMI values it is not imperative that each individual has the same number of observed BMI values, but if the extent of missing data is great then the curve fitting may not be able to provide estimates of the AR location. As the growth data in SWEDES are balanced, missing data may be dealt with using MI. The MI approach used is detailed in Section 8.3.2.

Due to the relatively complex approach to analysis, a schematic overview of the methods is provided in Section 8.3.3 for clarification.

8.3.1 Spline fitting

A theoretical background to smoothing splines is provided in Section 5.4.1.4. Here, more application-specific details such as data requirements, the selection of the smoothing parameters, the estimation of the AR location, and the software used in the spline fitting are discussed.

Data requirements

The 386 eligible individuals in the SWEDES dataset have between 1 and 10 non-missing annual BMI observations. Clearly attempting to fit a spline and derive from it the location of the minimum value with just a handful of points is unlikely to provide reliable results. Thus the following requirements are introduced which have to be satisfied in order for a spline to be fitted: a child must have 6 or more data points in total, at least 2 of which must be at age 6 years or younger, and at least 2 of which must be at age 6 years or older.

Selection of the smoothing parameters

Selection of the smoothing parameter for the splines is a key issue which can be done in a variety of ways. Green and Silverman [140] discuss two philosophical approaches to the question of choosing the smoothing parameter. The first regards the freedom of choice as an advantageous feature of the procedure whereby a variety of values can be experimented with and a subjectively optimal choice made. The second line of thought is that an automatic procedure is preferable so that the data themselves are choosing the value of the smoothing parameter.

In the present study both these elements seem to be of importance. The potential to vary the smoothing parameter between individuals in order to optimise the reliability of the identified AR is clearly essential. However, use of a common smoothing parameter, or at least a common method

by which to obtain it, would ensure comparability across individuals and remove the need to decide on smoothing parameters on an individual-by-individual basis. Whist this latter point may not be of too much concern with a sample size of 386, when using much larger datasets (for example the 100 imputed datasets each with 386 subjects) the procedure would become very time-consuming.

To assess methods by which the smoothing parameter could be chosen, either by the use of an existing automatic procedure or the selection of a global parameter for use across the dataset, an analysis is carried out on a subset of individuals. As the spline fitting procedure necessitates at least six non-missing BMI values for an individual, subjects in analysis of the original data can have either 6, 7, 8, 9 or 10 values to which the spline must be fit. As this is a relatively wide range, and because the selection of the smoothing parameter is often dependent on the number of data points, a stratified random subsample of 8 individuals (where possible) from each subgroup (i.e. those with 6 data points, those with 7, etc.) is taken. Each subject then has fitted several splines using cross-validation (CV), generalised cross-validation (GCV) and a variety of user-specified equivalent degrees of freedom (EDF) values (3, 4, 5, 6, 7 and 8).

Following this examination of different smoothing parameters an overall strategy for the smoothing parameter to use for each individual is devised. Subject-specific splines are then fitted to each individual meeting the previously defined data requirements.

Estimation of the adiposity rebound location

The estimated location of the AR for each individual is then defined as the minimum value of their fitted spline. Whilst all 10 BMI values between age 1 and 10 years are used in the spline-fitting procedure when present, the estimated AR is only searched for between age 2 and 9 years. This range of ages encompasses those over which the AR has generally been identified in previous studies. Identification of ARs outside of this interval would also be somewhat unreliable as ages would then be nearing the extremes of the interval over which the spline is fit.

The most simple criterion for identifying the minimum is as the point at which the first derivative of the fitted spline changes from negative to positive. However, this approach could easily identify situations which are either implausible or undesirable in the context of the AR, such as multiple minima and minor local minima or points of inflection which are of no real interest. To overcome the latter problem it is also necessitated that the value of the first derivative of the fitted spline be negative one year prior to the identified minimum and positive one year after. If after this stipulation there still exist multiple minima then the likelihood is that such minima are true features of the data. In these instances it is not possible to identify an AR, meaning that these individuals cannot contribute to any analysis which includes either dimension of the AR.

Software

Spline fitting is carried out using the `smooth.spline` package in R [155], a procedure for one-dimensional cubic spline fitting. The package allows user-specification of the degree of smoothing in terms of the smoothing parameter α, or in terms of the EDF, as well as automatic choice

via cross-validation (CV). With fewer than 50 distinct points, as is the case for all individuals in the study, the expression is minimised over cubic splines with knots allowed at all the data points, so the curve obtained is precisely the cubic spline smoother [140]. A further package, `predict.smooth.spline`, can then be used to calculate the estimated curve or its derivatives at any specified points.

8.3.2 Multiple imputation

A theoretical background to multiple imputation is provided in Section 5.2.4. In this section the variables to be included in the imputation model are discussed and the imputation specifications detailed. Some additions to the standard complete-data inference approach are also introduced.

Imputed variables

Many subjects in the SWEDES dataset have one or more missing BMI values (i.e. either missing height, weight or both for a given age) between age 1 and 10 years. When the number of missing BMI values for an individual is small this may result in a less reliable estimate of the location of the AR. When many BMI values are missing there may be insufficient non-missing BMI values for a spline to be fitted at all so that no estimate of the location of the AR can be made. This means that the individual cannot contribute to the analysis, reducing the effective sample size, and hence the precision of the estimates. Also, bias may be introduced if those not contributing to the analysis are not missing completely at random (MCAR) [120]. Imputation of the missing BMI values means that every individual has the full 10 data points so that a spline can be fitted, which should increase the proportion of subjects contributing to the analysis.

The adolescent outcomes — BMI and %BF at examination — have fewer missing values. Indeed, BMI at examination is fully observed and %BF has only 7 missing values (1.8%). Again, if subjects with missing %BF are excluded from the analysis the same concerns exist. Imputation of the missing %BF values ensures that all individuals can be included in the analysis, provided the the necessary explanatory variables can be derived from the fitted spline.

Schafer [124] suggests that for high-quality unbiased imputations to be obtained for a given variable it is important to include in the imputation model variables potentially related to either the variable of interest itself or its pattern of missingness. Whilst it has been suggested that the number of predictors in the imputation model should be as large as possible to make the MAR assumption more plausible [126], it may be impracticable to do so due to limitations in computing resources or in the data themselves [124]. As a result, potential explanatory variables are only included in the imputation model if doing so is deemed appropriate given Schafer's above conditions.

All height and weight variables (those at birth, 1, 3, 6 and 9 months, and 1 to 15 years) are included in the imputation model due to their relation with the missing height and weight values. Gestational age is also included for this reason. Height, weight, waist circumference, hip circumference, fat mass, fat-free mass, and systolic and diastolic blood pressure at examination

are included due to their relationship with %BF at examination, and also, to a lesser extent, with childhood BMI.

Several maternal variables (BMI and %BF at examination, type of employment, monthly income, hours worked per week, education level, civil status, country of origin, number of children and age at the birth of the child included in the study) are also included as they are judged to be likely to be related to either the missing values themselves (in the case of the anthropometric variables and country of origin) or to the patterns of missingness (in the case of indicators of socioeconomic status (SES)). As both the education level [94] and occupation [78] of the mother have previously been used as proxies for SES in published analyses using these data they are both taken to be reliable indicators.

Whilst the application of MI by Markov Chain Monte Carlo (MCMC), as described in Section 5.2.4, assumes multivariate normality, inferences may be robust to departures from this assumption if the amount of missing information is not large [124]. A number of the variables included in the imputation model are categorical, but as these are all virtually or entirely fully observed and only appear in the imputation model to improve the quality of the imputations (i.e. they do not appear in the analysis model), this should not cause any problems. Several continuous variables which *are* in the analysis model, and thus for which the quality of the imputed values is more important, exhibit slightly skewed distributions. In these cases a suitable transformation is applied prior to the MI procedure.

Imputation specifications

As the missing data pattern is intermittent (see Section 5.2.1), MCMC, as described in Section 5.2.4, is is an appropriate method by which to generate the imputed values.

In the MI procedure, the expectation-maximisation (EM) algorithm is used to derive a set of initial parameter values for the MCMC. The EM algorithm is a technique for maximum likelihood estimation (MLE) in parametric models for incomplete data. It is an iterative procedure, which repeats the same two steps until convergence. In the first (expectation) step, the conditional expectation of the complete-data log likelihood given the observed data and the present parameter estimates is calculated. In the second (maximisation) step, the parameter estimates which maximise the complete-data log likelihood calculated in the first step are found. These estimates are then fed back into the first step [120].

The EM algorithm can thus be used to compute the mean vector and covariance matrix for the variables prior to application of MCMC. The means and standard deviations (SDs) from the available cases are used as the initial estimates for the EM algorithm, with correlations set to zero. A noninformative Jeffreys prior is used to derive the starting values for the MCMC process from the EM algorithm [127].

Full-data imputation is carried out using a single chain for all imputations with 200 initial burn-in iterations before the first imputation and 100 iterations between each subsequent imputation.

Whilst in general relatively few imputations may be required to provide a relative efficiency

close to one (see Section 5.2.4), the large number of variables with missing data and the occasionally high proportions of missing data for a given variable mean that a greater number may be advisable here. A lack of constraints on computing power and the ease with which many datasets may be analysed also mean that there seems little point in risking having too few imputations. Thus one hundred imputed datasets are created.

Time-series and autocorrelation plots of parameters from iterations, as detailed in Section 5.2.4, are examined to ensure appropriate convergence of the MCMC process.

Complete-data inferences

Once suitably obtained, the imputed datasets can be analysed using standard procedures for complete data and the within-imputation results combined as described in Section 5.2.4. Regardless of the complete-data analysis approach used, the process of combining results across the imputed datasets is essentially the same, resulting in valid statistical inferences that properly reflect the uncertainly due to missing values.

In the present context, interest mainly lies in the location of the AR itself, and how this relates to measures of late-adolescent adiposity. Thus the within-imputation results to be combined will include summary statistics of age and BMI at AR, estimated correlation coefficients, and estimated coefficients in models relating the AR to later adiposity.

Whilst the combination of estimates of means and regression coefficients is simply achieved using the previously described approach, this requires a slight amendment when combining correlation coefficients across imputed datasets. A further issue is the extension of medians to the MI setting. Details regarding these somewhat non-standard approaches are provided below.

Correlation coefficients Correlation coefficients between several variables (age and BMI at AR, BMI and %BF at examination) are of interest in the present analysis. Whilst a sample correlation coefficient between a pair of variables *within* an imputed dataset can be calculated in the normal way, the distribution of these correlation coefficients *across* the imputed datasets will be skewed, making their combination less simple [127]. The distribution of the sample correlation coefficients r can, however, be normalised through Fisher's z transformation,

$$z = \frac{1}{2} \log\left(\frac{1+r}{1-r}\right).$$

The distribution of z is then approximately normal with mean

$$\log\left(\frac{1+\rho}{1-\rho}\right)$$

and variance

$$\frac{1}{n-3},$$

where ρ is the population correlation coefficient and n is the number of observations contributing to the calculation of the sample correlation coefficient. z can then be combined across the imputed

datasets in the usual manner and the associated variances used to calculate confidence limits if required. These values can then be transformed back to give an estimate of the correlation coefficient and confidence limits using

$$r = \tanh(z).$$

Mean medians In a simple setting, when the distribution of a variable is skewed it is often preferable to use the median as a measure of the 'average' of the variable as opposed to the mean. The generalisation of the median to a MI setting is, however, not so obvious. Thus the proposed statistic for use in this situation is the 'mean median', defined as the mean of the median values within each imputed dataset.

Software

MI is carried out using the `MI` procedure in SAS [174]. The 100 imputed datasets are then analyzed using standard SAS procedures (`CORR`, `REG`, `GENMOD`) and the `MIANALYZE` procedure used to combine the results and generate valid statistical inferences.

8.3.3 Diagrammatic overview of methods

The complex multi-stage nature of the present analysis means that it is not always easy to follow. Fig. 8.1 is provided as a diagrammatic summary of the analysis methods used.

The following comments relate to the labels in Fig. 8.1:

1. Start with the original data, which includes all 481 subjects. Of these, the 95 subjects with no observed childhood BMI values are excluded from the analysis, whilst the remaining 386 individuals are included.

2. These individuals have data which are subject to missingness. This can potentially be handled through MI.

3. If MI is not used, then the 'original dataset' is used. This can be partitioned into childhood BMI data and outcome variables.

4. Splines are fitted to those individuals with childhood BMI data meeting the data requirements, whilst those for whom this is not the case are excluded from the analysis at this point.

5. Of those subjects for whom a spline is fitted, not all will have a successfully identified AR. Those that do not are again excluded from the analysis at this point.

6. Individuals who do have a fitted spline and an estimated AR, however, are included in the final distal outcome model. Here, one or both dimensions of the AR are related to the outcome variables.

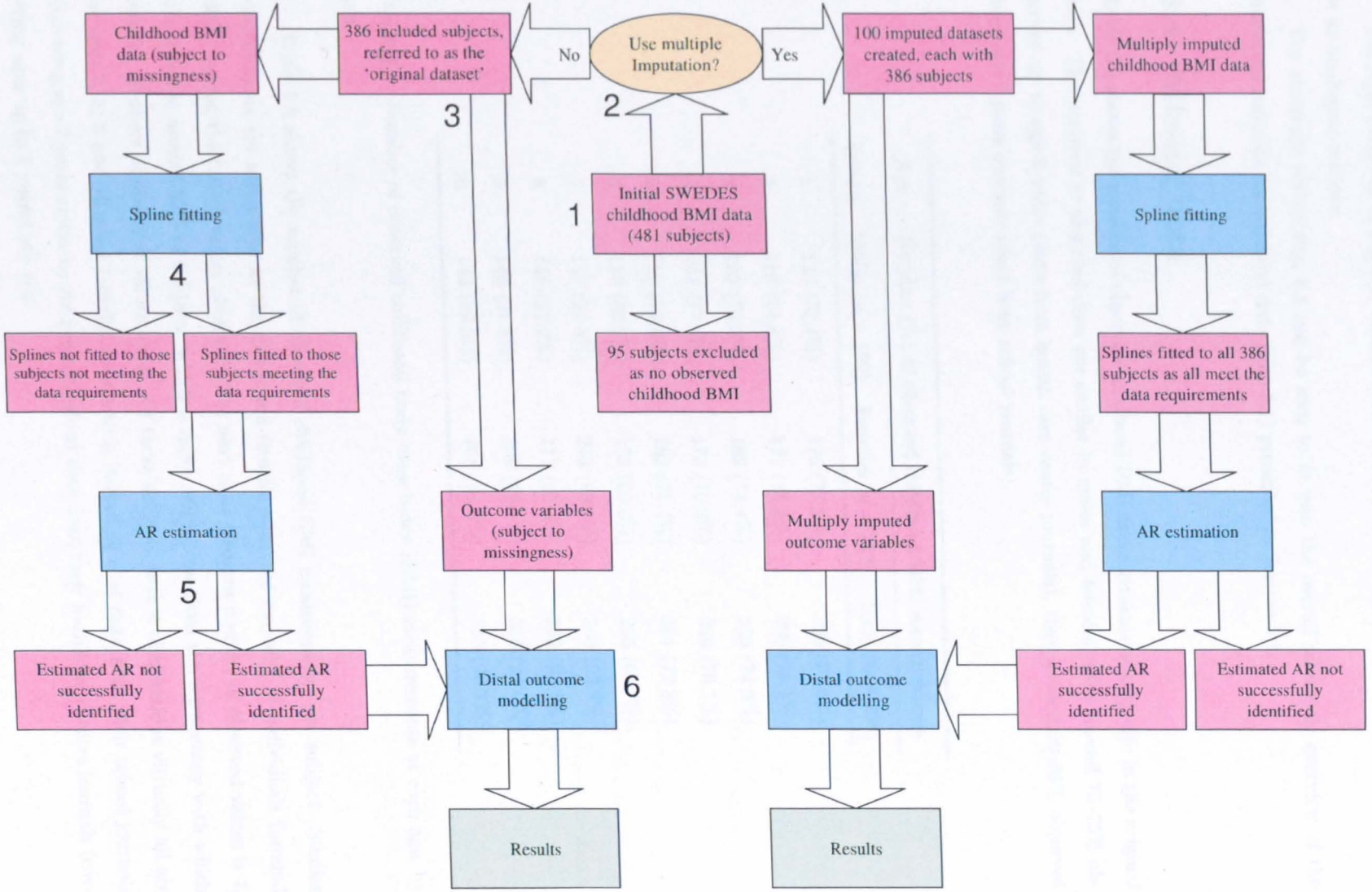

Fig. 8.1: Diagrammatic overview of methods.

The right hand side of Fig. 8.1, which describes the analysis when MI *is* used, could be explained in an analogous manner.

The structure within Fig. 8.1 can be seen to fit into the overall schematic overview of the statistical methods for balanced data (Fig. 5.1) presented in Section 5.5.

8.4 Missing data

Table 8.2 shows the number of observed childhood BMI measurements at each age in the original data. The patterns of observed data are similar in males and females, with around 70–75% observed up to age 6 years (data from health care centre journals), then around 90–95% observed from age 7 years onwards (data from school journals).

Age	Number (%) of observed childhood BMI measurements		
(years)	Males ($n = 159$)	Females ($n = 227$)	Total ($n = 386$)
1	123 (77.4%)	176 (77.5%)	299 (77.5%)
2	119 (74.8%)	171 (75.3%)	290 (75.1%)
3	120 (75.5%)	169 (74.4%)	289 (74.9%)
4	123 (77.4%)	173 (76.2%)	296 (76.7%)
5	119 (74.8%)	162 (71.4%)	281 (72.8%)
6	110 (69.2%)	155 (68.3%)	265 (68.7%)
7	147 (92.5%)	213 (93.8%)	360 (93.3%)
8	145 (91.2%)	217 (95.6%)	362 (93.8%)
9	146 (91.8%)	209 (92.1%)	355 (92.0%)
10	142 (89.3%)	208 (91.6%)	350 (90.7%)

Table 8.2: Number of observed childhood body mass index (BMI) measurements at each age, by sex.

Table 8.3 shows the number of observed childhood BMI measurements per subject. Similar distributions are again seen in the males and females with the majority of individuals (around 60%) having the full 10 values observed. The next most frequent number of observed values is 4, occurring in around 17% of subjects, which is quite anomalous given the infrequency with which similar numbers of observed values occur. Of those subjects with 4 observations virtually all are at ages 7, 8, 9 and 10 years (results not shown), indicating that data from their school journals (covering age 7 years onwards) are present, whilst data from their health care centre journals (covering ages up to 7 years) are not.

It is important to investigate whether those individuals with a sufficient number of observed BMI measurements between age 1 and 10 years to be included in the spline-fitting procedure

Number of observed childhood BMI measurements	Frequency (%)		
	Males (n = 159)	Females (n = 227)	Total (n = 386)
1	1 (0.6%)	0 (0.0%)	1 (0.3%)
2	7 (4.4%)	9 (4.0%)	16 (4.2%)
3	3 (1.9%)	4 (1.8%)	7 (1.8%)
4	27 (17.0%)	38 (16.7%)	65 (16.8%)
5	0 (0.0%)	4 (1.8%)	4 (1.0%)
6	1 (0.6%)	5 (2.2%)	6 (1.6%)
7	5 (3.1%)	1 (0.4%)	6 (1.6%)
8	10 (6.3%)	13 (5.7%)	23 (6.0%)
9	9 (5.7%)	20 (8.8%)	29 (7.5%)
10	96 (60.4%)	133 (58.6%)	229 (59.3%)

Table 8.3: Number of observed childhood body mass index (BMI) measurements per subject, by sex.

of Section 8.3.1, and thus potentially in any analyses, differ from those who do not. Table 8.4 summarises the distributions of a variety of variables at birth and at examination for subjects with different numbers of observed BMI measurements.

As individuals require at least 6 measurements to be eligible for the spline-fitting procedure, both those with 6–9 and 10 measurements will have splines fit when the original data are analysed and thus, potentially, contribute to any analysis. Those with 10 measurements, however, have no missing data so will remain identical in each of the 100 imputed datasets, whilst those with with 6–9 measurements will have 1–4 imputed values.

Subjects with 1–5 BMI measurements, however, have insufficient data points to allow subject-specific splines to be fitted thus will not contribute to any analysis using the original data. They will have 5–9 values imputed in the imputed datasets and thus, when these are analysed, will qualify for the spline-fitting procedure.

Differences in the variables examined in Table 8.4 between those subjects with varying degrees of observed BMI values are highly sex-specific. In females all of the variables appear to be relatively similarly distributed, regardless of the number of BMI values observed. In males, however, there are some clear trends. At birth, those with 5 or fewer observed values appear heavier than those with 6 or more. At examination this same group still have, on average, greater weight, and also greater BMI, waist and hip circumferences, and %BF.

8.5 Exploratory analyses

Exploratory analyses using the original data only (Section 8.5.1) and using the imputed datasets (Section 8.5.2) are presented.

Variable	Males (n = 159)								
	Number of observed childhood BMI measurements								
	1–5 (n = 38)			6–9 (n = 25)			10 (n = 96)		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
At birth									
Gest. age (weeks)	39.6	40	1.9	39.1	39	2.2	39.5	40	1.6
Weight (kg)	3.70	3.63	0.57	3.46	3.48	0.57	3.53	3.53	0.51
At examination									
Age (years)	16.9	16.9	0.4	16.9	17.0	0.4	16.8	16.9	0.4
Weight (kg)	72.7	70.4	12.4	68.5	63.2	13.0	66.3	64.5	9.4
Height (m)	1.82	1.82	0.06	1.82	1.83	0.06	1.79	1.79	0.06
BMI (kg/m^2)	21.9	20.9	3.3	20.6	19.2	3.1	20.6	20.0	2.4
Waist circ. (cm)	77.8	76	9.0	74.0	73	9.7	73.4	73	5.6
Hip circ. (cm)	95.8	94	8.5	92.2	91	8.6	91.3	91	5.9
%BF	18.2	16.3	8.3	16.1	14.4	6.9	14.3	13.9	5.6

Variable	Females (n = 227)								
	Number of observed childhood BMI measurements								
	1–5 (n = 55)			6–9 (n = 39)			10 (n = 133)		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
At birth									
Gest. age (weeks)	39.4	39	1.4	39.4	40	2.2	39.5	40	39.5
Weight (kg)	3.48	3.47	0.42	3.36	3.50	0.58	3.43	3.45	0.46
At examination									
Age (years)	16.9	16.9	0.4	16.8	16.8	0.4	16.8	16.8	0.4
Weight (kg)	59.6	59.1	8.7	60.7	59.4	10.3	58.8	58.5	8.4
Height (m)	1.66	1.65	0.06	1.67	1.67	0.06	1.67	1.67	0.06
BMI (kg/m^2)	21.8	21.1	3.3	21.8	20.9	3.8	21.0	20.8	2.5
Waist circ. (cm)	71.1	70	7.0	72.4	70	8.5	70.6	70	5.9
Hip circ. (cm)	92.8	93	6.5	91.9	92	7.4	91.6	91	6.3
%BF	29.1	28.2	6.6	30.1	29.3	6.8	28.6	28.2	6.0

Table 8.4: Distributions of variables at birth and at examination, by number of observed childhood body mass index (BMI) measurements and sex. Gest. age is gestational age, waist circ. is waist circumference, hip circ. is hip circumference and %BF is percentage body fat.)

8.5.1 Using the original data only

Fig. 8.2 illustrates both the univariate and bivariate distributions of BMI and %BF at examination. From the histograms it can be seen that both males and females have positively skewed distributions of both BMI and %BF at examination. Meanwhile, the scatterplots show a clear positive association between the two measures of adiposity. Indeed, the correlations between the two dimensions, calculated using the log-transformed variables due to the skew, are 0.57 and 0.63 in males and females respectively.

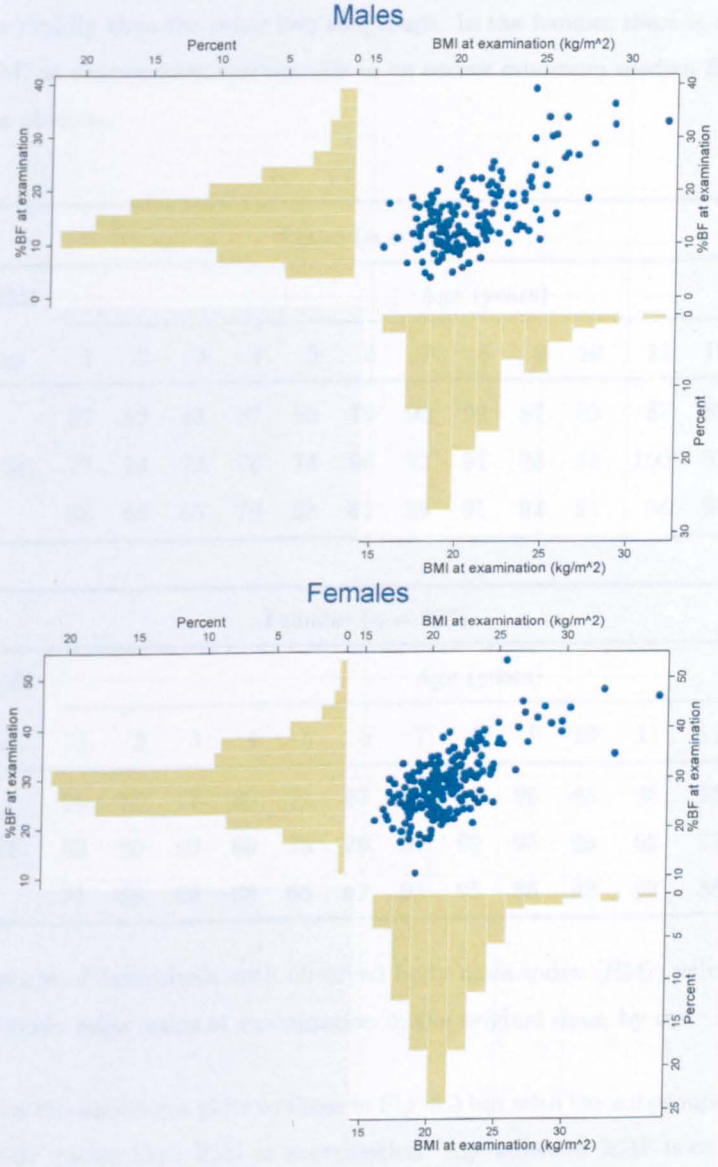

Fig. 8.2: Univariate and bivariate distributions of body mass index (BMI) and percentage body fat (%BF) at examination in the original data, by sex, for the 386 subjects included in the analysis.

Fig. 8.3 includes plots of median BMI through childhood in the three subgroups defined by the tertiles of age-adjusted BMI at examination (low, medium and high). A sex-specific simple linear regression of BMI at examination on age at examination is first fitted and the residuals used to

define the age-adjusted BMI at examination tertiles. The percentage of individuals with observed BMI values contributing to each plotted point is given in Table 8.5 and ranges between 62 and 100% in males and between 55 and 99% in females.

Whilst it is evident that the median BMI levels in the subgroups take the ordering that they do at examination well in advance of this point, providing evidence of BMI tracking, there are some sex-specific differences. In females, this ordering is established by age 1 year and the median BMI values in the subgroups diverge at a relatively constant rate from this age onwards. In the males, however, until age 6 years the median BMI within the medium and high BMI at examination subgroups are very similar, after which point those in the high BMI at examination tertile gain BMI much more rapidly than the other two subgroups. In the females there is also some evidence that a higher BMI at examination corresponds to an earlier minimum median BMI, though in the males this is less obvious.

Males ($n = 159$)															
Subgroup of BMI	Age (years)														
at examination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low ($n = 53$)	87	83	83	87	83	79	96	92	87	92	87	89	89	83	77
Medium ($n = 53$)	77	74	76	76	74	66	92	91	96	85	100	87	94	85	75
High ($n = 53$)	68	68	68	70	68	62	89	91	92	91	94	91	89	85	77

Females ($n = 227$)															
Subgroup of BMI	Age (years)														
at examination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low ($n = 75$)	79	77	77	80	71	67	91	93	96	88	95	95	97	92	71
Medium ($n = 76$)	83	80	80	80	78	76	97	99	95	95	92	93	93	87	63
High ($n = 76$)	71	68	66	68	66	62	93	95	86	92	80	88	86	83	55

Table 8.5: Percentage of individuals with observed body mass index (BMI) values at each age in each subgroup of body mass index at examination in the original data, by sex.

Fig. 8.4 includes the equivalent plots to those in Fig. 8.3 but with the subgroups defined in terms of age-adjusted %BF rather than BMI at examination. Age-adjusted %BF is calculated using an analogous method to that for age-adjusted BMI. The percentage of individuals with observed BMI values contributing to each plotted point is given in Table 8.6 and ranges between 64 and 98% in males and between 59 and 97% in females.

The median BMI trajectories seen in the plots are not dissimilar to those for the subgroups defined on BMI at examination, which, given the high levels of correlation between BMI and %BF

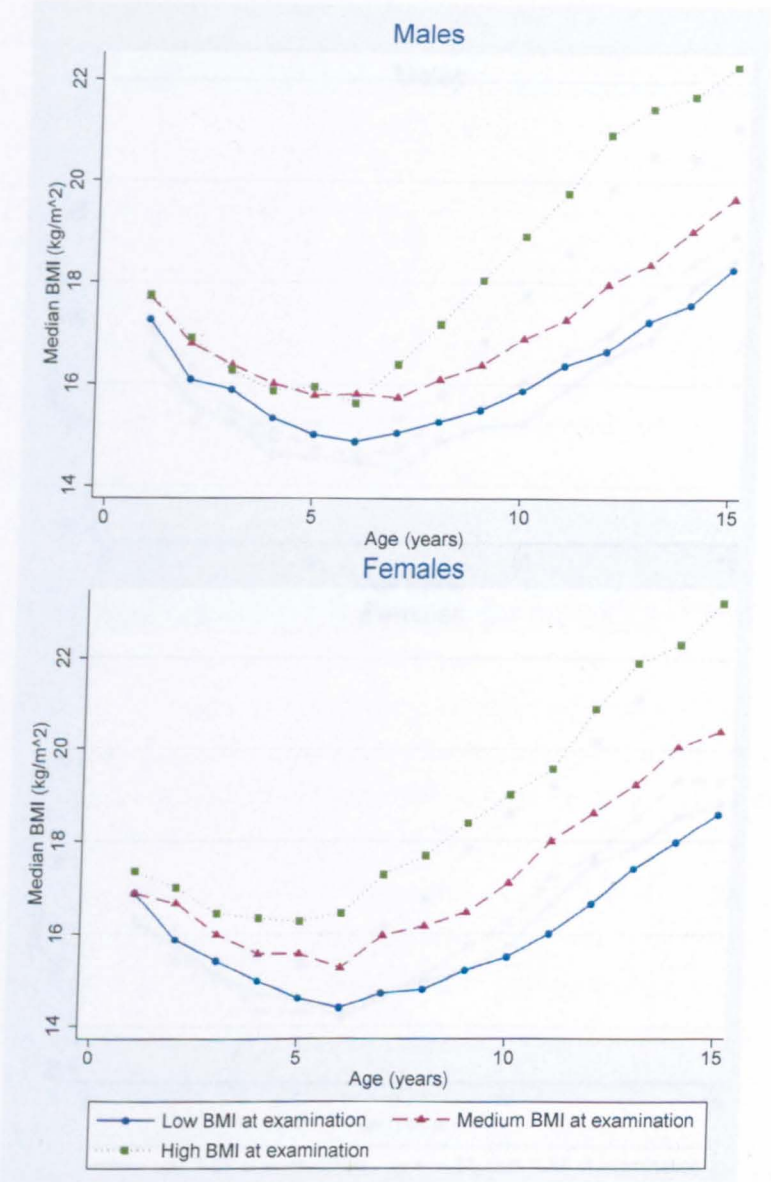

Fig. 8.3: Median body mass index (BMI) through childhood in the original data, by age-adjusted body mass index at examination and sex.

at examination, is not surprising. As in Fig. 8.3, the median BMI corresponding to the high %BF at examination subgroup is uppermost from age 1 year in females, whereas this is not the case until age 5 years in males. From Fig. 8.4, however, it is noticeable that this median trajectory then diverges more rapidly away from the others whilst, particularly in females, the trajectories corresponding to medium and low %BF at examination remain similar. There is only weak evidence in either sex of a negative association between age at minimum median BMI and %BF at examination.

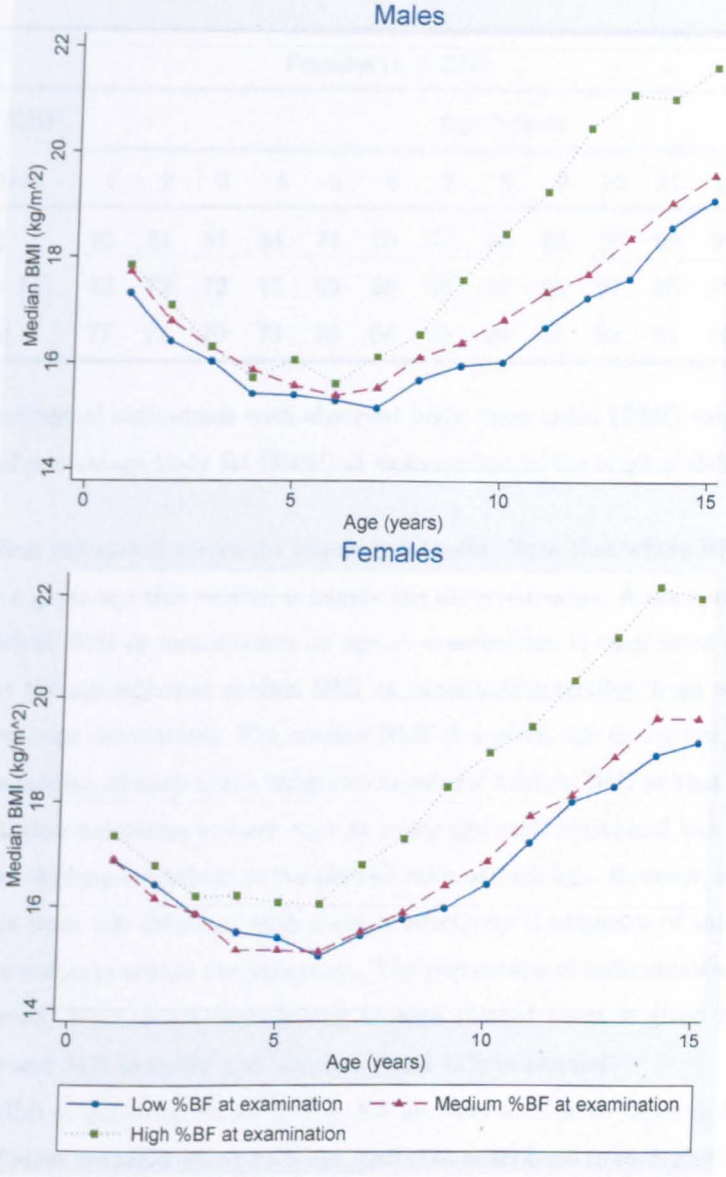

Fig. 8.4: Median body mass index (BMI) through childhood in the original data, by age-adjusted percentage body fat (%BF) at examination and sex.

8.5.2 Using the imputed datasets

Fig. 8.5 presents equivalent plots to Fig. 8.3 but using the 100 imputed datasets rather than only the original data. The median BMI in each individual at each age between 1 and 15 years and at

Males ($n = 157$)															
Subgroup of %BF	Age (years)														
at examination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low ($n = 52$)	85	81	81	83	77	71	98	94	92	90	94	94	87	85	75
Medium ($n = 52$)	79	75	77	79	79	71	94	90	94	87	96	87	96	85	81
High ($n = 53$)	68	68	68	70	68	64	87	91	89	92	91	87	89	85	74

Females ($n = 222$)															
Subgroup of %BF	Age (years)														
at examination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low ($n = 74$)	82	81	81	84	74	69	92	95	95	92	92	97	96	88	66
Medium ($n = 74$)	73	72	72	72	69	68	96	97	96	91	95	89	91	88	64
High ($n = 74$)	77	73	70	73	70	68	93	96	86	92	81	89	91	86	59

Table 8.6: Percentage of individuals with observed body mass index (BMI) values at each age in each subgroup of percentage body fat (%BF) at examination in the original data, by sex.

examination is first calculated across the imputed datasets. Note that where BMI is observed for an individual at a given age this median is merely the observed value. A sex-specific simple linear regression of median BMI at examination on age at examination is then fitted and the residuals used to calculate the age-adjusted median BMI at examination tertiles, from which the BMI at examination subgroups are defined. The median BMI at a given age in a given subgroup is then calculated as the median of each of the subgroup members' median BMI at that age.

As the imputation procedure ensures that at every age each individual has a BMI value, *all* subjects within a subgroup contribute to the plotted value at each age. However, as each individual contributes values from 100 datasets, each point is effectively a summary of values totalling 100 times the number subjects within the subgroup. The percentage of individuals with imputed (as opposed to observed) BMI values contributing to each plotted point is given in Table 8.7 and ranges between 0 and 34% in males and between 1 and 45% in females.

The median BMI trajectories shown in Fig. 8.5 are very similar to those in Fig. 8.3. As the majority of BMI values are observed at each age, and thus contribute to each plot in the same way, this is largely expected, though the degree of similarity suggests that the imputed BMI values at each age in each subgroup must be similar to those observed in other individuals. The only slight differences from Fig. 8.3 are that at age 1 year the median BMI values in the subgroups now take the same ordering as they do at examination, and at age 6 years the median BMI corresponding to a high BMI at examination is now greater than that corresponding to a medium BMI at examination.

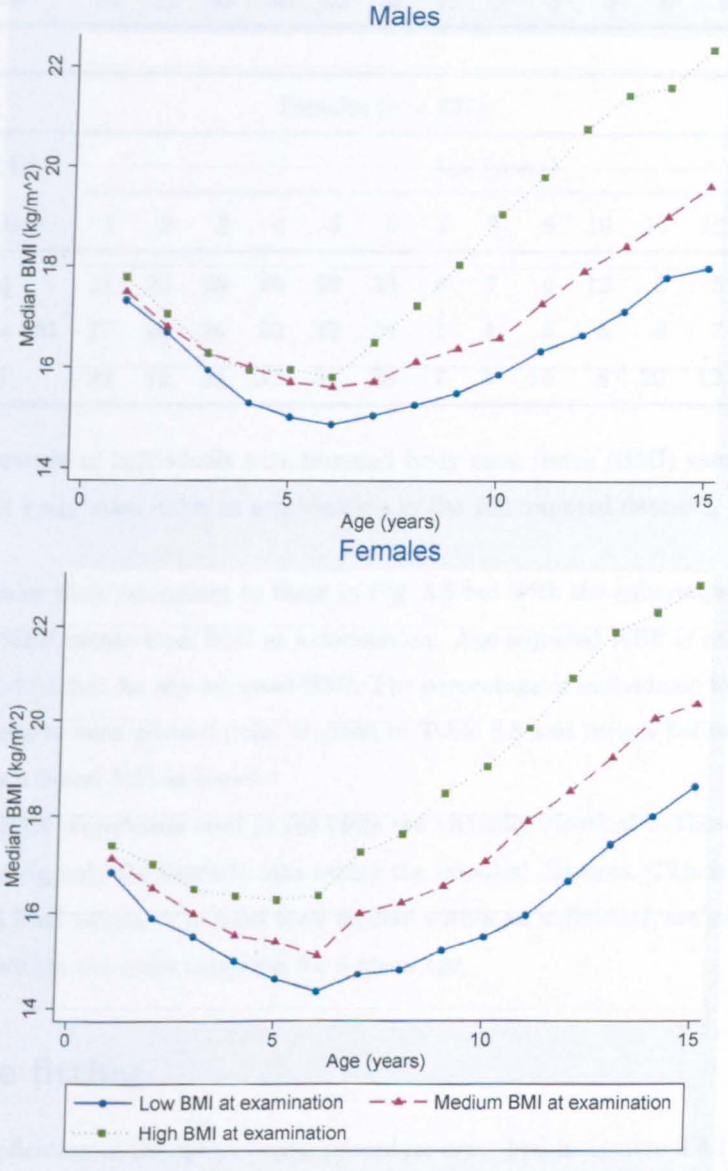

Fig. 8.5: Median body mass index (BMI) through childhood in the 100 imputed datasets, by age-adjusted body mass index at examination and sex.

Males (n = 159)															
Subgroup of BMI	Age (years)														
at examination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low (n = 53)	13	17	17	13	17	21	4	8	13	8	13	11	11	17	23
Medium (n = 53)	23	26	25	25	26	34	8	9	4	15	0	13	6	15	25
High (n = 53)	32	32	32	30	32	38	11	9	8	9	6	9	11	15	23

Females (n = 227)															
Subgroup of BMI	Age (years)														
at examination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low (n = 75)	21	23	23	20	29	33	9	7	4	12	5	5	3	8	29
Medium (n = 76)	17	20	20	20	22	24	3	1	5	5	8	7	7	13	37
High (n = 76)	29	32	34	32	34	38	7	5	14	8	20	12	14	17	45

Table 8.7: Percentage of individuals with imputed body mass index (BMI) values at each age in each subgroup of body mass index at examination in the 100 imputed datasets, by sex.

Fig. 8.6 includes plots equivalent to those in Fig. 8.5 but with the subgroups defined in terms of age-adjusted %BF rather than BMI at examination. Age-adjusted %BF is calculated using an analogous method to that for age-adjusted BMI. The percentage of individuals with imputed BMI values contributing to each plotted point is given in Table 8.8 and ranges between 4 and 36% in males and between 3 and 41% in females.

The median BMI trajectories seen in the plots are virtually identical to those in Fig. 8.4, the equivalent plot using only the original data rather the imputed datasets. This is again indicative that the imputed BMI values, or at least their median within an individual, are very similar to the observed values within the same subgroup for a given age.

8.6 Spline fitting

Details of the application of the spline-fitting procedure described in Section 8.3.1, using both the original data only (Section 8.6.1) and the imputed datasets (Section 8.6.2), follow.

8.6.1 Using the original data only

As can be seen from Table 8.2 in Section 8.4, 293 out of the 386 subjects (75.9%) have the required 6 observed BMI values between age 1 and 10 years to have splines fitted.

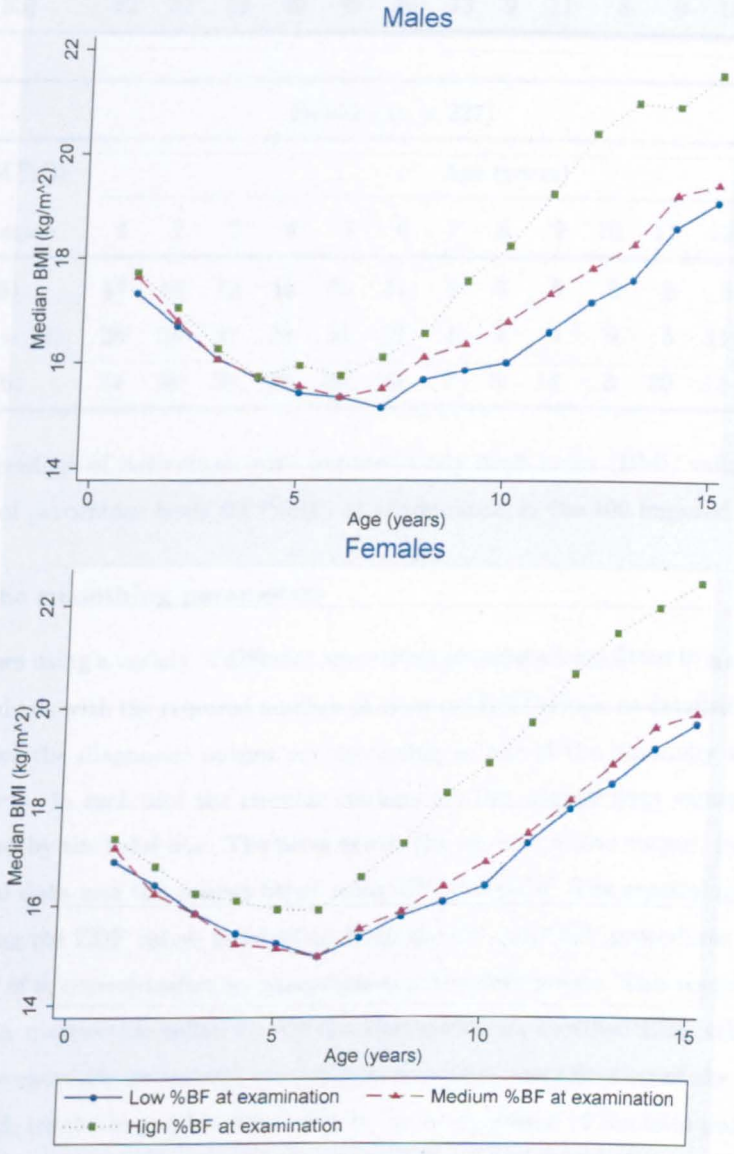

Fig. 8.6: Median body mass index (BMI) through childhood in the 100 imputed datasets, by age-adjusted percentage body fat (%BF) at examination and sex.

Males (n=159)															
Subgroup of %BF	Age (years)														
at examination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low ($n = 53$)	15	19	19	17	23	28	4	8	8	11	6	8	13	17	25
Medium ($n = 53$)	21	25	23	21	21	28	6	9	6	13	4	13	4	15	19
High ($n = 53$)	32	32	32	30	32	36	13	9	11	8	9	13	11	15	26

Females ($n = 227$)															
Subgroup of %BF	Age (years)														
at examination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low ($n = 75$)	17	19	19	16	25	31	8	5	5	8	8	3	4	12	33
Medium ($n = 76$)	26	28	28	28	30	32	4	3	4	9	5	11	9	12	37
High ($n = 76$)	24	28	30	28	30	33	7	5	14	8	20	11	11	14	41

Table 8.8: Percentage of individuals with imputed body mass index (BMI) values at each age in each subgroup of percentage body fat (%BF) at examination in the 100 imputed datasets, by sex.

Selection of the smoothing parameters

Smoothing splines using a variety of different smoothing parameters are fitted to a stratified random sample of individuals with the required number of observed BMI values, as detailed in Section 8.3.1.

Fig. 8.7 shows the diagnostic output corresponding to one of the randomly selected subjects with 9 data points. In each plot the circular markers are the original data values with the fitted spline represented by the solid line. The plots across the top row of the output, from left to right, show the original data and the splines fitted using CV and GCV. The remaining plots show the splines fitted using the EDF values as labelled. Both the CV and GCV procedures in this instance result in an EDF of 9, corresponding to interpolation of the data points. This results in insufficient smoothing and an undesirable spline fit. Of the alternative pre-specified EDF values, 3, 4 and 5 appear to provide excessive smoothing, resulting in unreliable identification of the AR. EDF of 8, on the other hand, results in a spline very close to an interpolation of the data points once more. The plots corresponding to EDF of both 6 and 7 display splines which are sufficiently smoothed to exclude minor deviations due to measurement error or other noise yet still appear to reliably identify the location of the AR. However, as these interpretations are subjective the assessment of the optimal degree of smoothness is clearly not definitive.

The procedure is repeated for all individuals in the stratified random sample and a subject-specific subjectively optimal EDF value, or range of EDF values when appropriate, selected for each, which are then analysed across the strata. It is immediately apparent that on the whole neither CV nor GCV provide a suitable degree of smoothing for the AR to be reliably identified.

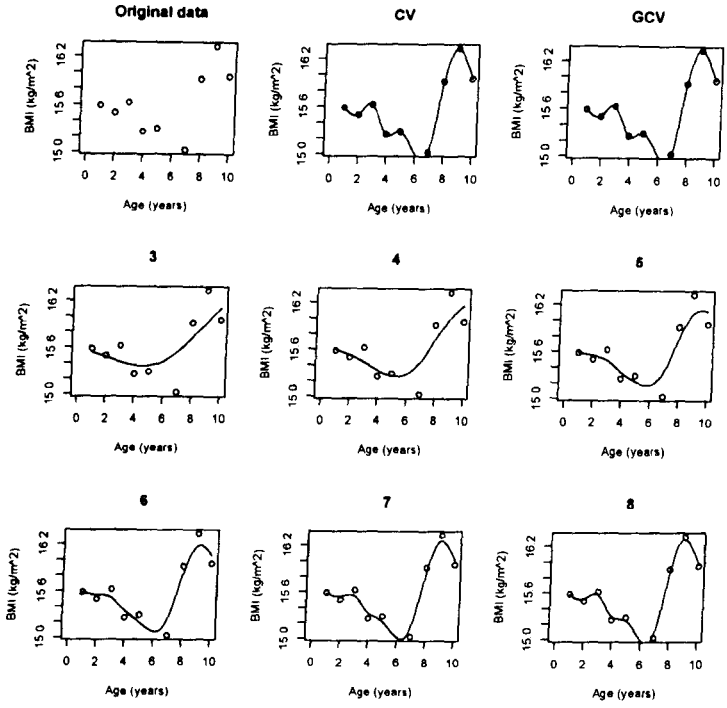

Fig. 8.7: Fitted splines for a randomly selected individual with the degree of smoothing defined by cross-validation (CV), generalised cross-validation (GCV), and equivalent degrees of freedom (EDF) ranging from 3 to 8.

Of the user-specified EDF values it emerges that for all subjects with 6 or 7 data points an EDF of 5 provides an optimal, or at least acceptable, level of fit to the data. Similarly, for those with 8, 9 or 10 data points an EDF of 6 is deemed appropriate. Thus it is decided that these values should be used across the dataset and a rule created in the spline-fitting routine specifying the EDF as a function of the number of non-missing BMI values.

Estimation of the adiposity rebound location

Fig. 8.8 shows examples of fitted splines for four individuals. In each plot the circular markers and solid lines again represent the original data values and fitted splines respectively. The upper-left plot shows a subject with data to which the fitted spline is in close agreement and an AR identified. The identified AR is signified by the square symbol, through which passes a vertical line corresponding to the age at AR and a horizontal line corresponding to the BMI at AR. Most of the fitted splines in the dataset are of this type.

The upper-right example again shows a well-fitting spline, though this time it is clearly not possible to identify an AR from the plot as it is monotone increasing. There are several individuals who have data of this type, and also some with monotone decreasing functions. As no AR can be identified, study members with this type of BMI trajectory cannot be included in any analyses which include either dimension of the AR.

The plot in the bottom-left shows a fitted spline with multiple minima. There are several examples of individuals with data corresponding to this type of spline. As no single AR location can be identified, subjects with this type of data pattern cannot contribute to analyses which

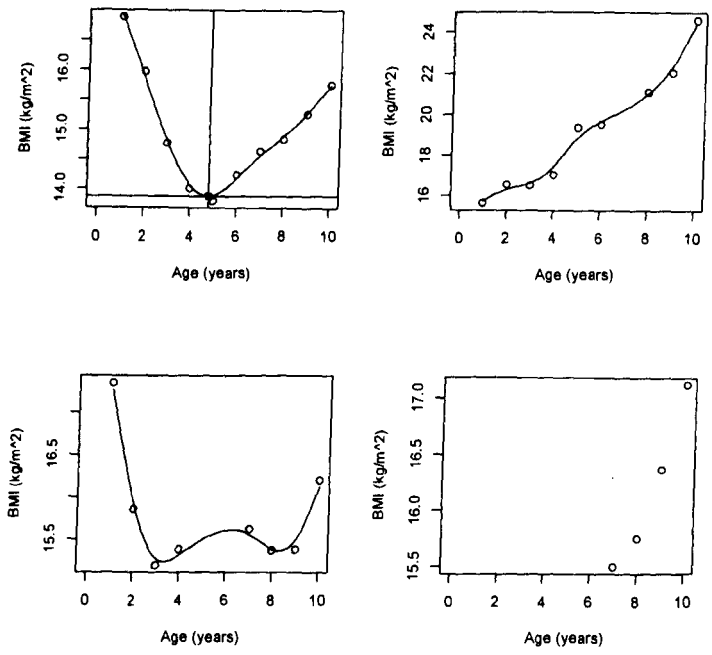

Fig. 8.8: Examples of fitted splines for four individuals.

include either age or BMI at AR.

The final example in the bottom-right plot is of an individual with insufficient data points for an AR to be reliably identified, thus no spline is even fitted to the data. As previously discussed, there are a substantial number of study members with insufficient data, often with a missing data pattern similar to that in the example whereby data are unobserved at younger ages then observed at later time points. Again, these subjects cannot be included in any analysis including either dimension of the AR.

8.6.2 Using the imputed datasets

In the 100 imputed datasets each individual has the full 10 non-missing BMI values between age 1 and 10 years, thus all 386 subjects meet the requirement of having 6 data points in order to have splines fitted. The spline-fitting procedure is identical to that followed using the original data only. As each individual has 10 data points, however, an EDF of 6 is used in each and every instance.

Whilst splines can be fitted to every subject when using the imputed datasets, there are still many individuals for whom AR location cannot be estimated from the fitted spline. The fitted splines for these subjects are generally either monotonic increasing or decreasing, or have multiple minima, as discussed in relation to Fig. 8.8 in Section 8.6.1.

8.7 Estimated adiposity rebound locations

The ages and corresponding BMI values at the ARs estimated from the fitted subject-specific splines, using both the original data only (Section 8.7.1) and the imputed datasets (Section 8.7.2), are examined.

8.7.1 Using the original data only

By fitting splines to the original data using the method outlined in Section 8.3.1, estimated ARs are identified for 261 individuals (67.6%). Table 8.9 summarises the age and BMI at which the identified ARs occur. As the distribution of BMI at AR exhibits a slight skew it is more reliable to use the median, as opposed to the mean, as a measure of the 'average' value. Median age at AR was found to be 5.7 years in the 111 males for which AR was identified and 5.5 years in the 227 females, with corresponding median BMI at AR values of 15.2 kg/m^2 and 15.0 kg/m^2.

Sex	Total number of subjects	Number (%) of subjects with AR identified	Age at AR (years)			BMI at AR (kg/m^2)		
			Mean	Median	SD	Mean	Median	SD
Males	159	111 (69.8%)	5.7	5.7	1.2	15.2	15.2	1.0
Females	227	150 (66.1%)	5.3	5.5	1.2	15.1	15.0	1.3

Table 8.9: Distributions of age and body mass index (BMI) at adiposity rebound (AR) in the original data, by sex.

Fig. 8.9 illustrates both the univariate and bivariate distributions of age and BMI at AR. From the histograms it can be seen that both males and females have slightly positively skewed distributions of BMI at AR. Meanwhile, the scatterplots show an association between the two dimensions of AR, with an earlier AR generally being associated with a higher BMI at AR and a later AR with a lower BMI, though this is more apparent in the females. Indeed, the correlations between the two dimensions (calculated using log-transformed BMI at AR) are -0.23 and -0.34 in males and females respectively.

Table 8.10 summarises the distributions of a variety of variables at birth and at examination for individuals with and without an estimated AR identified. Those individuals with an identified AR do seem to differ in some aspects to those with no identified AR. In particular, males with an identified AR appear to have lower weight at birth and lower weight, BMI, waist and hip circumferences, and %BF at examination. As all these variables are age-dependent it is important to observe that the average age at which examinations took place was very similar between those with and without identified ARs, meaning this is unlikely to be a factor in these discrepancies. Females display a similar difference in weight at birth, though those with AR identified appear similar to those with AR not identified in terms of the measurements at examination. Also of note is the greater variability in observed values for almost all variables among those with no identified

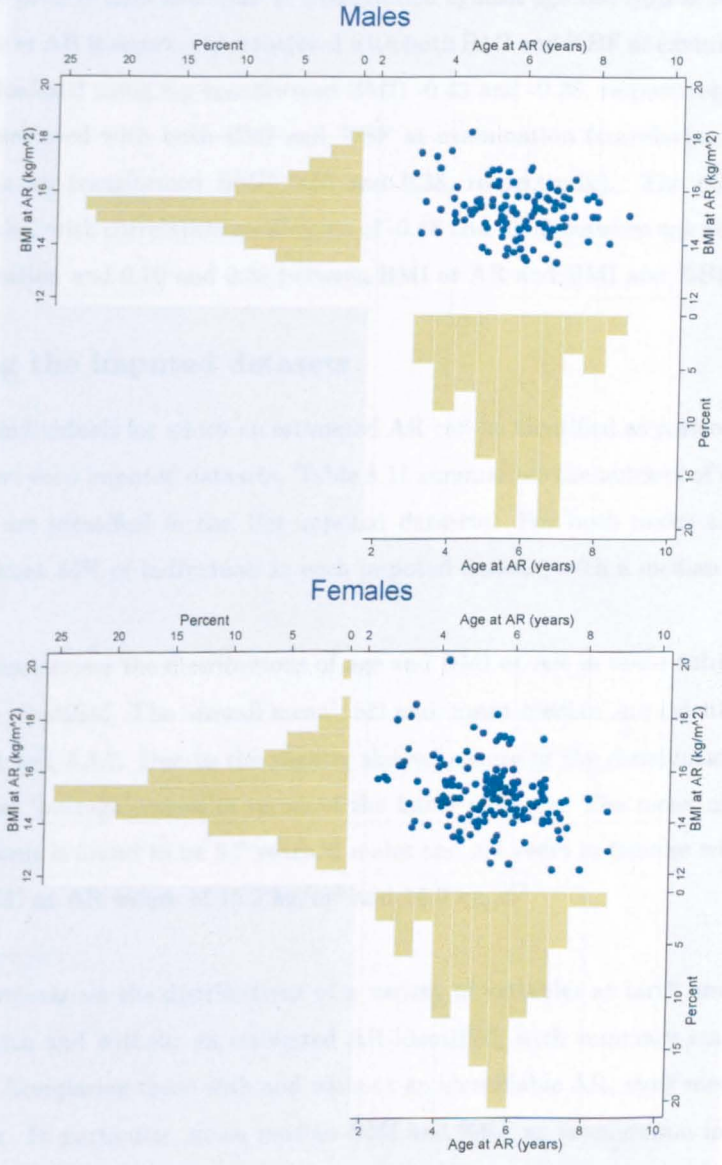

Fig. 8.9: Univariate and bivariate distributions of age and body mass index (BMI) at adiposity rebound (AR) in the original data, by sex.

AR, particularly the males.

In terms of comparisons between male and female anthropometry, these results are largely as would be expected. At birth there is little difference between the sexes, though at examination males have greater average weight and height and, whilst BMI distribution is similar between the sexes, females generally have greater %BF.

Fig. 8.10 are plots of BMI and %BF at examination against age and BMI at AR. In males, it can be seen that age at AR is negatively associated with both BMI and %BF at examination (correlation coefficients (calculated using log-transformed BMI) -0.43 and -0.36, respectively) and BMI at AR is positively associated with both BMI and %BF at examination (correlation coefficients (again calculated using log-transformed BMI) 0.56 and 0.38, respectively). The relationships appear similar for females, with correlation coefficients of -0.46 and -0.32 between age at AR and BMI and %BF at examination and 0.70 and 0.38 between BMI at AR and BMI and %BF at examination.

8.7.2 Using the imputed datasets

The number of individuals for which an estimated AR can be identified as part of the spline-fitting process differs between imputed datasets. Table 8.11 summarises the number of subjects for which estimated ARs are identified in the 100 imputed datasets. For both males and females AR is identified in at least 84% of individuals in each imputed dataset, with a median of over 88%.

Table 8.12 summarises the distributions of age and BMI at AR in those subjects for whom an estimated AR is identified. The 'overall mean', SD and 'mean median' are calculated as described in Sections 5.2.4 and 8.3.2. Due to the slightly skewed nature of the distributions it is again expedient to discuss 'average' values in terms of the latter measure. The mean median age at AR over all imputations is found to be 5.7 years in males and 5.4 years in females with corresponding mean median BMI at AR values of 15.2 kg/m^2 and 15.0 kg/m^2.

Table 8.13 summarises the distributions of a variety of variables at birth and at examination for individuals with and without an estimated AR identified, with summary statistics calculated as in Table 8.12. Comparing those with and without an identifiable AR, most mean median values are not dissimilar. In particular, mean median BMI and %BF at examination in both males and females are very similar, though there is greater variability among those with AR not identified. However, the relatively small sample sizes for both males and females with AR not identified may make the associated figures less reliable.

Whilst weight and height are generally greater in males than females, mean median BMI at examination is slightly greater in females, with mean median %BF at examination far greater.

Variable	Males ($n = 159$)					
	AR identified ($n = 111$)			AR not identified ($n = 48$)		
	Mean	Median	SD	Mean	Median	SD
At birth						
Gestational age (weeks)	39.5	40	1.8	39.5	40	1.8
Weight (kg)	3.53	3.52	0.52	3.62	3.62	0.58
At examination						
Age (years)	16.8	16.9	0.4	16.9	16.9	0.4
Weight (kg)	66.9	64.5	9.6	71.2	69.3	13.5
Height (m)	1.81	1.80	0.06	1.81	1.81	0.06
BMI (kg/m^2)	20.6	19.9	2.4	21.6	20.7	3.5
Waist circumference (cm)	73.5	73	5.7	77.1	74.5	10.0
Hip circumference (cm)	91.5	91	6.2	94.8	93	8.9
%BF	14.5	13.9	5.8	17.8	16.8	8.0

Variable	Females ($n = 227$)					
	AR identified ($n = 150$)			AR not identified ($n = 77$)		
	Mean	Median	SD	Mean	Median	SD
At birth						
Gestational age (weeks)	39.6	40	1.5	39.2	39	1.7
Weight (kg)	3.40	3.43	0.46	3.47	3.51	0.50
At examination						
Age (years)	16.8	16.8	0.4	16.9	16.9	0.4
Weight (kg)	59.4	58.9	8.5	59.2	59.1	9.4
Height (m)	1.67	1.67	0.06	1.66	1.65	0.06
BMI (kg/m^2)	21.2	20.8	2.9	21.6	21.1	3.2
Waist circumference (cm)	71.1	70	6.6	71.1	70	6.9
Hip circumference (cm)	91.8	91.5	6.4	92.2	92	6.9
%BF	29.1	28.9	6.2	28.7	28.0	6.6

Table 8.10: Distributions of variables at birth and at examination in the original data, by adiposity rebound (AR) identification and sex. SD is standard deviation, BMI is body mass index and %BF is percentage body fat.

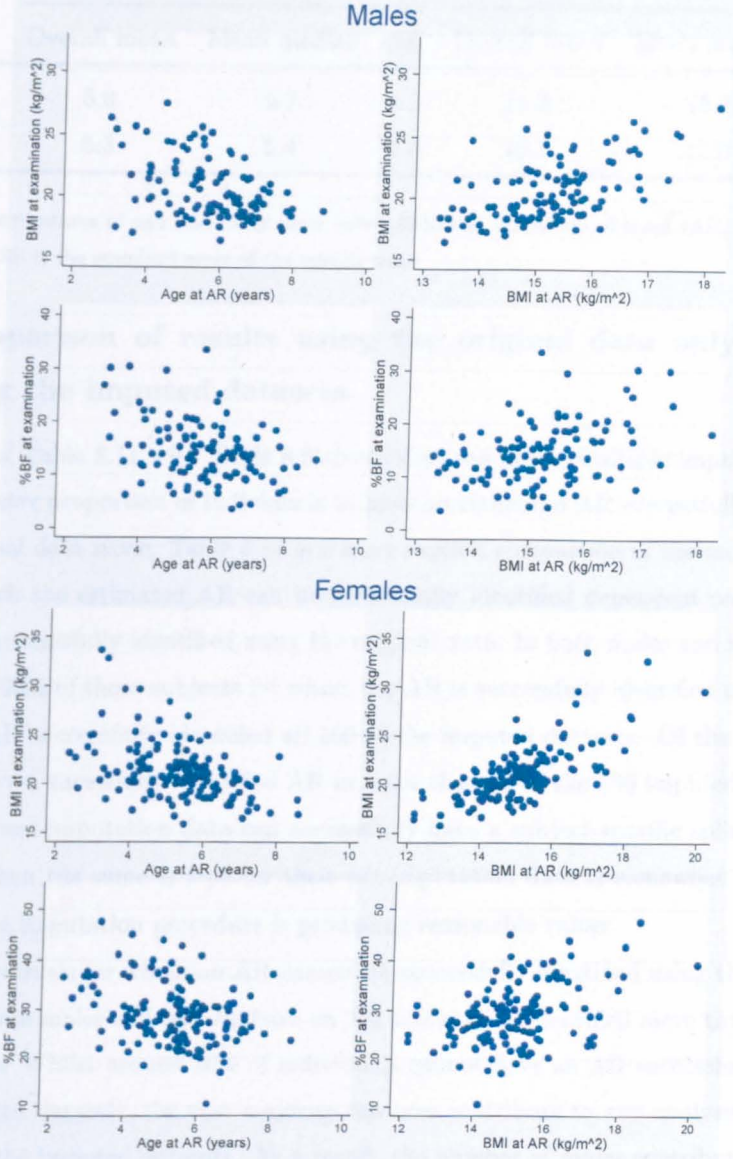

Fig. 8.10: Body mass index (BMI) and percentage body fat (%BF) at examination against age and body mass index at adiposity rebound (AR) in the original data, by sex.

Sex	Total subjects	Subjects with AR identified in each imputation (%)		
		Min.	Median	Max.
Males	159	134 (84.3%)	141 (88.7%)	146 (91.8%)
Females	227	193 (85.0%)	200 (88.1%)	207 (91.2%)
Total	386	331 (85.8%)	341 (88.3%)	351 (90.9%)

Table 8.11: Number of subjects with adiposity rebound (AR) identified in the 100 imputed datasets, by sex.

Sex	Age at AR (years)			BMI at AR (kg/m^2)		
	Overall mean	Mean median	SE	Overall mean	Mean median	SE
Males	5.6	5.7	1.2	15.3	15.2	1.0
Females	5.3	5.4	1.3	15.1	15.0	1.3

Table 8.12: Distributions of age and body mass index (BMI) at adiposity rebound (AR) in the 100 imputed datasets, by sex. SE is the standard error of the overall mean.

8.7.3 Comparison of results using the original data only and results using the imputed datasets

A comparison of Table 8.11 with Table 8.9 shows that use of the multiple imputation procedure allows a far greater proportion of individuals to have an estimated AR successfully identified than use of the original data alone. Table 8.14 is a more explicit comparison of the number of imputed datasets in which the estimated AR can be successfully identified dependent on whether or not the AR can be successfully identified using the original data. In both males and females it can be seen that about 90% of those subjects for whom the AR is successfully identified using the original data have the AR successfully identified all 100 of the imputed datasets. Of the remaining 10%, the majority have a sucessfully identified AR in more than 80 of the 100 imputed datasets. That an individual's post-imputation data can successfully have a subject-specific spline fitted and an AR identified when the same is true for their pre-imputation data is somewhat reassuring as it indicates that the imputation procedure is producing reasonable values.

Of those individuals for whom an AR *cannot* be successfully identified using the original data, around 60% of both males and females have an AR successfully identified more than 80 of the 100 imputed datasets. Whilst around 20% of individuals cannot have an AR successfully identified in *any* of the imputed datasets, the vast majority can now contribute to any analysis undertaken in at least some of the imputed datasets. As a result, the number of males contributing to analyses increases from 111 (69.8%) using the original data to a median of 141 (88.7%) using the imputed datasets. The equivalent increase in females is from 150 (66.1%) to a median of 200 (88.1%). This enlarged sample size should increase the power of any analysis.

Whilst the effective sample sizes are increased, a comparison of Table 8.12 with Table 8.9 shows that the distribution of identified ARs changes little. The 'mean median' ages at AR over the 100 imputation datasets of 5.7 and 5.4 years for males and females respectively are very similar to the medians of 5.7 and 5.5 years using the original data. The mean median BMI at AR values of 15.2 kg/m^2 (males) and 15.0 kg/m^2 (females) in the imputed datasets are identical to the medians using the original data.

The summaries of distributions of various anthropometric variables in the 100 imputed datasets (Table 8.13) show that the differences between those subjects with AR identified and not identified

Variable	Males (n = 159)					
	AR identified (n = 134–146)			AR not identified (n = 13–25)		
	Overall mean	Mean median	SE	Overall mean	Mean median	SE
At birth						
Gestational age (weeks)	39.5	40.0	1.8	39.2	39.9	1.7
Weight (kg)	3.57	3.53	0.53	3.48	3.60	0.63
At examination						
Age (years)	16.9	16.9	0.4	16.9	16.9	0.3
Weight (kg)	68.3	66.6	10.6	66.9	64.2	14.4
Height (m)	1.81	1.80	0.06	1.79	1.78	0.07
BMI (kg/m^2)	20.9	20.2	2.7	20.6	20.3	3.4
Waist circumference (cm)	74.6	73.7	6.9	74.2	71.5	11.0
Hip circumference (cm)	92.6	91.3	7.1	92.0	90.6	8.9
%BF	15.4	14.0	6.7	16.1	16.1	6.8

Variable	Females (n = 227)					
	AR identified (n = 193–207)			AR not identified (n = 20–34)		
	Overall mean	Mean median	SE	Overall mean	Mean median	SE
At birth						
Gestational age (weeks)	39.5	40.0	1.6	39.3	39.3	1.7
Weight (kg)	3.41	3.44	0.47	3.53	3.59	0.55
At examination						
Age (years)	16.8	16.8	0.4	16.8	16.8	0.4
Weight (kg)	59.4	59.1	8.5	58.7	56.6	10.9
Height (m)	1.67	1.67	0.06	1.65	1.65	0.05
BMI (kg/m^2)	21.3	20.9	2.8	21.4	20.6	3.8
Waist circumference (cm)	71.0	70.0	6.5	71.1	68.9	7.8
Hip circumference (cm)	92.0	92.0	6.3	91.2	89.2	8.1
%BF	29.1	28.8	6.3	28.3	27.0	7.6

Table 8.13: Distributions of variables at birth and at examination in the 100 imputed datasets, by adiposity rebound (AR) identification and sex. SE is the standard error of the overall mean, BMI is body mass index and %BF is percentage body fat.

Number of imputed datasets with AR identified	Males ($n = 159$)			Females ($n = 227$)			Total ($n = 386$)		
	AR identified using original data?		Total ($n = 159$)	AR identified using original data?		Total ($n = 227$)	AR identified using original data?		Total ($n = 386$)
	No ($n = 48$)	Yes ($n = 111$)		No ($n = 77$)	Yes ($n = 150$)		No ($n = 125$)	Yes ($n = 261$)	
0	10 (20.8%)	0 (0.0%)	10 (6.3%)	13 (16.9%)	0 (0.0%)	13 (5.7%)	23 (18.4%)	0 (0.0%)	23 (6.0%)
1–20	0 (0.0%)	1 (0.9%)	1 (0.6%)	0 (0.0%)	2 (1.3%)	2 (0.9%)	0 (0.0%)	3 (1.2%)	3 (0.8%)
21–40	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.3%)	1 (0.7%)	2 (0.9%)	1 (0.8%)	1 (0.4%)	2 (0.5%)
41–60	2 (4.2%)	0 (0.0%)	2 (1.3%)	6 (7.8%)	0 (0.0%)	6 (2.6%)	8 (6.4%)	0 (0.0%)	8 (2.1%)
61–80	8 (16.7%)	3 (2.7%)	11 (6.9%)	9 (11.7%)	4 (2.7%)	13 (5.7%)	17 (13.6%)	7 (2.7%)	24 (6.2%)
81–99	26 (54.2%)	8 (7.2%)	34 (21.4%)	46 (59.7%)	9 (6.0%)	55 (24.2%)	72 (57.6%)	17 (6.5%)	89 (23.1%)
100	2 (4.2%)	99 (89.2%)	101 (63.5%)	2 (2.6%)	134 (89.3%)	136 (59.9%)	4 (3.2%)	233 (89.3%)	237 (61.4%)

Table 8.14: Number and percentage (%) of imputed datasets in which the adiposity rebound (AR) can be successfully identified, by adiposity rebound identification in the original dataset and sex.

are generally reduced from those when using the original data (Table 8.10). However, these figures should be viewed with some caution due to the small sample sizes for those with AR not identified.

8.7.4 Comparison with previously published results

Median age at AR is found to be 5.7 years in those males for whom estimated AR is successfully identified when using both the original data only or the imputed datasets. In females, median age at AR is 5.5 years when using the original data only and 5.4 years when utilising multiple imputation. These values correspond reasonably well to previously published results.

Rolland-Cachera *et al* [82], in their initial AR paper concerning a sample of 151 French children from a longitudinal study of growth started in 1953, found 23 of the 79 males in their study (29.1%) and 23 of the 72 females (31.9%) to have AR at age less than or equal to 5.5 years. This compares to equivalent figures of 43.2% and 52.7% for SWEDES. Rolland-Cachera *et al* reported that 28 (35.4%) of the males and 25 (34.7%) of the females at age greater than or equal to 7.0 years, compared to 14.4% in both males and females in the present study.

Siervogel *et al* [84] fitted subject-specific cubic polynomials on log(BMI) for each of 496 children in the Fels longitudinal study. They reported mean ages at AR of 5.1 years and 5.3 years for males and females respectively. This finding of younger age at AR in males as opposed females does not agree with that seen using SWEDES and is rather anomalous when compared to other results.

Using a similar method to Siervogel *et al* for a USA study of 390 children born 1965–71 Whitaker *et al* [85] reported mean ages at AR of 5.8 years (males) and 5.4 years (females).

Williams *et al* [86] investigated age at AR using two different methods for a study of 922 New Zealand children born 1972–73. The first method, fitting subject-specific cubic polynomials on log(BMI), resulted in mean ages at AR of 6.3 years for males and 6.1 years for females. The second method, utilising a random coefficients model fitted on log(BMI) with two separate cubic polynomials for males and females and a different cubic polynomial for each individual, gave corresponding values of 6.0 years and 5.6 years.

Williams [172] fitted random coefficient cubic polynomials for log(height) and log(weight) for a study of 803 New Zealand children born 1972-3. Velocity curves were calculated by taking the first derivatives of the fitted curves and ARs identified as the point at which the velocity of log(weight) becomes greater than twice the velocity of log(height). Mean age at AR was reported to be 6.6 years for males and 6.0 years for females.

Skinner *et al* [165] visually determined the ARs of 70 white children born in 1992 in the USA. They reported mean ages at AR of 4.7 years (males) and 4.5 years (females).

For a contemporary dataset of 39 white girls in New Zealand Taylor *et al* [171], using a similar method to Williams [172], reported a mean age at AR of 5.1 years for the females in the study.

Clearly there is a certain amount of heterogeneity in the previously published results, as would be expected given the temporal, geographical and methodological differences between the studies. The present results using the SWEDES dataset are generally similar to those using datasets with comparable characteristics. One observable trend is that in the older datasets there is a tendency

towards later AR, whilst studies using the more contemporary datasets generally report younger ages at AR. This shift may be attributed to the acknowledged secular trends in increasing developmental tempo over recent years [22].

Few previously published results regarding the AR have reported the BMI as well as the age at AR. For the SWEDES dataset median BMI at AR was found to be 15.2 kg/m^2 for males and 15.0 kg/m^2 for females when using the original data only, with identical figures being obtained for the imputed datasets.

These values are again comparable to previously published results, with Siervogel *et al* [84] reporting BMI at AR values of 15.6 kg/m^2 for males and 14.8 kg/m^2 for females, and Williams *et al* [86] finding values of 15.8 kg/m^2 and 15.2 kg/m^2 using subject-specific cubic polynomials on log(BMI) and 15.7 kg/m^2 and 15.5 kg/m^2 using a random coefficients cubic polynomial model fitted on log(BMI).

Some previously published studies have also included calculated correlation coefficients between the two dimensions of AR and later outcome variables. Again, results using SWEDES appear largely comparable.

In the present study the correlation between age at AR and BMI at examination (mean age 16.8±0.4 years) is found to be −0.43 using the original data only and −0.47 using the imputed datasets for males, and −0.46 (original data only) and −0.44 (imputed datasets) for females. Siervogel *et al* [84] reported corresponding correlations of −0.46 for males and −0.54 for females, although their outcome was measured at age 18 years. Williams *et al* [86] found correlations of −0.59 (males) and −0.39 (females) for BMI age 18 years, and −0.56 and −0.43 for BMI age 21 years.

Using the SWEDES dataset, the correlation between BMI at AR and BMI at examination is found to be 0.56 (original data only) and 0.53 (imputed datasets) for males, with corresponding values of 0.70 and 0.64 for females. Siervogel *et al* [84] reported correlations of 0.51 and 0.58 for males and females respectively, whereas Williams *et al* [86] found them to be 0.61 (males) and 0.39 (females) for BMI at age 18 years and 0.48 and 0.43 at age 21 years.

8.8 Graphical exploration of the adiposity rebound

An initial exploration of the AR using graphical methods is enlightening. Only plots using the original data are examined as equivalent plots using the imputed datasets, taking the median value at each time point, are very similar.

First, plots of median BMI through childhood in different subgroups are examined. These plots are useful tools to informally assess any patterns in the data, but, as data are examined on a group level, the temptation to make inferences on an individual level must be avoided.

Fig. 8.11 is a plot of median BMI through childhood in the three subgroups defined by the sex-specific tertiles of age at AR. An 'early AR' corresponds to an age less than 5.24 years in males and 4.96 years in females, with a 'late AR' being an age greater than 6.30 years in males and 5.87 years in females. A 'middle AR' corresponds to the ages between these values. The percentage of individuals who contribute to each plotted point is given in Table 8.15 and ranges between 74 and 100% in males and between 58 and 100% in females.

Both males and females with an early AR appear to have the highest level of BMI at AR, but only in the females does a late AR correspond to the lowest BMI at AR. It can be seen for both males and females that at age 15 years those with an early AR have the highest median BMI and those with a late AR the lowest. In fact, by age 7 years in the males and 5 years in the females this ordering is already established, remaining the same throughout this period. This is evidence of BMI tracking.

In the period before the AR the levels of BMI in the subgroups are much more similar and the ordering of the tertiles more changeable. At age 1 year the ordering is the same as in adolescence with an early AR corresponding to the highest BMI and a late AR to the lowest BMI in both males and females. Those with an early AR then have a rapid reduction in BMI immediately before AR, so that at this point they in fact have the lowest median BMI.

Males ($n = 111$)															
Subgroup of	Age (years)														
age at AR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Early ($n = 36$)	100	97	100	100	94	92	97	97	89	92	94	92	92	86	89
Middle ($n = 38$)	100	95	97	100	97	84	97	95	87	95	87	87	87	74	79
Late ($n = 37$)	100	100	100	100	97	92	95	95	97	92	97	92	95	92	78

Females ($n = 150$)															
Subgroup of	Age (years)														
age at AR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Early ($n = 50$)	100	98	96	96	92	90	100	98	92	98	90	94	92	92	70
Middle ($n = 50$)	100	98	98	98	94	96	98	96	96	96	92	96	90	88	58
Late ($n = 50$)	98	98	98	100	98	88	96	96	96	94	86	96	94	96	64

Table 8.15: Percentage of individuals with observed body mass index (BMI) values at each age in each subgroup of age at adiposity rebound (AR) in the original data, by sex.

Fig. 8.12 is an identical plot to Fig. 8.11 but with both age and BMI centred about their median values in each subgroup. This allows a comparison of the shape of the median BMI trajectory within each subgroup separately from the effects of the displacement caused by the

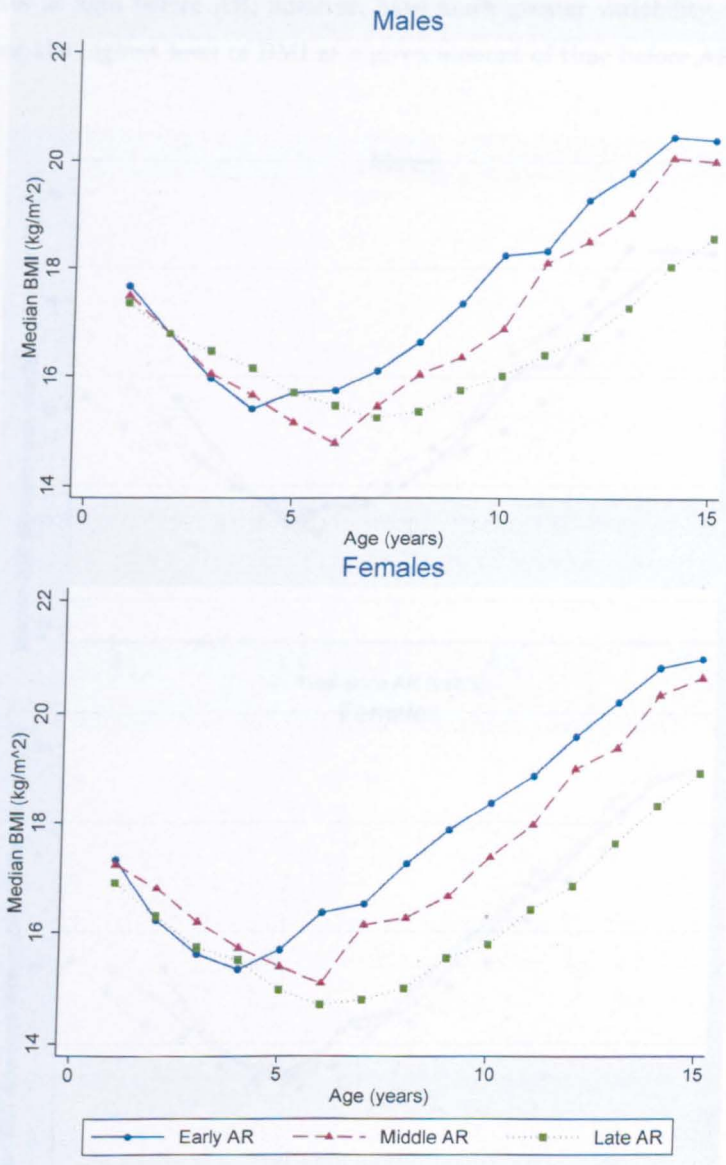

Fig. 8.11: Median body mass index (BMI) through childhood in the original data, by age at adiposity rebound (AR) and sex.

definition of the subgroups. As this definition remains the same, the number of observed BMI values contributing to each plotted point is the same as for Fig. 8.11, as given in Table 8.15.

From Fig. 8.12 it can be seen, especially in females, that the median trajectories are very similar in each subgroup from the age at AR (0 on the x-axis) onwards. This indicates that, besides the displacement in age caused by defining the subgroups on age at AR and the displacement in BMI caused by the association between age and BMI at AR, the BMI trajectories differ very little. The BMI trajectories at ages before AR, however, have much greater variability, with those with an early AR having the highest level of BMI at a given amount of time before AR.

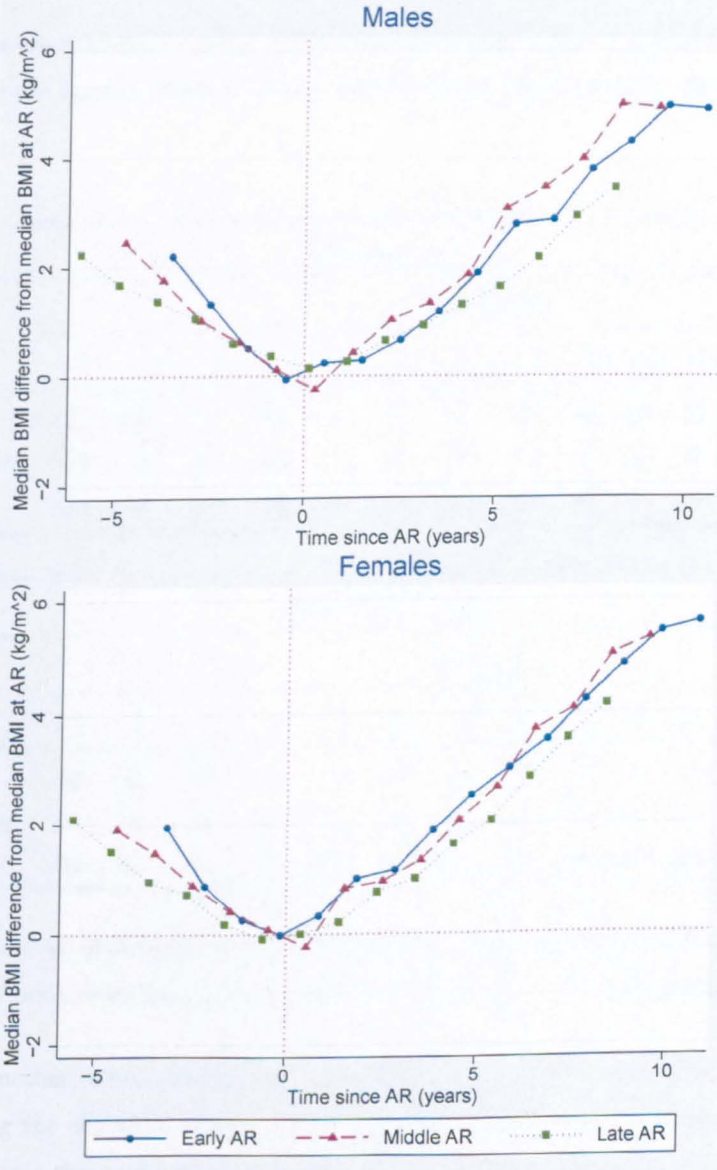

Fig. 8.12: Centred median body mass index (BMI) through childhood in the original data, by age at adiposity rebound (AR) and sex.

Fig. 8.13 is a plot of median BMI through childhood in the three subgroups defined by the tertiles of BMI at AR. A 'low BMI at AR' corresponds to a BMI of less than 14.84 kg/m^2 for

males and 14.56 kg/m^2 for females, with a 'high BMI at AR' being a BMI greater than 15.60 kg/m^2 in males and 15.53 kg/m^2 in females. A 'medium BMI at AR' corresponds to a BMI between these values. The number of observed BMI values contributing to each plotted point is given in Table 8.16 and ranges between 76 and 100% in males and between 56 and 100% in females.

It can be seen that for both males and females throughout the entirety of the age range examined those with a high BMI at AR have the highest median BMI and those with a low BMI at AR the lowest. Whilst the median BMI levels in the tertiles are slightly more similar at age 1 year, beyond this age the differences remain relatively constant. This shows that the differences in BMI evident at age 15 years are already established at much younger ages, again providing strong evidence of BMI tracking.

In the males the minimum median BMI observed in each tertile is at approximately the same age, whereas in the females there is some evidence that a higher BMI at AR corresponds to an earlier AR.

Males ($n = 111$)															
Subgroup of	Age (years)														
BMI at AR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low ($n = 37$)	100	100	97	100	95	89	97	95	86	95	86	89	89	81	76
Medium ($n = 37$)	100	95	100	100	97	86	97	92	95	89	95	89	92	84	86
High ($n = 37$)	100	97	100	100	97	92	95	100	92	95	97	92	92	86	84

Females ($n = 150$)															
Subgroup of	Age (years)														
BMI at AR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low ($n = 50$)	98	98	98	96	88	88	94	96	96	94	90	94	94	94	76
Medium ($n = 50$)	100	98	98	100	96	90	100	96	96	96	90	98	94	94	60
High ($n = 50$)	100	98	96	98	100	96	100	98	92	98	88	94	88	88	56

Table 8.16: Percentage of individuals with observed body mass index (BMI) values at each age in each subgroup of body mass index at adiposity rebound (AR) in the original data, by sex.

Fig. 8.14 is another plot of median BMI through childhood within subsets of the data, though this time showing the effects of interaction between age and BMI at AR. Rather than split each dimension of AR into three subgroups, providing nine interaction subgroups each with low membership, each is split about the median (as given in Table 8.9), resulting in four interaction subgroups (early AR and low BMI at AR, early AR and high BMI at AR, late AR and low BMI at AR, and late AR and high BMI at AR). The number of observed BMI values contributing to each plotted point is given in Table 8.17 and ranges between 77 and 100% in males and between 56 and 100%

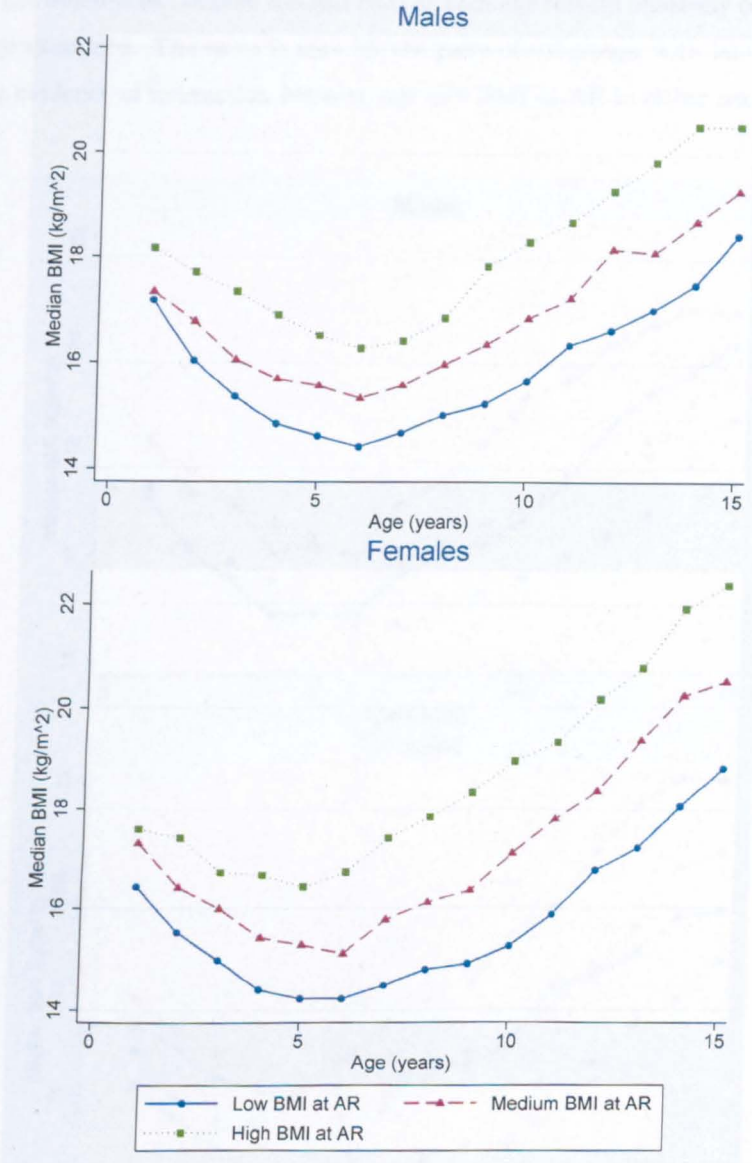

Fig. 8.13: Median body mass index (BMI) through childhood in the original data, by body mass index at adiposity rebound (AR) and sex.

in females.

At age 15 years for both males and females those with an early AR and high BMI at AR have the highest median BMI and those with a late AR and low BMI at AR the lowest. Indeed, this is true from age 7 years onwards in the males and age 6 years onwards in the females, whilst the remaining two subgroups have similar median BMI values through this period.

Considering, initially, the pairs of subgroups with early AR, it can be seen that in both males and females the differences between median BMI at each age remain relatively constant throughout the age range examined. The same is true for the pairs of subgroups with late AR, meaning that there is little evidence of interaction between age and BMI at AR in either sex.

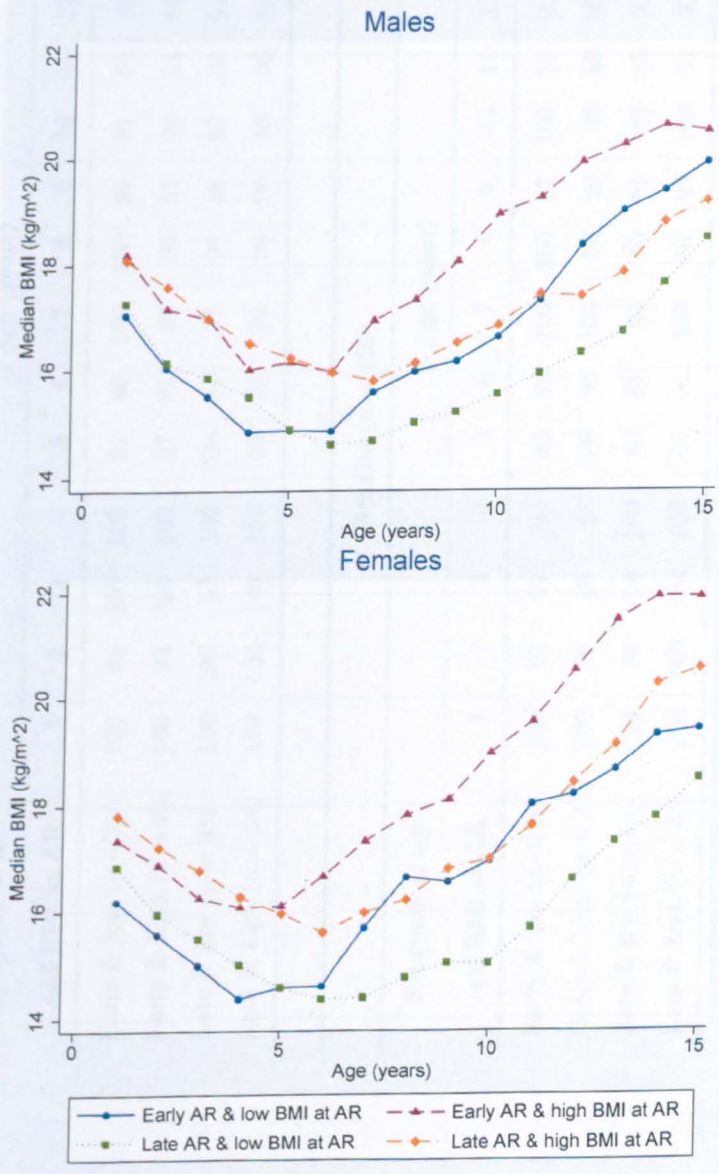

Fig. 8.14: Median body mass index (BMI) through childhood in the original data, by age at adiposity rebound (AR), body mass index at adiposity rebound and sex.

In Fig. 8.15 this comparison is made easier by the centring of each subgroup about its median

Males ($n = 111$)															
Subgroup of age	Age (years)														
and BMI at AR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Early & low ($n = 22$)	100	95	100	100	91	86	100	100	91	91	95	95	86	82	77
Early & high ($n = 32$)	100	94	100	100	97	91	97	94	91	94	94	88	94	81	91
Late & low ($n = 33$)	100	100	97	100	100	91	97	94	88	97	88	91	91	85	79
Late & high ($n = 24$)	100	100	100	100	96	88	92	96	96	88	96	88	92	88	79

Females ($n = 150$)															
Subgroup of age	Age (years)														
and BMI at AR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Early & low ($n = 34$)	100	97	97	94	85	88	100	100	97	100	97	97	94	91	65
Early & high ($n = 41$)	100	98	95	98	100	98	100	95	90	98	85	95	90	90	63
Late & low ($n = 41$)	98	98	98	100	93	88	93	95	98	88	85	93	93	95	71
Late & high ($n = 34$)	100	100	100	100	100	91	100	97	94	100	91	97	91	91	56

Table 8.17: Percentage of individuals with observed body mass index (BMI) values at each age in each subgroup of age and body mass index at adiposity rebound (AR) in the original data, by sex.

age and BMI at AR. After the median age at AR in each subgroup (0 on the x-axis), there is relatively little variability in the shape of the median trajectories in each subgroup, particularly in the females. However, some ordering does remain in both sexes, with those with an early AR and/or high BMI at AR generally having a higher median increase in BMI at a given time since AR.

One feature of the plot is that the trajectories often lie in pairs, with the two subgroups with an early AR having similar median BMI increases at a given time since AR and those with a late AR doing likewise. This suggests that, conditional on age at AR, BMI at AR has relatively little impact on later BMI.

Before the median age at AR in each subgroup there is some variability in the trajectory shapes, particularly in the males. Again, the trajectories lie largely in pairs with both early AR subgroups showing a more rapid decline in BMI prior to AR.

Fig. 8.16 is a plot of the correlations between age at AR and BMI through childhood and between BMI at AR and BMI through childhood. The dotted vertical line corresponds to the sex-specific median age at AR. The percentage of individuals with observed BMI values contributing to each plotted point is given in Table 8.18 and ranges between 82 and 100% in males and between 64 and 99% in females.

In both males and females the correlation between BMI at AR and BMI through childhood increases from around 0.6 at age 1 year to a peak of over 0.9 just prior to the median age at AR. The correlation then decreases with age until it reaches a plateau of around 0.6 from age 12 years onwards in males and around 0.7 from age 10 years onwards in females. Whilst clearly it would be expected that the BMI around the age of AR is highly correlated with the BMI at AR, the high levels of correlation remaining several years after AR illustrate a high level of BMI tracking.

The correlation between age at AR and BMI through childhood is, however, a little more difficult to interpret. In both sexes the correlation is close to zero through infancy, indicating that BMI at this age is not predictive of age at AR. A year or so before the median age at AR correlation begins to increase in magnitude. In females the correlation is nearly -0.5 at the median age at AR (5.5 years), though continues increasing in magnitude to around -0.6 at age 8 years. The correlation then gradually decreases in magnitude across the remain age range, though is still around -0.5 at age 15 years. In males the correlation is about -0.25 at the median age at AR (5.7 years), with a maximum magnitude of around -0.6 not reached until age 13 years (although magnitude increases little from age 7 years onwards). That the highest degree of correlation in females corresponds approximately to the age when the latest ARs occur seems reasonable as it is only at this age that all individuals are at a similar juncture of their BMI trajectory. For the peak correlation in the males to occur several years after the latest ARs is, however, somewhat surprising although, as has been noted, the level of correlation remains relatively constant for some while before this. Once again, the stable levels of correlation seen throughout adolescence are indicative of strong BMI tracking.

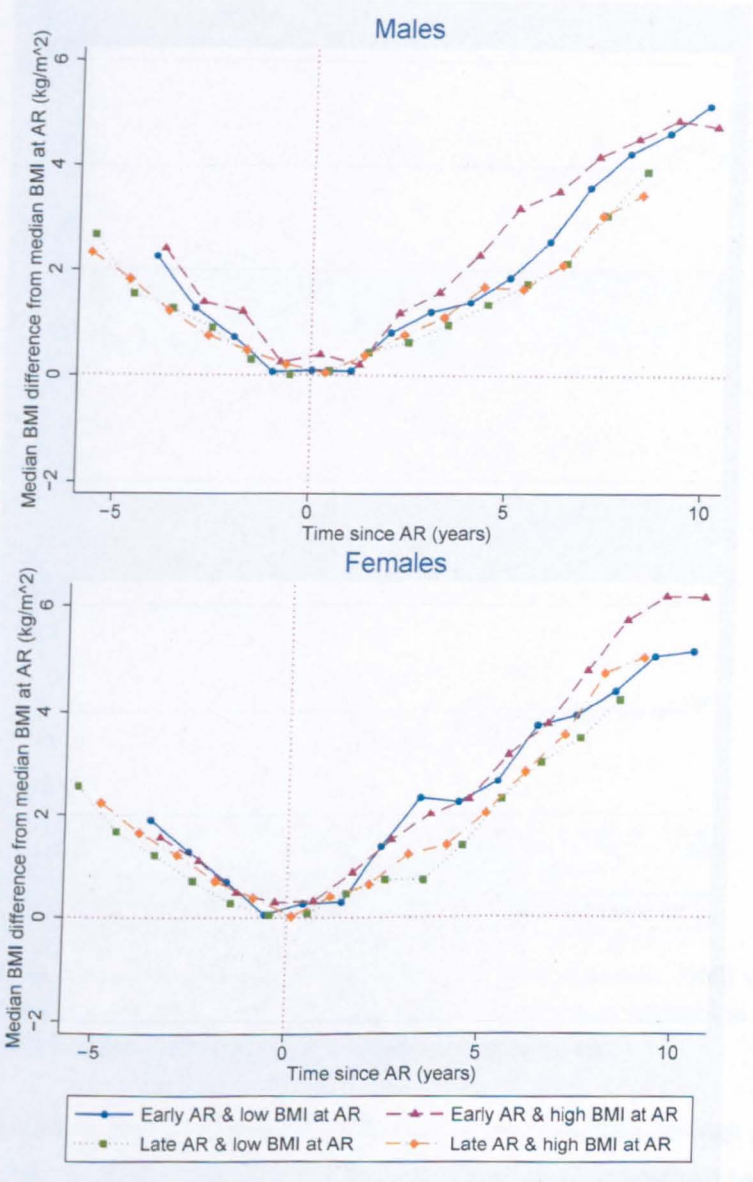

Fig. 8.15: Centred median body mass index (BMI) through childhood in the original data, by age at adiposity rebound (AR), body mass index at adiposity rebound and sex.

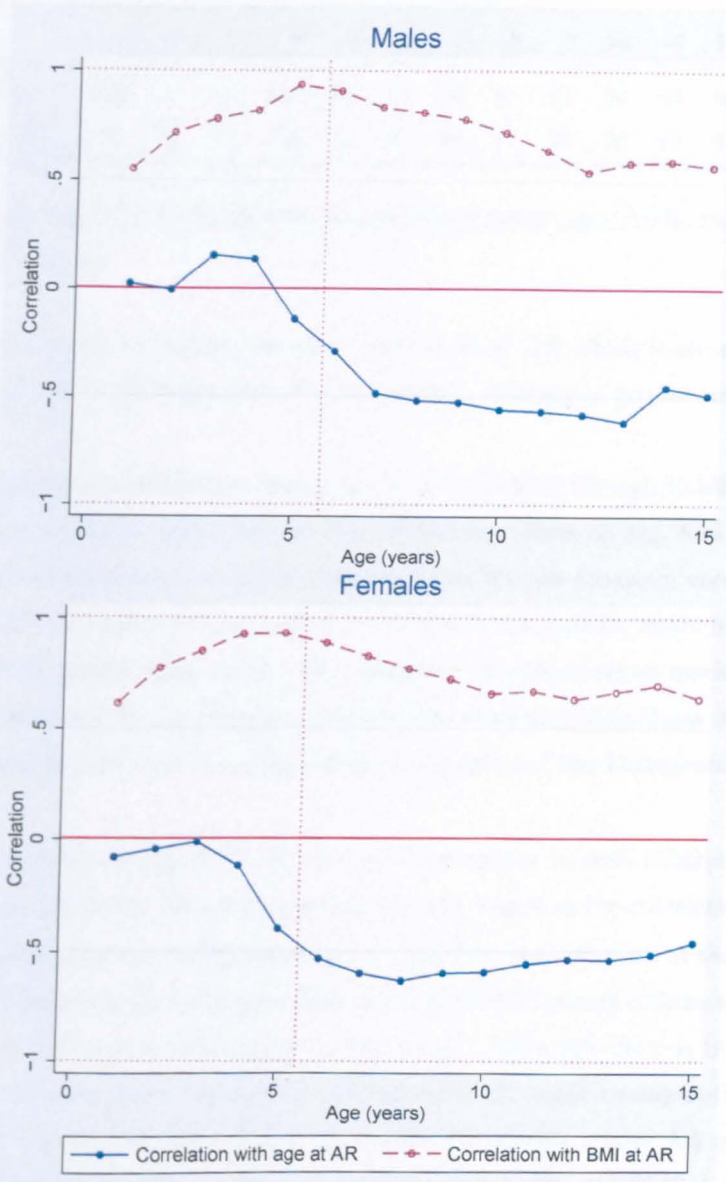

Fig. 8.16: Correlations between age at adiposity rebound (AR) and body mass index (BMI) through childhood and between body mass index at adiposity rebound and body mass index through childhood in the original data, by sex. The dotted vertical line corresponds to the sex-specific median age at AR.

Fig. 8.17 is a plot of the correlations between age at AR and BMI through childhood and between BMI at AR and BMI through childhood in the three subgroups defined by the tertiles of age at AR (early, middle and late AR) as in Fig. 8.11. The dotted vertical lines correspond to the sex-specific median age at AR in each subgroup. The percentage of individuals who contribute to each plotted point is the same as for Fig. 8.11, as given in Table 8.15.

Considering first the correlations between BMI at AR and BMI through childhood, it can be seen that in both sexes an earlier AR corresponds to an earlier peak in correlation. As correlation was shown to peak around the age of AR in the dataset as a whole in Fig. 8.16, this is somewhat expected. What is less so, however, is that, particularly in the males, towards the end of the age

Sex	Age (years)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Males ($n = 111$)	100	97	99	100	96	89	96	95	91	93	93	90	91	84	82
Females ($n = 150$)	99	98	97	98	95	91	98	97	95	96	89	95	92	92	64

Table 8.18: Percentage of individuals with observed body mass index (BMI) values at each age in the original data, by sex.

range examined it is the correlation for those with an early AR which is strongest when in this subgroup the time since AR is greatest. This is, perhaps, evidence of greater tracking in the early AR subgroup.

Considering now the correlations between age at AR and BMI through childhood, the patterns seen for males are similar to those seen for the dataset as a whole in Fig. 8.16 only the increase in the magnitude of correlation occurs at different times in each subgroup corresponding to the relevant age at AR. A similar pattern is largely evident in the females, apart from in the middle AR subgroup which, rather anomalously, has a negligible correlation across much of the age range. This is possibly explained by this subgroup being, as the remaining individuals once those at both extremes of the age at AR scale have been removed, a somewhat less homogeneous group.

Fig. 8.18 is identical to Fig. 8.17 only with the correlation in each subgroup centred about the median age at AR in the subgroup, meaning the the shapes of the correlation curves can be examined separately from the displacement effects caused by the definition of the subgroups.

It is clear that the correlations between BMI at AR and BMI through childhood peak at around the median age at AR in each subgroup (0 on the x-axis). After AR there is little variability in the correlation in females, but in males those with an early AR retain a comparatively higher level of correlation for a given time after AR. Prior to AR, those with a later AR appear to have a higher level of correlation between BMI at a given time before AR and BMI at AR, particularly in females.

In the males, the increases in magnitude of the correlation between age at AR and BMI in each subgroup are seen to occur at very similar times relative to the age at AR, as was suggested by Fig. 8.17. From about 2 years after the AR onwards there is little variability in the level of correlation in the subgroups. In females, the same features are displayed by the early and late AR subgroups, but those with a middle AR remain somewhat anomalous.

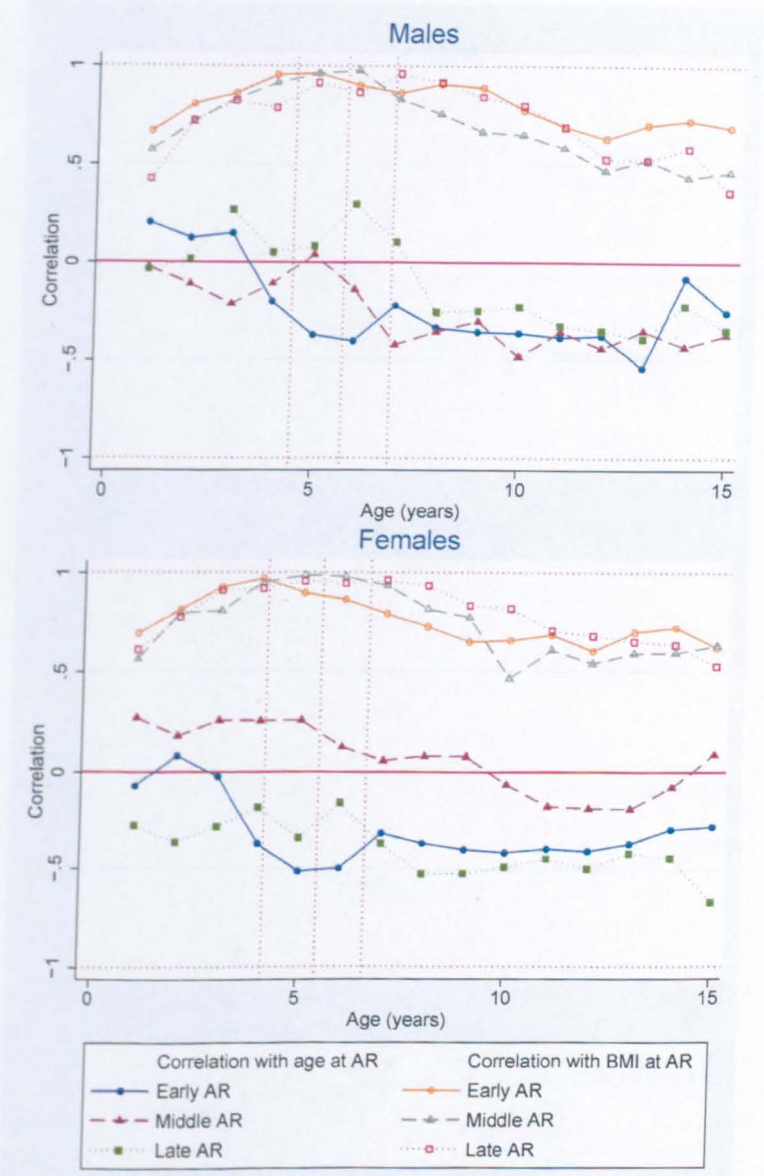

Fig. 8.17: Correlations between age at adiposity rebound (AR) and body mass index (BMI) through childhood and between body mass index at adiposity rebound and body mass index through childhood in the original data, by age at adiposity rebound and sex. The dotted vertical lines correspond to the sex-specific median age at AR in each subgroup.

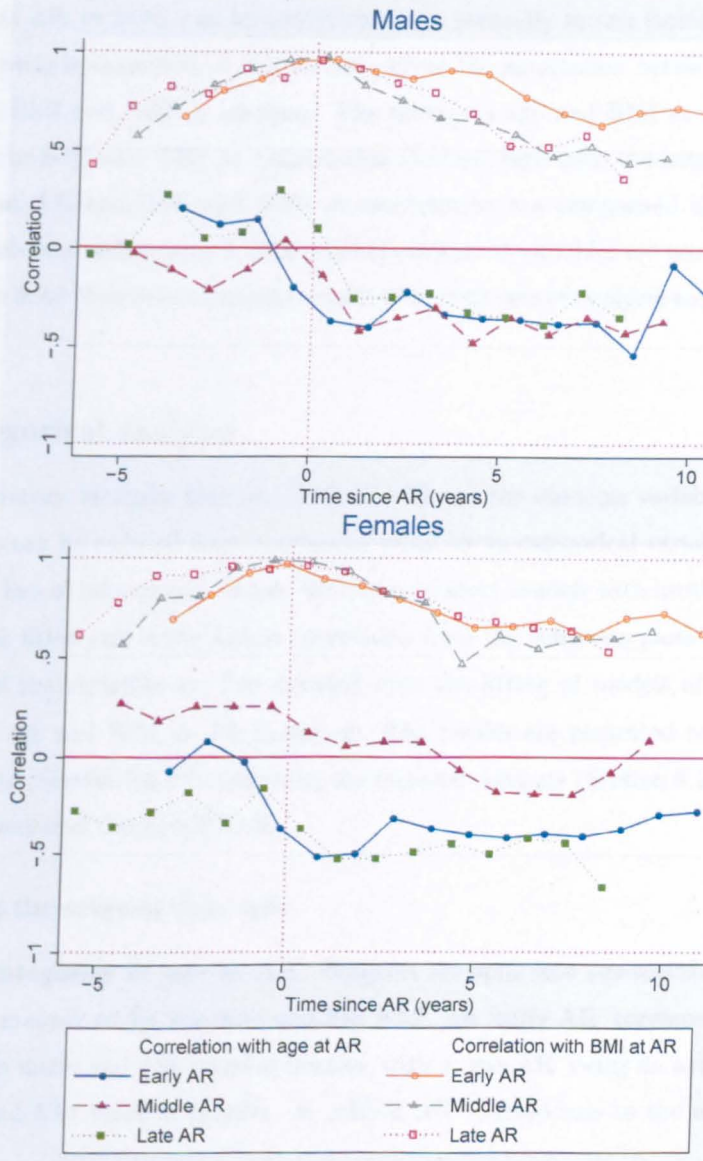

Fig. 8.18: Centred correlations between age at adiposity rebound (AR) and body mass index (BMI) through childhood and between body mass index at adiposity rebound and body mass index through childhood in the original data, by age at adiposity rebound and sex.

8.9 Are dimensions of the adiposity rebound associated with late-adolescent obesity?

The graphical exploration of the AR in Section 8.8 suggests that subjects with either an earlier AR, a higher BMI at AR, or both, are likely to have higher BMI at or before age 15 years than the other individuals in the dataset. These observations may lead to the hypothesis that an earlier AR, a higher BMI at AR, or both, can be considered more generally as risk factors for later obesity.

This hypothesis is examined in this section, where the association between the AR and both late-adolescent BMI and %BF is assessed. The estimated age and BMI at AR for each subject can be related to BMI and %BF at examination through regression modelling. In Section 8.9.1 age and BMI at AR and BMI and %BF at examination are categorised then used in logistic regression models, and in Section 8.9.2 the original continuous variables are used in linear regression models. Section 8.9.3 then draws together results from both sets of analyses to present some overall conclusions.

8.9.1 Categorical analysis

Both the explanatory variables (age and BMI at AR) and the outcome variables (BMI and %BF at examination) can be reduced from continuous variables to categorical variables. Whilst clearly this results in a loss of information, it also allows exploratory models with intuitively interpretable parameters to be fitted and is the logical progression from the subgroup plots in Section 8.8. The categorisation of the variables are first detailed then the fitting of models of BMI and %BF at examination on age and BMI at AR examined. The results are presented separately using the original data only (Section 8.9.1.1) and using the imputed datasets (Section 8.9.1.2), then the two sets of results compared (Section 8.9.1.3).

8.9.1.1 Using the original data only

Defining the categories of age at AR Subjects are split into sex-specific tertiles of age at AR in the same manner as for Fig. 8.11 and Fig. 8.17. An 'early AR' corresponds to an age less than 5.24 years in males and 4.96 years in females, with a 'late AR' being an age greater than 6.30 years in males and 5.87 years in females. A 'middle AR' corresponds to the ages between these values.

Table 8.19 summarises the distribution of BMI at examination by age at AR category for males and females separately. There is a clear trend in both males and females that as age at AR category moves from early to late both the mean and median BMI at examination are reduced. Also of note is the greater variability in BMI at examination corresponding to an earlier age at AR category, possibly due to the inclusion in the earlier AR categories of some individuals with unusually large BMI at examination values.

Age at AR	Males				Females			
category	n	Mean	Median	SD	n	Mean	Median	SD
Early	36	21.6	21.6	2.9	50	22.7	21.8	3.5
Middle	38	20.8	20.8	2.3	50	21.2	21.1	2.3
Late	37	19.4	19.4	1.3	50	19.8	19.8	1.7

Table 8.19: Distribution of body mass index (BMI) at examination in the original data, by category of age at adiposity rebound (AR) and sex.

An equivalent tabulation of %BF by category of age at AR (Table 8.20) shows a similar pattern, with earlier AR leading, on average, to greater %BF in both males and females. Again, there is greater variability in %BF for those with an earlier AR.

Age at AR	Males				Females			
category	n	Mean	Median	SD	n	Mean	Median	SD
Early	36	16.6	15.9	6.4	49	31.4	30.3	6.5
Middle	37	14.7	14.3	5.8	47	29.1	29.1	6.0
Late	36	12.2	12.4	4.1	50	26.8	27.1	5.1

Table 8.20: Distribution of percentage body fat (%BF) at examination in the original data, by category of age at adiposity rebound (AR) and sex.

Defining the categories of BMI at AR Subjects are also split by sex into tertiles of BMI at AR in the same manner as for Fig. 8.13. A 'low BMI at AR' corresponds to a BMI of less than 14.84 kg/m^2 for males and 14.56 kg/m^2 for females with a 'high BMI at AR' being a BMI greater than 15.60 kg/m^2 in males and 15.53 kg/m^2 in females. A 'medium BMI at AR' corresponds to a BMI between these values.

Table 8.21 summarises the distribution of BMI at examination by BMI at AR category for males and females separately. There is a clear trend in both males and females with BMI at AR category moving from low to high leading to both the mean and median BMI at examination being increased. Again there appears to be a corresponding trend in SD for BMI at examination with a higher BMI at AR category leading to increased SD, though this is far more marked in females.

Table 8.22 is the equivalent tabulation for %BF showing, again, both increased %BF and increased variability in %BF amongst those with a higher BMI at AR.

A cross-tabulation of the categories of age and BMI at AR, as shown in Table 8.23, illustrates the relationship between the two categorical variables, though this does seem to vary between

BMI at AR	Males				Females			
category	n	Mean	Median	SD	n	Mean	Median	SD
Low	37	19.3	18.8	2.0	50	19.3	19.3	1.5
Medium	37	20.6	19.9	2.0	50	21.1	21.2	2.0
High	37	21.9	21.6	2.6	50	23.3	22.5	3.2

Table 8.21: Distribution of body mass index (BMI) at examination in the original data, by category of body mass index at adiposity rebound (AR) and sex.

BMI at AR	Males				Females			
category	n	Mean	Median	SD	n	Mean	Median	SD
Low	36	12.8	12.6	4.2	49	26.5	27.0	4.9
Medium	36	13.9	13.7	5.8	48	29.2	29.0	6.0
High	37	16.9	16.8	6.3	49	31.6	30.7	6.5

Table 8.22: Distribution of percentage body fat (%BF) at examination in the original data, by category of body mass index (BMI) at adiposity rebound (AR) and sex.

males and females. In males, an early AR corresponds to a predominantly high BMI at AR and a middle age at AR to a low BMI at AR, with an even distribution of BMI at AR categories for a late AR. In females, an early AR also corresponds to a greater proportion of high BMI at AR, as does a middle age at AR, though a late AR is more associated with a low BMI at AR.

Defining the categories of BMI at examination The widely-used international standards for childhood overweight and obesity of Cole *et al* [60] are used to define the categories of BMI at examination. The BMI cut-off values vary with age — for example at age 17 years 'overweight' corresponds to a BMI of between 24.46 and 29.41 kg/m^2 in males and between 24.70 and 29.69 kg/m^2 in females, with 'obesity' defined as a BMI greater than 29.41 and 29.69 kg/m^2 in males and females, respectively. This results in 19 (12.0%) males and 17 (7.5%) females being classified as 'overweight' and 2 (1.3%) males and 5 (2.2%) females being classified as 'obese' at examination. As these categories are effectively adjusted for age it is not necessary to adjust for age at examination in any models with categorical BMI at examination as the outcome.

Cross-tabulation of categories of overweight at examination and age at AR, as in Table 8.24, illustrates how the distribution of subjects between overweight at examination categories differs by age at AR category. There are clear trends, with 25% of males with an early AR being overweight at examination but none of those with a late AR being so. Similarly, 16% of females with an early AR are overweight at examination with 6% obese, compared to only 2% overweight and none obese among those with a late AR. The distribution of overweight categories for those with no AR identified is also shown and is enlightening, with the distributions clearly not dissimilar to those

BMI at AR category	Age at AR category					
	Males			Females		
	Early	Middle	Late	Early	Middle	Late
Low	8	16	13	13	14	23
	22.2%	42.1%	35.1%	26.0%	28.0%	46.0%
Medium	13	12	12	15	16	19
	36.1%	31.6%	32.4%	30.0%	32.0%	38.0%
High	15	10	12	22	20	8
	41.7%	26.3%	32.4%	44.0%	40.0%	16.0%

Table 8.23: Cross-tabulation of categories of age and body mass index (BMI) at adiposity rebound (AR) in the original data, by sex. The top number in each case is the frequency and the bottom number is the corresponding column percentage.

for individuals with identified AR (perhaps with the exception of the two obese males) suggesting that these are not wholly disparate groups of subjects.

Table 8.25 examines the relationship between categories of BMI at AR and BMI at examination. It can be seen that of the males with low BMI at AR only 5% go on to be overweight at examination, whereas of those with high BMI at AR 22% do so. A similar pattern is evident in the females with nobody progressing to overweight or obesity following a low BMI at AR yet 18% being overweight and 6% obese following high BMI at AR. Again, the distributions among those subjects with no identified AR appear to be similar to those in the rest of the dataset.

Defining the categories of %BF at examination %BF at examination is also categorised using existing cut-off values, these developed by McCarthy *et al* [175]. Again, the cut-off values differ with age so that, for example, at age 17 years 'overfat' is defined as having a %BF of between 20.1 and 23.9 in males and between 30.4 and 34.4 in females. A %BF above the upper ends of these intervals is defined as 'obese' in each case. Categorisation results in 17 (10.8%) males being classified as overfat and 15 (9.6%) as obese. The corresponding figures for females are 41 (18.5%) and 40 (18.0%). For %BF at examination, unlike BMI, there are also several subjects (2 males and 5 females) with unobserved values. Whilst these prevalences, particularly among the females, do seem a little high, it should be noted that the reference data was taken from more affluent areas in an effort to obtain lower obesity rates. Additionally, the reference data were derived from data using a different body composition analysis system from the SWEDES data. Whilst there are thus potential cross-calibration issues, the use of these existing cut-offs remains more appealing than the alternative of splitting the data into arbitrary quantiles. As these categories are effectively adjusted for age it will not be necessary to adjust for age at examination in any models with

Males						
Overweight at examination category	AR identified			Total	AR not identified	Total
	Age at AR category					
	Early	Middle	Late			
Normal	27	34	37	98	40	138
	75.0%	89.5%	100.0%	88.3%	83.3%	86.8%
Overweight	9	4	0	13	6	19
	25.0%	10.5%	0.0%	11.7%	12.5%	12.0%
Obese	0	0	0	0	2	2
	0.0%	0.0%	0.0%	0.0%	4.2%	1.3%
Total	36	38	37	111	48	159

Females						
Overweight at examination category	AR identified			Total	AR not identified	Total
	Age at AR category					
	Early	Middle	Late			
Normal	39	47	49	135	70	205
	78.0%	94.0%	98.0%	90.0%	90.9%	90.3%
Overweight	8	3	1	12	5	17
	16.0%	6.0%	2.0%	8.0%	6.5%	7.5%
Obese	3	0	0	3	2	5
	6.0%	0.0%	0.0%	2.0%	2.6%	2.2%
Total	50	50	50	150	77	227

Table 8.24: Cross-tabulation of categories of overweight at examination and age at adiposity rebound (AR) in the original data, by sex. The top number in each case is the frequency and the bottom number is the corresponding column percentage.

Males						
Overweight at examination category	AR identified			Total	AR not identified	Total
	BMI at AR category					
	Low	Medium	High			
Normal	35	34	29	98	40	138
	94.6%	91.9%	78.4%	88.3%	83.3%	86.8%
Overweight	2	3	8	13	6	19
	5.4%	8.1%	21.6%	11.7%	12.5%	12.0%
Obese	0	0	0	0	2	2
	0.0%	0.0%	0.0%	0.0%	4.2%	1.3%
Total	37	37	37	111	48	159

Females						
Overweight at examination category	AR identified			Total	AR not identified	Total
	BMI at AR category					
	Low	Medium	High			
Normal	50	47	38	135	70	205
	100.0%	94.0%	76.0%	90.0%	90.9%	90.3%
Overweight	0	3	9	12	5	17
	0.0%	6.0%	18.0%	8.0%	6.5%	7.5%
Obese	0	0	3	3	2	5
	0.0%	0.0%	6.0%	2.0%	2.6%	2.2%
Total	50	50	50	150	77	227

Table 8.25: Cross-tabulation of categories of overweight at examination and age at adiposity rebound (AR) in the original data, by sex. The top number in each case is the frequency and the bottom number is the corresponding column percentage.

categorical %BF at examination as the outcome.

Cross-tabulation of categories of overfat at examination and age at AR, as in Table 8.26, illustrates how the distribution of subjects between overfat at examination categories differs by age at AR category. Similarly to the overweight at examination categories, there are greater proportions of overfat and obese subjects in the earlier AR categories: 17% of men with early AR are overfat and 11% obese compared to only 3% overfat and none obese for those with late AR. Likewise, following an early AR 16% of females go on to become overfat and 35% obese, compared to 18% overfat and 4% obese after a late AR. The inclusion of a column for subjects with no identified AR shows that the distribution between overfat categories in these individuals is not dissimilar to those with identified ARs in females. Males with no AR identified, however, show a greater prevalence of obesity than even those with an early AR.

Table 8.27 is an equivalent table to examine the relationship between categories of BMI at AR and %BF at examination. Of the males with low BMI at AR only 6% go on to be overfat with none obese at examination, whereas of those with high BMI at AR 14% become overfat and 14% obese. A similar pattern is evident in the females with 14% overweight and 4% obese following a low BMI at AR compared to 18% overweight and 37% obese following high BMI at AR. The distribution of females with no identified AR between categories of overfat is similar to the overall distribution among those with AR identified. Again, however, the prevalence of obesity amongst males with no AR identified is greater than even amongst those with high BMI at AR.

Logistic regression models Because of the scarcity of subjects within the obese category when considering overweight at examination, the overweight and obese categories are combined into one, which for simplicity will be referred to as 'overweight' and opposed to 'overweight or obese'. However, as can be inferred from Table 8.24, this leaves one age at AR category among the males (late AR) with no corresponding cases of overweight at examination. The presence of a zero cell count is problematic when fitting logistic regression models, with one solution to collapse the categories of the variable so as to eliminate it [131]. Thus, in this instance, late AR can be combined with middle AR so that the resulting category ('middle-late AR') has non-zero cases of overweight. Whilst the zero cell count for late AR only arises in males, to collapse the categories in this way amongst the males only would result in non-comparable male and females models, thus the same process is applied to the females. From Table 8.25 it can be seen that a similar issue exists for females with low BMI at AR. The solution is again the collapsing of this category into those with medium BMI at AR for both males and females to form a 'low-medium BMI at AR' category.

The overfat and obese categories are also combined to form a single 'overfat' category. Unlike for overweight there are both males and females who, following any given age or BMI at AR category, proceed to overfat at examination. This means that the problems caused by zero cell counts

Males						
Overfat at examination category	AR identified			Total	AR not identified	Total
	Age at AR category					
	Early	Middle	Late			
Normal	26	31	35	92	33	125
	72.2%	83.8%	97.2%	84.4%	68.8%	79.6%
Overfat	6	4	1	11	6	17
	16.7%	10.8%	2.8%	10.1%	12.5%	10.8%
Obese	4	2	0	6	9	15
	11.1%	5.4%	0.0%	5.5%	18.8%	9.6%
Total	36	37	36	109	48	157

Females						
Overfat at examination category	AR identified			Total	AR not identified	Total
	Age at AR category					
	Early	Middle	Late			
Normal	24	30	39	93	48	141
	49.0%	63.8%	78.0%	63.7%	63.2%	63.5%
Overfat	8	7	9	24	17	41
	16.3%	14.9%	18.0%	16.4%	22.4%	18.5%
Obese	17	10	2	29	11	40
	34.7%	21.3%	4.0%	19.9%	14.5%	18.0%
Total	49	47	50	146	76	222

Table 8.26: Cross-tabulation of categories of overfat at examination and age at adiposity rebound (AR) in the original data, by sex. The top number in each case is the frequency and the bottom number is the corresponding column percentage.

Males						
Overfat at examination category	AR identified			Total	AR not identified	Total
	BMI at AR category					
	Low	Medium	High			
Normal	34	31	27	92	33	125
	94.4%	86.1%	73.0%	84.4%	68.8%	79.6%
Overweight	2	4	5	11	6	17
	5.6%	11.1%	13.5%	10.1%	12.5%	10.8%
Obese	0	1	5	6	9	15
	0.0%	2.8%	13.5%	5.5%	18.8%	9.6%
Total	36	36	37	109	48	157

Females						
Overfat at examination category	AR identified			Total	AR not identified	Total
	BMI at AR category					
	Low	Medium	High			
Normal	40	31	22	93	48	141
	81.6%	64.6%	44.9%	63.7%	63.2%	63.5%
Overweight	7	8	9	24	17	41
	14.3%	16.7%	18.4%	16.4%	22.4%	18.5%
Obese	2	9	18	29	11	40
	4.1%	18.8%	36.7%	19.9%	14.5%	18.0%
Total	49	48	49	146	76	222

Table 8.27: Cross-tabulation of categories of overfat at examination and age at adiposity rebound (AR) in the original data, by sex. The top number in each case is the frequency and the bottom number is the corresponding column percentage.

do not occur here so there is no necessity to collapse any of the explanatory variable categories together. However, to allow for comparability between the overweight and overfat models the same combination of categories ('middle-late AR' and 'low-medium BMI at AR') is imposed in the models for overfat.

The following logistic regression models treat middle-late AR and low-medium BMI at AR as reference categories.

Table 8.28 details the estimated odds ratios (ORs) obtained when fitting the logistic regression models for overweight and overfat at examination on age and BMI at AR separately. The addition or removal of variables from the model can be tested via the likelihood ratio test (LRT) [116]. The LRT provides no evidence of effect modification of any of the relationships by sex (P=0.88 for the sex-age at AR interaction and P=0.28 for the sex-BMI at AR interaction in the models for BMI at examination, with corresponding P-values of 0.60 and 0.97 in the models for overfat), so common effect estimates for males and females are presented.

Outcome	Explanatory variable	n	OR	95% CI	P-value
Overweight at examination	**Age at AR**				
	Early vs. middle-late	261	6.35	2.66, 15.14	<0.001
	Sex				
	Female vs. male		0.81	0.36, 1.84	0.61
	BMI at AR				
	High vs. low-medium	261	6.20	2.60, 14.78	<0.001
	Sex				
	Female vs. male		0.82	0.36, 1.87	0.65
Overfat at examination	**Age at AR**				
	Early vs. middle-late	255	2.86	1.58, 5.17	0.001
	Sex				
	Female vs. male		3.25	1.72, 6.13	<0.001
	BMI at AR				
	High vs. low-medium	255	3.38	1.86, 6.14	<0.001
	Sex				
	Female vs. male		3.36	1.77, 6.38	<0.001

Table 8.28: Estimated odds ratios (OR), 95% confidence intervals (CI) and P-values for the logistic regression models for overweight and overfat at examination fitted separately on age and body mass index (BMI) at adiposity rebound (AR) in the original data.

An early AR is estimated to lead to over 6 times the odds of being overweight at examination and nearly 3 times of the odds of being overfat when compared to a middle-late AR. A high, as

opposed to low-medium, BMI at AR is associated with an estimated 6-fold increase in the odds of overweight at examination and over 3 times the odds of overfat. All four of these relationships are highly statistically significant.

In the fitted models for overweight at examination there is no real evidence of either males or females having greater odds of overweight for a given age or BMI at AR. In the overfat models, however, females have 3 times of the odds of overfat when controlling for either age or BMI at AR. This is probably largely due to the much higher proportion of females who are classified as overfat.

Table 8.29 details the estimated ORs from the logistic regression models for overweight and overfat at examination fitted jointly on age and BMI at AR.

Outcome	Explanatory variable	n	OR	95% CI	P-value
Overweight at examination	**Age at AR**				
	Early vs. middle-late		1.56	0.36, 6.81	0.55
	BMI at AR				
	High vs. low-medium	261	1.53	0.35, 6.64	0.57
	Age & BMI at AR				
	Interaction		8.67	1.19, 63.4	0.03
	Sex				
	Female vs. male		0.75	0.31, 1.83	0.34
Overfat at examination	**Age at AR**				
	Early vs. middle-late		2.54	1.38, 4.68	0.003
	BMI at AR	255			
	High vs. low-medium		3.06	1.66, 5.64	<0.001
	Sex				
	Female vs. male		3.48	1.81, 6.71	<0.001

Table 8.29: Estimated odds ratios (OR), 95% confidence intervals (CI) and P-values for the logistic regression models for overweight and overfat at examination fitted jointly on age and body mass index (BMI) at adiposity rebound (AR) in the original data.

In the fitted model for overweight at examination there is evidence of an interaction between age and BMI at AR (P=0.03) so this is included in the model. There is, however, no evidence of any sex-explanatory variable interactions (P=0.85 for sex-age at AR, P=0.44 for sex-BMI at AR and P=0.92 for sex-age at AR-BMI at AR — each interaction tested separately) so these parameters are not included.

The fitted model for BMI at examination can be interpreted as follows:

- Among subjects with a low-medium BMI at AR the estimated OR associated with an early as opposed to middle-late AR is 1.56.

- Among subjects with a middle-late AR the estimated OR associated with a high as opposed to low-medium BMI at AR is 1.53.
- Among subjects with a high BMI at AR the estimated OR associated with an early as opposed to middle-late AR is 13.53.
- Among subjects with an early AR the estimated OR associated with a high as opposed to low-medium BMI at AR is 13.27.

Clearly the effect of either explanatory variable is highly dependent of the value taken by the other explanatory variable, evidence of significant mutual effect modification. Also of note is the similarity in the magnitudes of the effects of age and BMI at AR.

In the fitted model for overweight at examination there is little evidence of a difference in the estimated odds of overweight at examination between males and females.

Evidence for an age at AR-BMI at AR interaction in the fitted logistic regression model for overfat at examination is limited (P=0.13) so in the interests of parsimony the parameter is excluded from the model. Again, there is also no evidence of effect modification by sex of either of the explanatory variables (P=0.55 for the sex-age at AR interaction and P=0.95 for the sex-BMI at AR interaction). It can be seen from the fitted model that, for a given BMI at AR, an early AR is estimated to lead to 2.5 times the odds of overfat at examination in both males and females. For a given age at AR a high BMI at AR is associated with a 3-fold increase in the odds of overfat. Finally, when controlling for both age and BMI at AR females are expected to have 3.5 times the odds of overfat when compared to males.

A comparison of the estimated ORs in Table 8.28 with their equivalent ORs in Table 8.29 can enable a crude assessment of the confounding of the relationships by the dimension of AR location which is present in the latter model but not the former. For example, if a relationship is found between one dimension of AR and an outcome at examination in the model containing only that dimension of AR as an explanatory variable, but in the model containing both dimensions of AR the magnitude of this relationship is diminished, then it could be suggested that the second dimension of AR is confounding the relationship between the first dimension of AR and the outcome at examination.

Although the estimated associations in the fitted models for overfat at examination do show some degree of attenuation (2.86 vs. 2.54 for age at AR and 3.38 vs. 3.06 for BMI at AR) the differences are small, suggesting that there is little confounding. That both explanatory variable parameters remain highly statistically significant in the model fitted jointly on them is evidence of the independent effects on overfat at examination of both age and BMI at AR.

Comparison of the fitted models for overweight at examination is somewhat more difficult due to the introduction of the interaction term in the latter model. However, a comparison of the estimated crude OR for an early AR of 6.35 in Table 8.28 with the strata-specific estimated ORs of 1.56 for a low-medium BMI at AR and 13.53 for a high BMI at AR illustrates the extent of the

interaction between the explanatory variables. A comparison of the estimated crude OR of 6.20 for a high BMI at AR with estimated ORs of 1.53 for subjects with a middle-late AR and 13.27 for those with an early AR shows a similar degree of effect modification.

8.9.1.2 Using the imputed datasets

Defining the categories of age at AR Subjects in the imputed datasets are split by sex into approximate tertiles of age at AR using the same cut-points as are used for subjects in the original dataset in Section 8.9.1.1. This categorisation results in, across the 100 imputed datasets, a mean of 35.2% of males being classified as early AR, 34.0% as middle AR and 30.8% as late AR. The corresponding figures for females are 33.8% early AR, 34.7% middle AR and 31.5% late AR. This signifies, particularly among the males, a shift towards greater a proportion of individuals exhibiting an early AR than in the original data.

Table 8.30 summarises the distribution of BMI at examination by age at AR category for males and females separately across the imputations. These values are obtained using Rubin's rules as described in Sections 5.2.4 and 8.3.2, as is the case for all the results in this section. The summary statistics are calculated as described in Sections 5.2.4 and 8.3.2. Due to the slightly skewed nature of the BMI at examination distribution the mean median is the preferred measure of the distributional 'average'. In both sexes there is a clear trend for mean median BMI at examination to reduce as age at AR category moves from early to late, though with much greater variability associated with earlier AR.

	Males				Females			
Age at AR category	Mean n	Overall mean	Mean median	SD	Mean n	Overall mean	Mean median	SD
Early	49.6	22.3	21.9	3.2	67.7	22.6	22.0	3.5
Middle	47.9	20.8	20.5	2.4	69.5	21.4	21.3	2.2
Late	43.4	19.5	19.4	1.5	63.1	19.9	19.8	1.9

Table 8.30: Distribution of body mass index (BMI) at examination in the 100 imputed datasets, by category of age at adiposity rebound (AR) and sex.

An equivalent tabulation of %BF by category of age at AR (see Table 8.31) shows a similar pattern, with earlier AR leading, on average, to greater %BF in both males and females. Again, there is greater variability in %BF for those with an earlier AR.

Defining the categories of BMI at AR As with age at AR, the same cut-offs as previously defined by the original data in Section 8.9.1.1 are used to categorise the subjects in the imputed datasets. This results in a mean of 33.3% of males classified as low BMI at AR, 32.7% as medium

	Males				Females			
Age at AR category	Mean n	Overall mean	Mean median	SD	Mean n	Overall mean	Mean median	SD
Early	49.6	17.9	16.6	7.7	67.7	31.3	30.4	6.6
Middle	47.9	15.2	14.3	6.2	69.5	29.2	28.8	6.3
Late	43.4	12.8	13.0	4.5	63.1	26.7	27.0	5.0

Table 8.31: Distribution of percentage body fat (%BF) at examination in the 100 imputed datasets, by category of age at adiposity rebound (AR) and sex.

and 34.0% as high. The corresponding figures for females are 33.8%, 33.2% and 33.0%.

Table 8.32 summarises the distribution of BMI at examination by BMI at AR category for males and females separately. In both sexes a low BMI at AR is seen to correspond to a lower BMI at examination and a high BMI at AR to a higher BMI at examination. There is also a pattern of increasing variability with increasing BMI at AR category, most noticeably among the females.

	Males				Females			
BMI at AR category	Mean n	Overall mean	Mean median	SD	Mean n	Overall mean	Mean median	SD
Low	46.9	19.6	19.1	2.1	67.7	19.5	19.5	1.6
Medium	46.0	20.8	20.3	2.2	66.5	21.1	21.1	2.1
High	47.9	22.3	21.8	3.1	66.1	23.4	22.7	3.2

Table 8.32: Distribution of body mass index (BMI) at examination in the 100 imputed datasets, by category of body mass index at adiposity rebound (AR) and sex.

Table 8.33 is the equivalent tabulation for %BF showing similar trends for both increased %BF and increased variability in %BF amongst those with a higher BMI at AR.

	Males				Females			
BMI at AR category	Mean n	Overall mean	Mean median	SD	Mean n	Overall mean	Mean median	SD
Low	46.9	13.7	13.1	5.4	67.7	26.7	27.1	4.8
Medium	46.0	15.0	13.9	6.6	66.5	28.9	28.5	6.2
High	47.9	17.5	16.9	7.3	66.1	31.7	30.9	6.7

Table 8.33: Distribution of percentage body fat (%BF) at examination in the 100 imputed datasets, by category of body mass index (BMI) at adiposity rebound (AR) and sex.

Defining the categories of BMI at examination BMI at examination is again categorised using the international standards of Cole *et al* [60], as for the original data in Section 8.9.1.1. As BMI at examination is completely observed, and thus has no values imputed as part of the MI procedure, the prevalences of overweight and obese are the same in each imputed dataset and the same as for the original data.

Defining the categories of %BF at examination As in the analysis of the original data in Section 8.9.1.1, %BF at examination is categorised according to the existing cut-offs of McCarthy *et al* [175]. In the original data there are a small number of unobserved %BF values so, unlike BMI at examination, overfat is affected by the MI procedure. As a result of different values being imputed into different datasets it is possible for the prevalence of overfat and obese to vary between the 100 imputed datasets. Between 17 and 18 males (10.7–11.3%) are classified as overfat and between 15 and 16 (9.4–10.1%) as obese in each imputation dataset. The corresponding figures for females are 41–44 (18.1–19.4%) and 40–44 (17.6–19.4%).

Logistic regression models In the analysis of the imputed datasets, assessing the extent of interactions involving either or both dimensions of the AR in the analysis models is not as straightforward as when dealing with the original data. As the AR locations are derived from what are often imputed data (i.e. estimation of AR location occurs *after* imputation) it is impossible to include in the imputation model any interactions between either dimension of the AR and any other variable (or, indeed, between the two dimensions of the AR). Generally, for a variable imputed under a no-interactions imputation model, if interactions are present then the MI estimates of them will be biased towards the null. Thus under normal circumstances the imputation model should reasonably preserve any features of the dataset which will be the subject of future analyses [123]. In this instance, however, it is impossible to do so, meaning that the potential biasing towards the null of the estimated interaction terms must instead simply be acknowledged. This is likely to lead to significance tests for the inclusion of such interaction terms underestimating their importance. As it is therefore impossible to accurately assess the evidence for the inclusion of interaction terms involving the AR when considering the imputed datasets, one possible approach is to include interaction terms in the analysis model if and only if they are deemed necessary when analysing the original data only (i.e. if and only if they are included in the analysis models in Section 8.9.1.1).

Table 8.34 details the estimated ORs from the logistic regression models for overweight and overfat at examination fitted separately on age and BMI at AR. Similarly to the analysis using the original data only, there is little evidence of sex-explanatory variable interactions in either model (P=0.95 for the sex-age at AR and P=0.29 for the sex-BMI at AR interaction in the model for overweight at examination, with equivalent P-values of 0.41 and 0.88 in the model for overfat at examination) so these parameters are not included in the models.

An early AR is estimated to lead to nearly 6 times the odds of overweight at examination and an almost 3-fold increase the in odds of overfat when compared to a middle-late AR in both males

Outcome	Explanatory variable	n per imputation	OR	95% CI	P-value
Overweight at examination	**Age at AR**				
	Early vs. middle-late	331–351	5.88	2.52, 13.71	<0.001
	Sex				
	Female vs. male		0.65	0.32, 1.34	0.25
	BMI at AR				
	High vs. low-medium	331–351	5.48	2.41, 12.46	<0.001
	Sex				
	Female vs. male		0.65	0.32, 1.33	0.24
Overfat at examination	**Age at AR**				
	Early vs. middle-late	331–351	2.88	1.63, 5.08	<0.001
	Sex				
	Female vs. male		2.42	1.42, 4.13	0.001
	BMI at AR				
	High vs. low-medium	331–351	2.98	1.75, 5.10	<0.001
	Sex				
	Female vs. male		2.42	1.42, 4.12	0.001

Table 8.34: Estimated odds ratios (OR), 95% confidence intervals (CI) and P-values for the logistic regression models for overweight at examination fitted separately on age and body mass index (BMI) at adiposity rebound (AR) in the 100 imputed datasets.

and females. From the fitted models a high, as opposed to low-medium, BMI at examination can be expected to increase the odds of overweight by 5.5 times and treble the odds of overfat. All four of these relationships are highly statistically significant (P<0.001). There is little evidence of sex affecting the odds of overweight at examination for a given age or BMI at AR, but the odds of overfat in females are estimated to be about 2.5 times those in males.

The estimated ORs from the logistic regression models for overweight and overfat at examination fitted jointly on age and BMI at AR are presented in Table 8.35. In neither model is there strong evidence of an age at AR-BMI at AR interaction to justify the inclusion of an interaction parameter (P=0.20 in the model for overweight at examination and P=0.23 in the model for overfat) although, as previously discussed, these P-values are likely to be biased away from significance. As there is reasonably strong evidence (P=0.03) of an age at AR-BMI at AR interaction in the model with BMI at examination as outcome when analysing the original data only, this interaction term is included here. There is also little evidence of any sex-explanatory variable interactions (P=0.87 for the sex-age at AR interaction and and P=0.30 for the sex-BMI at AR interaction in the model for overweight at examination, with equivalent P-values of 0.40 and 0.81 in the overfat model). These interactions are not included in the analysis model as they are not deemed necessary in the equivalent original data model.

Outcome	Explanatory variable	n per imputation	OR	95% CI	P-value
Overweight at examination	**Age at AR**				
	Early vs. middle-late	331–351	2.49	0.64, 9.78	0.19
	BMI at AR				
	High vs. low-medium		2.22	0.55, 8.97	0.26
	Age & BMI at AR				
	Interaction		3.37	0.52, 21.7	0.20
	Sex				
	Female vs. male		0.62	0.29, 1.35	0.23
Overfat at examination	**Age at AR**				
	Early vs. middle-late	331–351	2.56	1.43, 4.60	0.002
	BMI at AR				
	High vs. low-medium		2.67	1.54, 4.62	<0.001
	Sex				
	Female vs. male		2.55	1.47, 4.42	0.001

Table 8.35: Estimated odds ratios (OR), 95% confidence intervals (CI) and P-values from the logistic regression models for overweight at examination fitted jointly on age and body mass index (BMI) at adiposity rebound (AR) in the 100 imputed datasets.

The fitted model for BMI at examination can be interpreted as follows:

- Among subjects with a low-medium BMI at AR the estimated OR associated with an early as opposed to middle-late AR is 2.49.
- Among subjects with a middle-late AR the estimated OR associated with a high as opposed to low-medium BMI at AR is 2.22.
- Among subjects with a high BMI at AR the estimated OR associated with an early as opposed to middle-late AR is 8.39.
- Among subjects with an early AR the estimated OR associated with a high as opposed to low-medium BMI at AR is 7.48.

When controlling for both age and BMI at AR there is little evidence for sex altering the odds of overweight at examination.

From the fitted model for %BF at examination it can be seen that for a given BMI at AR an early AR is estimated to be associated with 2.5 times the odds of overfat when compared to a middle-late AR. Similarly, when controlling for age at AR a high rather than low-medium BMI at AR is estimated to increase the odds of overfat by about 2.5 times. Both of these relationships are highly statistically significant ($P \leq 0.002$). When controlling for both age and BMI at AR females are estimated to have 2.5 times the odds of overfat when compared to males.

A crude assessment of the confounding of the relationships in Table 8.34 by the dimension of AR location which is not present in each model is facilitated by a comparison of the estimated ORs in Table 8.34 with their equivalent ORs in the models of Table 8.35.

In the models for overfat at examination the ORs for both age at AR (2.88 vs. 2.56) and BMI at AR (2.98 vs. 2.67) are attenuated a little, providing evidence that each association is somewhat confounded by the other dimension. However, as both ORs remain highly statistically significant in the models fitted jointly on the explanatory variables it is clear that both explanatory variables are independently associated with being overfat.

Direct comparison of the fitted models for overweight at examination is not possible as the model fitted jointly on age and BMI at AR also includes an age at AR-BMI at AR interaction term.

8.9.1.3 Comparison of results using the original data only and results using the imputed datasets

A comparison of the fitted logistic models using the original data (Tables 8.28 and 8.29) and the imputed datasets (Tables 8.34 and 8.35) allows differences between the two analytical approaches to be examined.

From the models fitted separately on age and BMI at AR (Tables 8.28 and 8.34) it can be seen that the estimated ORs for overweight associated with both an early AR (6.35 vs. 5.88) and a high

BMI at AR (6.20 vs. 5.48) are reduced somewhat under the MI approach. Whilst the estimated OR for overfat associated with a high BMI at AR (3.38 vs. 2.98) is also reduced to some extent, that for an early AR (2.86 vs. 2.88) remains stable.

Comparing the model for overweight at examination fitted jointly on age and BMI at AR in Tables 8.29 and 8.35 is complicated slightly by the inclusion of the interaction term in both models. Estimated coefficients corresponding to age and BMI at AR are both seen to increase markedly when considering the imputed datasets (1.56 vs. 2.49 and 1.53 vs. 2.22, respectively), whilst the estimated interaction is attenuated dramatically (8.67 vs. 3.37). However, as has been detailed previously, the inability to include interaction terms involving the AR in the imputation model is likely to lead to an attenuation in the estimated interaction when considering the imputed datasets. This, in turn, is likely to lead to increased estimated age and BMI at AR coefficients, which may well explain the observed differences.

The model for overfat shows a reduced association with BMI at AR for a given age at AR (3.06 vs. 2.67), though little change in the estimated OR for age at AR when controlling for BMI at AR (3.54 vs. 3.56).

The reasons behind the reduced ORs under the MI approach are discussed in Section 8.11.

Whilst the estimated ORs in the fitted models may be reduced under the MI approach it is important to recognise that their associated CIs remain relatively wide and largely overlapping with those estimated for the corresponding ORs in the models using the original data only. Also, as all the estimated ORs remain highly statistically significant under the MI approach the evidence of the associations is little diminished by the use of MI.

8.9.2 Continuous analysis

Use of both the explanatory variables (age and BMI at AR) and the outcome variables (BMI and %BF at examination) as continuous as opposed to categorical variables retains the maximum amount of information. Multiple linear regression provides a framework for assessing the association between the two dimensions of the AR and later adiposity. Use of age- and sex-adjusted categoristaions of the measures of late-adolescent adiposity in Section 8.9.1 effectively controlled for the differing age at examination. When using the continuous variables, however, this controlling must be made more explicit by inclusion of age at examination in the regression models.

The results are presented separately using the original data only (Section 8.9.2.1) and using the imputed datasets (Section 8.9.2.2), then the two sets of results compared (Section 8.9.2.3).

8.9.2.1 Using the original data only

Linear regression models of BMI and %BF at examination on age and BMI at AR are fitted using the original data. During adolescence both of the outcome variables are age-dependent and are not measured at the same age in every subject, thus age at examination is included in every regression model to adjust for any potential confounding. Both age and BMI at AR are centred about their mean value to aid with model interpretation.

Table 8.36 details the linear regression models of BMI and %BF at examination fitted separately on age and BMI at AR. As there is little evidence of any interaction between sex and each explanatory variable (P=0.38 for the sex-age at AR interaction and P=0.43 for the sex-BMI at AR interaction in the models for BMI at examination, with corresponding P-values of 0.79 and 0.58 in the models for %BF at examination), combined-sex models with no interaction parameters are presented.

Outcome	Explanatory variable	n	Coefficient	95% CI	P-value
BMI at exam. (kg/m^2)	Age at AR (years)	261	−0.97	−1.21, −0.73	<0.001
	Female vs. male		0.29	−0.31, 0.89	0.34
	BMI at AR (kg/m^2)	261	1.52	1.30, 1.74	<0.001
	Female vs. male		0.85	0.34, 1.36	0.001
%BF at examination	Age at AR (years)	255	−1.63	−2.19, −1.07	<0.001
	Female vs. male		13.93	12.50, 15.36	<0.001
	BMI at AR (kg/m^2)	255	1.96	1.37, 2.55	<0.001
	Female vs. male		14.79	13.40, 16.18	<0.001

Table 8.36: Estimated coefficients, 95% confidence intervals (CI) and P-values for the linear regression models of body mass index (BMI) and percentage body fat (%BF) at examination fitted separately on age and body mass index at adiposity rebound (AR) using the original data. Models are adjusted for age at examination.

There is very strong evidence that both age and BMI at AR are associated with both BMI and %BF at examination. A one year delay in AR is estimated to lead to, on average, a 0.97 kg/m^2 decrease in BMI and a 1.63% decrease in %BF at examination, whilst a 1 kg/m^2 increase in BMI at AR is estimated to lead to a 1.52 kg/m^2 increase in BMI and a 1.96% increase in %BF. For a given BMI at AR females are expected to have a greater BMI and much greater %BF at examination than males. For a given age at AR females are expected to have greater %BF at examination, though there is no evidence of the same being true for BMI. This is perhaps explained by the distribution of age at AR being more sex-dependent than that of BMI at AR (see Table 8.9) meaning that the effect of sex acts via the age at AR parameter.

The results in Table 8.36 must be viewed with caution, however, due to the high correlation between age and BMI at AR, which has already been illustrated. This association means that, for example, the observed relationship between age at AR and BMI at examination could be wholly, or at least partially, explained by confounding due to BMI at AR.

Table 8.37 details the linear regression models for BMI and %BF at examination fitted *jointly* on age and BMI at AR. Again there is little evidence of interaction between sex and any of the other explanatory variables (P=0.64 with age at AR, P=0.66 with BMI at AR and P=0.81 with the age at AR-BMI at AR interaction in the model for BMI at examination, with corresponding P-values of 0.94, 0.47 and 0.46 in the %BF model), thus combined-sex models with no sex interactions are

presented. However, there is strong evidence of an interaction between age and BMI at AR in each model, making interpretation somewhat less trivial.

Outcome	Explanatory variable	n	Coefficient	95% CI	P-value
BMI at exam. (kg/m^2)	Age at AR (years)	261	−0.59	−0.78, −0.39	<0.001
	BMI at AR (kg/m^2)		1.25	1.04, 1.47	<0.001
	Interaction between age and BMI at AR		−0.27	−0.43, −0.12	0.001
	Female vs. male		0.52	0.05, 1.00	0.03
%BF at examination	Age at AR (years)	255	−1.16	−1.71, −0.62	<0.001
	BMI at AR (kg/m^2)		1.35	0.75, 1.95	<0.001
	Interaction between age and BMI at AR		−0.81	−1.24, −0.38	<0.001
	Female vs. male		14.05	12.71, 15.39	<0.001

Table 8.37: Estimated coefficients, 95% confidence intervals (CI) and P-values for the linear regression models of body mass index (BMI) and percentage body fat (%BF) at examination fitted jointly on age and body mass index at adiposity rebound (AR) using the original data. Models are adjusted for age at examination.

To aid interpretation it is beneficial to examine the fitted models more explicitly. For example, the model for BMI at examination is

$$\begin{aligned} \text{BMI}_{\text{exam}} = &- 0.59\ (\text{age}_{\text{AR}} - \text{mean}(\text{age}_{\text{AR}})) + 1.25\ (\text{BMI}_{\text{AR}} - \text{mean}(\text{BMI}_{\text{AR}})) \\ &- 0.27\ (\text{age}_{\text{AR}} - \text{mean}(\text{age}_{\text{AR}}))\ (\text{BMI}_{\text{AR}} - \text{mean}(\text{BMI}_{\text{AR}})) + 0.52\ \text{sex} \\ &+ \text{constant} \end{aligned} \tag{8.1}$$

where BMI_{exam} is predicted BMI at examination, age_{AR} and BMI_{AR} are age and BMI at AR, $\text{mean}(\text{age}_{\text{AR}})$ and $\text{mean}(\text{BMI}_{\text{AR}})$ are the mean age and BMI at AR across all subjects and sex is an indicator variable taking value 1 when female and 0 otherwise. It is possible to rewrite (8.1) in two ways to show more explicitly how changing each explanatory variable affects the outcome:

$$\begin{aligned} \text{BMI}_{\text{exam}} = &- 0.59\ (\text{age}_{\text{AR}} - \text{mean}(\text{age}_{\text{AR}})) \\ &+ (1.25 - 0.27\ (\text{age}_{\text{AR}} - \text{mean}(\text{age}_{\text{AR}})))\ (\text{BMI}_{\text{AR}} - \text{mean}(\text{BMI}_{\text{AR}})) \\ &+ 0.52\ \text{sex} + \text{constant} \end{aligned} \tag{8.2}$$

and

$$\begin{aligned} \text{BMI}_{\text{exam}} = &\ 1.25\ (\text{BMI}_{\text{AR}} - \text{mean}(\text{BMI}_{\text{AR}})) \\ &+ (-0.59 - 0.27\ (\text{BMI}_{\text{AR}} - \text{mean}(\text{BMI}_{\text{AR}})))\ (\text{age}_{\text{AR}} - \text{mean}(\text{age}_{\text{AR}})) \\ &+ 0.52\ \text{sex} + \text{constant} \end{aligned} \tag{8.3}$$

From (8.2) it can be seen that for a given age at AR a 1 kg/m^2 increase in BMI at AR is estimated to increase BMI at examination by $1.25 - 0.27\ (\text{age}_{\text{AR}} - \text{mean}(\text{age}_{\text{AR}}))$ kg/m^2. Thus

for an earlier AR the estimated increase in BMI at examination associated with an increase in BMI at AR is greater than for a later AR. Table 8.38 shows the estimated increase in BMI at examination for a 1 kg/m^2 increase in BMI at AR at different ages at AR corresponding to the range of observed values. The estimated increase is 3 times as great at age 3 years as it is at age 8 years and, whilst the increase is highly statistically significant (P<0.001) at younger ages, at age 8 years the evidence for an increase is somewhat lessened.

	Increase in BMI at examination (kg/m^2)		
Age at AR (years)	Estimate	95% CI	P-value
3	1.94	1.54, 2.34	<0.001
4	1.66	1.38, 1.95	<0.001
5	1.39	1.18, 1.60	<0.001
6	1.12	0.88, 1.36	<0.001
7	0.85	0.50, 1.19	<0.001
8	0.58	0.10, 1.05	0.02

Table 8.38: Estimates, 95% confidence intervals (CI) and P-values for the increase in body mass index (BMI) at examination (in kg/m^2) for a 1 kg/m^2 increase in body mass index at adiposity rebound (AR) at different age at AR levels using the original data.

Similarly, (8.3) shows that for a given BMI at AR a 1 year delay in AR is estimated to increase BMI at examination by $-0.59 - 0.27\ (\text{BMI}_{\text{AR}} - \text{mean}(\text{BMI}_{\text{AR}}))$ kg/m^2 (or equivalently to decrease it by $0.59 + 0.27\ (\text{BMI}_{\text{AR}} - \text{mean}(\text{BMI}_{\text{AR}}))$ kg/m^2). This means that for a greater BMI at AR the estimated decrease in BMI at examination associated with a later AR is greater than for a lower BMI at AR. Table 8.39 shows the estimated decrease in BMI at examination for a 1 year delay in AR at different BMI at AR levels. The estimated decrease is negligible at the lower end of the observed BMI at AR range but is almost 2 kg/m^2 at the upper end.

Rewriting the fitted model for %BF in Table 8.37 in the same way results in the values presented in Tables 8.40 and 8.41. It can be seen from Table 8.40 that, whilst a 1 kg/m^2 increase in BMI at AR is estimated to increase %BF at examination by over 3% when AR occurs at 3 years, if AR occurs later then there is an estimated *decrease* in %BF, albeit with a 95% CI which includes 0.

Table 8.41 illustrates a similarly interesting pattern, with a 1 year delay in AR associated with an estimated 5% decrease in %BF when corresponding to a BMI at AR of 20 kg/m^2, but associated with a slight *increase* in %BF when corresponding to a BMI at AR towards the lower end of the observed range. Again, however, there is little evidence that this estimate is truly less than 0.

Whilst these estimated anomalous results are perhaps plausible they also seem somewhat unlikely, and the associated levels of uncertainty surrounding them must be considered. It should also be borne in mind that the amount of data available for the fitting of these models are relatively small and thus the models obtained could potentially be greatly altered by one or two outlying values.

	Decrease in BMI at examination (kg/m^2)		
BMI at AR (kg/m^2)	Estimate	95% CI	P-value
13	0.00	−0.39, 0.39	0.99
14	0.27	0.00, 0.54	0.05
15	0.54	0.35, 0.74	<0.001
16	0.81	0.58, 1.04	<0.001
17	1.09	0.75, 1.42	<0.001
18	1.36	0.89, 1.83	<0.001
19	1.63	1.02, 2.25	<0.001
20	1.90	1.14, 2.66	<0.001

Table 8.39: Estimates, 95% confidence intervals (CI) and P-values for the decrease in body mass index (BMI) at examination (in kg/m^2) for a 1 year delay in adiposity rebound (AR) at different body mass index at adiposity rebound levels using the original data.

	Increase in %BF at examination		
Age at AR (years)	Estimate	95% CI	P-value
3	3.39	2.26, 4.52	<0.001
4	2.58	1.79, 3.37	<0.001
5	1.77	1.17, 2.36	<0.001
6	0.95	0.28, 1.63	0.01
7	0.14	−0.82, 1.01	0.78
8	−0.67	−2.01, 0.66	0.32

Table 8.40: Estimates, 95% confidence intervals (CI) and P-values for the increase in percentage body fat (%BF) at examination for a 1 kg/m^2 increase in body mass index (BMI) at adiposity rebound (AR) at different age at adiposity rebound levels using the original data.

	Decrease in %BF at examination		
BMI at AR (kg/m^2)	Estimate	95% CI	P-value
13	−0.61	−1.70, 0.49	0.28
14	0.21	−0.55, 0.96	0.59
15	1.02	0.47, 1.58	<0.001
16	1.83	1.19, 2.48	<0.001
17	2.65	1.70, 3.60	<0.001
18	3.46	2.13, 4.79	<0.001
19	4.27	2.55, 6.00	<0.001
20	5.09	2.95, 7.23	<0.001

Table 8.41: Estimates, 95% confidence intervals (CI) and P-values for the decrease in percentage body fat (%BF) at examination for a 1 year delay in adiposity rebound (AR) at different body mass index (BMI) at adiposity rebound levels using the original data.

The fitted models in Table 8.37 also estimate, for a given age and BMI at AR, much greater %BF at examination in females than males, though evidence of the same being true for BMI at exam is more limited. Again this is perhaps due to the effect of sex acting via the age at AR parameter.

A comparison between the models fitted separately on age and BMI at AR (Table 8.36) and those fitting jointly on age and BMI at AR (Table 8.37) is complicated by the interaction seen between the two explanatory variables. This means that the attenuation of an estimated coefficient between two comparable models cannot necessarily be ascribed to confounding by the other dimension of the AR. It may be the case that some of the association seen in the simpler model is merely acting via the interaction instead. However, in both models in Table 8.37 there remains strong evidence of relationships between each dimension of AR and the outcome conditional on both the other dimension of the AR and the interaction between the two dimensions of AR. This suggests that both age and BMI at AR are associated with both BMI and %BF at examination independently of each other and their interaction.

8.9.2.2 Using the imputed datasets

Linear regression models of BMI and %BF at examination on age and BMI at AR are fitted using the 100 imputed datasets. To maintain comparability with the models using the original data only, data from the imputed datasets are centred using the same values (the mean of the variable across all subjects in the original data). Age at examination is again included in each model to adjust for any potential confounding due to the relationship between age at examination and the outcome variables.

As with the logistic regression models in Section 8.9.1.2 there is likely to be a lack of power

when testing for the inclusion of any interaction terms which include either or both dimensions of the AR due to the impossibility of including interaction terms in the imputation model. Thus these interaction terms will again be included if and only if doing so was deemed necessary when considering the original data only in Section 8.9.2.1.

Table 8.42 details the linear regression models of BMI and %BF at examination on age and BMI at AR separately. Similarly to the analysis of the original data only there is no evidence of effect modification of these relationships by sex (P=0.97 for BMI at examination on age at AR and P=0.86 for BMI at examination on BMI at AR, with corresponding P-values of 0.59 and 0.87 for the %BF model models), so models are presented for males and females combined with no sex interactions.

Outcome	Explanatory variable	n per imputation	Coeff.	95% CI	P-value
BMI at exam. (kg/m^2)	Age at AR (years)	331–351	−1.00	−1.24, −0.75	<0.001
	Female vs. male		−0.09	−0.48, 0.67	0.74
	BMI at AR (kg/m^2)	331–351	1.43	1.19, 1.67	<0.001
	Female vs. male		0.64	0.14, 1.15	0.01
%BF at examination	Age at AR (years)	331–351	−1.67	−2.28, −1.06	<0.001
	Female vs. male		13.13	11.73, 14.53	<0.001
	BMI at AR (kg/m^2)	331–351	1.82	1.21, 2.43	<0.001
	Female vs. male		13.96	12.59, 15.34	<0.001

Table 8.42: Estimated coefficients (coeff.), 95% confidence intervals (CI) and P-values for the linear regression models of body mass index (BMI) and percentage body fat (%BF) at examination fitted separately on age and body mass index at adiposity rebound (AR) using the 100 imputed datasets. Models are adjusted for age at examination.

All four fitted models show highly significant relationships between the explanatory variable and the outcome, with age at AR inversely and BMI at AR directly related to both BMI and %BF at examination. A 1 year delay in AR is estimated to decrease BMI at examination by 1.00 kg/m^2 and %BF at examination by 1.67% for both males and females. A 1 kg/m^2 increase in BMI at examination leads to an expected increase of 1.43 kg/m^2 and 1.82% at examination. For a given age or BMI at AR females are expected to have a much greater %BF at examination. For a given BMI at examination there is some evidence that females have a greater BMI at examination whilst there is no evidence that females have greater BMI at examination conditional on age at AR.

Table 8.43 details the linear regression models of BMI and %BF at examination fitted *jointly* on age and BMI at AR. There is no evidence of sex-explanatory variable interactions either when considering the imputed datasets (P=0.58 for the sex-age at AR interaction, P=0.71 for the sex-BMI at AR interaction and P=0.81 for the sex-age at AR-BMI at AR interaction in the model with BMI at examination as outcome, with corresponding P-values of 0.58, 0.93 and 0.90 in the %BF at examination model) or the original data only, so combined-sex models with no sex interactions

are presented.

Outcome	Explanatory variable	n per imputation	Coeff.	95% CI		P-value
BMI at exam. (kg/m^2)	Age at AR (years)	331–351	−0.62	−0.85,	−0.40	<0.001
	BMI at AR (kg/m^2)		1.17	0.93,	1.40	<0.001
	Interaction between age and BMI at AR		−0.15	−0.35,	0.05	0.15
	Female vs. male		0.36	−0.12,	0.84	0.14
%BF at examination	Age at AR (years)	331–351	−1.22	−1.85,	−0.58	<0.001
	BMI at AR (kg/m^2)		1.23	0.59,	1.86	<0.001
	Interaction between age and BMI at AR		−0.50	−1.00,	0.00	0.05
	Female vs. male		13.36	11.99,	14.73	<0.001

Table 8.43: Estimated coefficients (coeff.), 95% confidence intervals (CI) and P-values for the linear regression models of body mass index (BMI) and percentage body fat (%BF) at examination fitted jointly on age and body mass index at adiposity rebound (AR) using the 100 imputed datasets. Models are adjusted for age at examination.

In both models there is some evidence of an age at AR-BMI at AR interaction, but in the BMI at examination model this is weak. However, as the evidence for both interaction terms is strong ($P \leq 0.001$) when considering the original data only, they are both retained in the model. As with the analysis using the original data only in Section 8.9.2.1, in order to assess the impact of this interaction it is easier to rewrite the model (as in (8.2) and (8.3)) and tabulate some appropriate values.

Table 8.44 shows the estimated increase in BMI at examination for a 1 kg/m^2 increase in BMI at AR for different ages at AR. This increase can be seen to be twice as great for an AR near the start of the observed range (3 years) as for an AR towards the end (8 years).

Age at	Increase in BMI at examination (kg/m^2)		
AR (years)	Estimate	95% CI	P-value
3	1.54	1.00, 2.07	<0.001
4	1.39	1.02, 1.75	<0.001
5	1.24	0.99, 1.49	<0.001
6	1.09	0.83, 1.36	<0.001
7	0.95	0.55, 1.35	<0.001
8	0.80	0.22, 1.38	0.01

Table 8.44: Estimates, 95% confidence intervals (CI) and P-values for the increase in body mass index (BMI) at examination (in kg/m^2) for a 1 kg/m^2 increase in body mass index at adiposity rebound (AR) at different age at adiposity rebound levels using the 100 imputed datasets.

Table 8.45 shows the estimated decrease in BMI at examination for a 1 year delay in AR at different BMI at AR levels. The decrease in BMI at examination is over 4 times as great for a BMI at AR near the top end of the range of observed values (20 kg/m^2) as for a BMI at AR towards the bottom (13 kg/m^2). However, although the decrease is highly statistically significant (P<0.001) towards the middle of the range of observed BMI at AR values, at either end of this range the evidence for it differing from 0 is reduced.

BMI at	Decrease in BMI at examination (kg/m^2)		
AR (kg/m^2)	Estimate	95% CI	P-value
13	0.30	−0.17, 0.78	0.21
14	0.45	0.14, 0.76	0.005
15	0.60	0.37, 0.82	<0.001
16	0.74	0.45, 1.04	<0.001
17	0.89	0.44, 1.34	<0.001
18	1.04	0.41, 1.67	0.001
19	1.18	0.36, 2.01	0.005
20	1.33	0.31, 2.35	0.01

Table 8.45: Estimates, 95% confidence intervals (CI) and P-values for the decrease in body mass index (BMI) at examination (in kg/m^2) for a 1 year delay in adiposity rebound (AR) at different body mass index at adiposity rebound levels using the 100 imputed datasets.

Table 8.46 is the equivalent table corresponding to the model with %BF as outcome. It can be seen that, although a 1 kg/m^2 increase in BMI at AR is estimated to correspond to around a 2.5% increase in BMI when AR occurs at a young age, when AR is towards the end of the range of observed values the estimated increase in %BF is negligible.

Age at	Increase in %BF at examination		
AR (years)	Estimate	95% CI	P-value
3	2.47	1.17, 3.77	<0.001
4	1.98	1.08, 2.87	<0.001
5	1.48	0.84, 2.12	<0.001
6	0.98	0.26, 1.70	0.01
7	0.48	−0.58, 1.54	0.37
8	−0.02	−1.51, 1.48	0.98

Table 8.46: Estimates, 95% confidence intervals (CI) and P-values for the increase in percentage body fat (%BF) at examination for a 1 kg/m^2 increase in body mass index (BMI) at adiposity rebound (AR) at different age at adiposity rebound levels using the 100 imputed datasets.

A similar pattern is observed in Table 8.47, with a 1 year delay in AR estimated to decrease

%BF by over 3% when BMI at AR is 20 kg/m^2, but when BMI at AR is 13 kg/m^2 there is virtually no decrease in %BF. Again, the estimated decrease in %BF at examination is highly statistically significant (P<0.001) when BMI at AR is towards the centre of the range of observed values, but the wider CIs at the more extreme BMI at AR values mean evidence of any decrease at all is markedly reduced.

BMI at	Decrease in %BF at examination		
AR (kg/m^2)	Estimate	95% CI	P-value
13	0.13	−1.08, 1.35	0.83
14	0.63	−0.20, 1.46	0.14
15	1.13	0.50, 1.76	0.001
16	1.63	0.85, 2.41	<0.001
17	2.12	0.98, 3.27	<0.001
18	2.62	1.04, 4.21	0.001
19	3.12	1.07, 5.18	0.003
20	3.62	1.08, 6.15	0.01

Table 8.47: Estimates, 95% confidence intervals (CI) and P-values for the decrease in percentage body fat (%BF) at examination for a 1 year delay in adiposity rebound (AR) at different body mass index (BMI) at adiposity rebound levels using the 100 imputed datasets.

The fitted model for BMI at examination in Table 8.43 also provides some evidence of greater BMI at examination in females for a given age and BMI at AR. The %BF at examination model, on the other hand, estimates a large and highly significant increase in %BF for females when compared to males.

A direct comparison of the models fitted separately (Table 8.42) and jointly (Table 8.43) on age and BMI at AR is again hampered by the interaction terms in the latter models. It can be seen from the highly statistically significant age and BMI at AR parameters in Table 8.43, however, that both dimensions of the AR remain strongly associated with both BMI and %BF at examination even when conditioning on the other dimension of AR and any potential interaction.

8.9.2.3 Comparison of results using the original data only and results using the imputed datasets

Comparison of the models using the 100 imputed datasets in Section 8.9.2.2 to those using the original data only in Section 8.9.2.1 allows an examination of how utilisation of the multiple imputation methodology impacts on the results obtained.

Comparing the models fitted separately on age and BMI at AR using the imputed datasets (Table 8.42) to those using the original data only (Table 8.36) shows the estimated models to be largely similar. The effects of BMI at AR on both BMI and %BF at examination are slightly

attenuated under multiple imputation, whilst the age at AR coefficients in both models remain almost identical. The effect of sex in each model is also attenuated somewhat. In the models for %BF at examination the estimated coefficients using the original data only are so highly significant that this attenuation has little impact. In the models for BMI at examination, however, this means that in the age at AR model using the imputed datasets there is no evidence at all for a sex effect and in the BMI at AR model the evidence for a sex effect is markedly weakened. The reasons behind the attenuated coefficients under the MI approach are discussed in Section 8.11.

The estimated models for BMI and %BF at examination fitted jointly on age and BMI at AR in Tables 8.37 and 8.43 show very similar patterns of coefficient attenuation for the age and BMI at AR coefficients to the models fitted separately on the explanatory variables, namely slight attenuation of the BMI at AR coefficients and stable age at AR coefficients.

The meaningful difference between the two approaches, however, is in the attenuated age at AR-BMI at AR interaction coefficients when analysing the imputed datasets. However, this can probably be explained by the imputation model lacking the equivalent interaction, as explained previously. The overall effect of this reduced interaction on the models is best investigated by comparison of the estimated increases or decreases in the outcome variables for different combinations of the explanatory variables. These are detailed in Tables 8.38, 8.39, 8.40, 8.41, 8.44, 8.45, 8.46 and 8.47, though a plot of the equivalent values using the original data only and using the imputed datasets on the same axes is more informative.

Fig. 8.19 plots the estimated increases in BMI (upper plot) and %BF (lower plot) at examination associated with a 1 kg/m^2 increase in BMI at AR for different ages at AR (see Tables 8.38, 8.39, 8.46 and 8.47). It can be seen that both relationships in the models using the imputed datasets are 'flatter' due to the smaller estimated interaction, meaning that the estimated increase in BMI or %BF at examination associated with increased BMI at AR is less dependent on age at AR. When considering %BF at examination the implications of this are somewhat more noticeable — in the fitted model using the original data only an increase in BMI at AR corresponding to a late AR is estimated to lead to a somewhat implausible *decrease* in %BF, but under the fitted model using the imputed datasets this is not the case, with increasing BMI at AR at a late AR merely seen to have little effect on %BF. However, the 95% CIs around this age are fairly wide under both approaches.

Fig. 8.20 shows the equivalent plots for the estimated decreases in the outcome variables associated with a 1 year delay in AR for different values of BMI at AR (corresponding to Tables 8.40, 8.41, 8.46 and 8.47). Once again use of the multiple imputation procedure results in a flattening of both relationships, meaning that the estimated decrease in BMI or %BF at examination associated with a delayed in AR is less dependent on BMI at AR. The lower plot shows a delayed AR corresponding to a low BMI at AR estimated to lead to *increased* %BF at examination in the fitted model using the original data only — but again this anomaly disappears in the model using the imputed datasets.

The effect of sex in the fitted models for BMI and %BF at examination fitted jointly on age and

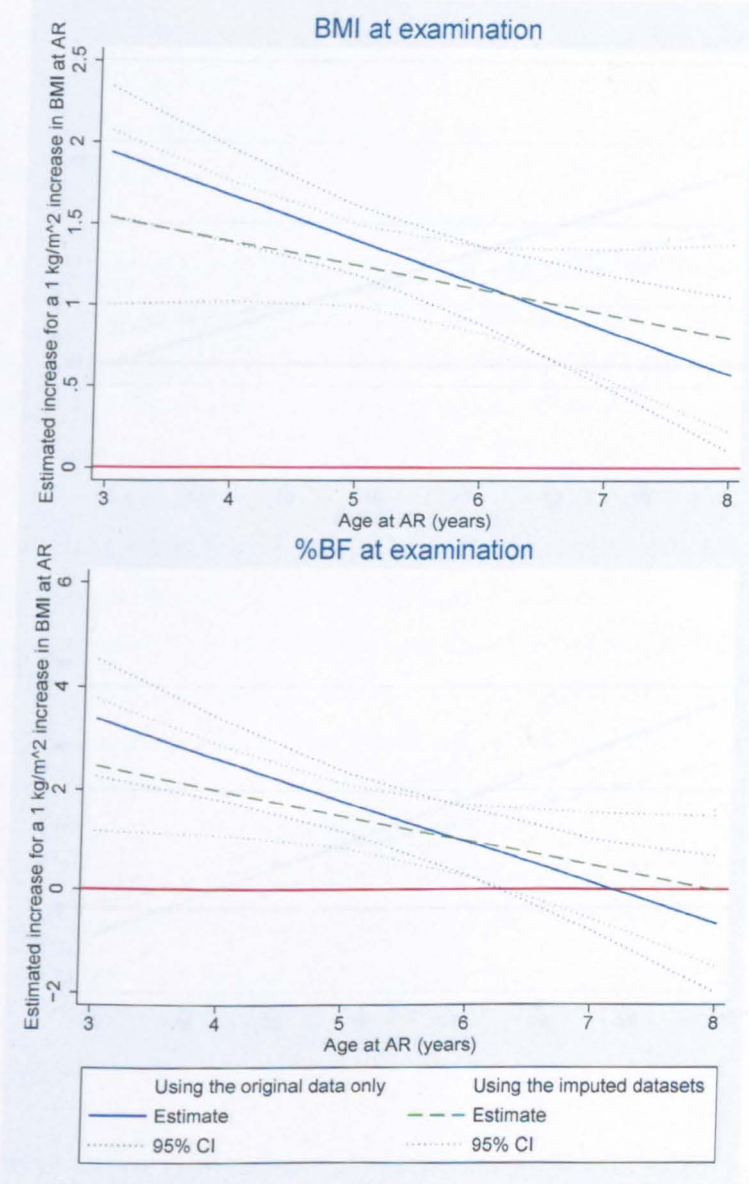

Fig. 8.19: Estimated increases in body mass index (BMI) and percentage body fat (%BF) at examination associated with a 1 kg/m^2 increase in body mass index at adiposity rebound (AR) for different ages at adiposity rebound using the original data only or the 100 imputed datasets.

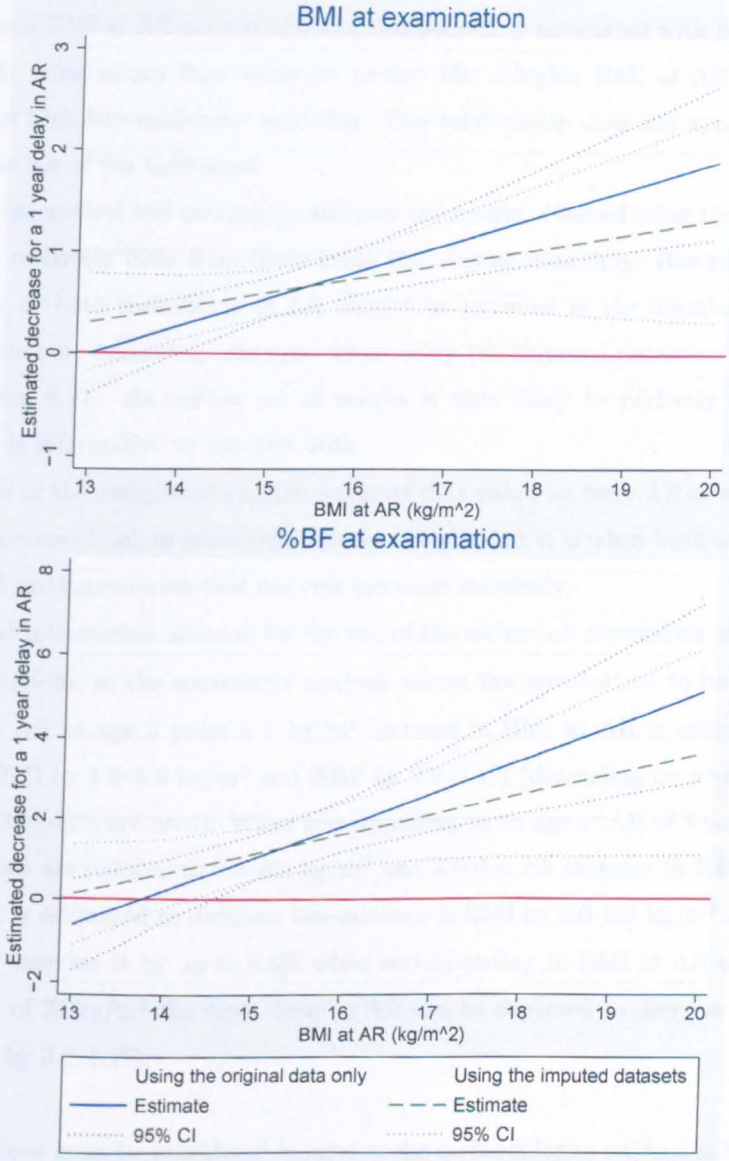

Fig. 8.20: Estimated decreases in body mass index (BMI) and percentage body fat (%BF) at examination associated with a 1 year delay in adiposity rebound (AR) for different values of body mass index at adiposity rebound using the original data only or the 100 imputed datasets.

BMI at AR in Tables 8.37 and 8.43 is also attenuated somewhat in the models using the imputed datasets. In the model for BMI at examination this means that there is no longer compelling evidence for the necessity of a sex parameter in the model.

8.9.3 Conclusions

It is clear from both the categorical analysis in Section 8.9.1 and the continuous analysis in Section 8.9.2 that age and BMI at AR are strongly and independently associated with both BMI and %BF at examination. This means that either an earlier AR, a higher BMI at AR or both increases the likelihood of high late-adolescent adiposity. This relationship does not appear to be modified according to the sex of the individual.

In both the categorical and continuous analyses the results obtained using the imputed datasets generally differ relatively little from those using the original data only. However, as interactions involving either or both dimensions of AR cannot be included in the imputation model, these interactions cannot be accurately explored when using the imputed datasets. This is discussed further in Section 8.11. As neither set of results is thus likely to perfectly describe the true relationships, it is informative to consider both.

Both versions of the categorical analysis suggests that either an early AR or a high BMI at AR will lead to an increased risk of late-adolescent overweight, but it is when both an early AR *and* a high BMI at AR are experienced that the risk increases massively.

The increased information afforded by the use of the estimated dimensions of AR, as opposed to categorised versions, in the continuous analysis allows the associations to be more closely examined. For an AR at age 3 years a 1 kg/m^2 increase in BMI at AR is estimated to increase late-adolescent BMI by 1.5–1.9 kg/m^2 and %BF by 2.5–3.4% (depending on whether the original data or the imputed data are used). When corresponding to an age at AR of 8 years, however, the estimated increases are reduced to 0.6–0.8 kg/m^2 and a 0.0–0.7% *increase* in %BF. Similarly, a 1 year delay in AR is estimated to decrease late-adolescent BMI by 0.0–0.3 kg/m^2 and may slightly decrease %BF or increase it by up to 0.6% when corresponding to BMI at AR of 13 kg/m^2, but for a BMI at AR of 20 kg/m^2 the same delay in AR can be expected to decrease BMI by 1.3–1.9 kg/m^2 and %BF by 3.6–5.1%.

These conclusions must be considered in light of the methodologies utilised in the analyses and the constraints of the data itself. However, as these issues are common to all analyses undertaken, this discussion is deferred until Section 8.11. In particular, as the majority of estimated ORs in Section 8.9.1 and estimated regression coefficients in Section 8.9.2 are attenuated when using the MI approach relative to an analysis of the original data only, it is important to consider whether using the original data only may result in an over-estimation of the associations or whether using the imputed datasets may result in under-estimation.

8.10 Is the adiposity rebound a critical period for late-adolescent obesity?

In Section 8.9 age and BMI at AR were shown to be significantly and independently associated with adolescent adiposity. What is not clear is whether there is anything 'special' about the AR. It is, by definition, an indicator of the level of BMI at the point in childhood when BMI stops decreasing and begins increasing once more (i.e. when BMI velocity is zero). However, if this is *all* that the AR is then there is little merit in using it as a predictor for later adiposity in preference to the value and velocity of BMI at *any* similar age in childhood. The issue here is really whether the AR can be considered a *critical period*, defined by Dietz [74] as 'a developmental stage in which physiologic alterations increase the later prevalence of obesity'.

The fitting of splines in the current dataset affords the opportunity for a closer examination of this issue. From the fitted splines it is possible to derive estimates for the BMI and BMI velocity at any given age. By including in the same linear regression model for adolescent adiposity both the age and BMI at AR and the BMI and BMI velocity at a given age in childhood it can be assessed whether, conditional on the BMI and BMI velocity at that age, knowledge of the AR provides any further information for the prediction of adolescent adiposity. If the AR is a critical period for adolescent adiposity then it should give additional information even when the BMI and BMI velocity included in the model correspond to post-AR ages. If, however, the AR is merely equivalent to BMI centile crossing at that age then this will only be the case when the BMI and BMI velocity correspond to pre-AR ages.

The results are presented separately using the original data only (Section 8.10.1) and using the imputed datasets (Section 8.10.2), then the two sets of results compared in Section 8.10.3. In each instance male and female results are presented separately as initial investigations highlighted sex-specific effects at some ages. Section 8.10.4 draws together the findings to present some overall conclusions.

8.10.1 Using the original data only

BMI and BMI velocity values at ages 4, 5, 6, 7 and 8 years are derived from the previously fitted subject-specific splines and incorporated into different models. 'Model 1' in each instance is a linear regression of the outcome (either BMI or %BF at examination) on the BMI and BMI velocity at each age in turn. 'Model 2' has the addition of the age and BMI at AR so can be used to assess whether knowledge of the location of the AR adds any further information given the prior knowledge of the BMI and BMI velocity at that age.

In all models only data from those subjects with identified ARs are used. Whilst this reduces the effective sample size somewhat, it ensures that those individuals with poorly fitted splines, either due to a lack of data or the available data displaying an unlikely growth trajectory, are not included in the analysis. Whilst BMI at examination is fully observed, %BF is not, meaning that in the models with %BF as outcome some subjects must also be excluded for this reason. As a

result the sample sizes differ between the models.

BMI at examination Table 8.48 details the regression models for BMI at examination in the males. At all five ages at which Model 1 is fitted there is no evidence of an interaction between BMI and BMI velocity (P>0.1 at all ages), thus an interaction term is not included in the models presented. When fitting Model 2 there is again little evidence of a BMI-BMI velocity interaction at any age, although there is limited evidence of an interaction between age and BMI at AR at age 6 years. However, in the interests of model comparability at each age Model 2 is fitted without interaction terms.

In Model 1 it can be seen that at all ages both BMI and BMI velocity are, conditional on each other, positively associated with BMI at examination. The coefficients for BMI are highly significant at all ages, whereas for BMI velocity this is only true at ages 6 and 7 years, with only weak evidence of any association at all at age 8 years. Given that the median age at AR in males was found to be 5.7 years (see Table 8.9 in Section 8.7.1), the BMI velocity between age 5 and 7 years will be indicative of whether or not AR has already been passed, thus this peak in coefficient significance may indicate the importance of the timing of the AR on later BMI.

The effect of the introduction of the age and BMI at AR in Model 2 is very much dependent on the age at which the BMI and BMI velocity values are considered. At ages 4 and 5 years (prior to the median AR) the age and BMI at AR coefficients are highly significant, reducing the BMI and BMI velocity coefficients to non-significance. Given that the variables corresponding to the location of the AR are temporally closer to the outcome and contain similar information it is unsurprising that they exert a greater influence. At age 6 years, however, whilst the BMI at AR coefficient remains significant, BMI velocity, rather than age at AR, is now highly significantly associated with BMI at examination. This is perhaps explained by the age under consideration being later than the median age at AR, though this would also lead to the expectation of BMI at age 6 exerting greater influence than BMI at AR, which is not the case. At ages 7 and 8 years, beyond the age at which AR occurs in most males, it is BMI and BMI velocity at that age, as opposed to age and BMI at AR, which have the greater effect on BMI at examination.

Table 8.49 details the equivalent models amongst the females. Again, there is a lack of evidence to support the inclusion of BMI-BMI velocity interactions at any ages in Model 1 and BMI-BMI velocity and age at AR-BMI at AR interactions at any ages in Model 2.

In Model 1 BMI is positively and highly significantly associated with BMI at examination at every age. BMI velocity is also exhibits a positive association, though the coefficient is only highly significant until age 6 years, with little evidence of any relationship after that age. This earlier non-significance of the BMI velocity coefficient in females when compared to males is perhaps attributable to the earlier AR (median 5.5 years) identified in females (see Table 8.9 in Section 8.7.1).

In Model 2 at age 4 years the AR variables have significant associations with BMI at examina-

Explanatory variable	Model 1			Model 2		
	Coeff.	95% CI	P-value	Coeff.	95% CI	P-value
BMI age 4 years	1.04	0.62, 1.47	<0.001	−0.01	−1.22, 1.20	0.98
BMI velocity age 4 years	1.03	−0.27, 2.32	0.12	−0.82	−2.10, 0.46	0.21
BMI at AR	-	-	-	1.25	0.01, 2.50	0.05
Age at AR	-	-	-	−0.68	−1.11, −0.24	0.003
BMI age 5 years	0.98	0.57, 1.38	<0.001	−0.61	−1.92, 0.71	0.37
BMI velocity age 5 years	1.52	0.38, 2.66	0.01	−0.30	−1.73, 1.12	0.67
BMI at AR	-	-	-	1.86	0.46, 3.26	0.01
Age at AR	-	-	-	−0.59	−1.03, −0.14	0.01
BMI age 6 years	0.97	0.65, 1.28	<0.001	−0.06	−1.11, 1.00	0.92
BMI velocity age 6 years	2.22	1.29, 3.15	<0.001	2.19	0.93, 3.45	0.001
BMI at AR	-	-	-	1.22	0.01, 2.42	0.05
Age at AR	-	-	-	−0.11	−0.52, 0.30	0.59
BMI age 7 years	1.12	0.84, 1.39	<0.001	0.81	−0.07, 1.69	0.07
BMI velocity age 7 years	1.31	0.51, 2.12	0.002	1.41	0.44, 2.38	0.01
BMI at AR	-	-	-	0.35	−0.67, 1.37	0.50
Age at AR	-	-	-	−0.11	−0.48, 0.27	0.57
BMI age 8 years	1.11	0.86, 1.36	<0.001	1.00	0.36, 1.65	0.003
BMI velocity age 8 years	0.66	−0.23, 1.55	0.14	0.72	−0.33, 1.76	0.18
BMI at AR	-	-	-	0.08	−0.68, 0.85	0.83
Age at AR	-	-	-	−0.12	−0.48, 0.24	0.52

Table 8.48: Estimated coefficients (coeff.), 95% confidence intervals (CI) and P-values for the linear regression models of body mass index (BMI) at examination (kg/m^2) on body mass index (kg/m^2) and body mass index velocity (kg/m^2/year) at a given age, and age (years) and body mass index (kg/m^2) at adiposity rebound (AR) in males using the original data. Models are adjusted for age at examination. 111 individuals in each model.

Explanatory variable	Model 1			Model 2		
	Coeff.	95% CI	P-value	Coeff.	95% CI	P-value
BMI age 4 years	1.35	1.05, 16.4	<0.001	−0.07	−1.12, 0.98	0.90
BMI velocity age 4 years	1.44	0.52, 2.36	0.002	0.15	−0.80, 1.10	0.76
BMI at AR	-	-	-	1.45	0.38, 2.51	0.01
Age at AR	-	-	-	−0.55	−0.92, −0.79	0.003
BMI age 5 years	1.33	1.08, 1.59	<0.001	−0.17	−1.02, 0.69	0.70
BMI velocity age 5 years	1.77	0.92, 2.63	<0.001	1.97	0.62, 3.32	0.01
BMI at AR	-	-	-	1.59	0.70, 2.48	0.001
Age at AR	-	-	-	−0.11	−0.52, 0.31	0.62
BMI age 6 years	1.29	1.07, 1.50	<0.001	0.82	0.20, 1.44	0.01
BMI velocity age 6 years	1.37	0.70, 2.04	<0.001	1.73	0.90, 2.56	<0.001
BMI at AR	-	-	-	0.57	−0.08, 1.23	0.09
Age at AR	-	-	-	0.08	−0.27, 0.43	0.66
BMI age 7 years	1.31	1.11, 1.51	<0.001	1.05	0.49, 1.61	<0.001
BMI velocity age 7 years	0.06	−0.64, 0.76	0.87	0.39	−0.48, 1.26	0.38
BMI at AR	-	-	-	0.34	−0.26, 0.95	0.26
Age at AR	-	-	-	0.02	−0.32, 0.35	0.91
BMI age 8 years	1.16	0.99, 1.33	<0.001	0.77	0.38, 1.15	<0.001
BMI velocity age 8 years	0.09	−0.68, 0.86	0.81	0.65	−0.20, 1.50	0.13
BMI at AR	-	-	-	0.63	0.18, 1.08	0.01
Age at AR	-	-	-	−0.03	−0.35, 0.30	0.88

Table 8.49: Estimated coefficients (coeff.), 95% confidence intervals (CI) and P-values for the linear regression models of body mass index (BMI) at examination (kg/m^2) on body mass index (kg/m^2) and body mass index velocity (kg/m^2/year) at a given age, and age (years) and body mass index (kg/m^2) at adiposity rebound (AR) for females using the original data. Models are adjusted for age at examination. 150 individuals in each model.

tion whereas there is no evidence of associations with BMI and BMI velocity. At age 5 years it is the BMI velocity and the BMI at AR are the only significant parameters. This is the same pattern as was identified in the males at age 6 (one year later), with this difference again perhaps due to the generally earlier AR in females. At age 6 years BMI and BMI velocity are more strongly associated with the outcome than the AR variables. At both ages 7 and 8 years BMI is highly significantly associated with BMI at examination but BMI velocity is non-significant. One slightly anomalous result is the return to significance of the BMI at AR coefficient at age 8 years in addition to the BMI variable at that age which would be expected to be carrying similar, though more recent, information.

From the trends observed in these models it is evident that for ages before the occurrence of AR in most individuals the location of the AR is more strongly associated with later BMI. When considering ages when AR has already been passed in the majority the opposite is true, with the BMI and BMI velocity at that age taking greater significance than the AR.

%BF at examination Table 8.50 and Table 8.51 show the corresponding models for %BF at examination in males and females respectively. When considering %BF as outcome there is no evidence of an interaction between BMI and BMI velocity at any age in Model 1 ($P>0.2$ at all ages for males and $P>0.3$ at all ages for females). There is also no evidence for a BMI-BMI velocity interaction in Model 2 ($P>0.3$ at all ages for males and $P>0.4$ at all ages for females), although there is some evidence of an interaction between age and BMI at AR at several ages. Again, however, to provide models which remain comparable with others these potential interactions are ignored.

From Model 1 in Table 8.50 it can be seen that BMI is positively and significantly related to %BF at all ages in males. The relationship with BMI velocity is also positive when significant, though this only occurs at ages 5 and 6 years. These results are very similar to those obtained in the models for BMI at examination.

Model 2 shows that %BF at examination is most strongly associated with age at AR at each age, with the relationship being highly significant at ages 4 and 5 years but only borderline-significant at later ages. This means that at, for example, age 8 years, even for given BMI and BMI velocity values at that age the location of the AR provides a significant amount of additional information. The pattern exhibited in the %BF at examination models for males is very different to that in the BMI models. Whilst %BF and BMI do inherently differ in what they are measuring, with %BF a more direct measure of adiposity and BMI a somewhat weaker proxy, these differences perhaps remain surprising.

Table 8.51 details models 1 and 2 for %BF at examination in females. Model 1, similarly to in males, shows %BF to be generally positively related to both BMI and BMI velocity at all ages considered. This association is highly significant for BMI at each age though only significant for

Explanatory variable	Model 1			Model 2		
	Coeff.	95% CI	P-value	Coeff.	95% CI	P-value
BMI age 4 years	1.61	0.52, 2.69	0.004	1.29	−2.01, 4.59	0.44
BMI velocity age 4 years	2.44	−0.85, 5.73	0.15	−0.69	−4.17, 2.80	0.70
BMI at AR	-	-	-	0.57	−2.83, 3.96	0.74
Age at AR	-	-	-	−1.82	−3.01, −0.64	0.003
BMI age 5 years	1.52	0.47, 2.56	0.01	0.97	−2.63, 4.57	0.59
BMI velocity age 5 years	3.46	0.51, 6.40	0.02	−0.60	−4.74, 3.27	0.76
BMI at AR	-	-	-	0.84	−2.97, 4.65	0.66
Age at AR	-	-	-	−1.64	−2.84, −0.43	0.01
BMI age 6 years	1.66	0.76, 2.55	<0.001	0.89	−2.11, 3.90	0.56
BMI velocity age 6 years	3.50	0.85, 6.15	0.01	1.59	−1.99, 5.18	0.38
BMI at AR	-	-	-	0.77	−2.66, 4.21	0.66
Age at AR	-	-	-	−0.98	−2.14, 0.19	0.10
BMI age 7 years	1.97	1.15, 2.79	<0.001	1.48	−1.10, 4.07	0.26
BMI velocity age 7 years	0.81	-1.60, 3.21	0.51	−0.01	−2.87, 2.86	0.99
BMI at AR	-	-	-	0.18	−2.82, 3.19	0.90
Age at AR	-	-	-	−0.96	−2.06, 0.13	0.08
BMI age 8 years	1.89	1.11, 2.66	<0.001	0.88	−1.07, 2.82	0.38
BMI velocity age 8 years	−0.65	−3.38, 2.08	0.64	−0.09	−3.25, 3.08	0.96
BMI at AR	-	-	-	0.83	−1.47, 3.14	0.47
Age at AR	-	-	-	−1.05	−2.14, 0.03	0.06

Table 8.50: Estimated coefficients (coeff.), 95% confidence intervals (CI) and P-values for the linear regression models of percentage body fat (%BF) at examination on body mass index (BMI) (kg/m^2) and body mass index velocity (kg/m^2/year) at a given age, and age (years) and body mass index (kg/m^2) at adiposity rebound (AR) for males using the original data. Models are adjusted for age at examination. 109 individuals in each model.

BMI velocity up to age 6 years.

Explanatory variable	Model 1			Model 2		
	Coeff.	95% CI	P-value	Coeff.	95% CI	P-value
BMI age 4 years	1.48	0.70, 2.25	<0.001	0.44	−2.59, 3.47	0.78
BMI velocity age 4 years	3.02	0.61, 5.44	0.02	1.24	−1.50, 3.98	0.37
BMI at AR	-	-	-	1.02	−2.06, 4.10	0.51
Age at AR	-	-	-	−0.94	−1.99, 0.11	0.08
BMI age 5 years	1.50	0.80, 2.21	<0.001	0.38	−2.11, 2.87	0.76
BMI velocity age 5 years	4.10	1.71, 6.50	0.001	4.72	0.76, 8.69	0.02
BMI at AR	-	-	-	1.21	−1.38, 3.81	0.36
Age at AR	-	-	-	0.12	−1.09, 1.33	0.85
BMI age 6 years	1.63	0.98, 2.28	<0.001	2.52	0.63, 4.41	0.01
BMI velocity age 6 years	2.45	0.42, 4.47	0.02	2.40	−0.13, 4.93	0.06
BMI at AR	-	-	-	−0.98	−2.96, 1.01	0.33
Age at AR	-	-	-	0.25	−0.80, 1.31	0.63
BMI age 7 years	1.71	1.11, 2.32	<0.001	2.57	0.85, 4.28	0.004
BMI velocity age 7 years	0.34	−1.81, 2.50	0.75	−0.53	−3.18, 2.13	0.70
BMI at AR	-	-	-	−1.04	−2.88, 0.80	0.27
Age at AR	-	-	-	0.15	−0.87, 1.16	0.78
BMI age 8 years	1.44	0.94, 1.94	<0.001	1.32	0.13, 2.51	0.03
BMI velocity age 8 years	1.70	−0.59, 4.00	0.14	1.87	−0.74, 4.48	0.16
BMI at AR	-	-	-	0.19	−1.20, 1.58	0.79
Age at AR	-	-	-	−0.02	−1.01, 0.97	0.97

Table 8.51: Estimated coefficients (coeff.), 95% confidence intervals (CI) and P-values for the linear regression models of percentage body fat (%BF) at examination on body mass index (BMI) (kg/m^2) and body mass index velocity (kg/m^2/year) at a given age, and age (years) and body mass index (kg/m^2) at adiposity rebound (AR) for females using the original data. Models are adjusted for age at examination. 150 individuals in each model.

The addition of age and BMI at AR in Model 2 of Table 8.51 gives results quite different to those for the males. At age 4 years %BF at examination is more strongly associated with the location of the AR, in particular the age at AR, though this is only borderline-significant. From age 5 years onwards %BF appears to be more influenced by the BMI and BMI velocity at that age than the location of the AR though in differing ways: at age 5 years BMI velocity has the greater effect, at age 6 years both BMI and BMI velocity, and from age 7 years onwards just BMI.

Unlike the models for BMI at examination, those for %BF do not unify across the sexes to provide an overarching pattern so readily. For both males and females the overall significance

of coefficients is less, making trends less discernible, though this is perhaps to be expected given that in this case the explanatory variables are not merely earlier measurements of the outcome variable. In females the pattern is similar to that with BMI at examination as the outcome, with whichever of the variables in the model were observed closest to the outcome exerting the greater influence. For the males, however, the location of, and in particular the age at, AR was shown to be of significance at all ages. Whilst the sample size is relatively small, that this pattern continues at all ages until age 8 years lends some gravitas to the observation. This could indicate that %BF in males may indeed have a more complex relationship with the AR than the reduction of AR to merely the relative level and rate of change of BMI at that age allows.

8.10.2 Using the imputed datasets

BMI and BMI velocity values at ages 4, 5, 6, 7 and 8 years are derived from the previously fitted subject-specific splines in each of the 100 imputed datasets in the same manner as for the original data in Section 8.10.1. Again, these values are incorporated into two different models, with 'Model 1' being a regression of the outcome (either BMI or %BF at examination) on the BMI and BMI velocity at each age in turn and 'Model 2' having the addition of the age and BMI at AR. The comparison of the estimated coefficients in these models then facilitates the assessment of whether knowledge of the location of the AR adds any further information given the prior knowledge of the BMI and BMI velocity at that age.

In all models only data from those subjects with identified ARs are used, providing a mechanism to ensure that only those individuals with well-defined splines contribute to the analysis. Although this reduces the effective sample size somewhat, the effect is not as marked as in the analysis using the original data as the MI procedure allows a greater proportion of splines to be fitted and thus AR to be identified. However, as the number of identified ARs differs between each imputed dataset so does the number of subjects contributing to each model: between 134 and 146 in those models for males and between 193 and 207 in those for females.

As discussed previously when using logistic (Section 8.9.1.2) or linear (Section 8.9.2.2) regression models to assess whether dimensions of the adiposity rebound are associated with late-adolescent obesity, the manner in which the AR location is estimated *after* imputation takes place means that no interactions involving either or both dimensions of the AR can be included in the imputation model. This is also true for interactions involving BMI and/or BMI velocity values derived from the fitted splines. If these interaction terms are then included in the analysis model using the imputed data, their estimated values will be biased towards the null [123]. This also means that P-values for significance tests for the inclusion of interaction terms in the analysis model are likely to be overestimated. As a result, these significance test are not conducted here. Instead, interaction terms are included if and only if they are deemed necessary in the equivalent model using the original data. Thus, as no interaction terms are included in any of the models in Section 8.10.1, none will be included here.

BMI at examination Table 8.52 details the regression models for BMI at examination in the males. In Model 1 it can be seen that at all ages both BMI and BMI velocity are, conditional on each other, positively associated with BMI at examination. Whilst the coefficients for BMI are highly significant (P<0.001) at all ages, those for BMI velocity peak in significance at ages 5 and 6 years, though the association remains significant across the range of ages. Given that the 'mean median' age at AR in males was found to be 5.7 years (see Table 8.12 in Section 8.7.2), the BMI velocity at ages 5 and 6 years will be indicative of whether or not AR has already been passed, thus this peak in coefficient significance may indicate the importance of the timing of the AR on later BMI.

Explanatory variable	Model 1			Model 2		
	Coeff.	95% CI	P-value	Coeff.	95% CI	P-value
BMI age 4 years	0.96	0.51, 1.41	<0.001	0.30	−1.06, 1.66	0.67
BMI velocity age 4 years	1.92	0.56, 3.27	0.01	0.10	−1.34, 1.55	0.89
BMI at AR	-	-	-	0.85	−0.59, 2.28	0.25
Age at AR	-	-	-	−0.81	−1.33, −0.29	0.002
BMI age 5 years	0.95	0.54, 1.36	<0.001	0.36	−1.03, 1.75	0.62
BMI velocity age 5 years	2.11	0.96, 3.25	<0.001	0.60	−1.01, 2.21	0.46
BMI at AR	-	-	-	0.74	−0.77, 2.25	0.34
Age at AR	-	-	-	−0.65	−1.19, −0.11	0.02
BMI age 6 years	1.04	0.72, 1.37	<0.001	0.77	−0.28, 1.81	0.15
BMI velocity age 6 years	2.37	1.36, 3.37	<0.001	2.21	0.72, 3.70	0.004
BMI at AR	-	-	-	0.32	−0.91, 1.54	0.61
Age at AR	-	-	-	−0.11	−0.61, 0.38	0.65
BMI age 7 years	1.15	0.87, 1.43	<0.001	1.19	0.34, 2.05	0.01
BMI velocity age 7 years	1.40	0.48, 2.31	0.003	1.23	0.11, 2.35	0.03
BMI at AR	-	-	-	−0.11	−1.12, 0.91	0.84
Age at AR	-	-	-	−0.10	−0.53, 0.32	0.63
BMI age 8 years	1.02	0.78, 1.27	<0.001	0.93	0.28, 1.57	0.005
BMI velocity age 8 years	1.71	0.67, 2.75	0.001	1.73	0.51, 2.96	0.01
BMI at AR	-	-	-	0.05	−0.74, 0.84	0.91
Age at AR	-	-	-	−0.15	−0.54, 0.24	0.45

Table 8.52: Estimated coefficients (coeff.), 95% confidence intervals (CI) and P-values for the linear regression models of body mass index (BMI) at examination (kg/m^2) on body mass index (kg/m^2) and body mass index velocity (kg/m^2/year) at a given age, and age (years) and body mass index (kg/m^2) at adiposity rebound (AR) for males using the 100 imputed datasets. Models are adjusted for age at examination.

The effect of the introduction of the age and BMI at AR in Model 2 is very much dependent on

the age at which the BMI and BMI velocity values are considered. At ages 4 and 5 years (prior to the mean median age at AR) the introduction of the AR parameters means that the BMI and BMI velocity coefficients are no longer significant, but that the age at AR coefficients are significantly inversely associated with later BMI. At age 6 years, however, it is the BMI velocity rather than the age at AR which has the greater association with later BMI, possibly due to the age under consideration now being greater than the mean median age at AR. At ages 7 and 8 years, beyond the age at which AR occurs in most males, it is both BMI and BMI velocity at that age which have the greater association with BMI at examination.

Table 8.53 details the equivalent models amongst the females. In Model 1 BMI, for a given BMI velocity, is positively and highly significantly associated with BMI at examination at every age. BMI velocity is also exhibits a positive association (conditional on the BMI), though the coefficient is only significant until age 6 years, after which there is little evidence for the relationship. This earlier non-significance of the BMI velocity coefficient in females when compared to males is perhaps attributable to the earlier AR (mean median 5.4 years) identified in females (see Table 8.12 in Section 8.7.2).

In Model 2 at age 4 years age at AR only has a significant association with BMI at examination (conditional on the other three variables) meaning that the introduction of the AR variables has removed the effect of the BMI and BMI velocity at that age. At age 5 years it is the BMI, as opposed to age, at AR that is the most significant variable. From age 6 years onwards (i.e. after the mean median age at AR), the BMI and BMI velocity variables become more strongly associated with the outcome than the AR variables — a pattern very similar to that exhibited in the males. At age 6 years the association with BMI velocity is highly significant, whereas with BMI itself it is borderline significant. At ages 7 and 8 years it is BMI only that has a significant effect. The non-significance of BMI velocity after age 6 years can perhaps be explained because at age 6 years many females are still to exhibit AR, thus BMI velocity is an important indicator of whether AR has been passed or not. At older ages very few females will still be pre-AR, making an evaluation of BMI velocity somewhat redundant.

From the trends observed in these models it is evident that for ages before the occurrence of AR in most individuals the location of the AR is more strongly associated with later BMI. When considering ages when AR has already been passed in the majority the opposite is true, with the BMI and BMI velocity at that age (in particular the BMI itself) taking greater significance than the AR. This indicates that later BMI is most strongly associated with whichever measures occurred more recently.

%BF at examination Table 8.54 and Table 8.55 show the corresponding models for %BF at examination in males and females respectively. From Model 1 in Table 8.54 it can be seen that BMI is positively associated with %BF at all ages in males, conditional on BMI velocity, though

Explanatory variable	Model 1			Model 2		
	Coeff.	95% CI	P-value	Coeff.	95% CI	P-value
BMI age 4 years	1.25	0.95, 1.56	<0.001	0.40	−0.62, 1.42	0.44
BMI velocity age 4 years	1.28	0.34, 2.21	0.01	−0.09	−1.13, 0.94	0.86
BMI at AR	-	-	-	0.87	−0.19, 1.93	0.11
Age at AR	-	-	-	−0.64	−1.01, −0.27	0.001
BMI age 5 years	1.22	0.93, 1.50	<0.001	0.26	−0.73, 1.26	0.60
BMI velocity age 5 years	1.37	0.42, 2.32	0.005	0.98	−0.50, 2.46	0.19
BMI at AR	-	-	-	1.00	−0.03, 2.03	0.06
Age at AR	-	-	-	−0.31	−0.77, 0.15	0.19
BMI age 6 years	1.13	0.87, 1.39	<0.001	0.71	−0.09, 1.52	0.08
BMI velocity age 6 years	1.35	0.52, 2.19	0.002	1.65	0.55, 2.74	0.003
BMI at AR	-	-	-	0.52	−0.32, 1.36	0.22
Age at AR	-	-	-	0.04	−0.37, 0.44	0.86
BMI age 7 years	1.14	0.90, 1.38	<0.001	0.87	0.17, 1.58	0.02
BMI velocity age 7 years	0.36	−0.49, 1.21	0.40	0.67	−0.39, 1.73	0.22
BMI at AR	-	-	-	0.36	−0.38, 1.10	0.34
Age at AR	-	-	-	0.00	−0.38, 0.38	0.99
BMI age 8 years	1.04	0.85, 1.24	<0.001	0.75	0.28, 1.22	0.002
BMI velocity age 8 years	0.33	−0.54, 1.20	0.46	0.76	−0.23, 1.75	0.13
BMI at AR	-	-	-	0.49	−0.04, 1.02	0.53
Age at AR	-	-	-	−0.01	−0.38, 0.35	0.94

Table 8.53: Estimated coefficients (coeff.), 95% confidence intervals (CI) and P-values for the linear regression models of body mass index (BMI) at examination (kg/m^2) on body mass index (kg/m^2) and body mass index velocity (kg/m^2/year) at a given age, and age (years) and body mass index (kg/m^2) at adiposity rebound (AR) for females using the 100 imputed datasets. Models are adjusted for age at examination.

this relationship is only significant from age 6 years onwards. The association with BMI velocity is also positive, though this relationship is only of any real significance at ages 5 and 6 years, around the period when the majority of males exhibit the AR. These results follow a similar trend to those obtained in the models for BMI at examination, albeit with relatively less significance.

Explanatory variable	Model 1			Model 2		
	Coeff.	95% CI	P-value	Coeff.	95% CI	P-value
BMI age 4 years	1.08	−0.15, 2.31	0.08	0.39	−3.62, 4.41	0.35
BMI velocity age 4 years	3.31	−0.46, 7.07	0.09	0.03	−4.19, 4.26	0.85
BMI at AR	-	-	-	0.99	−3.21, 5.20	0.64
Age at AR	-	-	-	−1.71	−3.23, −0.19	0.03
BMI age 5 years	1.07	−0.09, 2.23	0.07	1.00	−3.09, 5.09	0.63
BMI velocity age 5 years	4.92	1.56, 8.28	0.004	2.22	−2.44, 6.87	0.63
BMI at AR	-	-	-	0.19	−4.19, 4.58	0.93
Age at AR	-	-	-	−1.24	−2.81, 0.34	0.12
BMI age 6 years	1.48	0.50, 2.47	0.003	2.11	−1.00, 5.22	0.18
BMI velocity age 6 years	4.41	1.29, 7.52	0.01	2.87	−1.82, 7.56	0.23
BMI at AR	-	-	-	−0.91	−4.58, 2.76	0.63
Age at AR	-	-	-	−0.65	−2.18, 0.88	0.41
BMI age 7 years	1.79	0.91, 2.68	<0.001	2.50	−0.14, 5.15	0.06
BMI velocity age 7 years	1.85	−1.13, 4.82	0.22	0.38	−3.28, 4.03	0.84
BMI at AR	-	-	-	−1.28	−4.48, 1.92	0.43
Age at AR	-	-	-	−0.66	−2.02, 0.69	0.33
BMI age 8 years	1.43	0.63, 2.24	0.001	1.07	−1.03, 3.16	0.32
BMI velocity age 8 years	3.49	0.02, 6.96	0.05	3.42	−0.59, 7.43	0.09
BMI at AR	-	-	-	0.04	−2.54, 2.62	0.97
Age at AR	-	-	-	−0.80	−2.09, 0.50	0.23

Table 8.54: Estimated coefficients (coeff.), 95% confidence intervals (CI) and P-values for the linear regression models of percentage body fat (%BF) at examination on body mass index (BMI) (kg/m^2) and body mass index velocity (kg/m^2/year) at a given age, and age (years) and body mass index (kg/m^2) at adiposity rebound (AR) for males using the 100 imputed datasets. Models are adjusted for age at examination.

Model 2 shows that %BF at examination is most strongly associated with age at AR, conditional on the other three variables, at age 4 and 5 years. At age 6 years there is little association between %BF at examination and any of the four variables. This lack of association may be explained because, as this is the age closest to the mean median age at AR, the two pairs of variables effectively contain the same information in many cases (i.e. for the many males with AR around age 6 years, BMI at age 6 and BMI at AR will be similar and BMI velocity at age 6 years will be indicative, and thus highly correlated with, age at AR). At ages 7 and 8 years, BMI and BMI

velocity at that age become more important than the location of the AR. The pattern exhibited in the %BF at examination models for males is, whilst somewhat more diluted, similar to that in the BMI models, with the location of the AR seemingly more important prior to the mean median age at AR and the BMI and BMI velocity at a given age having greater significance at later ages.

Table 8.55 details models 1 and 2 for %BF at examination in females. Model 1 shows %BF to be uniformly highly significantly associated with BMI, for a given BMI velocity, though there is only evidence of an associated with BMI velocity (conditional on BMI) up to age 6 years. This pattern is identical to that observed for the corresponding BMI at examination models in Table 8.53.

Explanatory variable	Model 1			Model 2		
	Coeff.	95% CI	P-value	Coeff.	95% CI	P-value
BMI age 4 years	1.32	0.56, 2.07	0.001	−0.24	−2.70, 2.23	0.85
BMI velocity age 4 years	2.88	0.63, 5.12	0.01	0.90	−1.69, 3.49	0.50
BMI at AR	-	-	-	1.62	−0.88, 4.11	0.21
Age at AR	-	-	-	−0.84	−1.80, 0.12	0.08
BMI age 5 years	1.32	0.62, 2.02	<0.001	0.11	−2.13, 2.35	0.92
BMI velocity age 5 years	3.86	1.46, 6.25	0.002	4.08	0.57, 7.60	0.02
BMI at AR	-	-	-	1.31	−1.00, 3.62	0.27
Age at AR	-	-	-	−0.05	−1.14, 1.05	0.94
BMI age 6 years	1.45	0.81, 2.08	<0.001	1.87	−0.03, 3.78	0.05
BMI velocity age 6 years	2.54	0.46, 4.63	0.02	2.44	−0.18, 5.05	0.07
BMI at AR	-	-	-	−0.49	−2.51, 1.54	0.64
Age at AR	-	-	-	0.07	−0.97, 1.12	0.89
BMI age 7 years	1.59	1.01, 2.17	<0.001	2.06	0.37, 3.76	0.02
BMI velocity age 7 years	0.23	−1.92, 2.38	0.83	−0.39	−3.07, 2.29	0.78
BMI at AR	-	-	-	−0.67	−2.52, 1.19	0.48
Age at AR	-	-	-	−0.07	−1.09, 0.95	0.89
BMI age 8 years	1.37	0.90, 1.84	<0.001	1.15	−0.02, 2.32	0.05
BMI velocity age 8 years	1.10	−1.16, 3.35	0.34	1.30	−1.28, 3.88	0.32
BMI at AR	-	-	-	0.24	−1.16, 1.64	0.73
Age at AR	-	-	-	−0.19	−1.20, 0.82	0.71

Table 8.55: Estimated coefficients (coeff.), 95% confidence intervals (CI) and P-values for the linear regression models of percentage body fat (%BF) at examination on body mass index (BMI) (kg/m^2) and body mass index velocity (kg/m^2/year) at a given age, and age (years) and body mass index (kg/m^2) at adiposity rebound (AR) for females using the 100 imputed datasets. Models are adjusted for age at examination.

The addition of age and BMI at AR in Model 2 of Table 8.55 gives results not dissimilar to those for the males. At age 4 years %BF at examination is more strongly associated with the location of the AR, in particular the age at AR. From age 5 years onwards %BF appears to be more influenced by the BMI and BMI velocity at that age than the location of the AR though in differing ways. At age 5 and 6 years BMI velocity is quite strongly associated with late %BF, though the magnitude of this relationship declines as the proportion of subjects having passed AR increases. From age 6 years onwards it is BMI itself which has the greater influence. That the location of the AR appears to lose its influence on later %BF at an earlier age in females relative to males is again likely due to the generally earlier ARs exhibited in females.

As for the models concerning BMI at examination, there are common trends evident across the male and female models for %BF at examination. For both sexes, at ages prior to AR in most subjects (ages 4 and 5 years in males and age 4 years in females) it is the location of the AR, and more specifically the age at AR, which has greatest influence on later %BF. At ages when most subjects have already exhibited AR (age 7 years and onwards in males and age 6 years and onwards in females) the BMI and BMI velocity at that age have the stronger association. At the ages closest to the mean median age at AR in each sex (age 6 years in males and age 5 years in females) the models may not necessarily behave quite as expected due to the information in the pairs of variables being so similar, as previously noted.

8.10.3 Comparison of results using the original data only and results using the imputed datasets

A comparison of the results obtained using the imputed datasets in Section 8.10.2 with those obtained using the original data only in Section 8.10.1 can be informative as to the effects of the implementation of the MI procedure as part of the analysis.

The models for BMI at examination in males, presented in Table 8.48 and Table 8.52, show differing effects of the MI analysis. In Model 1 both the values and significance levels of the BMI parameters remain similar, whilst the estimated coefficients for BMI velocity are uniformly increased, in many cases leading to greater levels of significance. In Model 2 many of the coefficients change value to greater or lesser degrees, but when a coefficient is non-significant under the original data analysis it may well be indicative of instability in its estimation, meaning that a relatively small change in value in the MI analysis should not be over-interpreted. Thus the main effect of interest is the attenuation of the BMI at AR coefficient at ages 4 to 6 years, meaning that there is no longer evidence of an association with this parameter in these models under the MI analysis.

The equivalent models for females in Table 8.49 and Table 8.53 show a noticeable attenuation in the BMI coefficients in Model 1 when analysing the imputed datasets. BMI velocity coefficients in model 1 are attenuated at younger ages whilst increased at older ages. In Model 2 the age at AR coefficients show an amplification at younger ages with BMI at AR coefficients noticeably attenuated at most ages leading to reduced statistical significance. In particular, this removes the

somewhat anomalous result of BMI at AR having a significant association with later BMI even when BMI and BMI velocity at age 8 years are known. Also of note is the attenuation of the BMI velocity coefficient at age 5 years, reducing evidence of an association between this and BMI at examination.

In Model 1 for %BF at examination in males (Table 8.54 and Table 8.50), the BMI coefficients are uniformly attenuated across the age range when considering the imputed datasets, resulting in reduced statistical significance, particularly at younger ages. The BMI velocity coefficients, on the other hand, are all increased, often considerably, leading to greater significance. Of note in the corresponding Model 2 are the increases in both value and significance of the coefficients for BMI at age 7 years and BMI velocity at age 8 years. There is a uniform attenuation of the age at AR coefficients across all ages, leading to this variable becoming non-significant in several models where it was previously significant, notably those at older ages where the initial result may not have been expected.

The %BF models for females detailed in Table 8.51 and Table 8.55 show, in Model 1, a uniform attenuation of the BMI coefficients for the multiply imputed data, whilst those for BMI velocity are attenuated at younger ages and increased at older ages. In Model 2 the BMI coefficients are attenuated somewhat at older ages, with a corresponding decrease in significance.

8.10.4 Conclusions

Results obtained using the original data only and using the imputed datasets are generally relatively similar. From the trends observed in the models for BMI at examination it is evident that for ages before the occurrence of the AR in most individuals the location of the AR is more strongly associated with adolescent BMI. When considering ages when the AR has already been passed in the majority the opposite is true, with the BMI and BMI velocity at that age taking greater significance than the dimensions of the AR. These patterns appear equally strong in males and females.

In the equivalent models for %BF at examination the results are similar though the associations somewhat less strong, especially amongst the males. This is, however, to be expected given that in this case the explanatory variables are not merely earlier measurements of the outcome variable.

When the age being considered is either a long time before or after the expected age at AR it seems logical that the whichever event is the more temporally proximal to adolescence has the stronger association with adolescent adiposity, due to the widely acknowledged high levels of BMI and adiposity tracking through childhood. When the age at which BMI and BMI velocity are estimated is close to the age at AR in the majority of individuals, however, this is not so obvious.

The fact that age and BMI at AR appear to be no better predictors of adolescent adiposity than BMI and BMI velocity at a similar age implies that there is little extra information contained within these dimensions. Considering more explicitly the information gained from the two different cases is informative as to why this may be happening.

In case 1, the age (say a_1) and BMI (say b_1) at AR are known. This is effectively the same as

knowing that at age a_1 BMI is b_1 and BMI velocity (say c_1) is 0 as, by definition, BMI velocity at AR must be 0. In case 2, the BMI (say b_2) and BMI velocity (say c_2) for a given age (say a_2) are known. Thus at age a_2 it is known that the BMI is b_2 and the BMI velocity is c_2.

It can be seen that in each case the three elements of information are the same. In both cases knowledge of the BMI (b_1 or b_2) at that age (a_1 or a_2) may be considered loosely equivalent, and the BMI velocities (c_1 or c_2) are also related. Whilst in case 1 it is known that AR occurs at age a_1, much of this information is available in case 2 as if c_2 is negative then it must be the case that a_2 is prior to the AR. Similarly, if c_2 is positive then a_2 is later than the AR. Also, the closer c_2 is to 0 the closer a_2 is to the age at AR.

By examining the information available in each case it is perhaps no surprise that the AR is found to be no better a predictor of later adiposity that BMI and BMI velocity at a similar age. Whilst the logic followed here is perhaps restricted to cases where BMI neatly decreases to reach a single minimum value before immediately increasing once more, as this type of BMI trajectory is highly prevalent the implications are more widely applicable.

Thus, whilst the age and BMI at AR have been shown to be associated with adolescent adiposity in Section 8.9, it would appear that this relationship is more statistical than physiological. Perhaps the AR is therefore not 'a developmental stage in which *physiologic* alterations increase the later prevalence of obesity' [74], making its labelling as a critical period somewhat debatable. As a result of this, concentration of interventions to prevent obesity at or around the period of the AR are likely to be no more beneficial than similar interventions at other periods in childhood.

The question of the AR as a critical period has also been addressed by Cole [91] using an argument based on BMI centiles (BMI relative to others of the same sex and age) and centile crossing. Cole asserts that BMI centile and the rate of BMI centile crossing determines the age at AR for an individual. As a high BMI centile and/or upwards centile crossing around the period of AR are associated both with an early AR and with later high adiposity, early AR is often observed as a risk factor for later high adiposity. As these associations apply at all ages, not just at AR, it is posited by Cole that AR cannot be considered as a critical period for later adiposity.

When BMI and BMI velocity are considered at a given age and for each sex separately, as in the present analysis, there is a close relationship to BMI centile and rate of centile crossing. Whilst obviously on a different scale, the relative positions for BMI between individuals of the same age and sex will be the same as their relative BMI centile positions. Thus if one individual has a greater BMI than another they will also have a higher BMI centile. Similarly, for a given BMI (at a given age and for a given sex) the relative positions for BMI velocity will be the same as the relative rates of BMI centile crossing. Thus if one individual has a greater BMI velocity than another they will also exhibit the greater upwards (or lesser downwards) centile crossing.

In this way, when included as explanatory variables in a regression model, the effect of BMI and BMI velocity would be expected to be similar to that of BMI centile and the rate of centile crossing. Therefore, it is no surprise that the obtained results closely resemble what would be

predicted by the line of argument of Cole.

These conclusions must be considered in light of the methodologies utilised in the analyses and the constraints of the data itself. However, as these issues are common to all analyses undertaken, this discussion is developed in a separate section.

8.11 Discussion

Whilst the analyses undertaken in this chapter each have their own conclusions, there are many features which are common to all of them, including missing data, the MI procedure, spline-fitting and issues surrounding the data themselves, which are discussed here.

8.11.1 Diagrammatic overview of the analysis

The complex multi-stage nature of the present analysis means that subjects can be lost at a variety of different points. Some are lost before any analysis begins (see Section 8.2), some because they have insufficient data points for subject specific splines to be fitted (see Section 8.6) and some because AR cannot be estimated from their fitted spline, (see Section 8.7). This attrition is not always easy to follow, especially as it differs between analysis using the original data only and analysis using the imputed datasets. Fig. 8.21 summarises this information diagrammatically in an attempt to aid understanding of the various stages in the analysis. It is identical to Fig. 8.1 in Section 8.3.3 but with the addition of the number of subjects which are lost or retained at each stage.

8.11.2 Missing data

This application of MI is novel because imputation does not result in every individual within the dataset contributing to the final analyses, as would often be the case elsewhere. This is because of the two stage process in action here — whilst the imputation of childhood BMI values (stage 1) does mean that in each imputed dataset every child will, effectively, have all 10 BMI values present, the subsequent spline-fitting (stage 2) does not guarantee that every child will have an estimated AR, allowing them to contribute to any subsequent analyses. Thus it is important that at both these stages the role of missing data is examined. Any discussion of missing data is complicated further by the 95 subjects out of the initial 481 in the SWEDES dataset that are excluded from the analysis at an early stage (see Section 8.2) for having no observed childhood BMI values whatsoever. These different levels of missing data and the implications of each are discussed here in more detail in the order in which they occur in the analyses.

Excluded subjects

Of the 481 subjects in the SWEDES dataset 95 have no observed BMI values (i.e. no concurrent observed height and weight values) whatsoever between age 1 and 10 years. When using the

Fig. 8.21: Diagrammatic overview of the analysis.

original data only these individuals would clearly have no fitted-splines and thus no estimated AR, so would contribute nothing to any analyses. In each imputed dataset these subjects would necessitate their entire BMI trajectory to be imputed. As they would only have a small number of observed variables, and in particular no observed BMI values between age 1 and 10 years, to contribute information to the imputation model to do so would likely lead to unreliably imputed BMI trajectories. As a result these 95 subjects are excluded from the analysis at the very beginning, leaving 386 eligible subjects.

For this exclusion not to bias any results obtained on the reduced dataset, it is imperative that the excluded subset of individuals are no more than a random sample from the initial SWEDES dataset — or, to use the language of Rubin (see Section 5.2.1), that they are 'missing completely at random' (MCAR). Whilst it cannot be fully demonstrated that data are MCAR, a crude way to assess this is to compare the distributions of other more fully observed variables in the excluded subjects with those in the remaining 386 subjects. This is done in Section 8.2 for a variety of variables at birth and at examination and the conclusion reached that the distributions of virtually all variables are very similar in the two subgroups. Whilst it is not possible to fully demonstrate the missing data mechanism, these observations are indicative of the excluded subjects being MCAR, meaning that their exclusion should not bias the results obtained using the remaining 386 individuals in the dataset.

Unobserved childhood BMI

As detailed in Section 8.4, between ages 1 and 6 years around 25–30% of BMI values are missing from the dataset, with this figure reduced to 5–10% for ages 7 to 10 years. Approximately 60% are subjects are seen to have all 10 BMI value observed, with 75% having at least 6 non-missing values.

Analysis using the original data only In terms of the potential introduction of bias into the analysis it is not the missingness of the BMI values *themselves* which is of importance, but rather the missingness of the estimated AR locations (and the estimated BMI and BMI velocities in Section 8.10) which are derived from the fitted splines. However, as splines are only fitted to those individuals with at least 6 observed BMI values, and as the likelihood of a successfully fitted spline (and thus successfully derived growth features) is increased with a higher proportion of observed values, these two issues are clearly intertwined. Whilst differences between subgroups of the subjects with differing proportions of observed BMI values are investigated in Section 8.4, the *actual* missingness of the derived explanatory variables is examined here.

Analysis using the imputed datasets Clearly when conducting analysis using the 100 imputed datasets analogous concerns over unobserved childhood BMI do not exist as in each imputed dataset all 10 BMI values will be present, either because they are observed in the original dataset or because they are 'filled in' during the imputation procedure. Whilst this means that splines can be fitted to every individual, the above issues regarding the potential introduction of bias into

the analysis as a result of the nature of the missingness of the derived growth features are still of concern. Again, these are discussed below.

The missingness of the childhood BMI values is, however, of significance when considered in relation to the MI procedure itself. As seen in Section 5.2.4, one of the key assumptions underlying the validity of MI is that the data to be imputed must be 'missing at random' (MAR) [120]. This means that given the observed data the probability of an unobserved BMI value being missing *cannot* be dependent on the unobserved BMI value itself or any other unobserved covariates. Whilst it is conceivably plausible that, for example, those individuals with greater BMI at a given age try to avoid having their measurements taken, it is not possible to directly test this. It should be recalled, however, that the missing BMI values for many of those individuals with few observed values correspond exactly to those years covered by their health care centre journals (i.e. before age 7 years) (see Section 8.4). That *all* data from these journals are missing suggests that linkage to the journals was not possible. If missingness is due to an administrative issue then it is unlikely to be related to the BMI values themselves meaning that data in these cases are MAR.

Nevertheless it remains important to investigate the nature of the missingness more thoroughly, for example by examining the distributions of more fully observed variables between those subjects with observed BMI and those with unobserved BMI at each age between 1 and 10 years. In Section 8.4 this is looked at somewhat more crudely by separating subjects into categories dependent on their proportion of observed BMI measurements through childhood, rather than examining missingness at each age in turn. In females all of the variables are seen to be similarly distributed regardless of the number of BMI values observed, whereas in males some trends are observed. At birth, those with 5 or fewer observed values appear to be heavier than those with 6 or more and at examination this same group still have, on average, greater weight, and also greater BMI, waist and hip circumferences, and %BF. These differences are evidence that subjects with higher proportions of missing childhood BMI values may not be merely a random sample of the dataset as a whole, implying that the missingness is not MCAR. It is not possible, however, to distinguish whether the missing data are MAR or NMAR.

One recommended approach in order to make the MAR assumption more plausible is to make the number of predictors in the imputation model as large as possible [126], and this advice is followed in the present application (see Section 8.3.2).

No estimated adiposity rebound

Whether using the original data only, with missing childhood BMI values for many subjects, or the imputed datasets, with every childhood BMI value present, the spline fitting procedure does not guarantee that the estimated location of the AR can be obtained. Where this is not possible, these individuals do not contribute to any subsequent analyses. For the logistic and linear regression models used in the analyses to provide unbiased results it is necessary for these non-contributing subjects to be MCAR, although the potential extent of any bias is reduced as the proportion of non-contributors decreases. As the estimated BMI and BMI velocity at given ages in Section 8.10

are only calculated in those individuals from whom the spline fitting resulted in a successfully identified AR, these derived explanatory variables are subject to similar missingness.

Analysis using the original data only When using the original data only it is possible to estimate AR location in 261 (68%) of subjects, meaning that a sizeable proportion of individuals are unable to contribute to the analyses which follow. In Section 8.7.1 individuals with an identified AR are seen to differ in some respects from those with AR not identified (Table 8.10). In particular, males with an identified AR appear to have lower weight at birth and lower weight, BMI, waist and hip circumferences, and %BF at examination. Females display a similar difference for weight at birth, though those with an identified AR appear similar to those without in terms of the measurements at examination. These differences, particularly among the males, may indicate underlying differences in the two groups of subjects. It is then implicit that those individuals in whom AR location estimation is not possible, and thus non-contributors to the analyses, are not MCAR but are potentially, as their missingness appears to be related to some of the observed variables, MAR. If the missing data mechanism is indeed MAR then this would invalidate the MCAR assumption necessary for the logistic regression to provide unbiased estimates when based on complete subset.

The comparison of subjects with no identified AR, either due to having insufficient childhood BMI data to have a spline fitted or the AR not being identifiable from the fitted spline, with those with an identified AR is used as a crude assessment of whether those subjects who are excluded from the analysis can be considered as MCAR. A related issue is whether those subjects with sufficient childhood BMI data to have a fitted spline yet no identified AR differ from those with an identified AR. In particular, when comparing these two subgroups in terms of late-adolescent adiposity this is really assessing whether the unidentifiability of the AR can be itself considered as a risk factor for later obesity. Indeed, as has been previously discussed, the reason for the AR not being identifiable in some individuals is because their BMI trajectory continues to increase throughout childhood, which may be thought likely to result in higher adiposity.

This can be assessed by fitting linear regression models for the measures of late-adolescent adiposity on an indicator variable which signifies whether or not the AR can be identified. However, the data must be restricted to the subset for whom an AR could potentially have been identified (i.e. those with at least 6 childhood BMI observations). As both BMI and %BF at examination are age-dependent and are not measured at the same age in every subject, age at examination is included in each regression model to adjust for any potential confounding.

Table 8.56 details the fitted models. As there is no evidence of an interaction with sex when considering either BMI (P=0.91) or %BF (P=0.18) at examination, the models are fitted for both males and females together. Clearly there is no evidence of either increased expected late-adolescent BMI or increased expected late-adolescent %BF as a result of the unidentifiability of the AR.

Thus, whilst those subjects who are excluded from the analysis due to having no identified AR may appear to differ in terms of late-adolescent adiposity from those with an identified AR, the

Outcome	Explanatory variable	n	Coeff.	95% CI	P-value
BMI at exam. (kg/m^2)	Unidentified AR	293	−0.10	−1.12, 0.91	0.85
	Sex				
	Female vs. male		0.65	0.00, 1.29	0.05
%BF at examination	Unidentified AR	293	−0.17	−2.43, 2.08	0.88
	Sex				
	Female vs. male		14.23	12.79, 15.68	<0.001

Table 8.56: Estimated coefficients (coeff.), 95% confidence intervals (CI) and Wald test P-values for the linear regression models of body mass index (BMI) and percentage body fat (%BF) at examination fitted on identifiability of the adiposity rebound (AR) using the original data. Models are adjusted for age at examination.

same is not true when considered conditionally on having sufficient data for a spline to be fitted.

Analysis using the imputed datasets The use of MI allows estimation of the derived explanatory variables for a greater proportion of individuals, though the exact figure differs between 331 (86%) and 351 (91%) depending on the imputed dataset. Obtaining estimated AR for individuals for whom this is not possible using the original data only, and thus allowing them to contribute to the analysis, has several implications. Changes in the constituent members of the sample under analysis may affect both the regression coefficient estimates and the viability of the MCAR assumption underlying both linear and logistic regression. The increased proportion of subjects who are able to contribute to the analyses mean that, should the MCAR assumption be similarly violated in both cases, the bias in the results obtained using MI should be less. Finally, the increased sample size should increase the precision with which the parameters in the analysis models can be estimated (within each imputed dataset at least).

In each analysis there is generally attenuation in the parameter estimates when using the imputed datasets compared to when using the original data only. This is possibly suggestive of differences in the relative characteristics of the subsets of subjects who contribute to the analysis in each case. This is investigated further in Section 8.7.2 where the distributions of a variety of variables are compared between those with a successfully identified AR and those without a successfully identified AR in the imputed datasets (Table 8.13). The differences between these two subgroups of individuals are generally reduced from those seen when considering the original data (Table 8.10), especially for the key outcome variables of BMI and %BF in males. However, these figures for the imputed datasets should be viewed with some caution due to the small sample sizes involved. The greater similarities between the two subgroups is suggestive of those who are not contributing to the analyses being MCAR, meaning that bias is potentially reduced in the analysis using MI.

The greater precision achieved using MI is clear from the narrower confidence intervals for coefficient estimates generally observed. As these overall measures of precision include between-

as well as within-imputation variability, precision within each imputed dataset must certainly be increased.

As with the analyses using the original data, it is interesting to consider whether the unidentifiability of the AR is itself a risk factor for later obesity. This can again be assessed by fitting regression models for BMI and %BF at examination on an indicator variable for AR identification. Now, however, as every subject has a fitted spline, and thus could potentially have an identified AR, there is no need to place a restriction on the data used.

The resultant fitted models are presented in Table 8.57. It is not possible to accurately assess the interactions between AR being unidentified and sex, but as there is no evidence of these when using the original data only they are not included here either. Again, there is no evidence whatsoever of a relationship between AR identifiability and later adiposity. In fact, the similarity between the results using the original data (Table 8.56) and using the imputed datasets are quite remarkable, especially given that the former involves only a subset of the data used in the latter.

Outcome	Explanatory variable	n	Coeff.	95% CI	P-value
BMI at exam. (kg/m^2)	Unidentified AR	386	−0.07	−1.19, 1.05	0.90
	Sex				
	Female vs. male		0.48	−0.11, 1.07	0.11
%BF at examination	Unidentified AR	386	−0.18	−2.62, 2.26	0.88
	Sex				
	Female vs. male		13.55	12.21, 14.88	<0.001

Table 8.57: Estimated coefficients (coeff.), 95% confidence intervals (CI) and Wald test P-values for the linear regression models of body mass index (BMI) and percentage body fat (%BF) at examination fitted on identifiability of the adiposity rebound (AR) using the 100 imputed datasets. Models are adjusted for age at examination.

There is some variability in the number of subjects who have a successfully identified AR in each of the 100 imputed datasets. From Table 8.11 in Section 8.7.2 it can be seen that this figure varies between 86 and 91%. Thus there are clearly some individuals with a successfully identified AR in some, but not all, of the imputed datasets. Table 8.14 shows that the majority of subjects either have a successfully identified AR in none (6% of the total) or all (61% of the total) of the imputed datasets. However, this does mean that there is still a significant proportion (33%) of individuals who contribute in only some of the imputations (although it should be noted that two-thirds of this remaining 33% contribute in at least 81% of the imputed datasets). The reasons behind this and the ensuing implications should be considered.

As the only element of the analysis which changes between the imputed datasets is the imputed values themselves, it must be variability in the imputed values which causes the AR to be identifiable in some imputed datasets, but not in others. It should be remembered that some individuals have large proportions of missing childhood BMI data, meaning that many values are imputed. When this is the case, even with high quality imputations, there will be considerable variability

in the final BMI trajectories resulting from the imputed values. Thus it is no surprise that the identifiability of the AR also varies across imputations.

As subjects with no identified AR do not contribute to analyses with one or both dimensions of the AR as explanatory variables, individuals with identified ARs in some, but not all, of the imputed datasets make a down-weighted contribution to the final result when compared to those with a successfully identified AR in each and every imputation. If the proportion of imputed datasets in which an individual has a successfully identified AR can be thought to correspond to the probability of them actually having an AR given the observed data (prior to imputation), then the fact that they also contribute to the final results with the same probability appears reasonable.

Conclusions

These observations suggest that the analyses using the original data may be more susceptible to bias, so that the slightly attenuated coefficients often found when analysing the imputed datasets may be closer to the true relationships. Thus, if it is believed that the imputation model preserves every aspect of the structure of the data, it could be suggested that the coefficients found when analysing the imputed datasets should be the preferred values.

However, as has already been discussed, interactions involving either or both dimensions of the AR cannot be included in the imputation model, meaning that these interactions then cannot be reliably assessed in the analysis models. This issue is a direct result of the multi-stage nature of the analysis and would not occur in a simpler implementation of MI. For example, if the analysis models only included explanatory variables which were themselves in the original dataset then any interactions between these variables could be included in the imputation model, making the interaction terms in the analysis models fully assessable. This is a clear disadvantage of the MI approach in this application. Indeed, it should also be considered that there may be further associations which are not fully captured by the imputation model.

So, whilst the use of MI is likely to reduce bias by increasing the proportion of subjects who can contribute to analyses, there may also be problems due to the introduction of bias through (often unavoidable) deficiencies in the imputation model. As a result, is it perhaps wise to present results from both approaches. As the two sets of results generally differ relatively little, this does not seem like an unreasonable solution.

It should also be remembered that even though using MI does increase the proportion of individuals in the dataset contributing to the analyses, there are still subjects who do not. Thus the same considerations regarding underlying differences between those with successfully derived explanatory variables (and thus contributing to the analysis) and those without (and thus not) must be borne in mind.

Fig 8.22 summarises the above details regarding the proportion of subjects who contribute to the analyses when using either the original data only or the imputed datasets. The denominator used in the calculation of the percentages is the 386 subjects in the SWEDES dataset with at least one childhood BMI measurement. It can be seen that when using the original data 67.6% of these

subjects contribute to the analyses. Under MI only 61.4% of individuals contribute within all 100 of the imputed datasets, though 94.0% contribute at least once.

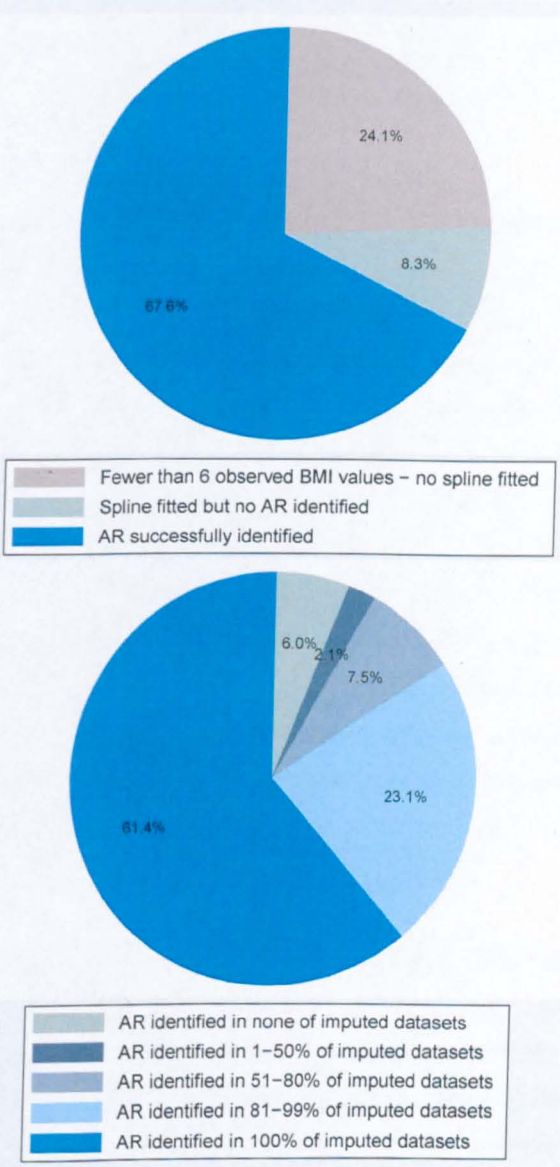

Fig. 8.22: Proportions of subjects contributing to the analysis when using either the original data only or the imputed datasets. Values are percentages of the 386 subjects in the Stockholm Weight Development Study dataset with at least one childhood body mass index measurement. BMI is body mass index and AR is adiposity rebound.

8.11.3 Spline fitting

Data requirements As detailed in Section 8.3.1, for a spline to be fitted for an individual they are required to have at least 6 BMI measurements between age 1 and 10 years with at least 2 of these being at age 6 years or younger and at least 2 being at age 6 years or older. Whilst these stipulations do not affect the spline-fitting procedure when using the imputed datasets as each individual will have all 10 BMI values present, when using the original data only the effective sample size is reduced. Relaxation of these requirements would mean that splines are fitted to

more subjects, but as the number of data points required for a spline to be fitted is reduced so is the likelihood of the resulting spline allowing reliable estimation of the location of the AR. Thus there is a trade-off which is largely subjective. As experimentation using the data suggests that the reliability of the estimated ARs would be potentially compromised by even a slight moderation of the data requirements, for example by requiring only 5 data points, it is perhaps unwise to do so (results not shown).

Selection of the smoothing parameters Selection of the smoothing parameters is informed by the use of a stratified random subsample of individuals taken from subgroups with different numbers of observed BMI values (i.e. 6, 7,...). Each subject is fitted with splines using different smoothing parameters then an overall strategy devised for deciding upon the smoothing parameter to use for any given individual. In the present application a rule is created so that the EDF of the spline is a function of the number of observed BMI values for that individual.

The resulting strategy allows the subject-specific splines to be fit quickly and easily as it eliminates the need for the smoothing parameters to be decided on a subject-by-subject basis. Whilst the fitted subject-specific splines obtained when using this general strategy all appear to be good fits to the data, without manipulation of the smoothing parameters on an individual level there remains the possibility that an improved fit could be achieved. Were time not a constraint then this would be an improved approach to the spline-fitting, but when dealing with large datasets, as is effectively the case when using the imputed datasets, this is not an option. Given the generally well-fitting curves obtained, especially when considering their intended use (i.e. reliable estimation of the AR), the strategy used does seem to be a good compromise.

Estimation of the adiposity rebound location Although all 10 BMI values between age 1 and 10 years are used in the spline-fitting procedure (when present), the estimated AR is only searched for between age 2 and 9 years. Whilst there could potentially be ARs outside of this range which remain unidentified due to this constraint being imposed, the number of individuals is likely to be negligible as the range of ages encompasses those over which the AR has generally been identified in previous studies. An attempt to locate ARs for ages outside of this interval could be made, but as the extremes of the range over which the spline is fitted are approached the estimated ARs will become less reliable. As for the data requirements above, there is a subjective trade-off to be considered.

8.11.4 Multiple imputation

Imputed variables As described in Section 8.3.2, the missing childhood BMI values affecting many individuals, as well as the %BF at examination values which are missing in only a handful of subjects, are imputed. All height and weight variables throughout childhood as well as many variables measured at examination are included in the imputation model. Further maternal variables relating to body size and socioeconomic status are also included. Although additional variables

are available in the SWEDES dataset they do not meet the criteria for inclusion in the imputation model, namely that they are potentially related to either the variable of interest itself or its pattern of missingness. Thus, although further variables could be included in the imputation model they would be expected to have little effect on the imputed values. Additionally, as discussed previously, whilst there is interest in interactions involving one or both dimensions of the AR it is not possible to include these in the imputation model.

Imputation specifications Markov chain Monte Carlo (MCMC) is used to generate 100 imputed datasets. Whilst this number of imputations is more than is widely advised as necessary the extra time and effort to create and maintain so many imputed datasets is minimal. With the large number of variables with missing data and the occasionally high proportions of missing data for a given variable encountered in the current application there seems little point having too few imputations and risking the adverse effects this could have on the results.

A single chain is used for all imputations with 200 initial burn-in iterations before the first imputation and 100 iterations between each subsequent imputation. Whilst these specifications could be modified, as the corresponding time-series and autocorrelation plots of parameters from iterations provide evidence of appropriate convergence of the MCMC process this would not appear to be necessary.

8.11.5 The Stockholm Weight Development Study

The SWEDES dataset provides a healthy contemporary birth cohort in which to investigate the relationships between AR and adolescent adiposity. There are, however, several issues and constraints associated with the dataset which require some discussion.

Data quality The standard of data collection in the SWEDES is generally very high, particularly so for the examinations when subjects were approximately 17 years old. As all the data were collected prospectively there is decreased risk of recall bias or unreliable measurements [78]. Whilst missing anthropometric data in a study of this kind is largely expected, the apparent problems with linking to health care centre journals for some subjects, as seen in Section 8.4, are somewhat unfortunate.

Sample size The already relatively small sample size of the SWEDES dataset (481 subjects) is reduced further by the exclusion of individuals with no observed BMI values between age 1 and 10 years. Whilst the small sample size may affect the precision of the estimated relationships, the power afforded by it is still sufficient to identify several important associations. However, replication of these analyses on larger datasets would be insightful as to the robustness of these associations.

Representativeness and generalisability It is important to examine whether these results are generalisable beyond the members of the dataset. This can be considered on two levels: the repre-

sentativeness of the members of SWEDES within the Swedish population and the generalisability from a Swedish dataset to subjects outside of Sweden.

Subjects were drawn from a population-based sample of the offspring of women who gave birth in 1984 or 1985 in Stockholm in a manner which has been seen to reasonably representative of the population in the Stockholm area [96]. It has been previously reported [94] that the prevalence of obesity at examination in the dataset is similar to that reported in Swedish adolescents and young adults generally but that BMI in is slightly lower in the males and higher in the females than in the Swedish reference datasets. However, the minimal differences were adjudged to indicate that body composition in the dataset is fairly representative of adolescent Swedes.

The conclusions reached here using data from Sweden, a developed European country, are likely to be able to be extrapolated relatively safely to similar populations. From the beginning of the 1980s to 2005 the percentage of obese people Sweden doubled from 5% to 10%, with prevalence increasing most among young women, non-manual workers and those who live outside of urban areas [176]. These are similar trends to those seen in many European countries, although the prevalence of obesity is not as high as that estimated in the UK [177].

The ages at which examinations occur (mean age 16.8±0.4 years) are sufficiently late so that BMI and %BF are approaching their stable adult levels [60, 175]. This means that, although the results obtaining in the analysis are concerned with measures of adiposity in late adolescence, the degree of extrapolation required to extend the conclusions through to adulthood is not particularly great.

Part III

General approaches

As stated before, this thesis focuses on relating childhood growth, in the form of repeated observations of an anthropometric variable for each child, to a later health outcome. An important distinction with regards to the analytical approaches which may be utilised in this scenario is between *balanced* and *unbalanced* childhood growth data. In Chapter 9 modelling strategies for use with unbalanced childhood growth data are explored, developed and implemented.

Unbalanced growth data, as defined in Section 5.1.2.1, are data which occur when there is no intention to observe the anthropometric variable at a common set of ages for each subject. When data are unbalanced many of the approaches detailed in Part II for use with balanced growth data cannot be used.

With unbalanced growth data, data are not 'missing' in the same sense as with balanced data as at no given time point for any individual are data 'expected'. Thus none of the approaches for handling missing data which can be used with balanced data are appropriate for use with unbalanced data. However, lack of data over a given time period for an individual is still problematic in unbalanced data, so methods for dealing with this are still required. This issue is referred to as data 'sparsity' rather than missingness.

It is not possible to use a single-stage analysis approach, for example a linear regression of a later health outcome on a childhood growth variable observed at several ages, with unbalanced data as this requires common ages at which the growth variable is observed. One solution to this is to interpolate between the observed measurements and estimate values at common times points so that the single-stage analysis methods can still be used. This involves fitting a *growth model* to the data, the simplest of which (linear interpolation) is effectively a piecewise linear model. However, this is clearly biologically implausible. Thus further more realistic growth models should be considered.

These concerns lead logically to the formulation of a two-stage analysis approach, as introduced in Section 5.4. In the first stage growth data for each individual is modelled. From the fitted models for each individual growth parameters can be derived. Whilst these parameters could include an estimate of the variable at a given age, as outlined above, others, such as growth velocity or acceleration at a given age, or the age at which maximum or minimum growth velocity or acceleration occurs, may also be of interest. These growth parameters can then be related to a later health outcome using similar methods to those when pursuing a single-stage analysis approach.

In Chapter 9, the unbalanced BMI growth data in the Uppsala Family Study (UFS) are modelled using penalised regression splines with random coefficients in a mixed model framework. From the fitted models, estimates of the location of the adiposity peak (AP) in infancy are derived for each subject. These derived growth features are then related to later BMI z-score using mixed models to take into account the structure of the dataset.

Whilst unbalanced data mean that many approaches for balanced data cannot be used, any approaches which are appropriate for unbalanced data can also be used for balanced data. Thus

the approaches described in this part of the thesis are 'general approaches' rather than 'approaches for unbalanced data'.

Chapter 9

Examining the relationship between the adiposity peak during infancy and later obesity in the Uppsala Family Study

9.1 Introduction

Whilst the adiposity rebound (AR) has been shown both here (see Chapter 8) and elsewhere (see Section 2.3.3) to be associated with later adiposity levels, other features of the BMI growth curve have been less well examined in this context. The AR, as a turning point, is a readily identifiable part of the typical BMI growth curve. So, however, is the BMI maximum usually reached between approximately age 6 months and 1 year, here referred to as the 'adiposity peak' (AP) during infancy. Fig. 9.1 shows a typical BMI growth curve with both the AR and AP identified. Unlike the AR, little research has been conducted into possible relationships between the timing of the AP and later adiposity.

Whilst the AP is here defined in terms of the BMI curve, as the AR generally is, a similar peak is also present during infancy for other measures of adiposity. For example, both triceps and subscapular skinfold thicknesses are seen to increase after birth before peaking, generally between age 6 months and 1 year [178].

The aim of the present analysis is to investigate the relationship between the timing of and BMI at the AP and BMI z-score in later childhood and adolescence in the Uppsala Family Study (UFS), described in Section 4.2. The analysis can be considered as a two stage process. First, infant BMI data are used to construct subject-specific BMI growth curves from which the AP can be identified. Then assessment is made of the relationship between later BMI z-score, calculated from BMI measured at physical examinations when the subjects were between 5 and 13 years old,

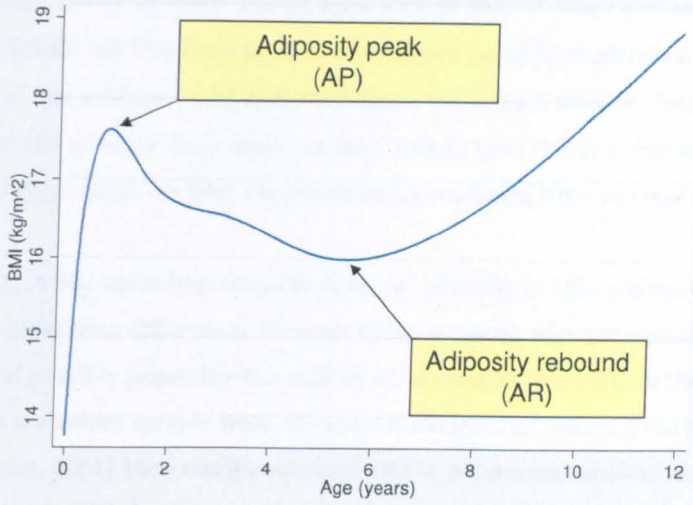

Fig. 9.1: A typical body mass index (BMI) growth curve through childhood with the adiposity peak (AP) and adiposity rebound (AR) identified.

and the derived AP locations (in terms of both age and BMI at AP). The subject-specific BMI growth curves are fitted using penalised regression splines with random coefficients in a mixed model framework. Multilevel modelling techniques are used to relate later BMI z-score to age and BMI at AP in order to incorporate the familial structure of the dataset.

9.2 Subjects

A general introduction to the Uppsala Family Study is provided in Section 4.2. The most relevant details for the present analysis are that the dataset includes 602 pairs of siblings from Uppsala, Sweden, born between 1987 and 1995. Only siblings both in the top or bottom quarter of the birthweight distribution ('concordant high birthweight' (CHB) or 'concordant low birthweight' (CLB)) or with a sex-adjusted difference in birthweight of 0.4 kg or more ('discordant birthweight' (DB)) were included. Children's postnatal growth data, including serial measurements of height and weight, were obtained from health records, kept by Child Health Centres and schools. All children had a physical examination between May 2000 and November 2001 when they were aged 5–13 years, at which several measurements, including height and weight, were recorded. From these BMI and age- and sex-adjusted BMI z-scores are calculated.

Preliminary exploratory analyses (not included here) estimate the AP to occur at an age of between 6 months and 1 year in the majority of individuals. To ensure that the AP is identified for as many subjects as possible in the UFS, a rather broad definition of the AP as 'the (main) BMI maximum between birth and age 2 years' is employed here. Whilst BMI maxima beyond age 2 years do not qualify as the AP under this definition, data up to age 3 years are included so that estimation of parameters is not conducted too close to the boundaries of the interval over which the curve is fitted. BMI values at birth are, however, excluded as these are often thought to be unreliable.

Once BMI observations at birth and at ages greater than 3 years are excluded, 1164 of the initial 1204 individuals (96.7%) have at least one remaining BMI observation. Whilst the mixed model structure of the analysis could deal with those individuals with no data points whatsoever by assigning them the relevant fixed effects as their fitted curve this is somewhat unappealing. As a result, the 40 subjects with no BMI observations are excluded from the rest of the analysis.

It is important when excluding subjects from an analysis in this manner to investigate the existence of any underlying differences between those subjects who are excluded and those who remain which could possibly jeopardise the validity of any results obtained. If the excluded subjects are no more than a random sample from the overall dataset (or 'missing completely at random' (MCAR), see Section 5.2.1) then results obtained on the remaining subjects should be unbiased.

Table 9.1 and Table 9.2 help to assess this, the former by displaying the number and percentage of subjects in various subgroups who are included in the analysis and the latter by comparing the distributions of several continuous variables between those who are included in the analysis and those who are excluded. From Table 9.1 it can be seen that very similar percentages of males and females are excluded from the analysis, implying that missingness is not related to sex. The percentage of excluded subjects is also similar in older and younger siblings and in the three birthweight groups.

Variable	Level	Number (%) of subjects included
Sex	Male	598 (96.5%)
	Female	566 (96.9%)
Sibling type	Older	581 (96.5%)
	Younger	583 (96.8%)
Birthweight group	CLB	260 (94.9%)
	CHB	267 (97.5%)
	DB	637 (97.1%)

Table 9.1: Number and percentage (%) of subjects with at least one body mass index observation between birth and age 3 years. CLB is concordant low birthweight, CHB is concordant high birthweight and DB is discordant birthweight.

In Table 9.2 the distributions are presented separately for males and females as these largely anthropometric variables would not necessarily be expected to take similar values in the two sexes. It can be seen that both weight and length at birth differ little between included and excluded subjects for males and females. Age at physical examination appears somewhat older in excluded males. Whilst the use of BMI z-score rather than an age-dependent variable reduces the consequence of this with regards to assessment of the association of interest, the difference in ages may indicate that a certain subset of individuals is being lost from the analysis. Of more note are the

observed differences in BMI z-score at physical examination, being higher in excluded males and lower in excluded females. However, care must be taken not to over-interpret these results given the small numbers of excluded subjects in both sexes.

Males ($n = 620$)						
Variable	Included ($n = 598$)			Excluded ($n = 22$)		
	Mean	Median	SD	Mean	Median	SD
At birth						
Weight (kg)	3.74	3.73	0.60	3.74	3.41	0.72
Length (cm)	51.7	51	2.2	52.3	51.5	2.5
At physical examination						
Age (years)	10.0	10.1	1.7	11.5	11.8	1.5
BMI z-score	0.26	0.12	1.20	0.58	0.26	1.15

Females ($n = 584$)						
Variable	Included ($n = 566$)			Excluded ($n = 18$)		
	Mean	Median	SD	Mean	Median	SD
At birth						
Weight (kg)	3.65	3.71	0.56	3.55	3.38	0.54
Length (cm)	50.9	51	2.2	50.8	50	2.3
At physical examination						
Age (years)	10.1	10.2	1.8	10.4	10.6	1.2
BMI z-score	0.36	0.30	1.09	−0.11	−0.20	0.91

Table 9.2: Distributions of variables at birth and at physical examination for subjects with/without at least one body mass index (BMI) observation between birth and age 3 years, by sex.

The number of BMI observations for each subject varies greatly between the remaining 1164 individuals. Whilst one subject has only one BMI observation, 96% have at least 7 and one has as many as 30. The distribution of the number of BMI observations for each subject is shown in Fig. 9.2.

The distribution of ages at which these BMI observations occur is far from uniform between birth and age 3 years, as illustrated in Fig. 9.3. It can be seen that over 50% of the data point are for ages less than 6 months and that data are markedly more sparse for ages greater than approximately 1.5 years.

The outcome in the present analysis is the sex- and age-adjusted BMI z-score, calculated from the BMI measurement taken for each subject at their physical examination. Physical examinations

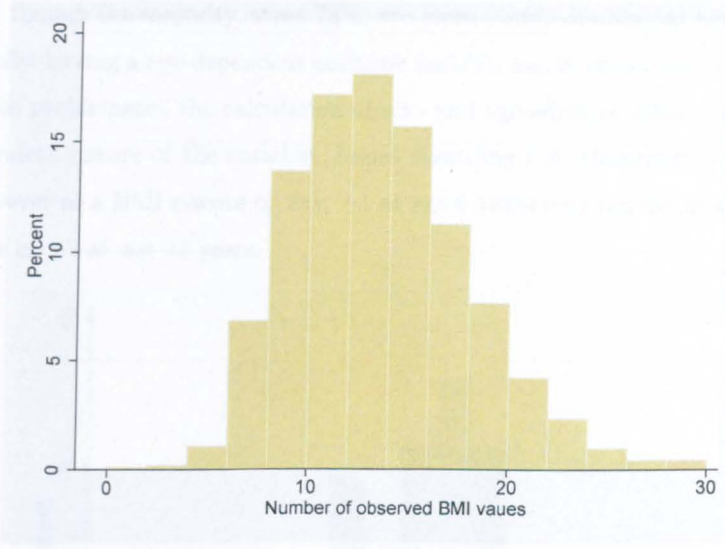

Fig. 9.2: Distribution of number of childhood body mass index (BMI) observations for the 1164 subjects with at least one body mass index observation. Total number of childhood body mass index observations is 15,296.

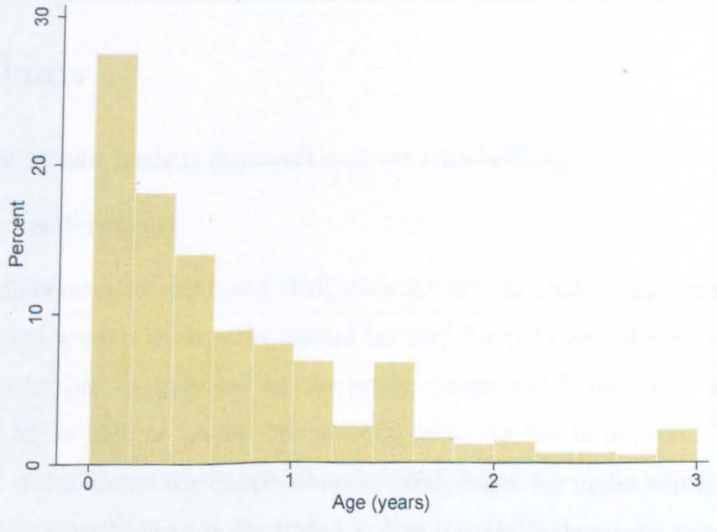

Fig. 9.3: Distribution of age at childhood body mass index observations for the 1164 subjects with at least one body mass index observation. Total number of childhood body mass index observations is 15,296.

were carried out over an 18 month period which, coupled with the fact that subjects are from sibling pairs, means that physical examinations were carried out across a wide range of ages, as shown in Fig. 9.4. Some individuals were as young as 5.5 years and some as old as 13.8 years at examination, though the majority, some 78%, are more evenly distributed between about 8.5 and 13 years. Whilst having a age-dependent outcome variable measured over such a wide range of ages would often be problematic, the calculation of sex- and age-adjusted BMI z-scores should remove the age-dependent nature of the variable. Issues regarding the interpretability of the variable do still exist however as a BMI z-score of, say, +1 at age 6 years may not be considered equivalent to a BMI z-score of +1 at age 13 years.

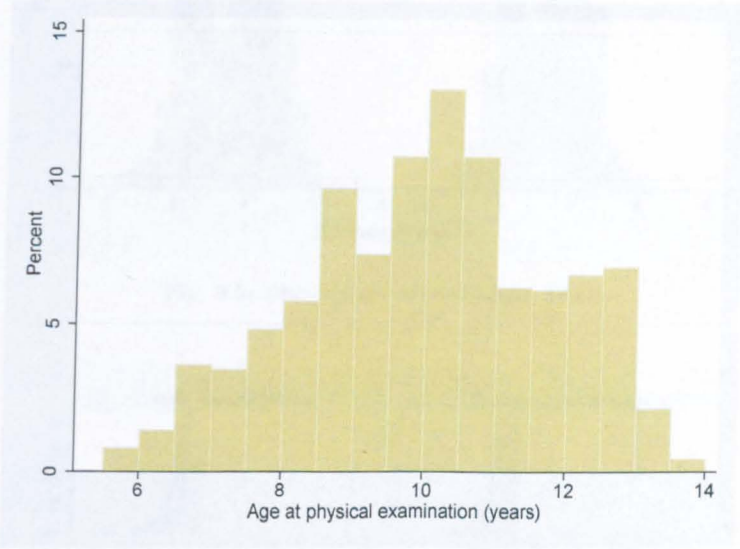

Fig. 9.4: Distribution of age at physical examination for the 1164 subjects with at least one body mass index observation.

9.3 Methods

9.3.1 Body mass index growth curve modelling

Preliminary considerations

Acknowledged differences in childhood BMI growth between males and females [68] mean that different underlying growth trajectories should be used for each sex. However, in the case of the UFS, analysis is further complicated by the study design which results in a preponderance of individuals with either high or low birthweight and relatively few in between. This can be clearly seen in Fig. 9.5, which shows the distributions of birthweight for males and females. The reason behind the bimodal distributions is illustrated in Fig. 9.6 which shows the same distributions but stratified by the birthweight group. As with sex, different growth patterns would be expected for individuals with different birthweights. As the subjects in this case form largely disparate groups for birthweight it may well be unwise to fit the same underlying growth trajectory for all individuals, even within the same sex. One approach to overcome this is to fit different underlying

growth trajectories for each birthweight group, so that a greater degree of homogeneity is achieved. Thus six different models are fitted, corresponding to CLB males, CHB males, DB males, CLB females, CHB females and DB females.

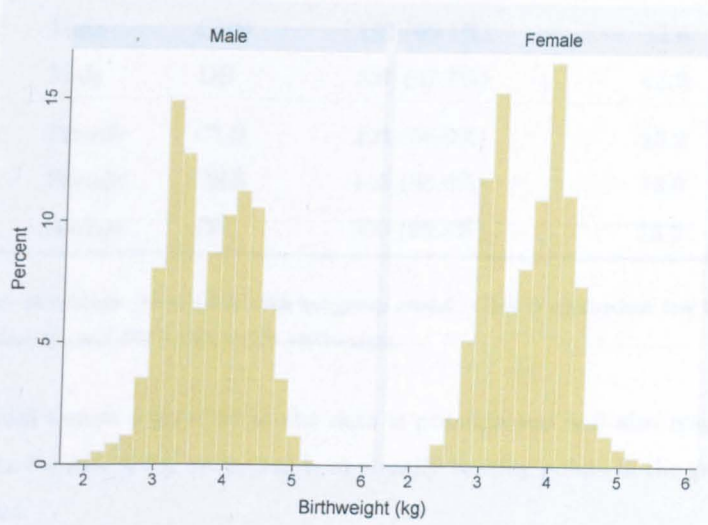

Fig. 9.5: Distributions of birthweight, by sex.

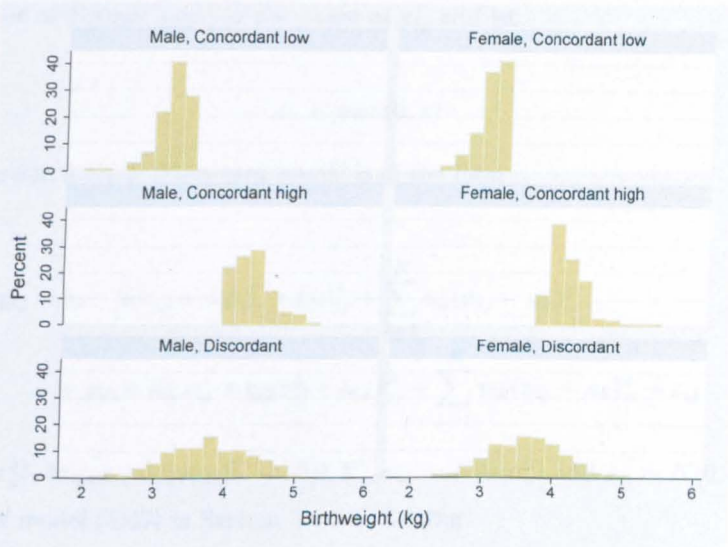

Fig. 9.6: Distributions of birthweight stratified by sibling group, by sex.

The subgroups of the dataset for which these models are fitted are summarised in Table 9.3. Within each subgroup at least 94% of the subjects have one or more BMI observations over the age range of interest and this equates to a mean of between 12 and 14 data points per individual.

Penalised regression spline models

Subject-specific BMI growth curves are fitted using penalised regression splines with random coefficients, as introduced in Section 5.4.1.5. Cubic penalised regression spline models, with both cubic fixed effects and cubic random effects, are used to model the BMI growth curves. The cubic nature

Subgroup model	Sex	Birthweight group	Number (%) of subjects included	Mean number of data points per subject
Model 1	Male	CLB	139 (93.9%)	13.9
Model 2	Male	CHB	122 (96.1%)	12.0
Model 3	Male	DB	337 (97.7%)	13.2
Model 4	Female	CLB	121 (96.0%)	13.2
Model 5	Female	CHB	145 (98.6%)	13.0
Model 6	Female	DB	300 (96.5%)	13.2

Table 9.3: Summary of subjects included in each subgroup model. CLB is concordant low birthweight, CHB is concordant high birthweight and DB is discordant birthweight.

of the models should ensure a good fit to the data is possible and will also result in continuous first derivatives. As the aim of the modelling is to identify turning points in the growth curves this second point is vital.

The models are fitted on log(BMI) rather than BMI itself to flatten the maxima and encourage a better fit. Let y_{ij} denote the log(BMI) of subject i, $i = 1, \ldots, m$, at age x_{ij}, $j = 1, \ldots, n_i$. Let $\kappa_1, \ldots, \kappa_K$ be a set of distinct knots in the range of x_{ij} and let

$$x_+ = \max(0, x)$$

as in (5.35) in Section 5.4.1.5. Then each model is of the form

$$\begin{aligned} y_{ij} =& \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \beta_3 x_{ij}^3 + \sum_{k=1}^{K} u_k (x_{ij} - \kappa_k)_+^3 \\ &+ a_{i0} + a_{i1} x_{ij} + a_{i2} x_{ij}^2 + a_{i3} x_{ij}^3 + \sum_{k=1}^{K} v_{ik} (x_{ij} - \kappa_k)_+^3 + \varepsilon_{ij} \end{aligned} \tag{9.1}$$

where $u_k \sim N(0, \sigma_u^2)$, $(a_{i0}, a_{i1}, a_{i2}, a_{i3})^T \sim N(0, \boldsymbol{\Sigma})$, $v_{ik} \sim N(0, \sigma_v^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, which is a simple extension of model (5.49) in Section 5.4.1.5. Letting

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_i \\ \vdots \\ \mathbf{y}_m \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix},$$

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{Z}_2 & \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_m & \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_m & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_m \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_K \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{pmatrix}$$

where

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} 1 & x_{i1} & x_{i1}^2 & x_{i1}^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{in_i} & x_{in_i}^2 & x_{in_i}^2 \end{pmatrix},$$

$$\mathbf{Z}_i = \begin{pmatrix} (x_{i1} - \kappa_1)_+^3 & \cdots & (x_{i1} - \kappa_K)_+^3 \\ \vdots & \ddots & \vdots \\ (x_{in_i} - \kappa_1)_+^3 & \cdots & (x_{in_i} - \kappa_K)_+^3 \end{pmatrix}, \quad \mathbf{a}_i = \begin{pmatrix} a_{i0} \\ \vdots \\ a_{i3} \end{pmatrix}, \quad \mathbf{v}_i = \begin{pmatrix} v_{i1} \\ \vdots \\ v_{iK} \end{pmatrix}$$

$$\text{and} \quad \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in_i} \end{pmatrix}$$

with

$$\mathbf{G} = \text{Cov}(\mathbf{u}) = \begin{pmatrix} \sigma_u^2 \mathbf{I} & 0 & 0 \\ 0 & (\text{blockdiagonal } \boldsymbol{\Sigma})_{1 \le i \le m} & 0 \\ 0 & 0 & \sigma_v^2 \mathbf{I} \end{pmatrix},$$

the model can be written in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \tag{9.2}$$

which is the general linear mixed model representation given in (5.24).

Knot selection

A simple method for choosing K, the number of knots, which usually works well in scatterplot smoothing [134] is

$$K = \min\left(\frac{1}{4} \times \text{number of unique } x_{ij},\ 35\right). \tag{9.3}$$

In the present dataset the number of unique x_{ij} in each of the six models ranges from 532 to 756, all of which give $K = 35$. However, as subject-specific curves are required rather than just a smooth of all the data with no regard paid to the structure, this is likely to be a vast overestimate for K. If, instead, models were fitted for each subject using only the data available for that subject then each model would be a smooth of a mean of 13.1 data points (see Table 9.3). Using (9.3) again, this would give $K = 3$. Therefore a sensible choice for K would appear to be somewhere between 3 and 35. Thus, somewhat arbitrarily, K is fixed at 12 which should provide a sufficient level of flexibility for the curve, especially given the relatively high degree (cubic), whilst avoiding the computational complications that a large number of knots would entail. This is a similar number of knots to that used for the fitting of subject-specific penalised regression splines with random coefficients elsewhere [179]. Other values for K slightly greater than or less than 12 were also examined, but were found to make little difference to the fit of the spline models.

A simple approach to selecting the knot locations, $\kappa_1, \ldots, \kappa_K$, which has also been used elsewhere [134, 179] is

$$\kappa_k = \left(\frac{k}{K+1}\right) \text{th sample quantile of the unique } x_{ij}.$$

This approach is utilised in the present analysis, giving knots which lie on the $(\frac{1}{13})^{\text{th}}, \ldots, (\frac{12}{13})^{\text{th}}$ centiles of the unique x_{ij}.

Whilst the number of knots ($K = 12$) is the same in all six models, as the knots locations are defined by the ages at which childhood BMI is observed (which are not common amongst the six subgroups) the knot locations are allowed to differ between the models. However, the knot locations are seen to be very similar across all six models.

The data points and knot locations for the CLB males are plotted in Fig. 9.7. The knot locations are clearly much closer together in the regions of the plot where there is a greater density of data points.

The *cubic spline basis* with knot locations as defined above is

$$1,\ x,\ x^2,\ x^3, (x-\kappa_1)_+^3, \ldots,\ (x-\kappa_{12})_+^3 \tag{9.4}$$

and is plotted in Fig. 9.8 for the CLB males model. Every subject-specific curve can be obtained as a linear combination of these basis functions.

The knot locations and resulting bases for the other five models only differ through minor relocations of the knots so are omitted for brevity.

Population average curves

The population average curves in each model are formed from the elements which do not vary between individuals. Thus, with reference to (9.1), the population average curve for a given model can be seen to be

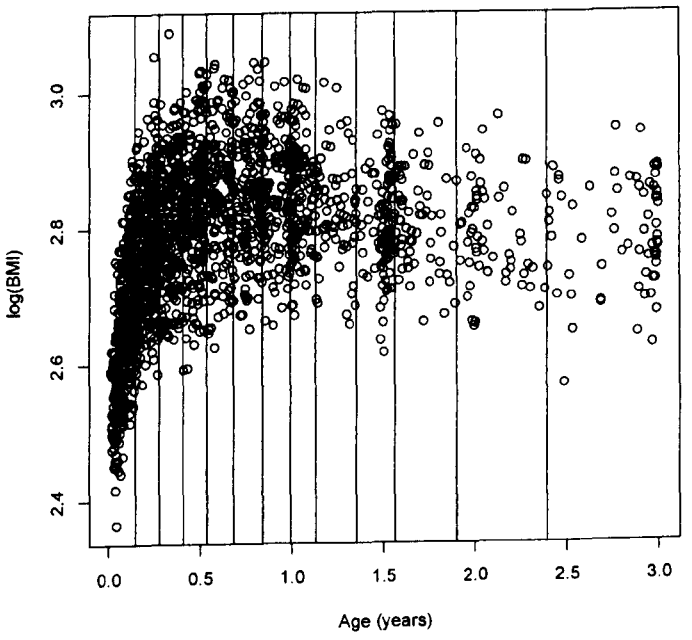

Fig. 9.7: Data points and knots locations for concordant low birthweight males.

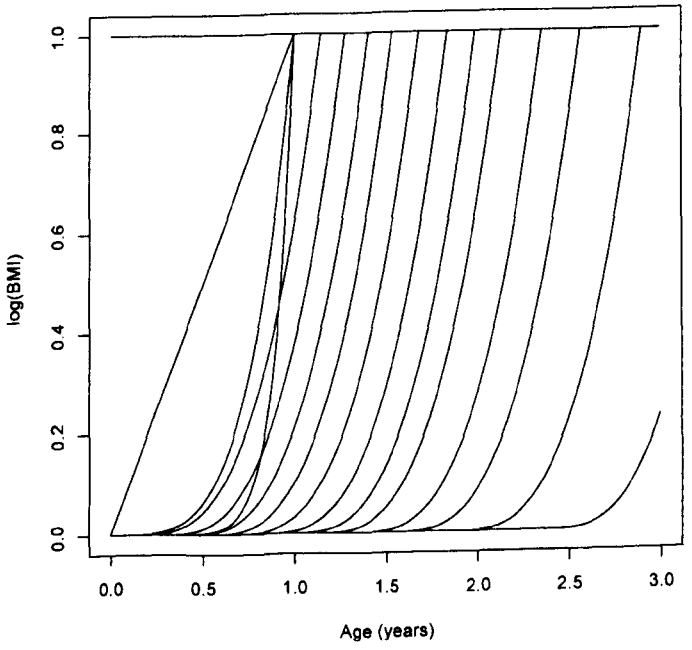

Fig. 9.8: Cubic spline basis for concordant low birthweight males.

$$\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{k=1}^{12} u_k (x - \kappa_k)_+^3 .$$

The population average curves for each model are plotted separately for males (Fig. 9.9) and females (Fig. 9.10) to aid clarity. The curves for males are as would perhaps be expected with the CLB subgroup having a lower trajectory right across the range of ages examined, the CHB subgroup having a higher trajectory and the DB subgroup (which is a complete mix of birthweights) being between the two. The observed trend in females is very similar, although in this case the trajectory of the DB subgroup much more closely mimics that of the CHB subgroup. It can be seen that for both males and females the ages at the maxima of the population average curves differ between the birthweight groups which gives some justification to the fitting of separate fixed effects for each subgroup.

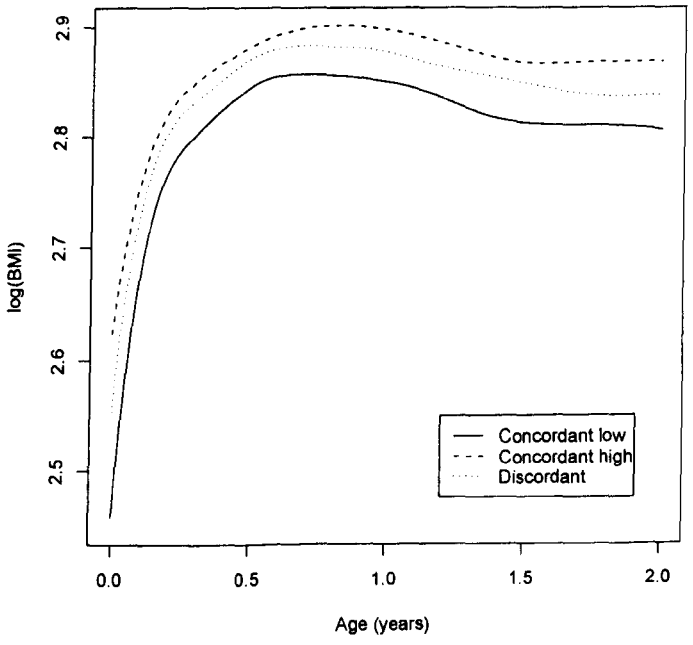

Fig. 9.9: Population average curves for males.

Subject-specific deviations

The subject-specific deviations from the population average curve in each model are the elements which vary between individuals in the model. For subject i they are given by

$$a_{i0} + a_{i1} x + a_{i2} x^2 + a_{i3} x^3 + \sum_{k=1}^{12} v_{ik} (x - \kappa_k)_+^3 .$$

The subject-specific deviations for the CLB males only are plotted in Fig. 9.11. The curves demonstrate a reasonable amount of between-subject variation, with a maximum deviation from the population average curve of around ± 0.2. The high levels of curvature in several of the subjects also justifies the inclusion of a cubic term to model the deviation from the population average curve.

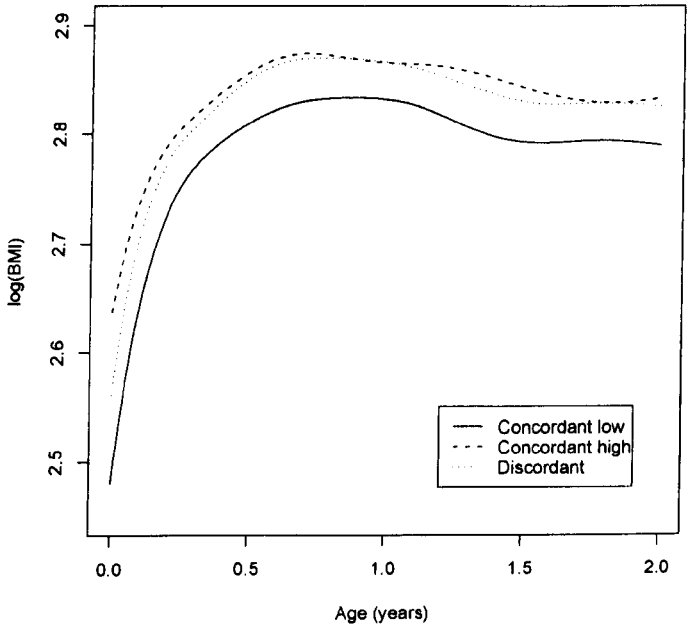

Fig. 9.10: Population average curves for females.

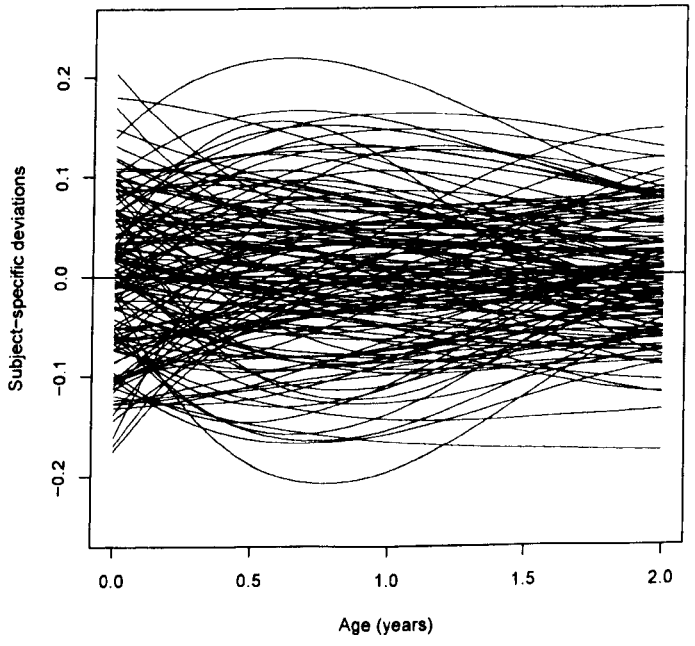

Fig. 9.11: Estimated subject-specific deviations for concordant low birthweight males.

Corresponding plots for the other sex and sibling group combinations are omitted here as they are largely similar.

Fitted growth curves

The combination of the population average curves and subject-specific deviations from them gives the overall fitted subject-specific log(BMI) curves. These are presented in Fig. 9.12 and Fig. 9.13 for several individuals selected in a stratified random manner. In each figure, the top row corresponds to the CLB model, the middle row to the CHB model and the bottom row to the DB model. Within each row, the left hand plot is for a randomly selected subject within the first quintile of number of observed childhood BMI values, the middle plot is for a randomly selected subject within the third quintile of number of observed childhood BMI values and the right hand plot is for a randomly selected subject within the fifth quintile of number of observed childhood BMI values. The collection of plots should then provide examples for each subgroup model when data are sparse and when data are plentiful. Population average curves (dashed lines) are also provided for reference. It can be seen that whilst the subject-specific curves all take the same general shape as the population average curves, the inclusion of the subject-specific deviations allow the subject-specific curves to, on the whole, provide excellent fits to the data. For individuals where data are more sparse, greater emphasis will be placed on the population average curve and in this way the fitted curves for these individuals will draw information from others. From just this small sample of individuals a variety of different subject-specific curve shapes are evident: the majority with obvious maxima, some with flatter sections and others which appear monotone increasing.

Residuals

The residual ε_{ij} is the difference between the fitted subject-specific curve and the observed data point for individual i at age x_{ij}. The residuals for the CLB males are plotted against age in Fig. 9.14. Whilst the residuals appear to have greater variability at younger ages this may be largely caused by the many more observations at these ages (see Fig 9.3), so given exactly the same variability more extreme values would be expected to be observed. However, as no subjects would be expected to have their AP within the first few months after birth, even if this is indicative of a slightly worse fit, the implications on the present analyses to follow are minimal. More encouragingly, there are no obvious systematic trends in this or the equivalent residual plots obtained for the other male and female subgroups.

Corresponding plots for the other sex and birthweight group combinations are again omitted due to their similarities with the plot shown.

Location of the adiposity peak

As the AP is a turning point in the BMI curve (and hence in the log(BMI) curve), one obvious approach to identifying the estimated location of the AP is by taking the first derivative (with

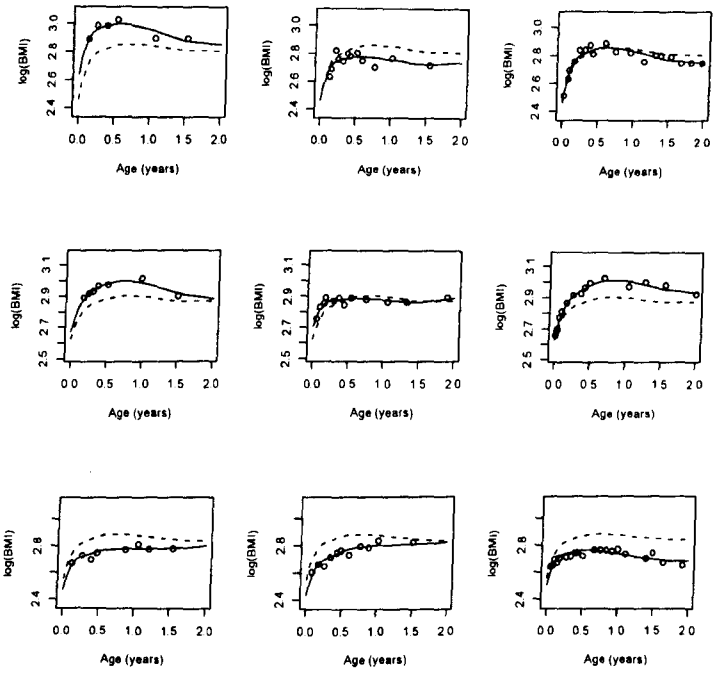

Fig. 9.12: Estimated population average curves (dashed lines) and fitted subject-specific curves (solid lines) for nine males. The top row corresponds to the concordant low birthweight model, the middle row to the concordant high birthweight model and the bottom row to the discordant birthweight model. Within each row the left hand plot is for a randomly selected subject within the first quintile of number of observed childhood body mass index (BMI) values, the middle plot is for a randomly selected subject within the third quintile of number of observed childhood body mass index values and the right hand plot is for a randomly selected subject within the fifth quintile of number of observed childhood body mass index values.

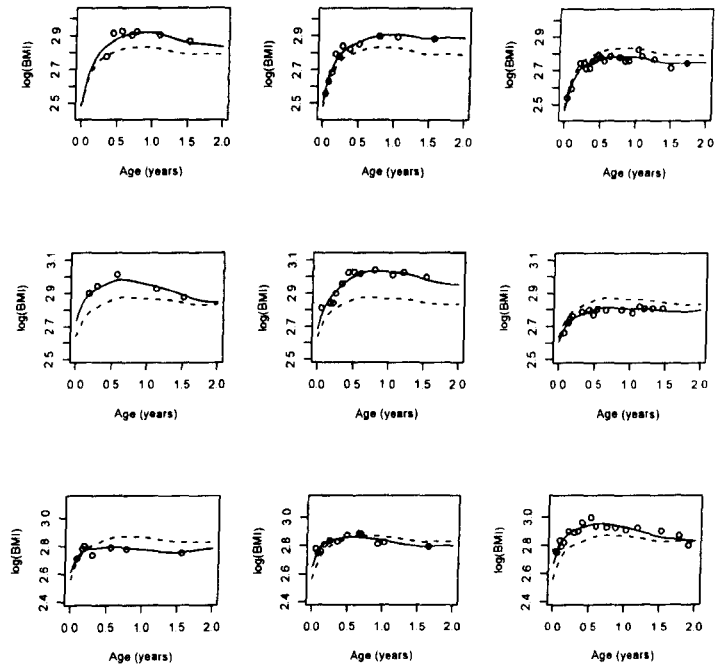

Fig. 9.13: Estimated population average curves (dashed lines) and fitted subject-specific curves (solid lines) for nine females. The top row corresponds to the concordant low birthweight model, the middle row to the concordant high birthweight model and the bottom row to the discordant birthweight model. Within each row the left hand plot is for a randomly selected subject within the first quintile of number of observed childhood body mass index (BMI) values, the middle plot is for a randomly selected subject within the third quintile of number of observed childhood body mass index values and the right hand plot is for a randomly selected subject within the fifth quintile of number of observed childhood body mass index values.

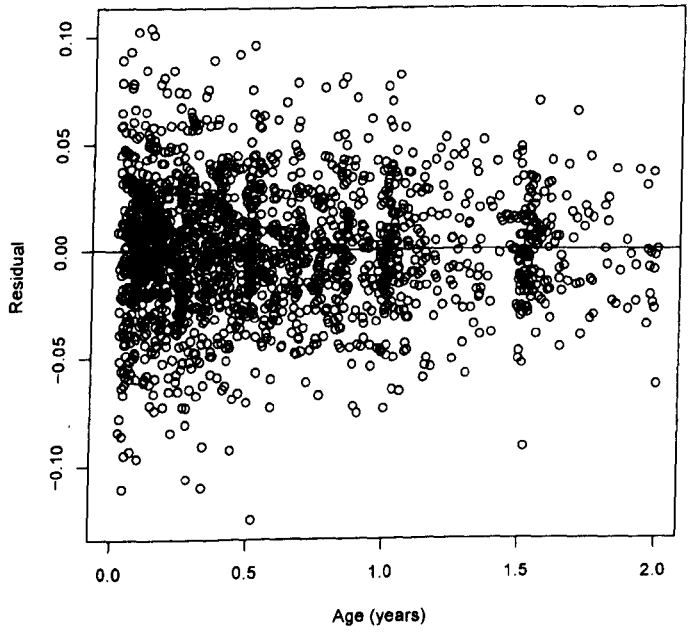

Fig. 9.14: Residuals for concordant low birthweight males.

respect to age) of the fitted curve and investigating at what point this crosses from positivity to negativity.

The first derivative of the fitted cubic penalised regression spline for an individual can be easily evaluated using the estimated fixed and random parameters. Let $\hat{y}_i$ be the fitted curve for subject i,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \hat{\beta}_3 x^3 + \sum_{k=1}^{12} \hat{u}_k (x - \kappa_k)_+^3 + \hat{a}_{i0} + \hat{a}_{i1} x + \hat{a}_{i2} x^2 + \hat{a}_{i3} x^3 + \sum_{k=1}^{12} \hat{v}_{ik} (x - \kappa_k)_+^3,$$

where $\kappa_1, \ldots, \kappa_{12}$ are the knot locations and $\hat{\beta}_0, \ldots, \hat{\beta}_3$, $\hat{u}_1, \ldots, \hat{u}_{12}$, $\hat{a}_{i0}, \ldots, \hat{a}_{i3}$ and $\hat{v}_{i1}, \ldots, \hat{v}_{i12}$ are estimates of the previously defined model parameters. Then $\hat{y}'_i$, the first derivative of the fitted curve with respect to age for subject i, is

$$\hat{y}'_i = \hat{\beta}_1 + 2\hat{\beta}_2 x + 3\hat{\beta}_3 x^2 + \sum_{k=1}^{12} 3\hat{u}_k (x - \kappa_k)_+^2 + \hat{a}_{i1} + 2\hat{a}_{i2} x + 3\hat{a}_{i3} x^2 + \sum_{k=1}^{12} 3\hat{v}_{ik} (x - \kappa_k)_+^2. \quad (9.5)$$

Evaluation of the first derivative can again be achieved using the general linear mixed model representation given in (9.2). To ensure that the first derivative is evaluated across the required range of ages it is perhaps preferable to use instead of the observed ages, x_{ij}, artificially assigned ages, x_l, $l = 1, \ldots, p$, occurring at regular intervals between birth and age 2 years, for example $(0.01, 0.02, \ldots, 2.00)$. As the ages at evaluation are common to all subjects the matrix notation now simplifies slightly so that for subject i

$$\hat{\mathbf{y}}'_i = \mathbf{X}' \hat{\boldsymbol{\beta}} + \mathbf{Z}' \hat{\mathbf{u}}_i \quad (9.6)$$

where

$$\hat{\mathbf{y}}'_i = \begin{pmatrix} \hat{y}'_{i1} \\ \vdots \\ \hat{y}'_{ip} \end{pmatrix}, \quad \mathbf{X}' = \begin{pmatrix} 0 & 1 & 2x_1 & 3x_1^2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 2x_p & 3x_p^2 \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix},$$

$$\mathbf{Z}' = \begin{pmatrix} 3(x_1 - \kappa_1)_+^2 & \cdots & 3(x_1 - \kappa_{12})_+^2 & 0 & 1 & 2x_1 & 3x_1^2 & 3(x_1 - \kappa_1)_+^2 & \cdots & 3(x_1 - \kappa_{12})_+^2 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 3(x_p - \kappa_1)_+^2 & \cdots & 3(x_p - \kappa_{12})_+^2 & 0 & 1 & 2x_p & 3x_p^2 & 3(x_p - \kappa_1)_+^2 & \cdots & 3(x_p - \kappa_{12})_+^2 \end{pmatrix}$$

and

$$\hat{\mathbf{u}}_i = \begin{pmatrix} \hat{u}_1 \\ \vdots \\ \hat{u}_{12} \\ \hat{a}_{i0} \\ \vdots \\ \hat{a}_{i3} \\ \hat{v}_{i1} \\ \vdots \\ \hat{v}_{i12} \end{pmatrix}.$$

Here $\hat{y}'_{il}$ denotes the first derivative of the fitted log(BMI) curve for subject i, $i = 1, \ldots, m$, at age x_l, $l = 1, \ldots, p$.

The evaluation of the first derivative of the fitted cubic penalised regression spline for a given individual, i, is used to identify the estimated location of the AP by finding an age x_q where $\hat{y}'_{iq} > 0$ but $\hat{y}'_{iq+1} < 0$. This signifies that there is a maximum in the fitted log(BMI) curve for subject i in the interval (x_q, x_{q+1}), so x_q is used as an estimate for the age at AP if $|\hat{y}'_{iq}| < |\hat{y}'_{iq+1}|$ and x_{q+1} is used otherwise. This value is then substituted into (9.1) and a corresponding estimate for BMI at AP obtained.

Whilst this simple approach to identifying the AP works well for most individuals, in some cases issues such as local non-AP maxima and multiple maxima mean that further criteria need to be included. Local non-AP maxima are of no interest in the present context so to avoid their detection a condition is included which states that for any 'true' AP the first derivative of the log(BMI) curve must be positive 3 months beforehand and negative 3 months afterwards. This is found to be an effective preventative measure, though brings with it the implication that no maxima can be found either prior to age 3 months or after age 1.75 years. However, as no subjects would be expected to have their AP outside of this range of ages then this should not cause any problems. This more stringent criterion for maxima coupled with the reduced range of ages over which the AP is sought also reduces the number of multiple maxima exhibited. In the negligible number of subjects where this is still an issue the problem is resolved by simply taking the first maximum to be the AP. The thinking behind this is that if two maxima exist between the ages of 3 months and 1.75 years then the first is far more likely to be within the expected range of ages at AP (age 6 months to 1 year) and thus more likely to be the true AP.

Fig. 9.15 illustrates the above described procedure for a randomly selected subject. The right hand plot shows the first derivative of the fixed effects (dashed line) and the first derivative of the subject-specific curve (solid line). The vertical line indicates the age at which the procedure locates the change from positivity to negativity of the first derivative of the subject-specific curve. The left hand plot shows the fixed effects of the fitted model (dashed line) and the fitted subject-specific curve (solid line). The vertical line passes through the age at which the change in sign of the

first derivative is detected and thus also through the maximum in the fitted subject-specific curve. The horizontal line passes through the value of the subject-specific curve which is calculated to correspond to this age.

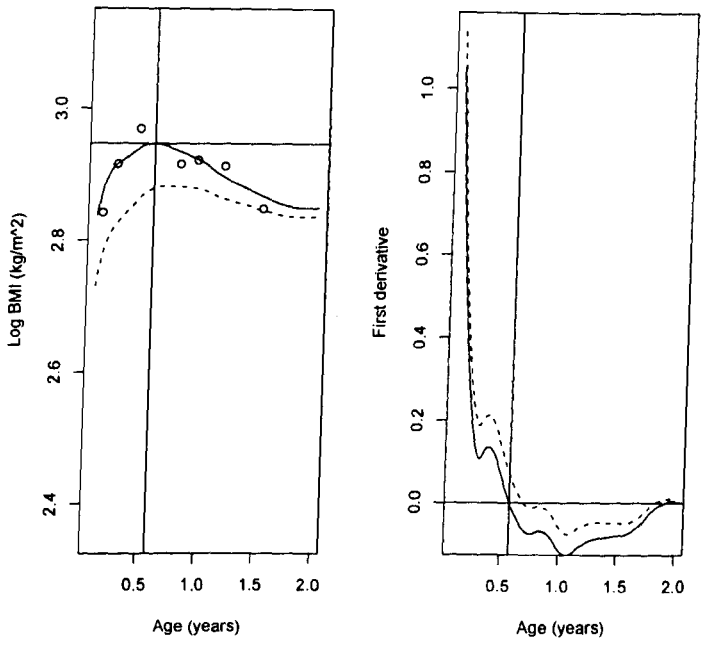

Fig. 9.15: Location of the adiposity peak for a randomly selected subject. Dashed lines represent the population average curve (left hand plot) or the first derivative of the population average curve (right hand plot) for the subgroup to which this subject belongs. Solid lines represent the fitted subject-specific curve (left hand plot) or the first derivative of the fitted subject-specific curve (right hand plot) for this individual. BMI is body mass index.

Similar plots are created for each subject in the dataset though, given the relatively large sample size, only a random sample of these can be visually checked.

Software

The mixed model representation of the penalised regression spline model, as shown in Section 5.4.1.5, means that model fitting can be easily implemented in standard statistical software. Thus the fitting of the BMI growth curve models is carried out using the `lme` procedure in R [155], which is a generic function for fitting linear mixed models.

9.3.2 Relating adiposity peak location to later body mass index z-score

Mixed model

In many situations an assessment of whether the AP is associated with later adiposity could be made by employing an ordinary least squares (OLS) regression of BMI z-score at physical examination on either age at AP, BMI at AP or both. OLS regression, however, assumes that each response is independent which, due to the inclusion of sibling pairs in the UFS, is unlikely to be

true in the present analysis. One approach to overcome this issue is to use multilevel (also known as random effect, hierarchical or mixed) modelling, as described in more detail in Section 5.3.3.

Taking as an example the model for BMI z-score at physical examination on age at AP, it seems reasonable that some families generally have higher BMI z-scores at physical examination than others, regardless of age at AP. This would necessitate the inclusion in the model of family-specific *random intercepts*. It could also be envisaged that in some families the relationship between age at AP and BMI z-score at physical examination differs to that in other families. For example, in one family the sibling with the later AP may have a greater BMI z-score than their sibling, whereas in another family the sibling with the later AP may have a lower BMI z-score. Incorporating this into the model requires family-specific *random slopes*.

Continuing with the same example, let BMI_{zij} and $\text{age}_{\text{AP}ij}$ be the BMI z-score at physical examination and age at AP for sibling $i = 1, 2$ in family $j = 1, \ldots, 602$. Let sex_{ij} be an indicator variable taking value 1 if the subject is female and 0 otherwise, and CLB_{ij} and CHB_{ij} be indicator variables taking value 1 if the birthweight group of the subject is, respectively, CLB or CHB and 0 otherwise, meaning that those who are DB are taken as the reference group. Then the random intercepts and slopes model can be written as

$$\text{BMI}_{zij} = \beta_{0j} + \beta_{1j}\text{age}_{\text{AP}ij} + \beta_2\text{sex}_{ij} + \beta_3\text{CLB}_{ij} + \beta_4\text{CHB}_{ij} + e_{ij} \tag{9.7}$$

where

$$\beta_{0j} = \beta_0 + u_{0j} \quad \text{and} \quad \beta_{1j} = \beta_0 + u_{1j},$$

with $(u_{0j}, u_{1j})^T \sim N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is an unstructured 2×2 covariance matrix, and $e_{ij} \sim N(0, \sigma_e^2)$. Here u_{0j} and $u_{0j'}$ are independent of each other for $j \neq j'$, u_{1j} and $u_{1j'}$ are independent of each other for $j \neq j'$, e_{ij} and $e_{i'j'}$ are independent of each other unless $i = i'$ *and* $j = j'$, and u_{0j} and u_{1j} are independent of e_{ij} for all j. However u_{0j} and u_{1j} may be correlated.

In (9.7) $\beta_0, \ldots, \beta_4$ are the *fixed effects* and the average relationship is given by $\beta_0 + \beta_1\text{age}_{\text{AP}} + \beta_2\text{sex} + \beta_3\text{CLB}_{ij} + \beta_4\text{CHB}_{ij}$. The *random intercepts*, $u_{0\cdot1}, \ldots, u_{0\cdot602}$, and *random slopes*, $u_{1\cdot1}, \ldots, u_{1\cdot602}$, correspond to family-specific differences from the average relationship. The *level 1 residuals*, e_{ij}, are the vertical distance between the observed BMI z-score at physical examination, $BMIz_{ij}$, and the corresponding fitted value, $\beta_{0j} + \beta_{1j}\text{age}_{\text{AP}ij} + \beta_2\text{sex}_{ij} + \beta_3\text{CLB}_{ij} + \beta_4\text{CHB}_{ij}$.

Further models for BMI z-score at physical examination on BMI at AP and for BMI z-score on both age and BMI at AP differ little from the above. The inclusion of additional covariates and interaction terms results in further fixed effects, but in each model the only random terms are the intercepts and slopes.

Software

The mixed models used to relate AP location to later BMI z-score are fitted using restricted maximum likelihood (REML) under the `xtmixed` procedure in Stata [147].

9.4 Results

9.4.1 Estimated age and body mass index at adiposity peak

Table 9.4 summarises the distributions of age and BMI at AP, along with the number and percentage of subjects with identified AP, by sex and birthweight group. The percentage of subjects with a successfully identified AP is generally high, although some differences between the birthweight groups are evident, with CLB subjects having the highest percentage of identified AP and CHB subjects the lowest in both males and females.

The AP appears to occur slightly later in CHB males and in CLB females than in the other birthweight groups, although the differences are not great so this should not be overinterpreted. Overall, both mean and median age at AP are slightly higher in females, a feature which is borne out by a simple t-test (ignoring the sibling pairs) providing a P-value of <0.001. The median age at AP is generally seen to be somewhat lower than the mean, suggesting a slightly skewed distribution.

Average BMI at AP is seen to be highest in CHB subjects and lowest in CLB subjects in both sexes, corresponding to the population average curves seen in Fig. 9.9 and Fig. 9.10. Generally, BMI at AP appears greater in males, which is again confirmed by a highly statistically significant (P<0.001) t-test. Mean and median are very similar in each group indicating a more symmetric distribution.

Sex	Birthweight group	Number (%) of subjects with identified AP	Age at AP (years)			BMI at AP (kg/m^2)		
			Mean	Median	SD	Mean	Median	SD
Males	CLB	126 (90.6%)	0.72	0.65	0.16	17.7	17.6	1.3
	CHB	102 (83.6%)	0.79	0.78	0.13	18.5	18.3	1.4
	DB	291 (86.4%)	0.72	0.67	0.17	18.1	18.1	1.3
	Total	519 (86.8%)	0.73	0.69	0.16	18.1	18.0	1.4
Females	CLB	118 (97.5%)	0.87	0.88	0.13	17.1	17.0	1.2
	CHB	121 (83.4%)	0.76	0.70	0.20	17.9	17.9	1.3
	DB	272 (90.7%)	0.79	0.75	0.17	17.9	17.7	1.2
	Total	511 (90.3%)	0.80	0.76	0.17	17.7	17.7	1.3

Table 9.4: Distributions of age and body mass index (BMI) at adiposity peak (AP), by sex and birthweight group. CLB is concordant low birthweight, CHB is concordant high birthweight and DB is discordant birthweight. Percentage of subjects with identified adiposity peak is calculated as a percentage of those included in each subgroup model (see Table 9.3).

Fig. 9.16 shows both the univariate and bivariate distributions of age and BMI at AP for males and females separately. Whilst BMI at AP appears to be normally distributed in both sexes, age at AP exhibits some positive skew. There is also, particularly amongst the males, some evidence of

bimodality. Plotting BMI at AP against age at AP provides little evidence of correlation between the two variables. This is reflected in the calculated correlation coefficients shown in Table 9.5, using log-transformed age at AP due to the skew, of 0.12 for males and 0.05 for females overall. Stratification by birthweight group, however, shows some degree of heterogeneity between the correlation coefficients, especially amongst the males, with CHB subjects showing a higher degree of correlation. Although associations are generally weak, these results do suggest that older ages at AP are more likely to correspond to a higher BMI at AP.

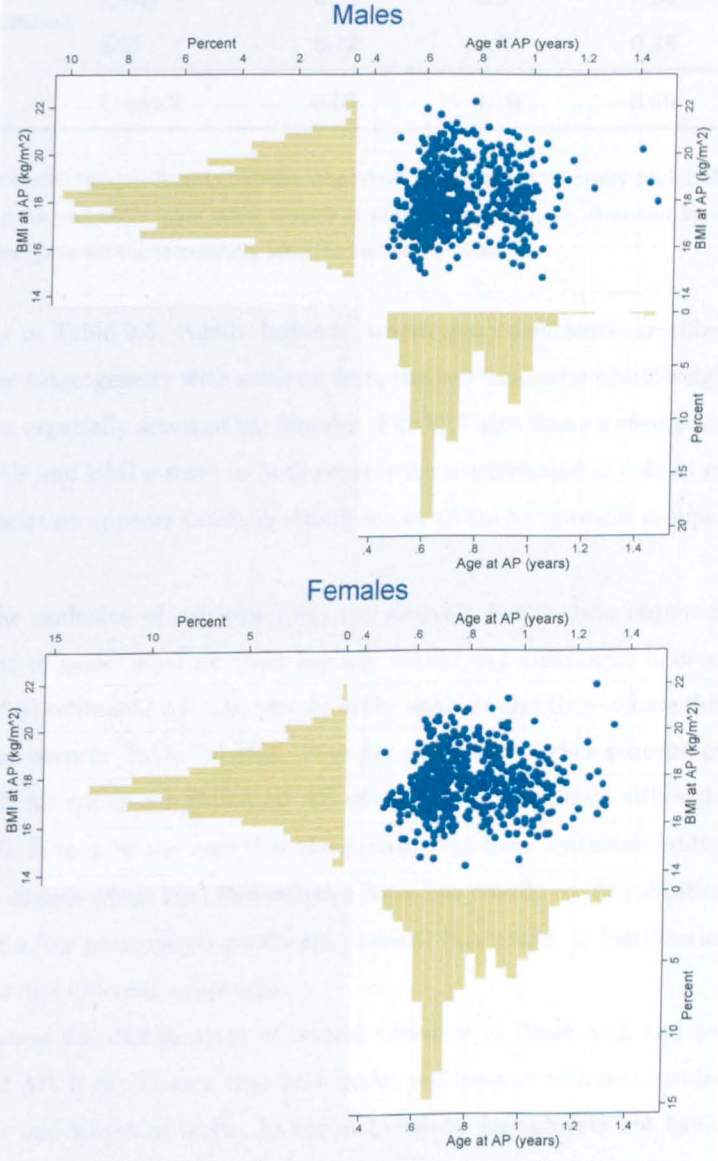

Fig. 9.16: Univariate and bivariate distributions of age and body mass index (BMI) at adiposity peak (AP), by sex, for the 1030 subjects with a successfully identified adiposity peak.

Fig. 9.17 shows scatterplots of BMI z-score at physical examination against age and BMI at AP for males and females separately. There appears to be little correlation between age at AP and later BMI z-score, with calculated correlation coefficients of 0.05 and 0.10 for males and fe-

Sex	Birthweight group	Age at AP BMI at AP	Age at AP BMI z-score	BMI at AP BMI z-score
Males	CLB	−0.02	0.08	0.45
	CHB	0.28	0.08	0.41
	DB	0.09	−0.02	0.40
	Overall	0.12	0.05	0.43
Females	CLB	0.06	0.23	0.34
	CHB	0.16	0.21	0.34
	DB	0.12	0.09	0.38
	Overall	0.05	0.10	0.39

Table 9.5: Pairwise correlation coefficients between (log transformed) age at adiposity peak (AP), body mass index (BMI) at adiposity peak and body mass index z-score at physical examination, stratified by sex and birthweight group, for the 1030 subjects with a successfully identified adiposity peak.

males respectively in Table 9.5. Again, however, correlation coefficients stratified by birthweight group exhibit some heterogeneity with subjects from the two concordant birthweight groups having greater correlation, especially amongst the females. Fig. 9.17 also shows a clear positive relationship between BMI at AP and BMI z-score in both sexes, with a correlation of 0.43 in males and 0.39 in females. This association appears similarly strong across all the birthweight groups within each sex.

Similarly to the exclusion of subjects from the analysis due to data requirements in Section 9.2, it is important to assess whether there are any underlying differences between subjects with successfully identified estimated AP who remain in the analysis and those where this is not possible. It has already been seen in Table 9.4 that there are somewhat higher percentages of males and subjects with CHB for whom an estimated AP could not be identified, although the differences are relatively small. It may be the case that these subgroups have marginally different underlying BMI growth curve shapes which lend themselves a little less readily to identification of the AP, for example by having a less pronounced maximum. Indeed, this would be justification for the use of separate models for the different subgroups.

Table 9.6 compares the distributions of several variables in those with and those without an identified estimated AP. It can be seen that both males and females with no identified AP generally have greater weight and length at birth. As the main reason for subjects not having an identified AP is that their BMI observations continue to increase over the first two years of life, this may indicate that this type of growth trajectory is more prevalent in those who are larger at birth. The age at physical examination, on the other hand, appears similarly distributed in those with and without identified AP, although males with no identified AP generally have a higher BMI z-score. This makes sense when coupled with the above observation that males with no identified AP are also larger at birth as tracking through childhood dictates that subjects who are larger at birth

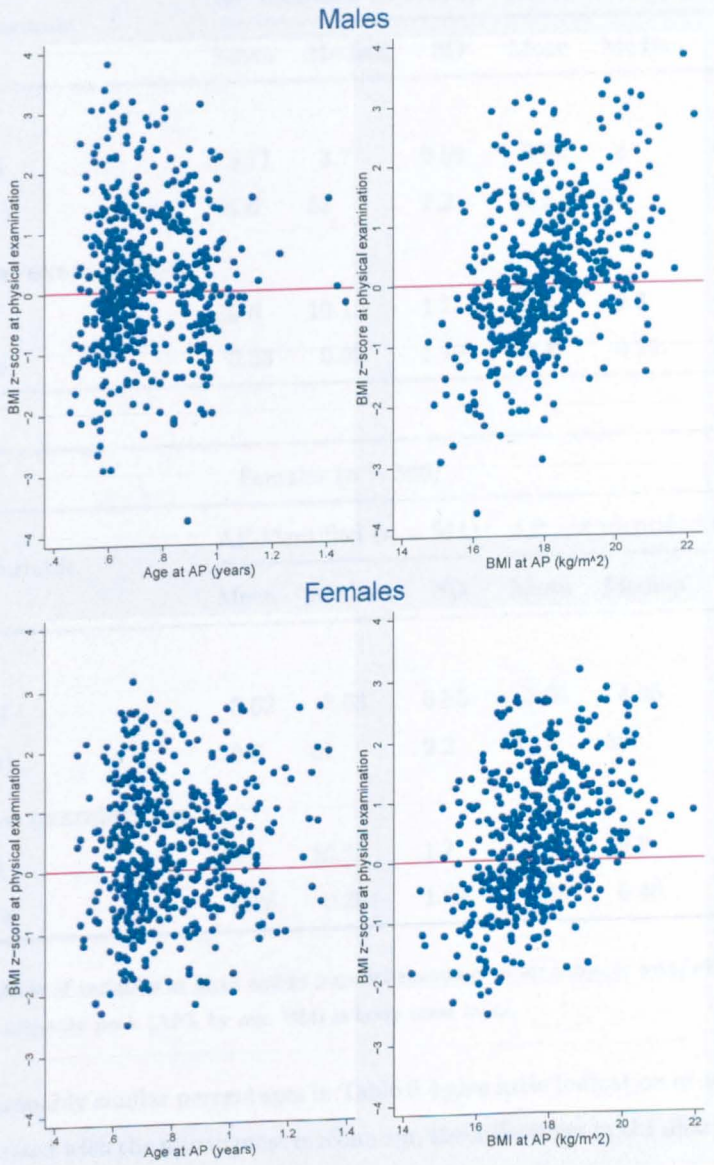

Fig. 9.17: Body mass index (BMI) z-score at examination against age and body mass index at adiposity peak (AP), by sex, for the 1030 subjects with a successfully identified adiposity peak.

are also likely to be larger at later ages. Also, as noted above, these subjects are more likely to have BMI observations which continue to increase through the first two years of life, which may well be likely to lead to higher later BMI than a trajectory which shows a marked decrease over this period.

Males ($n = 598$)						
Variable	AP identified ($n = 519$)			AP not identified ($n = 79$)		
	Mean	Median	SD	Mean	Median	SD
At birth						
Weight (kg)	3.71	3.7	0.59	3.90	4	0.63
Length (cm)	51.6	51	2.2	52.2	52	2.0
At physical examination						
Age (years)	10.0	10.1	1.7	9.9	9.8	1.7
BMI z-score	0.23	0.09	1.19	0.49	0.25	1.22

Females ($n = 566$)						
Variable	AP identified ($n = 511$)			AP not identified ($n = 55$)		
	Mean	Median	SD	Mean	Median	SD
At birth						
Weight (kg)	3.62	3.63	0.55	3.95	4.05	0.56
Length (cm)	50.8	51	2.2	51.6	52	2.1
At physical examination						
Age (years)	10.1	10.2	1.7	9.8	9.9	1.9
BMI z-score	0.36	0.30	1.09	0.38	0.40	1.09

Table 9.6: Distributions of variables at birth and at physical examination for subjects with/without a successfully identified estimated adiposity peak (AP), by sex. BMI is body mass index.

Whilst the reasonably similar percentages in Table 9.4 give little indication of sex or birthweight group being associated with the missingness mechanism, the differences in the distributions in Table 9.6, particularly at birth, are of more concern. These suggest that the subjects are possibly not MCAR, meaning that any results obtained are not necessarily extrapolatable to the dataset in general.

To prevent false conclusions being drawn, it could thus be claimed necessary to include the caveat that results are 'conditional on an AP being identifiable'. Hence it may be prudent to more formally investigate whether the unidentifiability of the AP is itself a 'risk factor' for higher adiposity in later childhood. If, as has been observed, many of those without a successfully identified AP have a BMI trajectory which continues to increase through infancy, perhaps this subgroup

would be expected to have relatively higher BMI at later ages.

This can be assessed by fitting a mixed model similar to those for evaluating the extent of the association between the AP and later adiposity, as described in Section 9.3.2. Now the exposure of interest is not one or both dimensions of the AP, but whether or not the AP is identified at all. For sibling i, $i = 1, 2$, in family j, $j = 1, \ldots, 602$, let BMI_{zij} be the BMI z-score at physical examination, sex_{ij} be an indicator variable for sex, and CLB_{ij} and CHB_{ij} be indicator variables for CLB and CHB, as in (9.7). Now let $\text{AP}_{\text{UI}ij}$ be an indicator variable taking value 1 if the AP cannot be successfully identified ('unidentified' (UI)) and 0 otherwise. Then a suitable random intercepts and slopes model can be expressed by

$$\text{BMI}_{zij} = \beta_{0j} + \beta_1 \text{AP}_{\text{UI}ij} + \beta_2 \text{sex}_{ij} + \beta_3 \text{CLB}_{ij} + \beta_4 \text{CHB}_{ij} + e_{ij} \tag{9.8}$$

where

$$\beta_{0j} = \beta_0 + u_{0j} \quad \text{and} \quad \beta_{1j} = \beta_0 + u_{1j},$$

with $(u_{0j}, u_{1j})^T \sim N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is an unstructured 2×2 covariance matrix, and $e_{ij} \sim N(0, \sigma_e^2)$. The dependencies and independencies between the parameters remain as detailed in Section 9.3.2.

Table 9.7 details the estimated fixed effects when (9.8) is fitted using REML. There is no evidence of an interaction between either sex (P=0.22) or birthweight group (P=0.50 for CLB subjects and P=0.65 for CHB subjects) and the identifiability of the AP, so the model includes both sexes and all three birthweight groups.

Explanatory variable	Coefficient	95% CI	P-value
Unidentified AP	0.11	−0.09, 0.31	0.28
Sex			
Female vs. male	0.08	−0.04, 0.20	0.18
Birthweight group			
CLB vs. DB	−0.24	−0.44, −0.05	0.01
CHB vs. DB	0.32	0.13, 0.51	0.001

Table 9.7: Estimated fixed effects, 95% confidence intervals (CI) and Wald test P-values for the random intercepts and slopes model of body mass index z-score at physical examination fitted on identifiability of the adiposity peak (AP), adjusted for sex and birthweight group. Model is fitted on all 1164 subjects. CLB is concordant low birthweight, CHB is concordant high birthweight and DB is discordant birthweight.

It can be seen from Table 9.7 that for a given sex and birthweight group, whilst there is a slightly greater expected BMI z-score at physical examination in those subjects with no identified AP, this relationship is far from statistically significant. Thus it would appear that unidentifiability of the AP does not lead to an increased propensity for higher BMI in later childhood. This indicates that, conditional on the observed covariates (sex and birthweight group), there is no relationship between data missingness (whether or not the AP can be identified) and the outcome.

9.4.2 Are dimensions of the adiposity peak associated with later adiposity?

Mixed models of the form (9.7) are used to relate the age and/or BMI at AP to BMI z-score at physical examination. All models are fitted using REML but use of ML was found to make negligible difference to the fitted models (results not shown).

Table 9.8 and Table 9.9 detail the estimated fixed effects for the random intercepts models of BMI z-score at physical examination fitted separately on age and BMI at AP. In neither model is there much evidence of an interaction between sex and the dimension of the AP (P=0.07 in the model for age at AP and P=0.36 in the model for BMI at AP), thus in both cases combined-sex adjusted models are used.

From Table 9.8 it can be seen that for a given sex and birthweight group a delayed age at AP is estimated to be associated with a positive and statistically significant increases in BMI z-score at examination. Conditional on age at AP and birthweight group there is no estimated difference in BMI z-score at examination between males and females, whilst for a given age at AP and sex CLB subjects are estimated to have a reduced BMI z-score at examination and CHB subjects an increased BMI z-score when compared to DB subjects.

Explanatory variable	Coefficient	95% CI	P-value
Age at AP (years)	0.64	0.24, 1.04	0.002
Sex			
Female vs. male	0.06	−0.07, 0.19	0.39
Birthweight group			
CLB vs. DB	−0.26	−0.46, −0.06	0.01
CHB vs. DB	0.31	0.10, 0.51	0.003

Table 9.8: Estimated fixed effects, 95% confidence intervals (CI) and Wald test P-values for the random intercepts and slopes model of body mass index z-score at physical examination fitted on age at adiposity peak (AP), adjusted for sex and birthweight group. Model is fitted on the 1030 subjects with a successfully identified adiposity peak. CLB is concordant low birthweight, CHB is concordant high birthweight and DB is discordant birthweight.

Table 9.9 shows that after adjustment for sex and birthweight group an increased BMI at AP is also estimated to be associated with a positive and highly statistically significant increases in BMI z-score at physical examination. Conditional on BMI at AP and birthweight group, females are estimated to have a significantly higher BMI z-score at examination than males. Whilst CHB subjects are estimated to have a higher BMI z-score than DB subjects for a given BMI at AP and sex, there is no evidence of a difference between DB and CLB subjects.

Table 9.10 details the estimated fixed effects for the random intercepts model of BMI z-score at examination fitted jointly on age and BMI at AP. There is a borderline statistically significant (P=0.03) interaction between age and BMI at AP. As the inclusion of this interaction is debatable,

Explanatory variable	Coefficient	95% CI	P-value
BMI at AP (kg/m^2)	0.35	0.30, 0.40	<0.001
Sex			
Female vs. male	0.25	0.13, 0.37	<0.001
Birthweight group			
CLB vs. DB	−0.03	−0.22, 0.15	0.74
CHB vs. DB	0.24	0.06, 0.43	0.01

Table 9.9: Estimated fixed effects, 95% confidence intervals (CI) and Wald test P-values for the random intercepts and slopes model of body mass index z-score at physical examination fitted on body mass index at adiposity peak (AP), adjusted for sex and birthweight group. Model is fitted on the 1030 subjects with a successfully identified adiposity peak. CLB is concordant low birthweight, CHB is concordant high birthweight and DB is discordant birthweight.

Table 9.10 includes two different versions of the model: 'Model 1' does not include this interaction term whereas 'Model 2' does. Both models presented are combined-sex models as there is little evidence of any interactions between the dimensions of AP and sex. In Model 1, P=0.07 for the addition of a sex-age at AP interaction term and P=0.39 for an interaction between sex and BMI at AP. In Model 2, P=0.13 for the addition of a sex-age at AP interaction term, P=0.66 for a sex-BMI at AP interaction, and P=0.20 for an interaction between sex, age at AP and BMI at AP.

Model 1 shows evidence of associations between both age and BMI at AP and BMI z-score at physical examination, even after mutual adjustment and adjustment for sex and birthweight group, although the evidence for the BMI at AP association is markedly stronger. This suggests that the association with age at AP seen in Table 9.8 is not merely an artifact of the correlation between age and BMI at AP (i.e. is not just due to confounding). In this model, for given age and BMI at AP and birthweight group, females are expected to have a higher BMI z-score at examination. Similarly adjusting for all other explanatory variables, CHB subjects tend to have a higher BMI z-score than DB subjects, though there is no evidence for a difference between DB and CLB subjects.

Due to the inclusion of an interaction term in Model 2 both age and BMI at AP are centred about their mean values (0.767 years and 17.90 kg/m^2, respectively). The presence of the interaction term makes interpretation somewhat more difficult, though this can be aided by examining the fixed effects of the fitted model more explicitly:

$$\begin{aligned} BMI_z = {} & 0.38\ (age_{AP} - 0.767) + 0.34\ (BMI_{AP} - 17.90) \\ & - 0.30\ (age_{AP} - 0.767)\ (BMI_{AP} - 17.90) \\ & + 0.22\ sex - 0.06\ CLB + 0.25\ CHB + constant \end{aligned} \tag{9.9}$$

where BMI_z, age_{AP}, BMI_{AP}, sex, CLB and CHB are as defined in (9.7). It is possible to rewrite (9.9) in two ways to show more easily how changing each explanatory variable of interest affects

Model	Explanatory variable	Coefficient	95% CI	P-value
Model 1	Age at AP (years)	0.42	0.06, 0.79	0.02
	BMI at AP (kg/m^2)	0.34	0.29, 0.39	<0.001
	Sex			
	Female vs. male	0.22	0.10, 0.34	<0.001
	Birthweight group			
	CLB vs. DB	−0.05	−0.23, 0.14	0.61
	CHB vs. DB	0.24	0.05, 0.43	0.01
Model 2	Age at AP (years)	0.38	0.01, 0.75	0.05
	BMI at AP (kg/m^2)	0.34	0.30, 0.39	<0.001
	Interaction between age and BMI at AP	−0.30	−0.57, −0.03	0.03
	Sex			
	Female vs. male	0.22	0.10, 0.34	<0.001
	Birthweight group			
	CLB vs. DB	−0.06	−0.25, 0.13	0.52
	CHB vs. DB	0.25	0.06, 0.44	0.01

Table 9.10: Estimated fixed effects, 95% confidence intervals (CI) and Wald test P-values for the random intercepts and slopes models of body mass index (BMI) z-score at physical examination fitted jointly on age and body mass index at adiposity peak (AP), adjusted for sex and birthweight group. Model is fitted on the 1030 subjects with a successfully identified adiposity peak. CLB is concordant low birthweight, CHB is concordant high birthweight and DB is discordant birthweight. Age and body mass index at adiposity peak are centred about their mean values in Model 2.

the outcome:

$$\begin{aligned} \text{BMI}_z = {} & 0.38\ (\text{age}_{\text{AP}} - 0.767) + (0.34 - 0.30\ (\text{age}_{\text{AP}} - 0.767))\ (\text{BMI}_{\text{AP}} - 17.90) \\ & + 0.22\ \text{sex} - 0.06\ \text{CLB} + 0.25\ \text{CHB} + \text{constant} \end{aligned} \tag{9.10}$$

and

$$\begin{aligned} \text{BMI}_z = {} & 0.34\ (\text{BMI}_{\text{AP}} - 17.90) + (0.38 - 0.30\ (\text{BMI}_{\text{AP}} - 17.90))\ (\text{age}_{\text{AP}} - 0.767) \\ & + 0.22\ \text{sex} - 0.06\ \text{CLB} + 0.25\ \text{CHB} + \text{constant} \end{aligned} \tag{9.11}$$

From (9.10) it can be seen that for a given age at AP (and sex and birthweight group) a 1 kg/m^2 increase in BMI at AP is estimated to increase BMI z-score at physical examination by $0.34 - 0.30\ (\text{age}_{\text{AP}} - 0.767)$. Thus for an earlier AP the estimated increase in BMI z-score at examination associated with an increase in BMI at AP is greater than for a later AP. Similarly, (9.11) shows that for a given BMI at AP (and sex and birthweight group) a 1 year delay in AP is estimated to increase BMI z-score at examination by $0.38 - 0.30\ (\text{BMI}_{\text{AP}} - 17.90)$, meaning that for a lower BMI at AP the estimated increase in BMI z-score at examination associated with a later AP is greater than for a higher BMI at AP.

Fig. 9.18 plots the estimated increases in BMI z-score at examination for a 1 year delay in age at AP (upper plot) or a 1 kg/m^2 increase in BMI at AP (lower plot) along with the estimated 95% CI for each across the ranges of values encountered. It can be seen from the upper plot that a delayed AP is estimated to be positively associated with increased BMI z-score at examination when BMI at AP is less than about 19.5 kg/m^2, although this relationship is only statistically significant (at the 5% level) when BMI at AP is less than approximately 18 kg/m^2. The lower plot, on the other hand, shows increased BMI at AP to be estimated to be positively and statistically significantly associated with BMI z-score at examination across virtually the entire range of observed ages at AP. Indeed, when the AP occurs at an age towards the younger end of this spectrum the relationship is *highly* statistically significant.

One way to compare Model 1 and Model 2 in Table 9.10 is to plot predicted BMI z-score values from each model for different combinations of explanatory variables. As there are five explanatory variables, this involves effectively plotting in six dimensions. However, by considering the different combinations of levels of the indicator variables separately and plotting the predicted values as contours on a plane, plotting becomes possible. Fig. 9.19 and Fig. 9.20 show the contour plots for DB males for Model 1 and Model 2, respectively.

As Model 1 does not include an interaction term between age and BMI at AP the contour lines in Fig. 9.19 are parallel. The region of highest predicted BMI z-score at physical examination is seen to correspond to a late AP and a high BMI at AP, although it is clear from the plot that it is BMI as opposed to age at AP which is exerting the greater influence. The lowest predicted BMI z-scores correspond to early AP and a low BMI at AP.

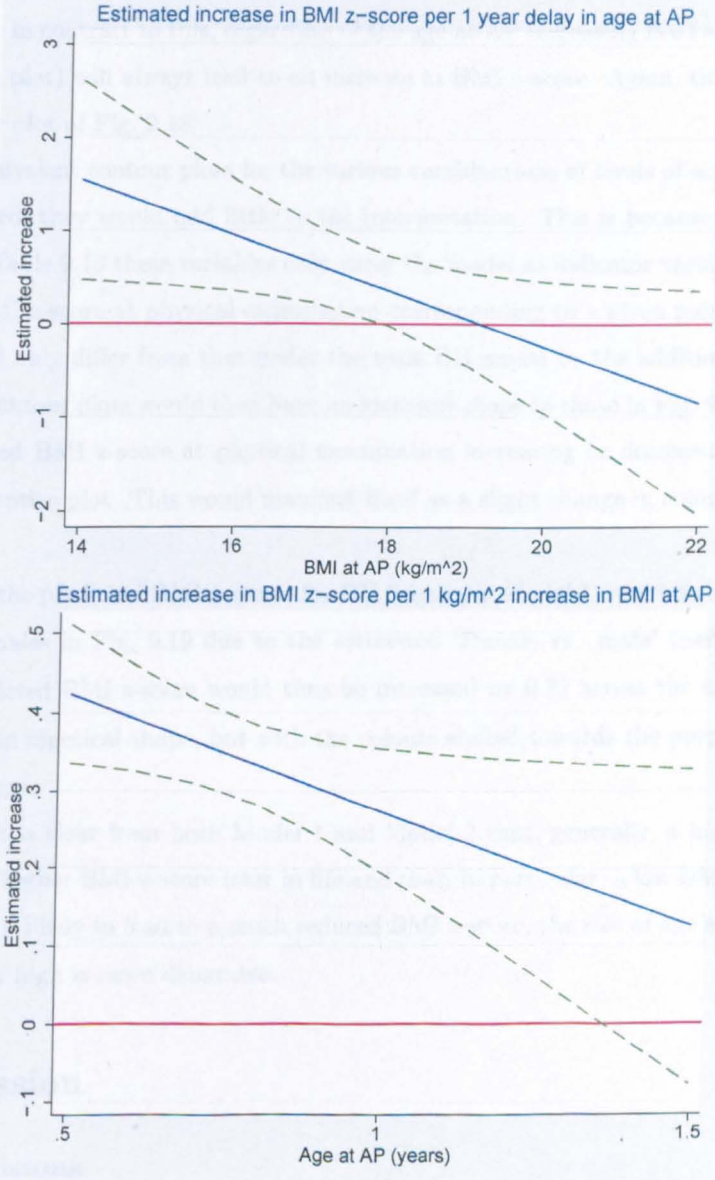

Fig. 9.18: Estimated increases in body mass index (BMI) z-score at physical examination for a 1 year delay in age at adiposity peak (AP) (upper plot) or a 1 kg/m^2 increase in body mass index at adiposity peak (lower plot) whilst the other dimension of adiposity peak and sex are held constant. Solid lines are estimated increases and dashed lines are their 95% confidence intervals.

In Fig. 9.20, the interaction term means that the observed pattern of predicted BMI z-score at physical examination is more complex. The region of highest predicted BMI z-score now corresponds to an *early* AP and a high BMI at AP, and the lowest predicted BMI z-score to early AP and a low BMI at AP. For a lower BMI at AP increasing age at AP (i.e. tracing horizontally across the plot) leads to an increase in BMI z-score, whilst for a higher BMI at AP this results in a slight *decrease* in BMI z-score. This corresponds precisely to the pattern exhibited in the upper plot of Fig. 9.18. In contrast to this, regardless of the age at AP increasing BMI at AP (i.e. tracing vertically up the plot) will always lead to an increase in BMI z-score. Again, this reflects what is seen in the lower plot of Fig. 9.18.

Although equivalent contour plots for the various combinations of levels of sex, CLB and CIIB could be produced, they would add little to the interpretation. This is because in both Model 1 and Model 2 in Table 9.10 these variables only enter the model as indicator variables. As a result, the predicted BMI z-score at physical examination corresponding to a given pair of age and BMI at AP values will only differ from that under the male DB model by the addition of one or more constants. The contour plots would then have an identical *shape* to those in Fig. 9.19 and Fig. 9.20 but with predicted BMI z-score at physical examination increasing or decreasing by a constant value across the entire plot. This would manifest itself as a slight change in colour scheme for the contour plot.

For example, the predicted BMI z-scores for DB females in Model 1 would be higher than those plotted for DB males in Fig. 9.19 due to the estimated 'Female vs. male' coefficient of 0.22 in Table 9.10. Predicted BMI z-score would thus be increased by 0.22 across the entire plot — the plot would have an identical shape, but with the colours shifted towards the purple (i.e. positive) end of the spectrum.

Thus, whilst it is clear from both Model 1 and Model 2 that, generally, a higher BMI at AP tends to lead to a higher BMI z-score later in life and that, in particular, a low BMI at AP coupled with an early AP is likely to lead to a much reduced BMI z-score, the role of age at AP when BMI at AP is relatively high is more debatable.

9.5 Discussion

9.5.1 Conclusions

The initial peak in BMI at around the age of 6 months to 1 year (the AP) has been shown to be a readily identifiable feature of the growth curve in the vast majority of subjects encountered using penalised regression splines with random coefficients.

Both age and BMI at AP have been found to be positively associated with later BMI z-score in this dataset. Whilst higher BMI at AP tends to result in relatively higher BMI in later childhood regardless of age at AP, the relationship with age at AP appears to be somewhat more complex. It is the first time that these associations have been reported.

The positive relationship generally seen between the timing of the AP and later BMI is in the

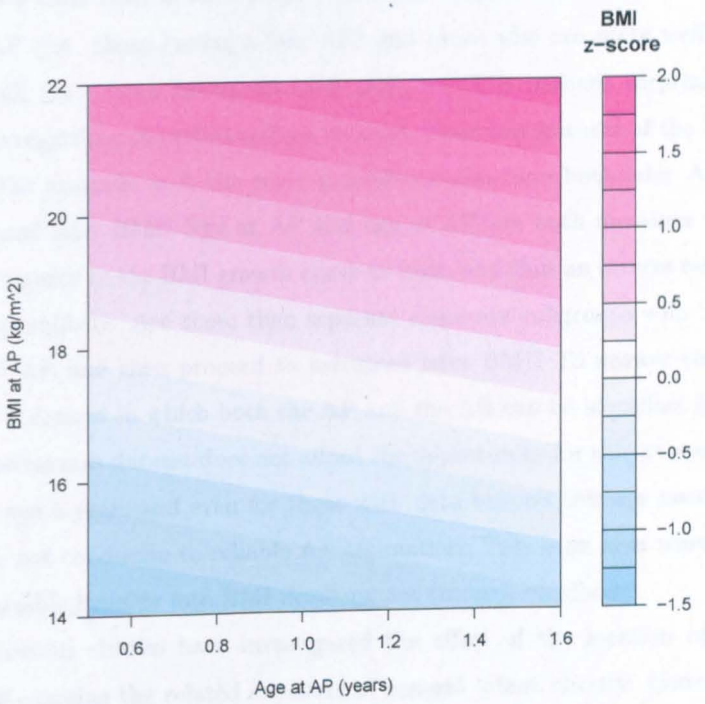

Fig. 9.19: Contour plot for predicted body mass index (BMI) z-score at physical examination from Model 1 for different combinations of age and body mass index at adiposity peak (AP) in discordant birthweight males.

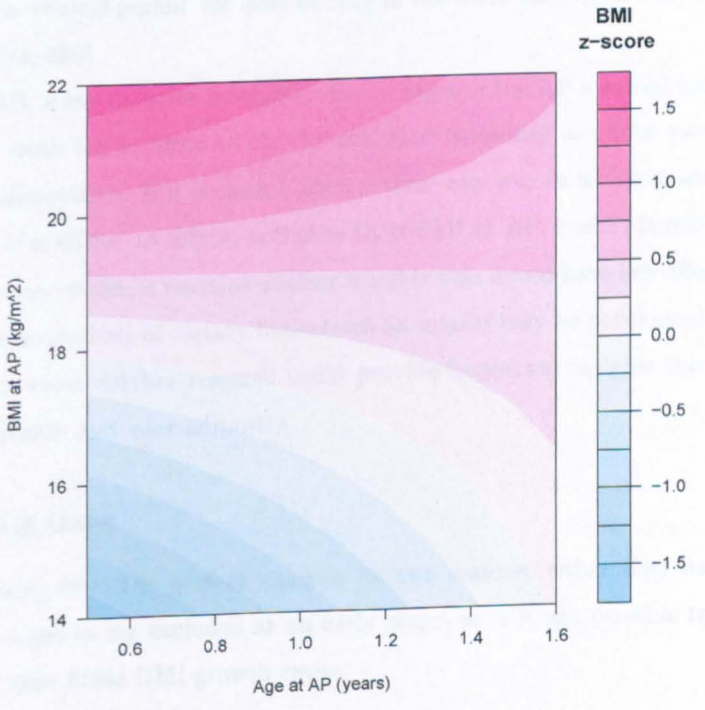

Fig. 9.20: Contour plot for predicted body mass index (BMI) z-score at physical examination from Model 2 for different combinations of age and body mass index at adiposity peak (AP) in discordant birthweight males.

opposite direction to that between the timing of the adiposity rebound (AR) and later BMI. This means that higher later BMI is associated with both those who are *less well developed* around the age of the AP (i.e. those having a late AP) and those who are *more well developed* around the age at the AR (i.e. those having an early AR), which is perhaps surprising. This leads to further questions regarding the relationships between these two features of the BMI growth curve and later BMI. For example, is it the same individuals who have both later AP and earlier AR, leading to increased later BMI? Age at AP and age at AR are both measures of development at that point, with regards to the BMI growth curve at least, and thus an inverse relationship between them would seem unlikely. Are there then separate disparate subgroups who have *either* a later AP *or* an earlier AR and then proceed to increased later BMI? To answer these questions it is essential to have a dataset in which both the AP and the AR can be identified for each individual. Unfortunately the current dataset does not afford the opportunity for this as some individuals only have data up to age 5 years and even for those with data beyond this age measurements become sparse and hence not conducive to reliable AR estimation. This is an area where further research could provide valuable insights into BMI development through childhood.

Whilst no previous studies have investigated the effect of the location of the AP on later adiposity, several examine the related exposure of general infant obesity. Conclusions are mixed, however, with some finding that there is little evidence that infant obesity is predictive of later obesity [76] and others suggesting that infant obesity correlates strongly with adult obesity [57].

The associations found in the present analysis may indicate that the AP is a meaningful feature of the infant BMI trajectory for prediction of later BMI. This may suggest that infancy needs to be considered as a 'critical period' for later obesity in the same manner in which the period around the AR often is [74, 180].

As with the AR, a key question is whether the location of the AP a causal factor later adiposity itself or whether both the location of the AP and later adiposity are both merely expressions of some genetic predisposition. If it is causal, then is there any way in which it can be manipulated? Whilst the level of BMI for an infant, and thus their BMI at AP, could plausibly be manipulated by changes in dietary intake, it remains unclear whether this would have any effect on the timing of the AP. Also, the imposition of dietary limitations on infants may be considered undesirable. This is clearly an area where further research could provide important insights into the relationships between infant growth and later adiposity.

9.5.2 Missing data

Subjects are missing from the present analysis for two reasons, either they have no data points over the relevant ages so are excluded at an early stage, or it is not possible to derive a location for the AP from their fitted BMI growth curve.

Excluded subjects

Whilst growth data for most subjects are available for much older ages, only those data from, but not including, birth to age 3 years are utilised in the present study. As the AP would not be expected to occur after age two years these data criteria seem appropriate as any maxima should occur sufficiently within the interval to be identifiable without the inclusion of data at older ages which would only serve to complicate the curve fitting procedure. There are, however, a small number of subjects (3.3%) having no BMI observations whatsoever over this period. These subjects are omitted completely from the analysis. Whilst they could remain in the analysis they would contribute little, having assigned as their fitted BMI growth curve the fixed effects from the relevant model. As the proportion of the dataset they make up is relatively small their omission seems a reasonable choice.

However, for any results obtained in the analysis to not be biased by their omission it is important that they are effectively just a random subset of the data, or that they are 'missing completely at random' (MCAR, see Section 5.2.1). In Section 9.2 the distribution of several variables are compared between those subjects with no observed BMI values who are excluded and those who remain. The distributions appear relatively similar, though due to the small numbers of excluded individuals it is important not to over-interpret any differences. It can be concluded that there is little evidence of the excluded subjects not being MCAR.

Subjects with no identified AP

Whilst an estimated AP is identified in the vast majority of individuals considered, there are still many for which this is not the case. Identification of these individuals and analysis of their data points and fitted subject-specific curves shows that the curves generally fit the observed values well and that their observed values really do not provide any evidence of an AP, usually because BMI appears to continue increasing throughout infancy. In the present analysis those subjects with no identifiable AP are merely excluded.

Again, to obtain unbiased results it is important for these excluded subjects to be MCAR. In Section 9.4.1 it is seen that there are small differences in the percentages of subjects for whom an AP can be identified in the different subgroup models (Table 9.4) and that subjects with no identified AP differ a little from the other subjects (Table 9.6). However, a more formal assessment (Table 9.7) concludes that there is no evidence of a relationship between AP identifiability and BMI z-score at physical examination. This suggests that those subjects with no identified AP who are excluded from the analysis do not differ significantly in terms of later BMI from those who are included.

9.5.3 Body mass index growth curve modelling

Penalised regression spline model

The use of penalised regression spline models with random coefficients to model BMI growth is very effective. This approach, as opposed to other spline approaches, has the attractiveness of being a relatively straightforward extension of linear regression modelling. The mixed model representation means that model fitting can be easily implemented in standard statistical software. The equivalence between a penalised smoother and the optimal predictor in a mixed model, as shown in Section 5.4.1.5, results in a unified approach to model estimation. The cubic population average curves and cubic subject-specific deviations from these allow sufficient flexibility to model a variety of different curve shapes and ensure that the derivative of each subject-specific curve is smooth and continuous, which is important when looking for turning points.

The subject-specific curves generally fit the data very well. For those individuals with few BMI observations overall, or with regions with few observations, this is still true. The approach allows a reasonable curve to be fitted by 'borrowing' information from the other subjects and fitting a subject-specific curve closer to the relevant population average curve.

However, there are always likely to be some individuals whose observed values lie on a sufficiently differently shaped trajectory from other subjects, and hence from the population average curve, so that their fitted curve does not fit their observed values as well as would be hoped. In the present analysis these cases are very few and their presence must be considered as a trade-off against the benefits of having a common underlying BMI trajectory in those individuals with sparse BMI data where fitting a truly subject-specific curve (i.e. using only the data points for that individual) would be problematic.

Improvements to the model

Six separate BMI growth models are fitted on the six subgroups defined by subjects' sex and birthweight group (CLB, CHB and DB). Whilst there is no evidence of this resulting in poorly fitting curves, it would be preferable to include *all* subjects in the same model with indicator variables for sex and birthweight group, similar to those used in (9.7) for relating AP location to later BMI z-score.

Let sex_i be an indicator variable taking value 1 if subject i is female and 0 otherwise and CLB_i and CHB_i be indicator variables taking value 1 if the birthweight group of subject i is, respectively, CLB or CHB and 0 otherwise. Then, if the effects of sex and birthweight group can be assumed to be additive, (9.1) becomes

$$
\begin{aligned}
y_{ij} = {} & \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \beta_3 x_{ij}^3 + \sum_{k=1}^{K} u_k (x_{ij} - \kappa_k)_+^3 \\
& + \text{sex}_i \left(\mu_0 + \mu_1 x_{ij} + \mu_2 x_{ij}^2 + \mu_3 x_{ij}^3 + \sum_{k=1}^{K} r_k (x_{ij} - \kappa_k)_+^3 \right) \\
& + \text{CLB}_i \left(\rho_0 + \rho_1 x_{ij} + \rho_2 x_{ij}^2 + \rho_3 x_{ij}^3 + \sum_{k=1}^{K} s_k (x_{ij} - \kappa_k)_+^3 \right) \\
& + \text{CHB}_i \left(\omega_0 + \omega_1 x_{ij} + \omega_2 x_{ij}^2 + \omega_3 x_{ij}^3 + \sum_{k=1}^{K} t_k (x_{ij} - \kappa_k)_+^3 \right) \\
& + a_{i0} + a_{i1} x_{ij} + a_{i2} x_{ij}^2 + a_{i3} x_{ij}^3 + \sum_{k=1}^{K} v_{ik} (x_{ij} - \kappa_k)_+^3 + \varepsilon_{ij}
\end{aligned}
$$

where $u_k \sim N(0, \sigma_u^2)$, $(a_{i0}, a_{i1}, a_{i2}, a_{i3})^T \sim N(0, \boldsymbol{\Sigma})$, $v_{ik} \sim N(0, \sigma_v^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ as before. Thus, for example, the fitted curve for a DB male would be

$$
\begin{aligned}
y_{ij} = {} & \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \beta_3 x_{ij}^3 + \sum_{k=1}^{K} u_k (x_{ij} - \kappa_k)_+^3 \\
& + a_{i0} + a_{i1} x_{ij} + a_{i2} x_{ij}^2 + a_{i3} x_{ij}^3 + \sum_{k=1}^{K} v_{ik} (x_{ij} - \kappa_k)_+^3 + \varepsilon_{ij},
\end{aligned}
$$

whilst the fitted curve for a CHB female would be

$$
\begin{aligned}
y_{ij} = {} & \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \beta_3 x_{ij}^3 + \sum_{k=1}^{K} u_k (x_{ij} - \kappa_k)_+^3 \\
& + \mu_0 + \mu_1 x_{ij} + \mu_2 x_{ij}^2 + \mu_3 x_{ij}^3 + \sum_{k=1}^{K} r_k (x_{ij} - \kappa_k)_+^3 \\
& + \omega_0 + \omega_1 x_{ij} + \omega_2 x_{ij}^2 + \omega_3 x_{ij}^3 + \sum_{k=1}^{K} t_k (x_{ij} - \kappa_k)_+^3 \\
& + a_{i0} + a_{i1} x_{ij} + a_{i2} x_{ij}^2 + a_{i3} x_{ij}^3 + \sum_{k=1}^{K} v_{ik} (x_{ij} - \kappa_k)_+^3 + \varepsilon_{ij}.
\end{aligned}
$$

Although this all-inclusive model is appealing in theory, the practicalities of fitting it in a dataset with even as many subjects and data points as the UFS may be troublesome. In the present analysis, each of the six subgroup models takes approximately three hours to fit. Fitting a model on the entirety of the dataset with a greater number of parameters needing to be estimated could therefore be expected to take a significantly longer amount of time, and perhaps even be beyond the capabilities of the computing power available.

In addition to the unification of the subgroup models into one overall model, further variables could be added to try and improve model fit. For example, as is acknowledged elsewhere, it may be expected that a given subject is likely to have a BMI growth curve more similar to that of his/her sibling than to that of another subject to whom they are not related. This expectation could be incorporated into the model by the introduction of one or more terms relating to an identifier for 'family'.

9.5.4 The Uppsala Family Study dataset

Features of the dataset

The structure of the UFS is a somewhat unusual, both in terms of being made up only of sibling pairs and, perhaps more importantly, the nature by which sibling pairs are selected for inclusion based on their birthweights relative to each other. Both of these issues are dealt with relatively satisfactorily in (9.7).

Allowing family-specific random effects acknowledges that subjects are likely to be more similar to their sibling than to other members of the dataset to whom they are not related. When considering the relationship between one or both dimensions of the AP and BMI z-score at physical examination, random intercepts allow for overall family-specific differences in BMI z-score, whilst random slopes allow for family-specific differences in the relationship between the dimension(s) of the AP and BMI z-score. This modelling approach would appear both appropriate and sufficient to deal with the structure of the dataset.

As birthweight is known to affect growth trajectories [45], the selection procedure of the study design may affect both the location of the AP and the later BMI z-score of an individual as well as, potentially, the relationship between the two. The issue of birthweight, or, at least, birthweight group, is handled in (9.7) through the inclusion of indicator variables which allow additive effects of different birthweight groups to be estimated and adjusted for. This approach appears to be adequate. Models including continuous birthweight instead of indicator variables for birthweight group were also fitted but the estimated coefficients changed little and the conclusions would be identical (results not shown).

A further unusual feature of the UFS is that the physical examinations, at which the outcome in the present analysis was observed, occur across a wide range of ages (see Fig. 9.4). However, this should not cloud the conclusions reached here to any great extent. Tracking of BMI throughout childhood is widely acknowledged [156] so that whether BMI z-score is at age 5 years or age 13 years it is not just a valid measure of BMI relative to others of the same sex at that precise age, but also highly indicative of relative BMI over a much wider range of ages. Whilst it may be preferable to be able to state that the AP is associated with 'BMI at age x', an outcome of this nature is not available in the present dataset.

Representativeness and generalisabiliy

It is important to examine whether the conclusions reached within the UFS can be extrapolated beyond the members of the dataset itself. Aside from the issues arising from the unusual selection approach based on birthweight, as outlined above, the representativeness of the members of the UFS within the Swedish population and the generalisability from a Swedish dataset to subjects outside of Sweden must also be considered.

The sampling frame from which the final UFS subjects are drawn is that of all families with at least two consecutive singleton children delivered at term and within 36 months of each other at

the Uppsala Academic Hospital between 1987 and 1995. It is thus a contemporary, healthy sample, which is likely to be representative of the wider Uppsala population. However, as participation rates were not particularly high [101], subjects in the UFS could potentially not be representative of this larger population. That the data are Swedish, being a developed European country, means that if the conclusions can be assumed to be representative of Sweden then they can be extrapolated relatively safely to similar populations.

Thus it is envisaged that whilst there are some issues which could plausibly reduce the generalisability of the results obtained, it is likely that they would be replicated in further datasets. Attempts to do so could prove valuable in improving understanding of BMI development through childhood.

Part IV

Discussion

Chapter 10

Discussion

This thesis explores, develops and implements modelling strategies for studying relationships between childhood growth and later health. The datasets used in the thesis are briefly summarised in Section 10.1 before the main epidemiological findings and conclusions are discussed in Section 10.2 and the methodological considerations deatiled in Section 10.3. Areas for future work are examined in Section 10.4.

10.1 Datasets

The two main datasets used in this thesis are the Stockholm Weight Development Study (SWEDES) and the Uppsala Family Study (UFS). Both datasets include longitudinal measurements of childhood growth, as well as several measures of later health outcomes, and thus correspond to the type of data structure on which the thesis concentrates. The salient features of both SWEDES and the UFS are briefly summarised below, although more detailed introductions to the datasets can be found in Chapter 4.

Three of the British birth cohorts (the National Survey of Health and Development (NSHD), the National Child Development Study (NCDS) and the British Cohort Study (BCS)) are also used in Chapter 6.2 to illustrate the standardardisation of childhood BMI data into age- and sex-specific z-scores. As their usage does not correspond to the main aims of the thesis, these datasets are not reviewed further here.

Stockholm Weight Development Study

SWEDES is a prospective longitudinal study of weight development in 481 children from Stockholm born over a 12 month period between 1984 and 1985. Comprehensive growth data from birth until age 15 years are available and a variety of anthropometric, metabolic, psychological and lifestyle variables were observed at follow-up when the subjects were approximately 17 years old.

Weight and length at birth were recorded from hospital records, and during infancy, height and weight were measured as part of routine visits to a child welfare centre. Measurements were taken three further times after birth during the first year (at 6, 9 and 12 months) and annually

thereafter until age 6 years. From age 7 years onwards annual measurements of height and weight were recorded in journals by the subjects' schools. As height and weight measurements occur at common ages for each individual, balanced growth data are available in SWEDES. The regular concurrent measurements of height and weight throughout childhood allow the calculation of BMI and thus permit the detailed exploration of childhood BMI development for each individual.

Of the available variables measured at the late-adolescent follow-up, it is only those pertaining to obesity that are utilised in this thesis. In particular, BMI is calculated from the observed values of height and weight, and percentage body fat (%BF) is derived using air-displacement plethysmography.

SWEDES thus provides balanced BMI growth data which can be related to the two measures of late-adolescent adiposity. In Chapter 7 this is accomplished directly using a single-stage analysis approach, whilst in Chapter 8 growth models are first fitted to the BMI growth data and estimated locations of the adiposity rebound (AR) derived, which are then related to late-adolescent adiposity in a two-stage analysis approach.

Uppsala Family Study

The UFS also provides longitudinally measured childhood growth data and outcome variables observed at a later follow-up, but differs from SWEDES in several key areas, including the overall data structure and the unbalanced nature of the childhood growth data.

The UFS is made up for 602 sibling pairs (1204 subjects) born within 36 months of each other in Uppsala, Sweden, between 1987 and 1995. The initial focus of the dataset was to study early and maternal effects on blood pressure and cardiovascular disease. To increase statistical efficiency only sibling pairs where both siblings had high birthweight, both had low birthweight, or where there was a large difference in birthweight were included.

Sampling was retrospective, so all childhood data were obtained via linkage to existing records: birth data from the mothers' obstetric records, and postnatal growth data, including serial measurements of height and weight, from health records, kept by Child Health Centres or at schools. The nature of this data collection means that the childhood growth data are not available for common ages across the subjects, resulting in unbalanced data. However, the concurrent measurements of height and weight again mean that BMI can be calculated, and, as growth data are often available on many occasions through childhood for each individual, detailed exploration of BMI development is again possible.

Follow-up in the UFS occurred between May 2000 and November 2001 when the subjects were aged 5–13 years. At a physical examination several anthropometric variables were observed, but again it is only the information regarding obesity (in this case BMI) which is used in the thesis. As physical examinations corresponded to a wide range of ages and as BMI is very much age-depended over this range, using BMI itself as an outcome is unwise. Instead, BMI z-scores are calculated using the Swedish population reference values [100].

As with SWEDES, the relationship of interest in the UFS is between childhood growth in

BMI and later obesity. The numerous BMI observations for each subject in infancy allow, via the explicit modelling of BMI growth curves, the identification of the adiposity peak (AP) in infancy. This is related to later BMI z-score using mixed models to account for the sibling pair structure of the dataset in Chapter 9.

10.2 Epidemiological conclusions

The main epidemiological conclusions in this thesis focus on how childhood growth, in particular the timing of features of the BMI growth curve, affects the development of obesity. The typical childhood BMI growth curve will increase from birth and reach a peak at around age 9 months before decreasing. At around age 6 years BMI generally begins increasing once more. Thus there are ordinarily two turning points in the BMI curve, the maximum in infancy, here referred to as the adiposity peak (AP), and the later minimum, generally referred to as the adiposity rebound (AR). Whilst there is an established literature regarding the relationship between the timing of the AR and later obesity, there is, to my knowledge, no corresponding literature for the AP. Thus, whilst some of the work in this thesis provides interesting new insights into the relationship between the AR and later obesity, it is the results regarding the AP which contribute entirely novel epidemiological findings.

Childhood BMI development and later obesity

In Chapter 7 a naive multivariable regression analysis approach is used to study the relationship between childhood BMI development (annually observed BMI from age 1 to 10 years) and late-adolescent adiposity (BMI and %BF at approximately age 17 years) in SWEDES. Whilst this approach has deficiencies due to missing data and collinearity, it does provide an initial exploratory analysis of this relationship.

It is seen that increased BMI velocity at any age during childhood, for given BMI velocities at all other ages, will tend to lead to higher late-adolescent adiposity. This relationship is found to be strongest between age 1 and 2 years in both sexes, and age 4 to 7 years in females and age 5 to 8 years in males. These observations suggest that rapid BMI development relative to others of the same age in infancy and around the period of the AR are associated with higher later adiposity, indicating that these periods could potentially be considered as critical periods for the development of obesity as suggested by Dietz [74, 180]. In particular, a BMI which increases rapidly relative to your peers during the period around the AR, which is equivalent to an earlier AR relative to your peers, is suggestive of an early AR being a risk factor for later obesity.

The adiposity rebound and later obesity

A more explicit investigation of the relationship between the AR and late-adolescent adiposity (BMI and %BF) in SWEDES is carried out in Chapter 8 using more robust analytical approaches. The AR is seen to be a feature of the childhood BMI growth curve which can be identified in

the majority of subjects. It occurs, on average, between age 5 and 6 years, although there is large between-subject variability. The AR is found to occur slightly later in males than females in SWEDES. This corresponds to the previous observation of the strongest relationship between BMI velocity and late-adolescent adiposity being seen at a slightly later age in males in Chapter 7. The observed ages at AR and between-sex differences correspond well to previously published results [82, 85, 86, 172, 165], although there are also examples of females having later AR than males [84].

When considering categorical age and BMI at AR, both dimensions of the AR are seen to be strongly and independently associated with late-adolescent adiposity in SWEDES. Either an earlier AR, a higher BMI at AR, or both, leads to a large increase in the odds of late-adolescent overweight (high BMI) and a smaller, though still sizeable, increase in the odds of overfat (high %BF). Whilst age and BMI at AR are seen to be negatively correlated, it is found that the inverse relationship between age at AR and later adiposity cannot be explained by confounding due to subjects with earlier AR having higher BMI at this age.

When using continuous age and BMI at AR there is some evidence of an interaction between the two dimensions of the AR. Increased BMI at AR is estimated to increase late-adolescent adiposity more when it corresponds to an early AR than when it corresponds to a late AR. Similarly, a delayed AR is estimated to decrease late-adolescent adiposity more when it corresponds to a high BMI at AR than when it corresponds to a low BMI at AR.

The adiposity rebound as a critical period for later obesity

These findings imply that, regardless of the size of an individual, the timing of their AR is important in the development of later obesity. This leads to the suggestion that the period around the AR may be considered as a critical period for later obesity — 'a developmental stage in which physiologic alterations increase the later prevalence of obesity' [74].

This is investigated in Section 8.10, where age and BMI at AR are considered as explanatory variables for later adiposity in models alongside estimated BMI and BMI velocity at different ages through childhood. At ages before the occurrence of the AR in most individuals, the two dimensions of the AR are seen to be more strongly associated with late-adolescent adiposity than BMI and BMI velocity at that age. At ages when the AR has already passed in the majority, the opposite is true, with BMI and BMI velocity taking greater significance. At ages near the average age at AR, there is often no discernible pattern. These observations are seen to be equally strong in males and females and suggest that age and BMI at AR are no better predictors of later adiposity than BMI and BMI velocity at a similar age. This implies there is little extra information contained within these dimensions and suggests that the relationship between the AR and later adiposity is more statistical than physiological. As a result, the AR cannot be considered as a critical period by the definition of Dietz [74].

Although there are complicated missing data issues (discussed in Section 8.11.2) in this analysis of the SWEDES dataset, the use of a principled missing data approach (multiple imputation (MI))

means that the the conclusions drawn should be sufficiently robust.

The adiposity peak and later obesity

In Chapter 9 the initial peak in BMI around age 6 month to 1 year (the AP) is seen to be a readily identifiable feature of the BMI growth curve in the vast majority of subjects in the Uppsala Family Study (UFS). Average age at AP is found to be marginally later in females, which is the opposite to the difference usually seen for the AR.

When considered separately, both age and BMI at AP are found to be strongly positively associated with BMI z-score in later childhood. However, whilst higher BMI at AP leads to higher BMI in later childhood regardless of the age at AP, the relationship between age at AP and later BMI conditional on BMI at AP is weaker and somewhat more complex. In particular, there is some suggestion of an interaction between age and BMI at AP, meaning that, whilst an early AP tends to lead to a lower BMI z-score in later childhood when it is coupled with a low BMI at AP, if BMI at AP is very high, an early AP may actually increase the expected BMI z-score. To my knowledge, it is the first time that these associations have been reported.

The novel growth curve fitting approach used in identifying the AP in the subjects of the UFS results in even those individuals with few data being able to contribute to the analysis. Consequently the proportion of subjects who are unable to contribute to the analysis is low, meaning that the findings are relatively robust to the effects of missing data. However, it is not possible to identify the AP for some individuals. Whilst this is often because their observed BMI continues to increase throughout infancy, these individuals are not found to have a significantly increased likelihood of high later BMI z-score.

Whilst there is some debate over the importance of infant growth with respect to later obesity [57, 76], these results show that there is a strong association with between size in infancy and later adiposity, and that development by this stage also plays a role. The AP is found to be a meaningful feature of the BMI curve for the prediction of later obesity. This suggests that, although the first year of life is already considered as a critical period for later obesity [74, 180], perhaps the AP should be more explicitly investigated in this context.

The adiposity peak and the adiposity rebound

The positive association seen between the age at AP and later BMI in Chapter 9 is in the opposite direction to that seen between the age at AR and later adiposity in Chapter 8 (and widely acknowledged elsewhere). Thus higher later adiposity appears, somewhat paradoxically, to be associated with both those individuals who are *less* well developed around the period of AP (in that they have a later AP) and those who are *more* well developed around the period of AR (in that they have an earlier AR). Whether or not it is the same individuals who have both a late AP and an early AR before progressing to higher later adiposity is an interesting question, although one that is, unfortunately, beyond the scope of the datasets used in this thesis. Analysis of these long-term patterns of growth, however, must remain an important future aim.

Both the timing of, and BMI at, AP and AR are seen to be associated with later adiposity. A positive relationship between BMI, as a proxy for adiposity, and another measure of adiposity at *any* two ages can be explained through adiposity tracking, thus it is the associations involving timing which are the more controversial. The statistical associations between age at AP and AR and later obesity appear robust, particularly for the AR given the existing literature, but it remains somewhat unclear whether these timings are truly causal factors for later obesity. Indeed, both the timing of one or both of these features of the BMI growth curve and the level of later adiposity may simply be expressions of some genetic and/or environmental predisposition. Only if a change in the timing of the AP or AR can be shown to affect later adiposity *within an individual* can the associations be though to be causal. This is discussed further in Section 10.4.

10.3 Methodological considerations

Naive multivariable analysis

When studying relationships between childhood growth and later health, if the longitudinal childhood growth data are balanced then one simple approach is to directly use the measurements at some or all of the ages as explanatory variables in a regression analysis.

However, as is seen in Chapter 7, this approach may be problematic. Firstly, when including many childhood measurements in a regression model may be difficult to interpret, especially if observations are close together in time, due to their respective conditioning. Further to this, measurements taken on the same individual at different ages are likely to be correlated, which can cause problems with collinearity. This may manifest itself as imprecise regression coefficient estimates, making interpretation difficult. Problems due to multiplicity and collinearity are likely to increase if the ages included in the model are close together or numerous.

A further difficulty is due to the use of a complete-case analysis approach to the handling of missing data. This means that any individual with missing data on one or more variables will not contribute to the analysis. Only if these excluded individuals are missing completely at random (MCAR) will the results remain unbiased. The proportion of excluded individuals generally increases with the number of explanatory variables (i.e. ages) included in the model. Even if the amount of missing data at any given age is small, if sufficient variables are included then the cumulative effect can be sizable.

Whilst interpretation of the estimated regression coefficients can be aided somewhat by plotting them against age to form a life course plot [130], which emphasises the dual nature of size and growth, the precision of the estimates may remain unsatisfactory.

One approach to overcoming the problems due to collinearity is to reparameterise the model so that childhood growth velocities (calculated from the observed growth data) are used as the explanatory variables. Velocities are generally far less susceptible to collinearity, allowing more reliable regression coefficient estimation, although this will not reduce the problems due to missing data. Indeed, as a greater number of BMI observations are required to calculate the same number

of BMI velocities, this approach may actually exacerbate the problems due to missing data.

A potential alternative approach would be to only use a subset of the ages at which the anthropometric variable is observed. This would be likely to reduce collinearity and also increase the proportion of subjects who can be included in the analysis, reducing problems due to missing data. However, the exclusion of variables may result in a loss of information.

Thus a naive multivariable analysis involving childhood growth data observed at many ages is unlikely to be an optimal approach. In particular, for datasets with even moderate amounts of missing data over many variables this approach is not recommended.

Multiple imputation

When faced with balanced longitudinal growth data a more robust approach to the handling of missing data is through multiple imputation (MI). Under this approach each missing value is replaced by a draw from the conditional distribution for the missing data given the observed data to create multiple completed datasets. Each dataset is analysed separately using standard complete data procedures, then the results combined.

MI is utilised in the analysis of the relationship between the AR and later obesity in SWEDES (Chapter 8). However, this application of MI is somewhat unusual as it does not result in every individual within the dataset contributing to the final analysis, as would generally be the case elsewhere. This is because of the three-stage analysis approach used. Firstly, missing values are imputed to create multiple completed datasets. Secondly, individual growth models are fitted to these completed datasets, and from these the location of the AR is estimated for each subject in each imputed dataset, where possible. Finally, the relationship between the AR and later obesity is examined in each imputed dataset and the results combined in the standard manner. So, although in each imputed dataset each individual effectively has a full set of BMI values present, the subsequent growth curve fitting may not successfully identify an estimated AR. These subjects are then excluded from the remainder of the analysis — thus the 'missingness' of the AR is effectively handled via a complete-case approach within the MI approach.

Whilst the use of MI does not completely eradicate missing data from the final analysis model, it still increases the proportion of subjects who can contribute relative to the equivalent analysis without the use of MI. Thus, if the subjects who are excluded from the analysis when MI is not used cannot be considered to be MCAR, the use of MI should reduce the extent of bias. However, as some subjects remain excluded from the analysis when using MI, if the missingness is not completely at random then there may remain residual bias. In particular, for individuals whose observed BMI values increase throughout childhood and for whom no AR thus occurs, missingness from the analysis is clearly dependent on the data and consequently cannot be considered MCAR. However, no relationship is found between AR unidentifiability and late-adolescent adiposity, indicating that those subjects with no AR who are excluded from the analysis do not differ significantly from the remainder with regards to the outcome variables.

A further issue with MI in this complex multi-stage setting is that interactions which involve

one or more of the derived growth features cannot be included in the imputation model by virtue of these features not being derived until after imputation. This means that these interactions then cannot be accurately assessed in the analyses. However, this situation is somewhat unusual and would not occur in a more standard MI analysis.

As well as the variables for which imputed values are required, the imputation model in Chapter 8 includes further variables which are not subject to missingness, and some of which are not even used in the later analysis models. The inclusion of these variables, which are all believed to be related to either the missing variables or the missing data mechanism, should help provide unbiased imputed values by making the MAR assumption more plausible [126].

A Markov chain Monte Carlo (MCMC) approach is used to create 100 imputed datasets. Although this is more than is generally advised as being necessary [120], this decision was taken in light of more recent research suggesting this to not always be the case [121]. Whilst increasing the number of imputed datasets in this manner has a small cost in terms of computing time, this more than made up for by the additional reassurance provided.

A joint multivariate normal distribution is assumed. Although the majority of variables included in the imputation model are continuous and can reasonably be assumed to follow a (marginally) normal distribution, perhaps after a transformation, some discrete or dichotomous variables, for example sex, are included by necessity. However, as these variables are all fully observed, the implausibility of the multivariate normality assumption is unlikely to be problematic [124].

In this particular application of MI, comparison of the results using the original data only and the results using MI shows only relatively minor differences. Certainly the conclusions drawn would be very similar. However, without conducting an analysis using MI this comparison would obviously be impossible. Thus, in the more general setting, it may be suggested that when analysing any datasets which are subject to missingness, a repetition of the analysis using a MI approach can provide a useful tool. If the initial analysis is conducted on a complete-case basis, under the assumption of MCAR, then comparison to the results using MI, under the more relaxed assumption of missing at random (MAR), allows an assessment of the extent to which results are robust to the missing data assumption.

Growth modelling

Growth modelling has been seen to be a useful method by which to summarise childhood growth data, and in particular to derive 'growth features' of interest for further analysis. When the growth data are balanced, the analysis of these derived growth features in relation to a later health outcome provides an alternative approach to the simple inclusion of some or all of the growth data in a multivariable regression model which, as previously discussed, may not be ideal due to the effects of collinearity and missing data. When the childhood growth data are not balanced, the option of multivariable regression modelling is not available, so growth modelling must often be used by necessity.

There are a vast array of both specifically developed growth models and more general modelling approaches which have been used to describe growth in various anthropometric dimension over different ages. Some of these are reviewed briefly in Section 6.1. Several existing growth models appear to provide good levels of fit to 'typical' growth in height and weight at young ages (Jenss-Bayley, Count A-curve, Berkey-Reed). There are also many existing models which handle height from birth or infancy right through to final adult height (Count, Bock-Thissen, Preece-Baines, Karlberg, JPPS) which appear to fit adequately. Polynomials are often suitable for modelling growth over a short period of time, but are not recommended generally. They are limited in the range of curves they can accommodate, cannot model data approaching asymptotes, for example height near maturity, and are also susceptible to 'edge effects'. Fractional polynomials (FPs) expand upon the range of curve shapes which conventional polynomials can provide so that asymptotes and points of infection better dealt with, but also suffer from many of the same deficiencies.

All of these modelling approaches impose a pre-determined algebraic form on the growth curve. In some instances, for example some of the well-specified multi-parameter models for height, this type of parametric approach may be perfectly suited to the specific application. However, in a more general situation the types of curve afforded by a parametric approach are often found to be unduly restrictive. As a result, the scope of the thesis is angled towards nonparametric growth modelling approaches, in particular the use of splines, which provide a greater degree of flexibility.

In Chapter 8 individual cubic smoothing splines are fitted to BMI growth data in the SWEDES dataset. When data are subject to missingness or sparsity, as is the case with this application, the fitting of subject-specific smoothing splines may require certain restrictions to be imposed on the amount of data required in order to obtain reliable fitted curves. In particular, when the objective is the identification of a specific feature of the growth curve, a reasonable density of data around the expected age of this growth feature should be ensured. However, whilst stronger data requirements should increase the likelihood of reliably fitted splines, this may also decrease the effective sample size. An assessment of this trade-off is one element of subjectivity which forms part of the model fitting process.

A further example of potential subjectivity is in the choice of the smoothing parameter. This determines how closely the fitted smoothing spline will follow the detail of the data and, for individual splines fitted only to the data of single subjects, need not take the same value for each individual. Indeed, allowing the smoothing parameter to vary across subjects permits the fitted curves to be 'fine-tuned' for each individual to give a, in some sense, 'optimal' fit. However, the level of subjectivity inherent in specifying 'optimal' subject-specific smoothing parameters makes automation of this process difficult, and to manually adjust the smoothing parameter for each subject would be very time consuming in a large dataset. In Chapter 8 a compromise approach is used whereby a stratified random subsample of the dataset is extracted, and for these individuals smoothing splines fitted using manually selected smoothing parameters. From observed trends

within the subsample, rules are created so that for the remainder of the dataset the smoothing parameter for each individual is specified as a function of the number of data points to which the curve is being fitted. Generally this approach is seen to work well. Whilst there will indubitably be some cases where subject-specific fine-tuning of the smoothing parameter would improve the fit of the curve somewhat, the benefits of the increased automatability of the process are great, and would be amplified further in larger datasets.

The ability to 'fine-tune' the smoothing parameter also means that curves with different degrees of smoothing can potentially be fitted to the same data points in order to meet differing objectives. Thus, for example, when fitting a smoothing spline to a given anthropometric variable, a certain degree of smoothing may be considered 'optimal' for the estimation of a feature at one age, whilst a differently smoothed curve may be thought preferable for estimating a different feature at a different age. This illustrates a further flexibility of the smoothing spline approach.

However, this level of subjectivity in the degree of smoothing may not always be desired. Although little attention is paid to them in this thesis, there do exist approaches, such as cross-validation, which allow automated smoothing parameter selection.

Once fitted, smoothing splines allow simple derivation of growth features. As smoothing splines are not restricted in the variables or age they can model, unlike many existing growth models, they can be used to model arbitrary anthropometric variables, affording great flexibility.

The related approach of regression splines is utilised in Chapter 9. Here, the knots at which the polynomial functions join, rather than being all the ages at which observations are made, are a smaller set of ages fixed in advance. Having common knot locations for each subject allows the regression spline fitting to be incorporated into a mixed model framework. The resulting semi-parametric mixed model approach can be easily implemented in standard statistical software and is found to be very effective in the fitting of subject-specific growth curves.

The equivalence between penalised smoothing and the optimal predictor in mixed modelling results in a unified approach to model estimation, but removes the previously discussed ability to 'fine-tune' smoothing. The best linear unbiased predictor (BLUP) approach to penalised regression spline fitting works well in the application of Chapter 9, but may not always provide an adequate level of smoothing for a given purpose.

Allowing a cubic population-average curve with cubic subject-specific deviations provides sufficient flexibility to model a wide range of curve shapes. Derivatives of both the population-average and subject-specific curves are easily calculated, and are themselves smooth continuous functions.

The fitted subject-specific regression splines generally fit the data very well. This even appears true for those subjects with sparse data, though obviously the goodness of fit in these instances must mainly be judged by conjecture. For these individuals, the model fitting process 'borrows' information from other subjects, so that the resultant curves are more strongly influenced by the population-average curve.

However, there is always the possibility of encountering individuals whose observed values lie

on a significantly different trajectory to other subjects, and hence from the population-average curve. In these situations the fitted subject-specific regression splines may not fit the data quite so well, although in the application in Chapter 9 there is little evidence of this being the case. More generally the presence of this issue must be viewed in terms of a trade-off against the benefits of having a common underlying trajectory in those with few observations where fitting a truly subject-specific curve would be problematic.

Thus both smoothing and regression splines are seen to be useful tools for the fitting of individual curves to growth data. This accords with the previous assertion of polynomials being adequate for modelling growth over short periods, as spline functions are effectively a series of 'polynomials modelling growth over short periods' joined together.

It is difficult to directly compare the two spline methods utilised in the thesis as the applications in Chapters 8 and 9 differ in terms of the ages at which growth is examined, the objective of the curve fitting, and the data which are used. However, it would be very interesting to apply the MI and cubic smoothing spline approach of Chapter 8 and the mixed model penalised regression spline approach of Chapter 9 to the same data and compare the fitted curves. One advantage the regression spline approach has over the smoothing spline approach is that the latter becomes less practical as sample size increases as it uses all the observations as knots, whereas the former uses a fixed number of knots.

10.4 Areas for future work

There are many ways in which the work presented in this thesis could be further extended.

As discussed previously, the naive multivariable analysis in Chapter 7 encounters problems due to collinearity between the childhood growth measurements at different ages and the complete-case analysis approach resulting in the exclusion of many subjects. In particular, if these individuals are not MCAR then bias may be introduced.

One simple approach to counter the issue of excluded subjects would be to use MI to impute the missing childhood growth data. Several completed datasets would be created, analysed individually, then the results combined, as has been described previously. The MI procedure could be conducted in a similar manner to that in Chapter 8.

Using MI would allow every subject to contribute to the analysis, so the reduced precision would be overcome. The main assumption for a MI analysis to provide unbiased results is for the missing observations to be MAR. As this is a weaker assumption than that required for the complete-case analysis to provide unbiased results, its validity is more likely.

A comparison of the results using MI to those reported in Chapter 7 would be illustrative as to the effect the missing data had on the original results. However, if the analysis model is parameterised in terms of childhood BMI (as opposed to BMI velocity) then it is likely that the

previously acknowledged presence of collinearity would remain. Thus it may be more informative to examine the model reparameterised in terms of BMI velocity.

A variety of models which have been developed to describe growth are briefly described in Section 6.1. Whilst these models are not used in the later applications in the thesis which involve the modelling of growth, they remain in widespread use elsewhere. Thus a more detailed and formal comparison of the different growth models may be propitious.

This would necessitate a much larger sample of subjects to whom the various growth curves would be fitted. The sample would need to include subjects with a wide variety of curve shapes. As a first stage, an attempt could be made to assess and categorise the curve shapes of individuals, then a stratified random sample could be taken based on this categorisation.

The goodness of fit of each growth curve for each individual could then be assessed more formally using the deviance of the model. For nested models, significance tests can be used to examine the importance of the extra parameter(s). Otherwise, the trade-off between reduced deviance and the extra degrees of freedom in models can be assessed using the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) to examine whether the extra complexity can be justified.

However, goodness of fit is not the only criterion for assessment of a model, and an objective comparison considering only this could be misleading. More formal approaches to the comparison of ease of fitting, data requirements, the interpretability of parameters, and the automatability of the procedure should also be considered.

Previously published studies, for example those of Berkey [30] (Jenss-Bayley and Count A-curve), Berkey and Reed [34] (Jenss-Bayley and Berkey-Reed), Jolicoeur *et al* [151] (Preece-Baines and JPPS) and Ledford and Cole [152] (Preece-Baines and JPPS), have formally compared selected models to each other. However, as far as I am aware, there are no published formal comparisons between so many of the available growth models.

The main epidemiological findings of the thesis involve the relationships between the AP, the AR and later obesity. Whilst broadly similar findings regarding the AR have previously been published elsewhere [82, 83, 84, 85, 86, 87, 88, 165, 172], there appear to be no equivalent studies concerning the AP. As there is thus no means of comparison for the results found in the thesis, it is imperative that further analyses of the relationship between the AP and later obesity are conducted in order to examine the robustness of the association. These studies should ideally consider individuals from across a range of geographical locations.

Further to this is interest in the relationship between the AP and the AR, and how interaction between the two may affect later obesity. Due to restrictions in the data, this cannot be properly examined in the thesis. To do this would require a dataset where both the AP and the AR can be identified for each individual, as well as one or more measures of later obesity. The relative timings of these two features of the BMI growth curve could then be investigated. Of particular interest is

the previously described observation of both subjects with a later AP and subjects with an earlier AR being the most likely to exhibit later obesity when the two growth features are considered separately. Use of a dataset in which both the AP and the AR are identified within the same subject would allow an assessment of whether or not it is the same individuals with both a later AP and an earlier AR. Further research of this nature concerning the AP could provide valuable insight into BMI development through childhood.

Whilst relationships have been seen between the AP, the AR and later obesity, the transferal of these findings into interventions to reduce obesity remains difficult. In particular, it remains unclear whether it is possible to manipulate the timing of either the AP or the AR. Additionally, assuming this manipulation is possible, it is unclear whether the, say, artificially delayed AR would lead to a reduced risk of later obesity in the same manner in which a naturally occurring AR at that age would.

It is acknowledged that adiposity within an individual can be manipulated somewhat by alterations in their energy balance [181], through either the consumption of fewer calories, the expenditure of a greater number of calories, or both. However, research specifically into factors affecting the timing of the AR [90] found no association between any of the measured dietary variables (protein, fat, carbohydrates and energy) and timing of the AR. Instead, parental obesity was found to be an associated with an earlier AR, which perhaps lends itself less favourably to use as an intervention. To my knowledge there is no corresponding research into factors affecting the timing of the AP, so further research is thus required regarding factors affecting the timing of both features of the BMI growth curve. Of particular interest with regards to the timing of the AP is the developmental stage of the infant. It is plausible that the decrease in adiposity seen following the AP may be influenced by the progression to a more mobile developmental stage.

If an intervention was found which was believed to have the potential to manipulate the timing of the AP or the AR, as the timings of both features naturally differ between subjects it would be impossible to assess on an individual level the effect of the intervention on the timing. The ideal approach to examining this would be via a randomised controlled trial, where subjects are randomised to having their AP or AR either artificially accelerated, delayed or neither. This would likely necessitate near-continuous monitoring of BMI and, for example, appropriate modification of the energy balance for each individual. Timing of the growth feature being considered could then be compared across intervention groups to assess the short-term effect of the intervention, and a measure of later obesity could be compared across intervention groups to assess the long-term effect. Whilst this approach could be fruitful, whether such a precise manipulation of the growth trajectory is possible remains debatable.

Even if this level of manipulation is possible, it may be considered undesirable. Interventions whereby children are encouraged to eat more healthily or to exercise more are commonplace and widely accepted, but one in which the aim is explicitly to alter the trajectory of growth, even if it is only using the same tools of reduced calorific intake and enhanced calorific expenditure, may

seem somewhat less palatable. In particular, the potential to manipulate growth in infancy around the period of the AP should be considered only with the utmost caution.

10.5 Concluding comments

The scenario considered in this thesis, of relating childhood growth to a later health outcome, can be seen as just one example of relating longitudinal data to some distal outcome. Further examples of this include relating systolic blood pressure profiles to risk of myocardial infarction, or occupational exposures over a working lifetime to risk of various lung conditions. This setting need not even be confined to health, and similar scenarios could be envisaged across a range of alternative subject areas. For example, it may be wished to examine the relationship between repeated measures of educational attainment though childhood and adult income.

The same issues of balanced or unbalanced data structure, collinearity between measurements, and missing or sparse data would be present in these applications. As the statistical approaches used throughout this thesis are not health-specific, there is no reason that this work cannot be used to inform the approach to analysis in alternative settings. In particular, when growth curves are fitted in Chapters 8 and 9, using individual cubic smoothing splines and cubic penalised regression splines within a mixed model framework respectively, the decision to use nonparametric modelling approaches makes the overall analysis approach far more generalisable. Clearly, models which have been developed specifically to describe the growth of some human dimension over some period of childhood, such as those discussed in Section 6.1, are unlikely to be suitable in this more general setting. As splines can be used to model arbitrary variables they provide a reasonable solution to many such problems.

Appendix: Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size

There follows a statistical opinion article written with Tim Cole for Ultrasound in Obstetrics and Gynecology entitled 'Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size' [3].

Ultrasound Obstet Gynecol 2007; 29: 6–13
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/uog.3911

Statistical Opinion

Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size

INTRODUCTION

Many fetal size variables, for example head measurements, abdominal measurements and femur length, increase over the course of gestation. Reference intervals (RIs) and centile charts provide a means of assessing these measurements, at a given gestational age (GA) or across a range of GAs, respectively, and are tools of great importance in clinical medicine.

RIs (sometimes, misleadingly, called 'normal ranges') represent the interval between a pair of symmetrically placed extreme centiles (such as the 5th and 95th for a 90% interval) of a size variable, denoted y, at a given GA. Centile charts plot the values of y corresponding to one or more RIs against the relevant GA over a range of GAs. In the field of fetal size, values which lie outside the RI are regarded as extreme and may indicate the presence of a disorder such as intrauterine growth restriction[1] or macrosomia[2]. More informative, however, than this forced dichotomy is the calculation of a value's centile position, or Z-score, relative to the reference population, estimated from knowledge of the distribution of y at a given GA. For a given observation, the proximity of the centile position to 0% or 100% (alternatively the magnitude and sign of the Z-score) is then a measure of how extreme the observation is compared to the reference data at that GA. A centile position above 50% (equivalently a positive Z-score) signifies a measurement greater than average for that GA, and a centile position below 50% (or a negative Z-score) one less than average.

While recent years have seen the publication of a variety of strategies for the construction of RIs, incorrect methods have still been used for fetal measurements of all kinds[1]. The choice of suitable methodology in this field is especially crucial as inaccurate centiles may lead to false conclusions regarding the development of the fetus, resulting in suboptimal clinical care.

In an article in this issue of the Journal, Sherer *et al.*[3] construct centile charts of the axial cerebellar hemisphere circumference (CHC) and area (CHA) through gestation using one such method, based upon regression modelling of both the mean and the standard deviation (SD) across GA, as detailed by Altman and Chitty[4] and Royston and Wright[1].

It is the aim of the present article to further examine the statistical approach used by Sherer *et al.*[3], while taking a more general look at the problem of constructing GA-related RIs and considering alternative approaches to this problem. Techniques for longitudinal data, where each subject contributes repeated observations, as opposed to cross-sectional data, where they contribute only one, require a different approach and are not considered here. Further information on this area can be found in, for example, Royston and Altman[5] and Royston[6].

While many of the techniques explored here could be, and indeed have been, used in the context of anthropometric measurements, the focus here is on applications in the field of fetal size.

THE GENERAL PROBLEM

Prior to the statistical analysis, many RIs and charts for fetal size are already flawed by weaknesses in study design. As with any study, the choice of an appropriate sample is of great importance. While some published studies use routinely collected data, resulting in the inclusion of multiple observations on some fetuses, Altman and Chitty[4] note that these fetuses are likely to be those with clinical indications, introducing bias to the sample. They advocate collecting data specifically for the purpose of developing the RI, with each fetus being included only once. Within this framework it is important to have as unselected a sample as possible because reference data should relate to 'normal' fetuses. Altman and Chitty[4] suggest that it is reasonable to exclude fetuses subsequently found to have a congenital abnormality, though they recommend the inclusion of neonatal deaths and fetuses large or small for dates at birth where this is not the case. Maternal conditions which could affect fetal growth are also deemed reasonable exclusion criteria.

While imprecise estimates of the RI will be obtained when the sample size of the dataset is too small[1], it is not easy to accurately specify appropriate sample sizes. In particular, when interest is focused on the extreme centiles, as is often the case, several hundred observations may be necessary to obtain estimates at an appropriate level of precision.

There are a variety of available statistical approaches for the calculation of RIs, the most important of which are to be reviewed presently. The method needs to produce reference centiles which change smoothly with GA and provide a good fit to the data. While clearly these requirements are essential, it is also preferable, for the sake of general usability and accessibility, to maintain as simple a statistical model as possible. Accordingly, the choice of approach must strike a balance between these conditions. It is also desirable that tools are available for

calculating the relevant centile positions and Z-scores for any further measurements, which again should be as user-friendly as possible in their application. Not only is the calculation of Z-scores useful on an observation-specific level, it has also been shown to be instrumental in the assessment of chart comparison[7] and quality control[8].

MEAN AND SD MODEL

The statistical approach followed by Sherer *et al.*[3], here referred to as the 'mean and SD model', is one which has been found to be sufficiently general to cope with a wide range of fetal measurements available from ultrasound scanning[1]. Generally, under the assumption that at each GA the measurement of interest has a Gaussian (or normal) distribution with mean and SD that vary smoothly with GA, the centile curve at a given GA may be calculated by:

$$\text{centile}_{GA} = \text{mean}_{GA} + K \times \text{SD}_{GA} \quad (1)$$

where mean_{GA} and SD_{GA} are, respectively, the mean and SD at the required GA, and K is the desired normal equivalent deviate (NED). The NED takes a value corresponding to the proportion of the standard normal distribution (with mean of 0 and SD of 1) lying to the left of it. For example, the 50th centile (with a proportion of 0.5 of the standard normal distribution to the left of it) has an NED of 0, while the determination of a 90% reference range (i.e. the 5th and 95th centile curves) would require $K = \pm 1.645$.

The 'mean and SD model' approach aims to find functions that adequately represent how the mean and SD change with GA, allowing any desired centile curve to be readily calculated by appropriate choice of K.

Firstly the mean is modeled by fitting a polynomial curve to the raw data by means of least squares regression analysis. Royston and Wright recommend the initial use of a cubic polynomial $(a + bt + ct^2 + dt^3$, where, for simplicity, GA is represented by $t)$[1]. If the cubic coefficient, d, is not significantly different from zero (approximately if d is less than twice its SD), a quadratic polynomial $(a + bt + ct^2)$ should be fitted with the same assessment made of the quadratic coefficient, c. The process should be repeated until no further removal of terms is possible. While quadratic or cubic curves will often give a good fit to the data, Altman and Chitty[4] suggest the linear-cubic model $(a + bt + dt^3)$ as a good alternative for fetal size data. It is advocated that the choice of curve be based not only on statistical significance, but also that the quality of fit to the data and esthetic appearance, especially at the extremes of GA, should be taken into account. Sherer *et al.* found a linear model $(a + bt)$ to be sufficient for the CHC curve and a quadratic polynomial to be suitable for CHA[3].

Once a suitable mean model has been decided upon, attention can turn to the variability in the data. Residuals from the fitted mean model (observed value minus predicted value) should be calculated and plotted against GA to show if and how variability changes with GA[4].

Previously, modeling of the variability was not often considered, even though in the field of fetal size SD almost always changes with GA[9]. While other methods have been proposed[10], the approach most frequently used is that of Altman[9]. It follows – from the assumption that the variable under consideration is normally distributed at all GAs – that the residuals from the mean model should also be normally distributed. This in turn means that the absolute residuals (residuals with the sign removed) have a half normal distribution. As the mean of a half standard normal distribution is $\sqrt{(2/\pi)}$, the mean of the absolute residuals multiplied by $\sqrt{(\pi/2)}$ is an estimate of the SD of the residuals. Hence if the SD is not reasonably constant over GA, predicted values from a regression of the absolute residuals on age multiplied by $\sqrt{(\pi/2)}$ will give age-specific estimates of the SD of the residuals, and hence of y.

An alternative formulation for Altman's approach favored by Royston and Wright[1], and employed in this instance by Sherer *et al.*[3], is to produce 'scaled absolute residuals' (SARs) by multiplying the absolute residuals by $\sqrt{(\pi/2)}$. The SARs are then regressed on GA, the predicted values from which again estimate the SD of the residuals.

Under either formulation, if the absolute residuals, be they scaled or unscaled, show no trend with GA, the SD is estimated as the SD of the unscaled original residuals (observed value minus predicted value). If there is a trend, polynomial regression is needed to estimate an appropriate curve in the same way as for the mean. Altman suggests that it is unlikely that a curve more complex than quadratic is required for a satisfactory fit to the SD[9]. Superimposing $\pm 1.645 \times$ SD on the residual plot is useful to see how well the SD has been modeled, as approximately 90% of the observed residuals should fall within these limits. Sherer *et al.*[3] found the CHC SARs to be suitably represented by a linear relationship with GA, while those for CHA required a cubic polynomial.

As the regression analysis to estimate the mean should really take into account any increase in SD with GA, at this juncture the mean model can be refitted using the reciprocal of the square of the estimated SD as weights. However, Altman and Chitty report that the effect of refitting is almost always rather small[4].

A useful tool in assessing model fit are Z-scores (also known as SD scores), defined as:

$$Z = \frac{\text{observed } y \text{ value} - \text{mean}_{GA}}{\text{SD}_{GA}} \quad (2)$$

where mean_{GA} and SD_{GA} are, respectively, the mean and SD given by the model for the GA at which the observation is made. Hence Z-scores represent the observed values expressed on a standard normal scale (with a mean of 0 and SD of 1), with the mean and SD adjusted for GA.

Altman and Chitty[4] recommend three methods of evaluation for the goodness of fit, all of which Sherer

et al.[3] appear to have carried out. These methods will be illustrated using data on fetal biparietal diameter (BPD). A subset of 850 of the 19 647 fetuses analyzed by Salomon *et al.*[11] were fitted with a 'mean and SD model' in the standard manner, as outlined above, resulting in a cubic mean model and a linear SD model. Firstly, a plot of the Z-scores against GA should be checked for the existence of any patterns. The Z-scores should be randomly scattered about zero at all GAs, with any deviation from this indicating that the mean curve may require modification. This is shown in Figure 1 for the example dataset, with the BPD Z-scores appearing to adhere to this stipulation.

Secondly, a normal plot (essentially a scatterplot of the actual data values plotted against the 'ideal' values from a normal distribution) can be used to check that the Z-scores have a close to normal distribution. This is signified by a roughly straight line but can be confirmed more formally using the Shapiro–Wilk W test or Shapiro–Francia W' test. Figure 2 shows that in the example dataset the BPD Z-scores do have a close to normal distribution and this is corroborated by both the Shapiro–Wilk W and Shapiro–Francia W' tests having P of 0.998.

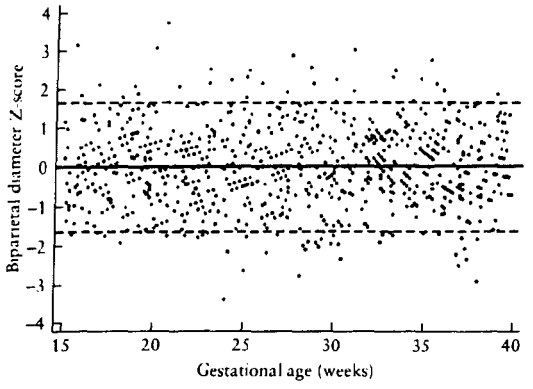

Figure 1 Plot of calculated Z-scores against gestational age in the example dataset.

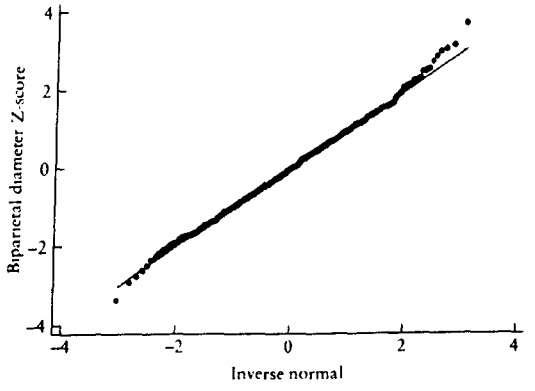

Figure 2 Normal plot of calculated Z-scores in the example dataset.

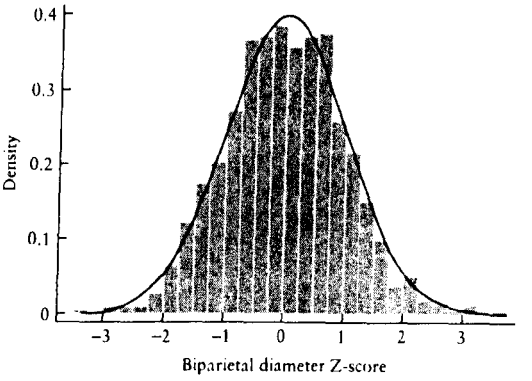

Figure 3 Histogram of calculated Z-scores in the example dataset with overlaid standard normal distribution.

Finally, the appropriate proportion of observations should fall between and outside fitted centiles, for example approximately 90% of Z-scores should lie between $Z = -1.645$ and $Z = +1.645$. Deviation from this may imply that a higher-order polynomial curve for the SD is needed. For the example dataset, lines corresponding to a BPD Z-score of ± 1.645 have been plotted on Figure 1. A brief examination suggests that approximately 90% of the data lie between the lines, with calculations confirming that 4.9% of the data lie below $Z = -1.645$ and 4.2% above $Z = +1.645$ (compared to an expected 5% for each). It is unlikely that the values will both be exactly 5%, so figures such as these indicate an adequate level of fit.

This aspect of the data can be further examined in a plot such as in Figure 3, a histogram of the Z-scores with an overlaid standard normal distribution. If the model fits well then the histogram should match up with the standard normal distribution, meaning that the expected and observed centiles lie at the same values. Given the sample size of the dataset, the histogram for the BPD data shows a close to standard normal distribution, indicating an adequate model fit.

Once a satisfactory model has been determined, the centile curves for the desired reference interval may be calculated by substituting the expressions for the mean and SD into equation (1). The Z-score for any new individual may be calculated using equation (2) and its centile obtained using the inverse normal distribution. Finally, the calculated centiles should be superimposed on the scatter diagram of observed values against GA to ensure a suitable fit.

Besides the study currently under consideration, this approach to the construction of RIs has been widely used in the field of fetal measurements. Altman illustrated his absolute residual approach by developing reference centiles of fetal foot length[9]. Chitty *et al.* constructed new charts for fetal head circumference, BPD and other head dimensions[12], fetal abdominal circumference and area[13], and fetal femur length[14]. Royston and Wright[1] estimated RIs for fetal head circumference (using the same

data as Chitty *et al.*[12]), hemoglobin concentration and kidney volume. Salomon *et al.* constructed new reference charts and equations for fetal biparietal diameter, head circumference, abdominal circumference and femur length[11].

Extensions to the mean and SD model

Several extensions to the basic 'mean and SD model' approach described above have been posited as ways to improve the performance of the method. The use of logarithmic transformations and fractional polynomials is described below.

Mean and SD model with logarithmic transformation

Many size measurements tend to follow a skewed normal distribution at a given GA, usually a positive skew where the right tail of the distribution is longer than the left. While this clearly conflicts with the assumption that at each GA the data come from a population with a normal distribution, it can often be overcome by the application of a logarithmic transformation. This same solution will also increase the ease with which a model can be fitted if the SD of the original measurements increases rapidly with GA.

Royston suggests initially attempting to fit the mean model to the original measurements[10]. If the residuals from this model show a positive skew then a logarithmic transformation should be performed on the original values, y, and the model refitted on $\log(y)$. If residuals from the refitted model are once again skewed, it is then recommended to try using a modified logarithmic transformation of the form $\log(y + C)$, where C is positive if the new residuals are negatively skewed, and negative otherwise. A polynomial model of the same degree as the optimal model for $\log(y)$ is then repeatedly fitted, with the value of C varied until the highest (i.e. least significant) P-value for the normality test of the residuals is reached. Often a value of C will be found that makes the distribution of residuals satisfactorily normal.

Once acceptable residuals from the mean model have been obtained, the rest of the 'mean and SD model' fitting procedure is continued as before. However, it is important to back-transform the curves once the model has been finalized using the antilog (exponential if a natural logarithmic transformation was used), also remembering to subtract C for a modified logarithmic transformation. While this simple procedure can easily cope with the problem of skewed data, Altman and Chitty report that very few fetal size measurements require transformation[4].

The effect of the logarithmic transformation is illustrated here using data on birth weight in 58 940 neonates as analyzed by Salomon *et al.*[15]. Figure 4, a scatterplot of birth weight against GA at birth, shows a marked increase in variability with GA and also suggests a slight positive skew to the data at a given GA. In Figure 5 the birth weights have undergone a natural logarithmic transformation, resulting in a more constant variance over

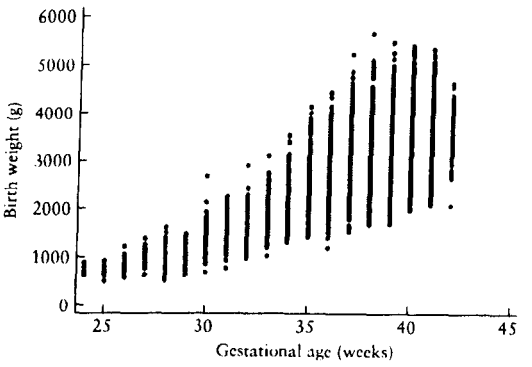

Figure 4 Scatterplot of fetal birth weight against gestational age in the example dataset.

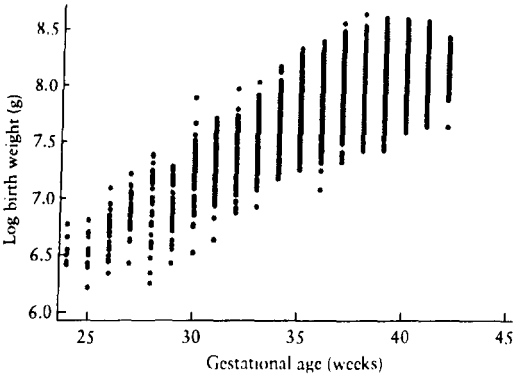

Figure 5 Scatterplot of logarithmically transformed fetal birth weight against gestational age in the example dataset.

GA with any evidence of skew being removed. The fitting of a 'mean and SD model' to this transformed data should now be relatively more simple.

The modification of the 'mean and SD model' by the addition of a logarithmic transformation is somewhat less common than the unmodified version in the fetal size literature. Royston used a modified logarithmic transformation in an example concerning fetal triglycerides[10]. After fitting an initial quadratic mean model, positive skew was identified in the residuals. A logarithmic transformation was performed on the original values and a quadratic mean model fitted on $\log(y)$. However, this introduced negative skewness, so a modified logarithmic transformation was utilized. Wright and Royston, in an example regarding fetal abdominal circumference, also used a logarithmic transformation[16].

Mean and SD model using fractional polynomials

Fractional polynomials (FPs), formalized by Royston and Altman[17], extend the range of models afforded by conventional polynomials by allowing parameters to also

take fractional powers. Whilst a conventional polynomial is of the form

$$a + bt + ct^2 + dt^3 + \ldots.$$

FPs are defined as

$$a + bt^{p_1} + ct^{p_2} + dt^{p_3} + \ldots.$$

where p_1, p_2, p_3.... are chosen from a predetermined set, usually taken to be {−2, −1, −0.5, 0, 0.5, 1, 2, 3}. Here a value of −1 represents the inverse of t and 0.5 the square root of t. By convention the power 0 is defined to be $\log(t)$. If one or more power(s) in the model is/are duplicated then the model will include 'repeated powers', whereby the second term is multiplied by $\log(t)$. As an example, an FP of degree 3 with powers (0, 2, 2) (i.e. $p_1 = 0$, $p_2 = 2$ and $p_3 = 2$) is of the form

$$a + b\log(t) + ct^2 + dt^2\log(t).$$

Estimation of the best fitting FP for a given dataset involves both a systematic search for the best power or combination of powers from the permitted set, and estimation of the associated parameter coefficients. This selection process includes fitting a model for each combination of powers in the permitted set. This means, for example, that fitting a fractional polynomial of degree 2 (i.e. of the form $a + bt^{p_1} + ct^{p_2}$) using the standard set detailed above would involve fitting a different model for each of the 36 permissible combinations of powers. From these models the one with the lowest residual standard deviation is chosen to be optimal.

FPs give at least as good a fit to data as a conventional polynomial of corresponding degree and often offer a better fit than conventional polynomials of higher degree. Royston and Wright recommend the use of FPs for modeling the mean or SD curve if a quartic or quintic polynomial is required for an adequate fit to the data[1].

Over recent years the use of FPs in the construction of RIs has become more popular. Kurmanavicius *et al.*[18] created ranges for BPD, occipitofrontal diameter, head circumference and cephalic index using this method, although in each case, bar the cephalic index SD, the best fitting fractional polynomial was found to be a conventional polynomial. Kurmanavicius *et al.*[19] also modeled mean abdominal diameter, abdominal circumference and femur length using FPs, with only femur length SD taking a fractional model. Size charts for fetal bones (radius, ulna, humerus, tibia, fibula, femur and foot) were presented by Chitty and Altman after fitting FPs, with all but one mean model, though none of the SD models, taking fractional form[20].

ALTERNATIVE METHODS

Besides the 'mean and SD model', Wright and Royston[16] report the other most widely applied statistical approaches for estimating GA-specific reference intervals in practice to be those of smoothed crude centiles[21] and LMS[22–24], as detailed below.

Centile curves based on direct centile estimates

For a sufficiently large dataset (several hundred observations at each week of gestation, according to Altman and Chitty[4]), one intuitive approach is to calculate empirical estimates for each desired centile at a given GA. While the curves produced by joining these values will be rough, even for large sample sizes, smoother curves can be obtained by considering 'windows' of GAs instead of each GA separately. Here, increasing window size will increase smoothness, though information can easily be lost through oversmoothing[16].

A more formalized version of this approach, with a second stage involving centile smoothing based on the technique of Cleveland[25], is presented by Healy *et al.*[21]. This approach makes no assumption about the nature of the distribution of measurements at a given GA but takes advantage of the knowledge that both the centiles themselves and the intervals between centiles at a fixed GA should behave smoothly.

In the first stage, observations are ordered by GA and the first k, where k usually represents 5–10% of the total data, selected. Initial empirical centile estimates at the required values, for example 5%, 10%, 25%, 50%, 75%, 90% and 95%, are calculated from these k measurements by sorting and counting, and then plotted against the median GA of the k observations. This 'window' of k observations is then moved on to encompass measurements 2 to $k+1$, then 3 to $k+2$, etc., with the same estimation procedure repeated on each occasion, until all observations have been included.

The initial centile estimates will be irregular, so the second stage smoothes them to provide more usable centile curves. It is first assumed that each centile curve can be approximated by a polynomial of degree p, so that y_i, the smoothed value of the ith centile, is given by

$$y_i = a_{0i} + a_{1i}t + a_{2i}t^2 + \ldots + a_{pi}t^p \qquad (3)$$

where t again represents GA. Now consider the proportion corresponding to the ith centile (for example 0.5 for the 50th centile) and define z_i as its NED, similarly to previously.

The coefficients a for a fixed j are then modeled as a polynomial in z_i, so that

$$a_{ji} = b_{j0} + b_{j1}z_i + \ldots + b_{jq_j}z_i^{q_j} \qquad (4)$$

where the degree q_j of the polynomial may differ from one value of j to another. This restricts the distance between centiles and prevents the resulting curves from crossing. Combining equations (3) and (4) gives a linear model for the centile values which can be fitted by least squares regression. It follows that for any observation a corresponding Z-score can be calculated by solving a polynomial equation, though the order of the polynomial

may realistically prohibit this. Goodness of fit should be judged by counting the points falling between adjacent centiles. This method was applied by Wright and Royston to measurements of fetal abdominal circumference and provided an adequate fit[16].

LMS

The LMS method, introduced by Cole[22,23] and refined by Cole and Green[24], provides a general method for fitting smooth centile curves to reference data. It utilizes the power transformation family of Box and Cox[26] to allow the skewness of the measurement distribution, as well as the median and variability, to vary with age. These three features of the distribution are summarized by the parameters λ, μ and σ, the initials of which (L, M and S) give rise to the name of the method. The original form[22,23] necessitated age to be split into groups – an arbitrary procedure whereby different groupings would produce different centile curves. This subjective stage was removed by Cole and Green[24] through the addition of a nonparametric aspect. Owing to the superiority of the later version, only this is detailed here.

As previously asserted, many size measurements follow a skewed normal distribution. The use of a suitable power transformation, which stretches one tail of the distribution and shrinks the other, can remove this skewness and 'normalize' the data. One such family of transformations, proposed by Box and Cox[26], is used in the LMS method, with the optimal power at a given GA calculated from the data to completely remove skewness in the distribution. As skewness changes with GA, the calculated power also changes.

Given a variable of interest y with median μ and a power transformation so that y^λ (or $\log(y)$ if $\lambda = 0$) is normally distributed, we consider the transformed variable

$$x = \begin{cases} \dfrac{(y/\mu)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \quad (5a) \\ \log\left(\dfrac{y}{\mu}\right) & \text{if } \lambda = 0 \quad (5b) \end{cases}$$

based on the Box–Cox transformation[26]. This transformation maps the median μ of y to $x = 0$ and is continuous at $\lambda = 0$. For $\lambda = 1$ the SD of x is the coefficient of variation (CV) of y, and this remains approximately true for all moderate values of λ[24]. The optimal value of λ now minimizes the SD of x.

Denoting the SD of x (and CV of y) by σ, the Z-score (or SD score) of x (and hence y) is given by:

$$z = \frac{x}{\sigma} = \begin{cases} \dfrac{(y/\mu)^\lambda - 1}{\lambda\sigma} & \text{if } \lambda \neq 0 \quad (6a) \\ \dfrac{\log\left(\dfrac{y}{\mu}\right)}{\sigma} & \text{if } \lambda = 0 \quad (6b) \end{cases}$$

and is assumed to take a standard normal distribution.

Assume that the distribution of y varies with GA, t, and that λ, μ and σ at t are read off smooth curves $L(t)$, $M(t)$ and $S(t)$. Then

$$z = \begin{cases} \dfrac{\left(\dfrac{y}{M(t)}\right)^{L(t)} - 1}{L(t)S(t)} & \text{if } L(t) \neq 0 \quad (7a) \\ \dfrac{\log\left(\dfrac{y}{M(t)}\right)}{S(t)} & \text{if } L(t) = 0 \quad (7b) \end{cases}$$

Rearranging equation (7) shows that centile 100_α of y at t is given by

$$C_{100\alpha}(t) = \begin{cases} M(t)[1 + L(t)S(t)z_\alpha]^{\frac{1}{L(t)}} & \text{if } L(t) \neq 0 \quad (8a) \\ M(t)\exp(S(t)z_\alpha) & \text{if } L(t) = 0 \quad (8b) \end{cases}$$

where z_α is the normal equivalent deviate of size α. This shows that if L, M and S are smooth, then so are the centile curves.

Cole and Green then introduce a penalized likelihood function, derived from equation (7), with three integrals providing roughness penalties for the curves $L(t)$, $M(t)$ and $S(t)$[24]. The extent of these penalties, and hence the smoothness of the curves, are controlled by three smoothing parameters, and these are the only parameters requiring specification in order to fit the model. However, 'equivalent degrees of freedom' (EDFs), calculated for each fitted curve as a function of these smoothing parameters, give a more usable measure of the extent of the smoothing.

The illustrative examples of Cole and Green[24], although not from the field of fetal measurements, show values of the L curve falling well below zero. This indicates the presence of considerably more skew that a log transformation would remove and the extent of variability of the L curves with age reinforces the notion that transformation using a single power for all ages is inappropriate.

While examples of the application of the LMS method for fetal size do not abound, using the same fetal abdominal circumference data as Chitty *et al.*[13], Wright and Royston[16] used this approach to fit centile curves to good effect.

DISCUSSION

There are several viable methods available, of varying complexity, for constructing age-related RIs and centile charts. Ideally, methods should be understandable by clinicians, and the results easy to use, even without a statistical computer package. It is desirable that any published method should provide the potential user with the means of calculating the corresponding Z-score and centile for a given measurement. The mere provision of a mean model or centile chart, regardless of the quality, is not really adequate. Any approach must also be sufficiently flexible to be applicable successfully to many sets of data. Unfortunately, none of the methods currently

available fulfills all these criteria, so it is unlikely that any one would be appropriate in all circumstances.

In the simplest setting, if it is plausible that the observed measurements at each GA do indeed come from a population with a normal distribution and, in addition, the variance across the age range is constant, then the use of conventional polynomial regression may be justified. However, the strict adherence to these assumptions is unlikely, meaning that the model may not produce sufficiently reliable reference intervals. Slightly more realistic is the acknowledgment that variance is likely to change over the age range. This feature can be included by fitting the 'mean and SD model' as described previously, though again the assumption of an underlying normal distribution is not always tenable. This issue can often be dealt with by the addition of a (modified) logarithmic transformation prior to the model fitting to correct any skew (distribution asymmetry). However, this approach still suffers from the well-known limitations of polynomial curve shapes. This last hurdle can be overcome by the relaxation of the restrictions imposed on the powers of the polynomial, allowing the use of FPs. As FPs give at least as good a fit to data as a conventional polynomial of corresponding degree, and as the fitting of FPs with most basic statistical software is relatively straightforward, there seems little reason not to adopt them as standard.

All of these variations on the 'mean and SD model' benefit from being relatively conceptually simple and easy to use, with the necessary techniques available in most basic statistical packages. The resulting centile curves and *Z*-scores can be expressed as explicit formulae, meaning that the centile position of any individual is easily obtainable. While the method as described here is adequate for most fetal measurements, there are some cases that cannot be handled properly by this approach. It is important to emphasize the strong assumption that at each GA the data come from a population with a normal distribution. While skewed data may sometimes be corrected by a log transformation, this is not always successful, with time-varying skewness especially difficult to accommodate. Even after transformation, kurtosis (a non-normal distribution shape) may remain in the data, again in contravention of the assumption. Variables with a complex curve shape beyond those available from conventional (or even fractional) polynomials may also require alternative techniques.

The method of producing centile curves based on empirical centile estimates as described by Healy *et al.* makes no assumption about the nature of the distribution of measurements at a fixed GA, which is an appealing feature[21]. This approach provides a flexible way of constructing centile curves that is capable of handling many patterns of growth due to the lack of a pre-specified functional form. However, there are some drawbacks. Experience is needed to find the best ways of choosing the values of the adjustable parameters involved, and clearly there is some degree of subjectivity here. The estimation of the centile values of further observations is not simple unless a very basic model has been fitted. There is also some vulnerability to outlying values affecting the derived centile values. We agree with the conclusion of Altman and Chitty that this is not a suitable method for the derivation of fetal size charts, except when other methods are unsuccessful[4].

The LMS method with penalized likelihood[24] is extremely flexible and widely applicable[16]. It is usually easy to produce convincing centile curves, regardless of the complexity of the curve shape, and time-varying skewness is easily dealt with. It also has the appealing by-product of the L, M and S curves which completely summarize the measurement's distribution over the age range and facilitate further investigation into the underlying structure of the data. Penalized likelihood provides an elegant solution for ridding the earlier method of its arbitrary categorization, with the smoothing of the three curves becoming an integral part of the likelihood maximization. Now the only arbitrariness in the procedure is the choice of the three smoothing parameters.

There are, however, some general problems with the smoothing approach. Where data are more sparse near the ends of the age range, 'edge effects' (spurious changes in the centiles) may be observed, though this can be avoided by truncating the data at each end. One major drawback of non-parametric estimators is the lack of a succinct formula with which to estimate further centile values. This means that centiles may only be displayed graphically or in tabular form. Finally, the assumption of normality following the Box–Cox transformation may be violated by the presence of kurtosis, for which the transformation does not adjust.

A more recently proposed generalization of the LMS approach, the LMSP method of Rigby and Stasinopoulos[27], uses the Box–Cox power exponential (BCPE) distribution to try to overcome the issue of kurtosis. A fourth parameter is introduced in the power transformation in order to account for the observed kurtosis in the distribution, and centile estimation proceeds in a manner not dissimilar to that of the conventional LMS method.

While for the first-time user application of the LMS method may appear a daunting task, the advent of specially designed programs such as the LMSChartmaker of Cole and Pan[28], as well as packages for the widely used general statistical programs, mean that with brief instruction this need not be the case.

Wright and Royston advise that a 'simple formula' to allow estimation of centile position for an individual is extremely valuable[16]. If, when considering the statistical approach to follow in light of requirements specific to the data under analysis, this requirement is deemed to be essential, then this would exclude both the LMS method and any approach based on empirical centile estimates. Of the methods examined here, this leaves only the parametric approach of the 'mean and SD model'. So the choice of approach is really reduced to the trade-off between the simplicity, usability and accessibility of the

inferior model provided by the parametric approach, and the superior but less user-friendly model provided by the LMS method.

CONFLICT OF INTEREST STATEMENT

T. J. C. provides the lmsChartMaker Light program as a free download, and earns royalties on the Pro version of the program, which contains more facilities.

R. J. Silverwood*† and T. J. Cole‡
†*Medical Statistics Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT and* ‡*MRC Centre for Epidemiology of Child Health, UCL Institute of Child Health, London, UK*
**Correspondence.*
(e-mail: Richard.Silverwood@lshtm.ac.uk)

REFERENCES

1. Royston P, Wright EM. How to construct 'normal ranges' for fetal variables. *Ultrasound Obstet Gynecol* 1998; **11**: 30–38.
2. Sokol RJ, Chik L, Dombrowski MP, Zador IE. Correctly identifying the macrosomic fetus: improving ultrasonography-based prediction. *Am J Obstet Gynecol* 2000; **182**: 1489–1495.
3. Sherer DM, Sokolovski M, Dalloul M, Pezzullo JC, Osho JA, Abulafia O. Nomograms of the axial fetal cerebellar hemisphere circumference and area throughout gestation. *Ultrasound Obstet Gynecol* 2007; **29**: 31–36.
4. Altman DG, Chitty LS. Charts of fetal size: 1. Methodology. *Br J Obstet Gynaecol* 1994; **101**: 29–34.
5. Royston P, Altman DG. Design and analysis of longitudinal studies of fetal size. *Ultrasound Obstet Gynecol* 1995; **6**: 307–312.
6. Royston P. Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal measurements. *Stat Med* 1995; **14**: 1417–1436.
7. Salomon LJ, Bernard JP, Duyme M, Buvat I, Ville Y. The impact of choice of reference charts and equations on the assessment of fetal biometry. *Ultrasound Obstet Gynecol* 2005; **25**: 559–565.
8. Salomon LJ, Bernard JP, Ville Y. Analysis of Z-score distribution for the quality control of fetal ultrasound measurements at 20–24 weeks. *Ultrasound Obstet Gynecol* 2005; **26**: 750–754.
9. Altman DG. Construction of age-related reference centiles using absolute residuals. *Stat Med* 1993; **12**: 917–924.
10. Royston P. Constructing time-specific reference ranges. *Stat Med* 1991; **10**: 675–690.
11. Salomon LJ, Duyme M, Crequat J, Brodaty G, Talmant C, Fries N, Althuser M. French fetal biometry: reference equations and comparison with other charts. *Ultrasound Obstet Gynecol* 2006; **28**: 193–198.
12. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 2. Head measurements. *Br J Obstet Gynaecol* 1994; **101**: 35–43.
13. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 3. Abdominal measurements. *Br J Obstet Gynaecol* 1994; **101**: 125–131.
14. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 4. Femur length. *Br J Obstet Gynaecol* 1994; **101**: 132–135.
15. Salomon LJ, Bernard JP, De Stavola BL, Kenward MG, Ville Y. Birth weight and size in France. Charts and equations. *J Gynecol Obstet Biol Reprod (Paris)* 2007; (in press).
16. Wright EM, Royston P. A comparison of statistical methods for age-related reference intervals. *J R Statist Soc A* 1997; **160**: 47–69.
17. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Statist* 1994; **43**: 429–467.
18. Kurmanavicius J, Wright EM, Royston P, Wisser J, Huch R, Huch A, Zimmermann R. Fetal ultrasound biometry: 1. Head reference values. *Br J Obstet Gynaecol* 1999; **106**: 126–135.
19. Kurmanavicius J, Wright EM, Royston P, Zimmermann R, Huch R, Huch A, Wisser J. Fetal ultrasound biometry: 2. Abdomen and femur length reference values. *Br J Obstet Gynaecol* 1999; **106**: 136–143.
20. Chitty LS, Altman DG. Charts of fetal size: limb bones. *BJOG* 2002; **109**: 919–929.
21. Healy MJ, Rasbash J, Yang M. Distribution-free estimation of age-related centiles. *Ann Hum Biol* 1988; **15**: 17–22.
22. Cole TJ. Fitting smoothed centile curves to reference data. *J R Statist Soc A* 1988; **151**: 385–418.
23. Cole TJ. The LMS method for constructing normalized growth standards. *Eur J Clin Nutr* 1990; **44**: 45–60.
24. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992; **11**: 1305–1319.
25. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979; **74**: 829–836.
26. Box GEP, Cox DR. An analysis of transformations. *J R Statist Soc B* 1964; **26**: 211–252.
27. Rigby RA, Stasinopoulos DM. Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. *Stat Med* 2004; **23**: 3053–3076.
28. Cole TJ, Pan H. LMSChartmaker. 2005; http://www.healthforallchildren.co.uk [Accessed 1 December 2006].

Bibliography

[1] G. E. P. Box. Robustness in the strategy of scientific model building. In R. L. Launer and G. N. Wilkinson, editors, *Robustness in Statistics*. Academic Press, New York, 1979.

[2] *LaTeX 2ε*. LaTeX3 Project Team, 2005.

[3] R. J. Silverwood and T. J. Cole. Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size. *Ultrasound. Obstet. Gynecol.*, 29(1):6–13, 2007.

[4] D. Kuh and Y. Ben-Shlomo. Introduction. In D. Kuh and Y. Ben-Shlomo, editors, *A Life Course Approach to Chronic Disease Epidemiology*. Oxford University Press, Oxford, 2004.

[5] D. Kuh, Y. Ben-Shlomo, J. Lynch, J. Hallqvist, and C. Power. Life course epidemiology. *J. Epidemiol. Community Health*, 57(10):778–83, 2003.

[6] Y. Ben-Shlomo and D. Kuh. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int. J. Epidemiol.*, 31(2):285–93, 2002.

[7] B. L. De Stavola, I. dos Santos Silva, V. McCormack, R. J. Hardy, D. J. Kuh, and M. E. Wadsworth. Childhood growth and breast cancer. *Am. J. Epidemiol.*, 159(7):671–82, 2004.

[8] D. J. P. Barker. *Mothers, Babies and Health in Later Life*. Churchill Livingstone, Edinburgh, 1998.

[9] A. Lucas, M. S. Fewtrell, and T. J. Cole. Fetal origins of adult disease - the hypothesis revisited. *Br. Med. J.*, 319(7204):245–9, 1999.

[10] I. dos Santos Silva, B. L. De Stavola, V. Mann, D. Kuh, R. Hardy, and M. E. Wadsworth. Prenatal factors, childhood growth trajectories and age at menarche. *Int. J. Epidemiol.*, 31(2):405–12, 2002.

[11] B. L. De Stavola, D. Nitsch, I. Dos Santos Silva, V. McCormack, R. Hardy, V. Mann, T. J. Cole, S. Morton, and D. A. Leon. Statistical issues in life course epidemiology. *Am. J. Epidemiol.*, 163(1):84–96, 2006.

[12] Y. Ben-Shlomo. Rising to the challenges and opportunities of life course epidemiology. *Int. J. Epidemiol.*, 36(3):481–3, 2007.

[13] R. C. Hauspie. Curve-fitting. In S. J. Ulijaszek, F. E. Johnston, and M. A. Preece, editors, *The Cambridge Encyclopedia of Human Growth and Development.* Cambridge University Press, Cambridge, 1998.

[14] F. E. Johnston. Somatic growth of the infant and preschool child. In F. Falkner and J. M. Tanner, editors, *Human Growth: A Comprehensive Treatise*, volume 2: Postnatal Growth; Neurobiology. Plenum, New York, 1986.

[15] S. J. Ulijaszek, F. E. Johnston, and M. A. Preece. General introduction. In S. J. Ulijaszek, F. E. Johnston, and M. A. Preece, editors, *The Cambridge Encyclopedia of Human Growth and Development.* Cambridge University Press, Cambridge, 1998.

[16] W. H. Dietz and T. N. Robinson. Use of the body mass index (BMI) as a measure of overweight in children and adolescents. *J. Pediatr.*, 132(2):191–3, 1998.

[17] J. C. K. Wells. Body composition in childhood: effects of normal growth and disease. *Proc. Nutr. Soc.*, 62(2):521–8, 2003.

[18] M. F. Rolland-Cachera. Prediction of adult body composition from infant and child measurements. In P. S. W. Davies, editor, *Body composition techniques in health and disease.* Cambridge University Press, Cambridge, 1995.

[19] J. M. Tanner. *Foetus into Man.* Castlemead Publications, Ware, 1989.

[20] J. Karlberg. The human growth curve. In S. J. Ulijaszek, F. E. Johnston, and M. A. Preece, editors, *The Cambridge Encyclopedia of Human Growth and Development.* Cambridge University Press, Cambridge, 1998.

[21] S. J. Ulijaszek. The secular trend. In S. J. Ulijaszek, F. E. Johnston, and M. A. Preece, editors, *The Cambridge Encyclopedia of Human Growth and Development.* Cambridge University Press, Cambridge, 1998.

[22] T. J. Cole. The secular trend in human physical growth: a biological view. *Econ. Hum. Biol.*, 1(2):161–8, 2003.

[23] T. J. Cole. Weight-stature indices to measure underweight, overweight, and obesity. In J. H. Himes, editor, *Anthropometric assessment of nutritional status.* Wiley-Liss, New York, 1991.

[24] T. J. Cole and P. J. Green. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med*, 11:1305–19, 1992.

[25] R. T. Benn. Some mathematical properties of weight-for-height indices used as measures of adiposity. *Br. J. Prev. Soc. Med.*, 25(1):42–50, 1971.

[26] T. J. Cole. Weight/height^p compared to weight/height^2 for assessing adiposity in childhood: influence of age and bone age on p during puberty. *Ann. Hum. Biol.*, 13(5):433–51, 1986.

[27] R. Lazarus, L. Baur, K. Webb, and F. Blyth. Adiposity and body mass indices in children: Benn's index and other weight for height indices as measures of relative adiposity. *Int. J. Obes. Relat. Metab. Disord.*, 20(5):406–12, 1996.

[28] M. J. R. Healy. Growth curves and growth standards - the state of the art. In J. M. Tanner, editor, *Auxology 88: Perspectives in the Science of Growth and Development.* Smith-Gordon, London, 1989.

[29] M. A. Preece and M. J. Baines. A new family of mathematical models describing the human growth curve. *Ann. Hum. Biol.*, 5(1):1–24, 1978.

[30] C. S. Berkey. Comparison of two longitudinal growth models for preschool children. *Biometrics.* 38(1):221–34, 1982.

[31] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis.* Springer, New York, 1997.

[32] R. M. Jenss and N. Bayley. A mathematical method for studying the growth of a child. *Hum. Biol.*, 9:556–63, 1937.

[33] E. W. Count. Growth patterns of human physique: an approach to kinetic anthropometry. *Hum. Biol.*, 15:1–32, 1943.

[34] C. S. Berkey and R. B. Reed. A model for describing normal and abnormal growth in early childhood. *Hum. Biol.*, 59(6):973–87, 1987.

[35] R. D. Bock and D. M. Thissen. Fitting multi-component models for growth in stature. *Proceedings of the 9th International Biometric Conference*, 1:431–42, 1976.

[36] J. Karlberg. On the modelling of human growth. *Stat. Med.*, 6(2):185–92, 1987.

[37] World Health Organisation. *Obesity: preventing and managing the global epidemic.* WHO Technical Report Series. World Health Organisation, Geneva, 2000.

[38] The Information Centre. *Statistics on Obesity, Physical Activity and Diet: England, 2006.* The Information Centre, 2006.

[39] Foresight. *Tackling Obesities: Future Choices.* Department for Innovation, Universities and Skills, London, 2007.

[40] BBC News. Obesity 'as bad as climate risk'. `http://news.bbc.co.uk/1/hi/health/7043639.stm`, viewed 14 October 2007.

[41] H. B. Hubert, M. Feinleib, P. M. McNamara, and W. P. Castelli. Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. *Circulation*, 67(5):968–77, 1983.

[42] A. Must, J. Spadano, E. H. Coakley, A. E. Field, G. Colditz, and W. H. Dietz. The disease burden associated with overweight and obesity. *JAMA*, 282(16):1523–9, 1999.

[43] E. E. Calle and R. Kaaks. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat. Rev. Cancer*, 4(8):579–91, 2004.

[44] G. K. Reeves, K. Pirie, V. Beral, J. Green, E. Spencer, and D. Bull. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *Br. Med. J.*, 335(7630):1134, 2007.

[45] K. K. Ong, M. L. Ahmed, P. M. Emmett, M. A. Preece, D. B. Dunger, and the Avon Longitudinal Study of Pregnancy and Childhood Study Team. Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. *Br. Med. J.*, 320(7240):967–71, 2000.

[46] T. Cole. The life course plot in life course analysis. In A. Pickles, B. Maughan, and M. Wadsworth, editors, *Epidemiological Methods in Life Course Research.* Oxford University Press, Oxford, 2007.

[47] C. L. Ogden, M. D. Carroll, L. R. Curtin, M. A. McDowell, C. J. Tabak, and K. M. Flegal. Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA*, 295(13):1549-55, 2006.

[48] J. J. Reilly, J. Armstrong, A. R. Dorosty, P. M. Emmett, A. Ness, I. Rogers, C. Steer, and A. Sherriff. Early life risk factors for obesity in childhood: cohort study. *Br. Med. J.*, 330(7504):1357–9, 2005.

[49] T. J. Parsons, C. Power, S. Logan, and C. D. Summerbell. Childhood predictors of adult obesity: a systematic review. *Int. J. Obes.*, 23(Suppl 8):S1–107, 1999.

[50] R. L. Sjöberg, K. W. Nilsson, and J. Leppert. Obesity, shame, and depression in school-aged children: a population-based study. *Pediatrics*, 116(3):e389–92, 2005.

[51] J. Sandhu, Y. Ben-Shlomo, T. J. Cole, J. Holly, and G. Davey Smith. The impact of childhood body mass index on timing of puberty, adult stature and obesity: a follow-up study based on adolescent anthropometry recorded at Christ's Hospital (1936-1964). *Int. J. Obes.*, 30(1):14–22, 2006.

[52] J. J. Reilly, E. Methven, Z. C. McDowell, B. Hacking, D. Alexander, L. Stewart, and C. J. Kelnar. Health consequences of obesity. *Arch. Dis. Child.*, 88(9):748–52, 2003.

[53] J. L. Baker, L. W Olsen, and T. I. A. Sørensen. Childhood body-mass index and the risk of coronary heart disease in adulthood. *N. Engl. J. Med.*, 357(23):2329–37, 2007.

[54] M. K. Serdula, D. Ivery, R. J. Coates, D. S. Freedman, D. F. Williamson, and T. Byers. Do obese children become obese adults? A review of the literature. *Prev. Med.*, 22(2):167–77, 1993.

[55] J. Eriksson, T. Forsen, J. Tuomilehto, C. Osmond, and D. Barker. Size at birth, childhood growth and obesity in adult life. *Int. J. Obes.*, 25(5):735–40, 2001.

[56] R. C. Whitaker, J. A. Wright, M. S. Pepe, K. D. Seidel, and W. H. Dietz. Predicting obesity in young adulthood from childhood and parental obesity. *N. Engl. J. Med.*, 337(13):869–73, 1997.

[57] E. Charney, H. C. Goodman, M. McBride, B. Lyon, and R. Pratt. Childhood antecedents of adult obesity. Do chubby infants become obese adults? *N. Engl. J. Med.*, 295(1):6–9, 1976.

[58] BBC News. England's children getting fatter. `http://news.bbc.co.uk/1/hi/health/4496947.stm`, viewed 29 April 2005.

[59] J. C. K. Wells. A critique of the expression of paediatric body composition data. *Arch. Dis. Child.*, 85(1):67–72, 2001.

[60] T. J. Cole, M. C. Bellizzi, K. M. Flegal, and W. H. Dietz. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ*, 320(7244):1240–3, 2000.

[61] A. M. Prentice. Body mass index standards for children. Are useful for clinicians but not yet for epidemiologists. *Br. Med. J.*, 317(7170):1401–2, 1998.

[62] C. Bouchard. BMI, fat mass, abdominal adiposity and visceral fat: where is the 'beef'? *Int. J. Obes.*, 31:1552–3, 2007.

[63] L. M. Maynard, W. Wisemandle, A. F. Roche, W. C. Chumlea, S. S. Guo, and R. M. Siervogel. Childhood body composition in relation to body mass index. *Pediatrics*, 107(2):344–50, 2001.

[64] D. M. Hall and T. J. Cole. What use is the BMI? *Arch. Dis. Child.*, 91(4):283–6, 2006.

[65] A. Pietrobelli, M. S. Faith, D. B. Allison, D. Gallagher, G. Chiumello, and S. B. Heymsfield. Body mass index as a measure of adiposity among children and adolescents: a validation study. *J. Pediatr.*, 132(2):204–10, 1998.

[66] T. J. Cole, M. S. Faith, A. Pietrobelli, and M. Heo. What is the best measure of adiposity change in growing children: BMI, BMI %, BMI z-score or BMI centile? *Eur. J. Clin. Nutr.*, 59(3):419–25, 2005.

[67] J. J. Reilly. Diagnostic accuracy of the BMI for age in paediatrics. *Int. J. Obes.*, 30(4):595–7, 2006.

[68] T. J. Cole, J. V. Freeman, and M. A. Preece. Body mass index reference curves for the UK, 1990. *Arch. Dis. Child.*, 73(1):25–9, 1995.

[69] R. J. Kuczmarski, C. L. Ogden, L. M. Grummer-Strawn, K. M. Flegal, S. S. Guo, R. Wei, Z. Mei, L. R. Curtin, A. F. Roche, and C. L. Johnson. CDC growth charts: United States. *Advance Data from Vital and Health Statistics*, 314, 2000.

[70] M. Preece, T. Cole, and T. Fry. Body mass index standards for children. 1990 data will remain available. *Br. Med. J.*, 319(7202):122, 1999.

[71] WHO Multicentre Growth Reference Study Group. *WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development.* World Health Organisation, Geneva, 2006.

[72] T. J. Cole. Babies, bottles, breasts: is the WHO growth standard relevant? *Significance*, 4(1):6–10, 2007.

[73] S. S. Guo and W. C. Chumlea. Tracking of body mass index in children in relation to overweight in adulthood. *Am. J. Clin. Nutr.*, 70(1):145S–8S, 1999.

[74] W. H. Dietz. Critical periods in childhood for the development of obesity. *Am. J. Clin. Nutr.*, 59(5):955–9, 1994.

[75] M. W. Gillman. A life course approach to obesity. In D. Kuh and Y. Ben-Shlomo, editors, *A Life Course Approach to Chronic Disease Epidemiology.* Oxford University Press, Oxford, 2004.

[76] E. M. Poskitt and T. J. Cole. Do fat babies stay fat? *Br. Med. J.*, 1(6052):7–9, 1977.

[77] J. Baird, D. Fisher, P. Lucas, J. Kleijnen, H. Roberts, and C. Law. Being big or growing fast: systematic review of size and growth in infancy and later obesity. *Br. Med. J.*, 331(7522):929, 2005.

[78] U. Ekelund, K. Ong, Y. Linne, M. Neovius, S. Brage, D. B. Dunger, N. J. Wareham, and S. Rossner. Upward weight percentile crossing in infancy and early childhood independently predicts fat mass in young adults: the Stockholm Weight Development Study (SWEDES). *Am. J. Clin. Nutr.*, 83(2):324–30, 2006.

[79] J. M. Tanner. Growth as a target-seeking function: catch-up and catch-down growth. In F. Falkner, editor, *Human growth: a comprehensice treatise.* Plenum, New York, 1986.

[80] N. Stettler. Nature and strength of epidemiological evidence for origins of childhood and adulthood obesity in the first year of life. *Int. J. Obes.*, 31(7):1035–1043, 2007.

[81] L. Ibáñez, K. Ong, D. B. Dunger, and F. de Zegher. Early development of adiposity and insulin resistance after catch-up weight gain in small-for-gestational-age children. *J. Clin. Endocrinol. Metab.*, 91(6):2153–8, 2006.

[82] M. F. Rolland-Cachera, M. Deheeger, F. Bellisle, M. Sempe, M. Guilloud-Bataille, and E. Patois. Adiposity rebound in children: a simple indicator for predicting obesity. *Am. J. Clin. Nutr.*, 39(1):129–35, 1984.

[83] M. F. Rolland-Cachera, M. Deheeger, M. Guilloud-Bataille, P. Avons, E. Patois, and M. Sempe. Tracking the development of adiposity from one month of age to adulthood. *Ann. Hum. Biol.*, 14(3):219–29, 1987.

[84] R. M. Siervogel, A. F. Roche, S. M. Guo, D. Mukherjee, and W. C. Chumlea. Patterns of change in weight/stature^2 from 2 to 18 years: findings from long-term serial data for children in the Fels longitudinal growth study. *Int. J. Obes.*, 15(7):479–85, 1991.

[85] R. C. Whitaker, M. S. Pepe, J. A. Wright, K. D. Seidel, and W. H. Dietz. Early adiposity rebound and the risk of adult obesity. *Pediatrics*, 101(3):E5, 1998.

[86] S. Williams, G. Davie, and F. Lam. Predicting BMI in young adults from childhood data using two approaches to modelling adiposity rebound. *Int. J. Obes.*, 23(4):348–54, 1999.

[87] S. S. Guo, C. Huang, L. M. Maynard, E. Demerath, B. Towne, W. C. Chumlea, and R. M. Siervogel. Body mass index during childhood, adolescence and young adulthood in relation to adult overweight and adiposity: the Fels Longitudinal Study. *Int. J. Obes.*, 24(12):1628–35, 2000.

[88] D. S. Freedman, L. Kettel Khan, M. K. Serdula, S. R. Srinivasan, and G. S. Berenson. BMI rebound, childhood height and obesity among adults: the Bogalusa Heart Study. *Int. J. Obes.*, 25(4):543–9, 2001.

[89] C. Corvalán, C. Gregory, M. Ramirez-Zea, R. Martorell, and A. Stein. Size at birth, infant, early and later childhood growth and adult body composition: a prospective study in a stunted population. *Int J Epidemiol*, 36(3):550–7, 2007.

[90] A. R. Dorosty, P. M. Emmett, S. Cowin, J. J. Reilly, and the ALSPAC Study Team. Factors associated with early adiposity rebound. *Pediatrics*, 105(5):1115–8, 2000.

[91] T. J. Cole. Children grow and horses race: is the adiposity rebound a critical period for later obesity? *BMC Pediatr.*, 4:6, 2004.

[92] H. S. Sachdev, C. H. Fall, C. Osmond, R. Lakshmy, S. K. Dey Biswas, S. D. Leary, K. S. Reddy, D. J. Barker, and S. K. Bhargava. Anthropometric indicators of body composition in young adults: relation to size at birth and serial measurements of body mass index in childhood in the New Delhi birth cohort. *Am. J. Clin. Nutr.*, 82(2):456–66, 2005.

[93] A. Öhlin and S. Rössner. Maternal body weight development after pregnancy. *Int. J. Obes.*, 14(2):159–73, 1990.

[94] U. Ekelund, M. Neovius, Y. Linne, S. Brage, N. J. Wareham, and S. Rossner. Associations between physical activity and fat mass in adolescents: the Stockholm Weight Development Study. *Am. J. Clin. Nutr.*, 81(2):355–60, 2005.

[95] W. E. Siri. Body composition from fluid spaces and density: analysis of methods. 1961. *Nutrition*, 9(5):480–91; discussion 480, 492, 1993.

[96] K. Elfhag and Y. Linne. Gender differences in associations of eating pathology between mothers and their adolescent offspring. *Obes. Res.*, 13(6):1070–6, 2005.

[97] V. Mann, B. L. De Stavola, and D. A. Leon. Separating within and between effects in family studies: an application to the study of blood pressure in children. *Stat. Med.*, 23(17):2745–56, 2004.

[98] Uppsala University. Uppsala family study. `http://www.pubcare.uu.se/family`, viewed 16 October 2007.

[99] I. Koupil. The uppsala studies on developmental origins of health and disease. *J. Intern. Med.*, 261(5):426–36, 2007.

[100] J. Karlberg, Z. C. Luo, and K. Albertsson-Wikland. Body mass index reference values (mean and SD) for Swedish children. *Acta. Paediatr.*, 90(12):1427–34, 2001.

[101] D. A. Leon, I. Koupil, V. Mann, T. Tuvemo, G. Lindmark, R. Mohsen, L. Byberg, and H. Lithell. Fetal, developmental, and parental influences on childhood systolic blood pressure in 600 sib pairs: the Uppsala Family Study. *Circulation*, 112(22):3478–85, 2005.

[102] I. Koupil, V. Mann, D. A. Leon, U. Lundberg, L. Byberg, and D. Vagero. Morning cortisol does not mediate the association of size at birth with blood pressure in children born from full-term pregnancies. *Clin. Endocrinol.*, 62(6):661–6, 2005.

[103] Medical Research Council NSHD team. University College London. National Survey of Health and Development (NSHD). `http://www.nshd.mrc.ac.uk`, viewed 5 July 2007.

[104] E. Ferri, J. Bynner, and M. Wadsworth, editors. *Changing Britain, Changing Lives: Three Generations at the Turn of the Century.* Institute of Education, London, 2003.

[105] M. Wadsworth, D. Kuh, M. Richards, and R. Hardy. Cohort profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *Int. J. Epidemiol.*, 35:49–54, 2006.

[106] C. S. Peckham, O. Stark, V. Simonite, and O. H. Wolff. Prevalence of obesity in British children born in 1946 and 1958. *Br. Med. J.*, 286:1237–42, 1983.

[107] C. Power and J. Elliott. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol*, 35:34–41, 2006.

[108] Centre for Longitudinal Studies. National Child Development Study. `http://www.cls.ioe.ac.uk/studies.asp?section=000100020003`, viewed 5 July 2007.

[109] UK Data Archive. National Child Development Study datasets. `http://www.data-archive.ac.uk/findingData/ncdsTitles.asp`, viewed 5 July 2007.

[110] Centre for Longitudinal Studies. National Child Development Study Data Dictionary. `http://www.cls.ioe.ac.uk/studies.asp?section=0001000200030014`, viewed 5 July 2007.

[111] J. Elliott and P. Shepherd. Cohort profile: 1970 British Birth Cohort (BCS70). *Int. J. Epidemiol.*, 35:836–43, 2006.

[112] Centre for Longitudinal Studies. British Cohort Study. http://www.cls.ioe.ac.uk/studies.asp?section=000100020002, viewed 5 July 2007.

[113] R. M. Viner and T. J. Cole. Adult socioeconomic, educational, social, and psychological outcomes of childhood obesity: a national birth cohort study. *Br. Med. J.*, 330:1354, 2005.

[114] UK Data Archive. 1970 British Cohort Study datasets. http://www.data-archive.ac.uk/findingData/bcsTitles.asp, viewed 5 July 2007.

[115] Centre for Longitudinal Studies. British Cohort Study Data Dictionary. http://www.cls.ioe.ac.uk/studies.asp?section=0001000200020012, viewed 5 July 2007.

[116] B. R. Kirkwood and J. A. C. Sterne. *Essential Medical Statistics.* Blackwell, Oxford, 2003.

[117] P. Armitage, G. Berry, and J. N. S. Matthews. *Statistical Methods in Medical Research.* Blackwell Science, Malden, MA, 2001.

[118] P. J. Diggle, K-Y. Liang, and L. Z. Scott. *Analysis of Longitudinal Data.* Oxford University Press, New York, 2002.

[119] P. Clarke and R. Hardy. Methods for handling missing data. In A. Pickles, B. Maughan, and M. Wadsworth, editors, *Epidemiological Methods in Life Course Research.* Oxford University Press, Oxford, 2007.

[120] R. J. A. Little and D. B. Rubin. *Statistical Analysis With Missing Data.* Wiley, New York, 2002.

[121] M. G. Kenward and J. Carpenter. Multiple imputation: current perspectives. *Stat. Methods. Med. Res.*, 16:199–218, 2007.

[122] G. Molenberghs. Editorial: what to do with missing data? *J. R. Statist. Soc. A*, 170(4):861–863, 2007.

[123] J. L. Schafer. Multiple imputation: a primer. *Stat. Methods Med. Res.*, 8(1):3–15, 1999.

[124] J. L. Schafer. *Analysis of Incomplete Multivariate Data.* Chapman and Hall/CRC, Boca Raton, 1997.

[125] D. B. Rubin. Multiple imputation after 18+ years. *J. Am. Stat. Assoc.*, 91(434):473–489, 1996.

[126] S. van Buuren, H. C. Boshuizen, and D. L. Knook. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat. Med.*, 18(6):681–94, 1999.

[127] SAS Institute Inc. *SAS OnlineDoc 9.1.3.* Cary, NC, 2004.

[128] D. B. Rubin. *Multiple imputation for nonresponse in surveys.* John Wiley, New York, 1987.

[129] A. J. Dobson. *An introduction to generalized linear models.* Chapman and Hall/CRC, Boca Raton, FL, 2002.

[130] T. J. Cole. Modeling postnatal exposures and their interactions with birth size. *J. Nutr.*, 134(1):201-4, 2004.

[131] D. W. Hosmer and S. Lemeshow. *Applied Logistic Regression.* Wiley, New York, 2000.

[132] J. R. Carpenter. Multilevel models. In B. S. Everitt and C. R. Palmer, editors, *Encyclopaedic companion to medical statistics.* Hodder Education, London, 2005.

[133] G. K. Robinson. That BLUP is a good thing: the estimation of random effects. *Statistical Science*, 6(1):15–51, 1991.

[134] D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression.* Cambridge University Press, New York, 2003.

[135] P. Royston and D. G. Altman. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl. Statist.*, 43(3):429–467, 1994.

[136] Stata Corporation. *Stata Base Reference Manual, Volume 1, A-F, Release 8.* Stata Press, College Station, Texas, 2003.

[137] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–74, 1982.

[138] I. Kreft and J. De Leeuw. *Introducing Multilevel Modelling.* Sage, London, 1999.

[139] H. Goldstein, M. J. Healy, and J. Rasbash. Multilevel time series models with applications to repeated measures data. *Stat. Med.*, 13(16):1643–55, 1994.

[140] P. J. Green and B. W. Silverman. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* Chapman and Hall, London, 1994.

[141] B. W. Silverman. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. R. Statist. Soc.*, 47(1):1–52, 1985.

[142] M. A. Preece and I. Heinrich. Mathematical modelling of individual growth curves. *Br. Med. Bull.*, 37(3):247–52, 1981.

[143] J. Deming and A. H. Washburn. Application of the Jenss curve to the observed pattern of growth during the first eight years of life in forty boys and forty girls. *Hum. Biol.*, 35:484–506, 1963.

[144] A. H. Manwani and K. N. Agarwal. The growth pattern of Indian infants during the first year of life. *Hum. Biol.*, 45(3):341–9, 1973.

[145] P. van Dommelen, M. C. de Gunst, A. W. van der Vaart, and D. I. Boomsma. Genetic study of the height and weight process during infancy. *Twin. Res.*, 7(6):607–16, 2004.

[146] T. J. Cole. Assessment of growth. *Best Practice & Research Clinical Endocrinology and Metabolism*, 16(3):383–398, 2002.

[147] *Stata 10.0.* StataCorp, College Station, TX, 2007.

[148] E. W. Count. A quantitative analysis of growth in certain human skull dimensions. *Hum. Biol.*, 14(2):143–165, 1942.

[149] R. D. Bock, H. Wainer, A. Petersen, D. Thissen, J. Murray, and A. Roche. A parameterization for individual human growth curves. *Hum. Biol.*, 45(1):63–80, 1973.

[150] A. Roche and S. Shumei. *Human Growth: Assessment and Interpretation.* Cambridge University Press, Cambridge, 2003.

[151] P. Jolicoeur, J. Pontier, M. O. Pernin, and M. Sempe. A lifetime asymptotic growth curve for human height. *Biometrics*, 44(4):995–1003, 1988.

[152] A. W. Ledford and T. J. Cole. Mathematical models of growth in stature throughout childhood. *Ann. Hum. Biol.*, 25(2):101–15, 1998.

[153] E. Marubini. Mathematical handling of long-term longitudinal data. In F. Falkner and J. M. Tanner, editors, *Human Growth*, volume 1: Principles & Prenatal Growth. Baillière Tindall, London, 1978.

[154] R. H. Largo, T. Gasser, A. Prader, W. Stuetzle, and P. J. Huber. Analysis of the adolescent growth spurt using smoothing spline functions. *Ann. Hum. Biol.*, 5(5):421–34, 1978.

[155] *R 2.2.1.* The R Development Core Team, 2005.

[156] Y. Wang, K. Ge, and B. M. Popkin. Tracking of body mass index from childhood to adolescence: a 6-y follow-up study in China. *Am. J. Clin. Nutr.*, 72(4):1018–24, 2000.

[157] Centers for Disease Control and Prevention. CDC Growth Charts: United States. `http://www.cdc.gov/nchs/about/major/nhanes/growthcharts/background.htm`, viewed 5 July 2007.

[158] T. J. Cole and H. Pan. *lmsGrowth version 2.12.* Medical Research Council, 2005.

[159] Centers for Disease Control and Prevention. BMI-for-age charts, 2 to 20 years, LMS parameters and selected smoothed BMI (kilograms/meters squared) percentiles, by sex and age. `http://www.cdc.gov/nchs/data/nhanes/growthcharts/bmiage.txt`, viewed 5 July 2007.

[160] A. R. Tate, C. Dezateux, and T. J. Cole. Is infant growth changing? *Int J Obes*, 30:1094–6, 2006.

[161] M. E. Wadsworth, S. L. Butterworth, R. J. Hardy, D. J. Kuh, M. Richards, C. Langenberg, W. S. Hilder, and M. Connor. The life course prospective design: an example of benefits and problems associated with study longevity. *Soc Sci Med*, 57:2193–205, 2003.

[162] D. F. Hollingsworth. The changing patterns in British food habits since the 1939–45 war. *Proc. Nutr. Soc.*, 20:25–30, 1961.

[163] M. Eriksson, F. Rasmussen, and T. Nordqvist. Changes in shape and location of BMI distributions of Swedish children. *Acta Paediatr*, 94:1558–65, 2005.

[164] L. S. Adair and T. J. Cole. Rapid child growth raises blood pressure in adolescent boys who were thin at birth. *Hypertension*, 41(3):451–6, 2003.

[165] J. D. Skinner, W. Bounds, B. R. Carruth, M. Morris, and P. Ziegler. Predictors of children's body mass index: a longitudinal study of diet and growth in children aged 2-8 y. *Int. J. Obes.*, 28(4):476–82, 2004.

[166] W. H. Dietz. Overweight in childhood and adolescence. *N. Engl. J. Med.*, 350(9):855–7, 2004.

[167] R. J. Silverwood, M. Neovius, B. L. De Stavola, U. Ekelund, and Y. Linn. Analysis of critical periods of childhood BMI development for late-adolescent adiposity (both BMI and percentage body fat) using standard multiple regression, structural equation modelling, and spline-based approaches. *Early Human Development*, 82(8):561–2, 2006.

[168] W. A. Marshall and J. M. Tanner. Puberty. In F. Falkner and J. M. Tanner, editors, *Human Growth: A Comprehensive Treatise*, volume 2: Postnatal Growth; Neurobiology. Plenum, New York, 1986.

[169] T. Chandola, P. Clarke, J. N. Morris, and D. Blane. Pathways between education and health: a causal modelling approach. *J. R. Statist. Soc. A*, 169:337–59, 2006.

[170] Y. Ben-Shlomo and D. Kuh. Conclusions. In D. Kuh and Y. Ben-Shlomo, editors, *A Life Course Approach to Chronic Disease Epidemiology.* Oxford University Press, Oxford, 2004.

[171] R. W. Taylor, A. Goulding, N. J. Lewis-Barned, and S. M. Williams. Rate of fat gain is faster in girls undergoing early adiposity rebound. *Obes. Res.*, 12(8):1228–30, 2004.

[172] S. M. Williams. Weight and height growth rate and the timing of adiposity rebound. *Obes. Res.*, 13(6):1123–30, 2005.

[173] A. Kroke, S. Hahn, A. E. Buyken, and A. D. Liese. A comparative evaluation of two different approaches to estimating age at adiposity rebound. *Int. J. Obes.*, 30(2):261–6, 2006.

[174] *SAS 9.1.3.* SAS Institute Inc., Cary, NC, 2006.

[175] H. D. McCarthy, T. J. Cole, T. Fry, S. A. Jebb, and A. M. Prentice. Body fat reference curves for children. *Int. J. Obes.*, 30(4):598–602, 2006.

[176] Statistics Sweden. Every tenth Swede is obese, 2007. `www.scb.se/templates/pressinfo____195661.asp`.

[177] T. J. Lobstein, W. P. James, and T. J. Cole. Increasing levels of excess weight among children in England. *Int. J. Obes.*, 27(9):1136–8, 2003.

[178] J. M. Tanner and R. H. Whitehouse. Revised standards for triceps and subscapular skinfolds in British children. *Arch. Dis. Child.*, 50(2):142–5, 1975.

[179] M. Durbán, J. Harezlak, M. P. Wand, and R. J. Carroll. Simple fitting of subject-specific curves for longitudinal data. *Stat. Med.*, 24(8):1153–67, 2005.

[180] W. H. Dietz. Periods of risk in childhood for the development of adult obesity - what do we need to learn? *J. Nutr.*, 127(9):1884S–1886S, 1997.

[181] J. O. Hill and F. L. Trowbridge. Childhood obesity: future directions and research priorities. *Pediatrics*, 101:570–4, 1998.