A Critique of Models for Body Composition and Energy-balance Components in Childhood and Adolescence

Caroline E. Haig

A dissertation submitted to the

University of Glasgow

for the degree of

Doctor of Philosophy

School of Mathematics and Statistics

College of Science and Engineering

February 2013

Typeset using LaTeX

Abstract

Keywords: Statistical modelling; missing data; MICE; longitudinal simulation; energy-balance; energy intake; energy expenditure; overweight and obesity; ALSPAC; DEXA; BIA; pubertal staging; healthy growth.

It is well known that, in Western countries, people of all ages and both sexes are becoming 'fatter' in general. In a 'healthy' population, we arbitrarily consider cut-offs to be that 10% of people should be 'overweight' and 5% 'obese', as there is limited evidence that these cut-off points are related to ill-health [1]. However, we are seeing a dramatic rise in the numbers of people in each of these categories [2, 3]. The mechanism behind weight gain is energy-imbalance. At energy-balance for adults - i.e. where weight is expected to remain stable over time, we know that:

$$energy\ \ intake\ \ (EI) = energy\ \ expenditure\ \ (EE) \tag{1}$$

This equation is far less straightforward than it first appears. The first important issue is that EE has several different components (e.g. resting EE). The second issue is to do with measurement - how do we measure energy intake and energy expenditure? Another is down to physiological differences between people - how do things vary between individuals and do they differ systematically between males and females, adults and children? The above equation applies to adults, but we know that children and adolescents actually require a positive imbalance for healthy growth - what is not known is what degree of positive imbalance is healthy.

This thesis is particularly concerned with energy-balance and imbalance during puberty, at which time the human body goes through extreme changes. We investigate how these changes are measured, and how energy-imbalance and the

modelling thereof must change across this time. We will show that the proportions of children who are overweight and obese are higher than we would expect; commonly used models for body composition are not in agreement; commonly used models for resting energy expenditure are not in agreement; children do not need a high energy-imbalance for normal growth; and those girls with early menarche are more likely to become overweight than their counterparts.

Declaration

This thesis has been prepared by myself and no section of it has been submitted previously as part of any application of a degree. The work herein reported was carried out by myself unless stated otherwise.

This work has been presented at the Research Students' Conference (2008) and at University of Glasgow seminars (2007, 2008, 2009, 2010).

All analysis has been carried out using SPSS version 15 [4] or 18 [5] and R versions between 2.6.0 and 2.15.2 [6].

January 2013

Acknowledgements

First and foremost, I would like to express my sincere thanks to the College of Science and Engineering (formerly the Faculty of Information and Mathematical Sciences) and the School of Mathematics and Statistics (formerly the Department of Statistics) at the University of Glasgow. Throughout the years I spent on this project, the department was invaluable - from providing support through difficult times to sharing in the good times - and of course for providing me with the funding and opportunity to undertake this research.

I extend thanks to Professor John H. McColl, my main supervisor and mentor, for his never-ending support, advice, knowledge, patience and wisdom. There are no words to describe the extent of my gratitude. I am very thankful for the guidance of Professor John Reilly who acted as my second supervisor, for providing valuable expertise and insight.

Thanks must also go to the providers of the data used throughout this project. The Avon Longitudinal Study of Parents and Children, based at Bristol University - in particular Andrea Sherriff and Jon Heron, both previously of ALSPAC, Laura Miller and Professor Andy Ness. I also thank Professor Jonathan Wells (Institute of Child Health, London) - an undisputed expert in the field - for his expertise, support and data access, and Dr. Dympna Gallagher (New York Obesity Research Center) for graciously providing data that became crucial to certain parts of this work.

I would also like to give thanks to The Robertson Centre for Biostatistics for the use of their resources nearing the end of this project, and allowing me to spend a proportion of my working hours completing this thesis.

Last but certainly not least, I must thank my friends - those here in Scotland as well as those further afield, family and partner Benn for their continued support. Between them, these wonderful people spent these years dealing with the trials and tribulations, tears and tantrums, and joys and celebrations that a PhD inevitably brings to all those brave enough to undertake it.

Contents

List of Tables

List of Figures

List of Abbreviations

2-C 2 component

3-C 3 component

4-C 4 component

ALSPAC Avon Longitudinal Study of Parents and Children

BA Bland-Altman

BCU Body Composition Unit

BIA Bioelectrical impedance analysis

BMI Body mass index

BMM Bone mineral mass

CHD Coronary heart disease

DEXA Dual-energy x-ray absorptiometry

DIT Diet-induced thermogenesis

EI Energy expenditure

EI Energy intake

FAO Food and Agriculture Organization

FFM Fat-free mass

FM Fat mass

h Hydration constant

HB Harris and Benedict

ICC Intraclass correlation coefficient

IOM Institute of Medicine

kcal 1 kilocalorie = 1000 calories

MAR Missing at random

MCAR Missing completely at random

MICE Multiple imputation by chained equations

MNAR Missing not at random

MRI Magnetic resonance imaging

NEAT Non-exercise activity thermogenesis

NHANES National Health and Nutrition Examination Survey

NLS Nonlinear least squares

OLS Ordinary least squares

P() Probability

PA Physical activity

PAEE Physical activity energy expenditure

PAL Physical activity level

PM Air-displacement plethysmography

PMM Predictive mean matching

REE Resting energy expenditure

RI Rohrer index

RMR Resting metabolic rate

SDS Standard deviation score

TBW Total body water

TEE Total energy expenditure

TEF Thermic effect of food

UNU United Nations University

WC Waist circumference

WHO World Health Organisation

WHR Waist-to-hip ratio

Z Impedance (ohm)

Chapter 1

Introduction, background and outline

1.1 Introduction

In recent years, prevalence of overweight and obesity has increased at an astonishing rate. In 1980, it was estimated that 6% of adult men and 8% of adult women were obese in the UK. Twenty-two years later, published figures reported that these figures had risen to 23% and 25%, respectively [3]. Perhaps even more worrying than the increasing trends in adult obesity are those now emerging in childhood [7]. The potential consequences of being overweight or obese are immense. On the individual level, overweight adults are at risk of numerous co-morbidities including (but not limited to) diabetes, coronary heart disease (CHD) and cancer, as well as mental health difficulties such as depression and reduced self-esteem. On the population level, these "side effects" of the obesity epidemic have serious economic implications. Fry and Finley reported in 2005 that the total direct and indirect annual costs of obesity in the EU were €31.8 bn [8].

With such a rising trend in obesity, more and more research is being carried out into obesity-related health implications. It is now known that an individual who is overweight or obese in childhood is at increased risk of lifelong physical and mental health problems. For example, type 2 diabetes has only been seen in UK children since the overweight proportion began its steep increase [9]. Extremely worryingly, suicidal ideation has been shown to be increased among overweight children as a result of reduced self-esteem, depression and weight-related bullying [10].

Energy-balance occurs when

$$energy\ \ intake\ \ (EI) = energy\ \ expenditure\ \ (EE) \tag{1.1}$$

Energy-imbalance occurs when the right-hand side of equation (1.1) does not equal the right-hand side. This PhD project is concerned with modelling energy-imbalance during adolescence. Energy-imbalance (where energy intake is not equal to energy expenditure) is known to lead to weight gain when the imbalance is positive (more intake than expenditure) and weight loss when the imbalance is negative. In today's climate, it is known that, while intake of energy dense foods is increasing, physical activity among children is decreasing in favour of more sedentary activities such as watching television or playing video games. In terms of energy-balance and weight management with children, there is an additional issue as we know that children need some degree of positive imbalance to allow normal growth and development [11]. What is not known, however, is how much imbalance is necessary and sufficient for such development. The study of obesity itself and of obesity prevention and cure is an enormous field with numerous unanswered questions and unresolved issues. This is perhaps more true for the study of childhood obesity than adult as a result of increasing trends in childhood obesity emerging more recently. This thesis aims to draw together and address a great many of the methodological questions in this field that have arisen in the literature published in recent years. For example, how do we measure the amount of fat a person has in his or her body? Then, how much is too much? How should we define obesity? However, what has

become abundantly clear over the years working on this project is that many questions cannot yet be answered with certainty. Many are, in-fact, confused and complicated by conflicting sources in the literature.

Further questions arise when studying trends over adolescence. How should we be defining puberty? What stages are involved in the pubertal process? Are physical changes alone well enough defined to be confidently used in modelling? Should we, in addition, be considering emotional or chemical changes? What are the actual measurement methods of such changes? How do puberty and pubertal effects on energy-imbalance compare between the sexes? This list of potential issues is by no means exhaustive.

1.2 Background

1.2.1 Body Composition

It is known that too much or too little fat mass can have serious adverse effects on health for both adults and children alike [12, 13]. However, body composition cannot be directly measured in living subjects, since the only "direct" method of accurate measurement is chemical analysis of human cadavers [14]. For this reason, it is necessary to estimate body composition using "indirect" methods.

1.2.1.1 Models of body composition

The traditional model of body composition is the two component model, which states that the body comprises fat mass (FM) and fat-free mass (FFM). This model is widely used in practice and has the advantage of being relatively easy to work out and understand. However, the two component model makes the assumption that protein, water and minerals exist in fixed proportions in fat-free mass [15]. This assumption may not be valid, particularly in children, due to changes in factors such as hydration during maturation [16]. For this reason, researchers have developed models with more than two components. The three component model takes into account variation in either hydration or bone mineral content [17], by considering body composition to consist of either: (1) fat mass, total body water (TBW) and fat-free dry mass or (2) fat mass, bone mineral mass (BMM) and lean soft tissue. Further, these aspects of body composition can be combined into the four-

component model comprising FM, TBW, BMM and dry fat-free soft mass [18, 19]. Studies have been carried out to compare the accuracy of these models and have generally reported improved accuracy with both three and four component models when compared to the two component model [18, 16, 19]. It has also been reported that, although the improvement may be minimal, the four component model does provide improved accuracy over the three component model [19]. Baumgartner et al. suggest that the four component model should be used as the gold standard for measuring body composition in adults [20].

1.2.1.2 Measures estimating body composition

There are various ways to estimate the body composition of individuals, which may be useful in identifying people who are at risk of obesity and associated co-morbidities. These measures range from height and weight indices which can be calculated with little difficulty by most people, to far more complicated methods and techniques which require complicated equipment and highly-trained technicians. Some of these methods can be used to identify abdominal adiposity (fatness), which is known to be a major predictor of heart disease [21, 22]. Others can, in conjunction with 2, 3 or 4 component models, be used to estimate the proportion of each type of mass in the body. The aims of this section are to introduce each method and discuss the "pros and cons" of each in terms of tolerability, accuracy and accessibility.

1.2.1.2.1 Measures based on height and weight

Body "fatness", estimated using calculations with height and weight, can be used as a proxy indicator of body composition. These calculations have the advantage that height and weight can be easily measured without any particularly sophisticated or expensive equipment. Weight-for-height is a standard that is generally used for children of between 2 and 5 years of age (with weight-for-length being used for infants younger than 2 years old). Tables and charts of z-scores and percentiles for boys and girls separately are available from the World Health Organisation [23]. Poustie et al (2000) found weight-for-height to be an unreliable measure of nutritional status in children because of inter- and intra-examiner variation in measurement [24], while Mei et al found no difference between weight-for-height and other measures of body composition [25].

Body Mass Index (BMI) (also referred to as Quetelet's Index [26, 27]) is widely used in a variety of scenarios, from individuals "watching their weight" to clinical settings. BMI requires no complicated measurements, simply weight and height, and is calculated from the following equations:

For metric measurements:

$$BMI = \left(\frac{weight(kg)}{height(m)^2} \right) \tag{1.2}$$

For imperial measurements:

$$BMI = \left(\frac{weight(lbs)}{height(in)^2}\right) \times 703 \tag{1.3}$$

Overweight and obesity in adults can be determined objectively from BMI using generally-agreed guidelines. These guidelines do vary slightly from source to source. Guidelines established by the World Health Organisation (WHO) in 1997 and published in 2000 [28] are shown in Table 1.1.

BMI	Classification
< 18.5	underweight
18.5 to 24.9	healthy weight
25.0 to 29.9	overweight (or "pre-obese")
30.0 to 34.9	class I obesity
35.0 to 39.9	class II obesity
> 40.0	class III obesity

Table 1.1: WHO criteria for overweight and obesity in adults

Although BMI is relatively simple to calculate and use, its reliability is questionable because BMI doesn't take into account proportions of FM and FFM in the body. For any given volume, FFM is heavier than FM and, as a result, people can be misclassified. For example, an athlete may be classified as overweight due to having a higher proportion of FFM than, and therefore being heavier than, the general population. Additionally, there exist differences in body composition as a result of age, gender and possibly race that are not currently considered when interpreting BMI [27]. For example, a World Health Organisation expert consultation on appropriate BMI for Asian populations suggested that the acceptable range be narrowed to $18.5\frac{kg}{m^2}$ - $23\frac{kg}{m^2}$ [29].

Determining the risk of overweight and obesity from BMI is not as straightforward for children as for adults. The BMI calculations are the same, as shown above (equations (1.2) and (1.3)), but further calculations must be carried out. One common method currently, in keeping with WHO guidelines, is to use BMI to calculate percentile rankings from gender-specific BMI-for-age growth charts (for 2 - 20 years of age, available from the World Health Organisation (WHO)) [28]. These percentiles are then categorised as follows:

Percentile	Category
$<$ 5th percentile	Underweight
80th to $<$ 95th percentile	Overweight ("pre-obese")
$\geq$ 95th percentile	Obese

Table 1.2: WHO criteria for overweight and obesity from gender-specific BMI-for-age growth charts for children and adolescents

As an alternative to the WHO percentiles, BMI standard deviation scores (SDS) for UK children can be relatively easily calculated using the LMS method developed by Tim Cole in 1990 [30]. This method, described in detail in section 2.2, adjusts BMI for height, sex and age, and allows BMI to be expressed as an exact centile or age-and-sex specific SD score relative to a reference population. It can also be used to calculate a standardised measure of the difference between an individual's BMI SD Score at two ages, using a model published by Cole in 1997 [31].

It appears that using BMI as a simple proxy or surrogate measure of body composition presents fewer problems with children and adolescents than with adults. The Centers for Disease Control and Prevention website tells us that "BMI is a reliable indicator of body fatness for most children and teens" [32]. Additionally, Dietz and Bellizzi state that "the body mass index (BMI; in kg/m^2) offered a reasonable measure with which to assess fatness in children and adolescents" [33]. Many researchers, however, disagree with these ideas, and conclude that BMI is not the most effective method of determining body composition [34]. While the issue of increased lean mass may not be such a problem during the early stages of life, covariates such as age, gender and race should be taken into consideration when interpreting BMI as a measure of obesity-related health risk in children [26].

There do exist some alternatives to BMI, such as the Rohrer Index (RI):

$$\text{RI} = \frac{\text{weight(kg)}}{\text{height(m)}^3} \tag{1.4}$$

As with BMI, the Rohrer Index is classified for children using RI-for-age charts (although these do not appear to be publicly available). However, while BMI is widely used in practice, RI does not appear to be. Mei et al (2002) found that, of weight-for-height, BMI and RI, the latter was the least reliable [25].

Other widely used proxy indicators of adiposity in both adults and children are measures based on circumference. These measures include the waist-to-hip ratio, waist circumference, and the conicity index [35].

The waist-to-hip ratio (WHR) is calculated as

$$\frac{\text{circumference of the waist}}{\text{circumference of the hips}} \tag{1.5}$$

For adults, it is generally accepted that a ratio of below 0.8 is healthy for women, and below 1.0 is healthy for men [36]. A ratio higher than these may indicate that the individual concerned is at an increased risk of cardiovascular illness. Taylor et al (1998) found that WHR is not as effective at assessing body composition as BMI [37], while some other researchers and medical professionals consider WHR to be very useful in adults for determining risk of cardiovascular illness resulting from excess abdominal adiposity [38]. Alternative (though not as widely used) ratios are waist-to-height and waist-to-arm. WHR and other such ratios are not widely used to determine adiposity in children.

Calculating waist circumference (WC) alone generally seems to be more effective as a gauge of truncal adiposity than WHR in adults [34] and has also been accepted for use with children [35]. Chan et al [34] consider WC to be a more effective measure of abdominal fat than both WHR and BMI - with neither adding any significance to predictions abdominal fat from WC.

Current guidelines for WC in adult men are increased risk at 94cm and substantially increased risk at 104cm. For women, increased risk is considered to be 80cm with 88cm indicating substantially increased risk. However, it seems unlikely that a man of height 200cm with a waist circumference of 94cm would be at the same level of risk as a man of height 165cm with the same waist measurement. For children, risk determined by waist circumference, as with BMI, is determined using age-and-sex specific population percentiles. In an editor's note to a 2004 paper by Fernández et al [39], it is stated that those children with WCs above the 90th percentile (adjusted for age, sex and ethnicity) could be considered at significant risk for obesity-related co-morbidities. However, there does not appear to be a widely accepted cut-off point.

Developed by Dr. Margaret Ashwell, the Ashwell chart (suitable for both male and female adults) [40, 41] adjusts waist circumference for height in order to assess health risk from abdominal adiposity. Some research has been carried out assessing the potential for similar charts for children and adolescents [42] but these do not appear to be widely used in medical practice.

The conicity index [35, 43] uses waist circumference adjusted for height and weight to determine body composition from the formula shown in equation (1.6).

$$C = \frac{\text{abdominal girth (m)}}{0.109\sqrt{\frac{\text{weight (kg)}}{\text{height(m)}}}} \tag{1.6}$$

Findings published by Taylor et al in 2000 show that the conicity index, too, is not as effective as WC alone at highlighting the potential problems and risks of overweight or obesity. [35]

The surrogate measures of body composition presented so far are relatively straightforward to obtain and can be used by either skilled professionals in a clinical setting or people at home (though individuals should be trained to be able to get accurate height and weight measurements). However, the interpretation on the individual level in terms of actual body composition - that is, proportions of fat and fat-free mass in the body - is extremely vague. While they provide some idea of whether or not an individual may be relatively underweight or overweight, or at particular risk of cardiovascular illness, they don't give a precise measure of actual body fatness. It would perhaps be sensible to make use of a combination of measures such as BMI and WC where actual measures (to be discussed) of the distribution of body fat itself are not available.

1.2.1.2.2 More complicated measures

A more complicated but highly accurate [44] method of measuring body composition is hydrodensitometry (also called hydrostatic or underwater weighing). Underwater weighing is considered to be one of the most accurate methods for body composition analysis [45, 46]. This procedure is based on Archimedes' Principle of Displacement, and uses the two component (2C) model of FM+FFM. As a result of bone and muscle being more dense than

water, a person with a lot of FFM would, compared with someone of the same height with less FFM, displace more water and therefore weigh more and have a lower percentage body fat. On the other hand, fat is less dense than water and so a large amount of FM will result in a lesser displacement of water, indicative of a higher percent body fat [47]. However, hydrodensitometry is a complicated procedure involving expensive equipment and may not be suitable for children - Chan et all stated in 1998 that "Underwater weighing is laborious to perform and it can be frightening for young children" [48]. Additionally, this method does not give any information about the distribution of body fat (for example, specific abdominal adiposity), simply a proportion for the whole body.

Air-displacement plethysmography (PM) works on a similar principle to hydrodensitometry, but with displacement of air rather than water. Various studies have found air displacement plethysmography to be accurate in both adults and children [49, 50, 16], with Fields and Goran stating in a 2000 study that "PM was the only technique that could accurately, precisely and without bias determine FFM in 9- to 14- yr old children" [16]. This method involves an individual, wearing tight-fitting clothing and with a swim hat on in order to prevent pockets of air, sitting in a chamber with a fixed air volume. Body volume is measured from the air reduction in the chamber, and body density and body fat are then determined from equations [51]. It has been noted that this method is preferable to hydrodensitometry for children, elderly and obese individuals [50] - as air is more easily tolerated than water!

Body composition can also be estimated using X-rays. This method is known as Dual Energy X-ray Absorptiometry (DEXA), and is becoming increasingly popular in the field of body composition research. DEXA uses a three component (3-C) model: fat mass, fat-free soft tissue, and bone mineral content [52]. Although the process does involve some exposure to radiation from x-rays, the University of Alberta's Human Nutrition Research Unit reports that the exposure for a whole body scan (both in adults and children) "is 1000 times less than the limit for trivial exposure, and is classified as a negligible individual dose according to the standards of the National Council of Radiation Protection and Measurements" [53]. The same source states that the radiation for a bone density scan is higher than a full body scan and is "similar to the radiation exposure during one commercial flight across Canada in adults, and is $\frac{1}{6}$th of the radiation exposure during one commercial flight across Canada for children." Van loan and Mayclin [54] found DEXA to be reliable, easy to use, and to give accurate values for the estimation of FFM for both men and women. As with the methods previously discussed, however, DEXA is not without disadvantages. Equipment is expensive and the procedure must be carried out by highly trained staff. A study published in 2005 found that DEXA was not suitable for subjects of greater than 140kg, making it unsuitable for the very obese population [55]. It is also, as a result of radiation exposure, unsuitable for women who are (or have any possibility of being) pregnant.

DEXA is often referred to as a "gold standard", but estimates of soft tissue derived from DEXA, when validated against 3-C or 4-C models, have

been shown to be of limited accuracy [56, 57]. Pending further studies into the use of DEXA alongside an accurate determination of body volume (i.e. from hydrodensitometry), DEXA cannot be considered to be a gold standard method of determining body composition!

Magnetic resonance imaging (MRI) scanners can be used to determine body composition by taking scans of the whole body. However, this involves highly sophisticated and expensive equipment and trained operators. This method of body composition analysis does not appear to be widely used.

Bioelectrical impedance analysis (BIA) is a method of determining the FFM in the body by measuring the body's reactance and resistance to electric current. Full-body bioelectrical impedance should be carried out in a controlled setting [58] and involves passing a small electrical current through the body with the use of electrodes. Reactance and resistance is then measured and from these values, the proportion of water in the body can be estimated. This is then used to estimate the amount of FFM in the body. Models for both calculations (known as resistivity and hydration) often vary between researchers, and are frequently being tested, updated and developed.

For the purposes of this project, we will make use of the following BIA models, developed in Glasgow in 2008 [59, 60].

$$Z = \sqrt{R^2 + X_i^2} \tag{1.7}$$

where Z=impedance (ohm); R=resistance (ohm), X_i=reactance (ohm).

Children

$$TBW = 0.61 \times \frac{height^2}{Z} - 0.63 \tag{1.8}$$

$$FFM = \frac{TBW}{h} \tag{1.9}$$

$$FM = weight - FFM \tag{1.10}$$

where h, the *hydration constant*, is given by:

$$BOYS : h = -0.000083 \times age + 0.77 \tag{1.11}$$

$$GIRLS : h = -0.0001667 \times age + 0.794 \tag{1.12}$$

Adults

$$TBW = 0.66 \times \frac{height^2}{Z} \tag{1.13}$$

$$FFM = \frac{TBW}{0.732} \tag{1.14}$$

$$FM = weight - FFM \tag{1.15}$$

where TBW is total body water in litres, weight is in kilograms, height is in centimetres and age is in months. It can be noted that, for adults, the hydration constant does not change according to age or sex.

This is a useful approach but has not been widely used since its publication.

There are potential issues with, and limitations of, this approach, which is based on a simplified model of the human shape as a cylinder. While the method has been shown to be reliable in population studies, it is said to have limited accuracy in individuals [61], and results can vary depending on several factors including food intake before the measurement, hydration and temperature. An important issue with these BIA models is that of spurious correlation, described as the correlation that exists between ratios even if the component variables of the ratios are uncorrelated. This problem is known to occur when ratios are used as independent variables in linear regression [62]. By using $\frac{height(cm)^2}{impedance}$, it is possible that we are forcing correlation where none exists. Nevertheless, we will continue to use the above model throughout this project.

Kyle et al (2004) find that BIA is an accurate method of estimating body fat when used in conjunction with appropriate population, age or pathology-specific equations and established procedures [63]. It has also been shown that BIA is easily tolerated by children [64, 18]. Chan et al, however, raise some question about accuracy, stating that "Bioelectrical impedance had a low correlation with total body fat and its use alone in estimating total body fat is not recommended" [48]. There are various factors which are known to result in the possibility of false impedance readings. These include under- or over-hydration, recent exercise and temperature at the time of the readings. There are also underlying assumptions of BIA which have not been thoroughly explored, "For example, traditional BIA methods employ a geometric model that assumes the component of interest is homogeneous in compo-

sition and uniform in cross-sectional area. The typical electrical pathways used with BIA in humans fail to conform to such idealized conditions." [58].

Note that, while BIA estimates FM and FFM in terms of body mass percentage, it does not determine particular areas of adiposity which may be vital in detecting cardiovascular risk - these may be estimated using a simpler method such as WHR.

A relatively recent development in "user-friendly" body composition estimation is the body-fat analysing scale. This scale, which can be used by individuals at home without trained personnel, uses methods similar to BIA to calculate the percentage body fat of the user. The adiposity (largely in the legs) is estimated, and this is used to estimate the percent body fat of the user. This raises the question - how representative of the entire body are the legs alone? Would this method be as accurate for a runner as for the general public? Naturally, effectiveness of these scales may vary considerably between both models and manufacturers. A 2000 study by Jebb et al found that the results obtained from the Tanita body-fat analyser were not significantly different from more traditional impedance techniques [65]. The study did show, however, that this and other impedance techniques were not as accurate at measuring body fat as other methods such as DEXA.

Skinfold measurements are a 'simpler' way of assessing body composition in that they do not require expensive equipment or highly technical procedures.

They do, however, require highly trained professionals. Skinfold calipers are used by technicians, and body composition is then estimated from established equations. It is very important that technicians are well trained in precisely how and where to take skinfold measurements. This method of determining body composition has been found to be accurate and reliable, but there is a high possibility of error due to inter- and intra-examiner variation [66]. There can also be differences in results due to the sex of the subject or the site being measured, among other factors. Therefore, great care must be taken when using skinfold measurements. A study by Cyrino et al in 2003 showed that maximised measurement error could be caused simply by using different calipers [46]. Kravitz and Heyward [45] suggest use of strict measurement guidelines for ensuring measurements are as accurate as possible. It should also be considered that skinfold measurements may not be well tolerated, particularly by children, since subjects must remain as still as possible while measurements are taken.

Of all the methods discussed, BMI is perhaps the simplest and a reasonably valid index of overweight and obesity regarding children. When more detailed determination of body composition is required, DEXA appears at least fairly precise compared to other measures and so could be considered acceptable as a measurement of changes in body composition over time.

In summary, no single method of body composition analysis is ideal for all situations, and few are suitable for large epidemiological and longitudinal studies. Each method should therefore be used with both caution and awareness of limitations and potential inaccuracies.

1.2.1.3 From childhood to adulthood

Often, one will hear parents, when discussing children, dismiss potential overweight by stating that children inevitably have 'growth spurts'. It would appear that people are generally unconcerned about childhood overweight or obesity leading to similar or more serious problems in later life. It has recently been reported by the Institute of Medicine that "growth spurts do occur at several points throughout childhood and adolescence, but it cannot be assumed that a child will lose his or her excess weight at those times" [67]. The publication then goes on to report that those children who are at a certain BMI percentile at age 4 can be expected to remain at a similar BMI percentile throughout childhood. In other words, those children who are overweight at 4 will most likely remain overweight throughout childhood.

Furthermore, an increasing number of studies are finding clear links between childhood weight-related problems and similar adult problems. A 2008 review by Singh et al considered more than 20 recent studies into such relationships, and found that "all studies... reported increased risk for overweight or obese youth to become overweight or obese in adulthood" [68].

With the ever-increasing prevalence of overweight and obese adults threatening serious health problems in the UK such as the diabetes epidemic currently seen in America [69], it seems clear that as well as tackling adult obesity with diet and activity measures in adulthood, society should be working to tackle it where it begins: in childhood and adolescence.

1.2.2 Energy Expenditure

When one talks about **energy** in terms of human nutrition, one is generally referring to the fuel that is used by the body for everything from sleeping to running. This fuel, or energy, taken from food and used by the body over the course of the day, is most commonly referred to in units called kilocalories (also, and more commonly, called "Calories"). One kilocalorie, or Calorie, contains 1000 calories, and can be described as the amount of heat energy needed to raise the temperature of 1 kilogram of water by 1 degree Celsius. An alternative, increasingly common way of expressing this energy is in kilojoules. The relation between these units can be shown as follows:

1000 calories = 1 kilocalorie = 1 Calorie = 4184 joules = 4.184 kilojoules

Total energy expenditure (TEE) is the amount of energy used by a person in the space of one day (often expressed as kcal/d), and can be broken down into the following components:

- Resting energy expenditure (REE) (*Also called Basal Metabolic Rate (BMR) or Resting Metabolic Rate (RMR) under some circumstances*)
- Diet-induced thermogenesis (DIT) (*Also known as the thermic effect of food (TEF)*)
- Physical activity (PA) thermogenesis

Ravussin stated, in 1992, that "The RMR generally accounts for $\sim$ 70% of the daily energy expenditure in sedentary individuals. DIT is considered to

be $\sim 10\%$ of the daily energy expenditure. The energy cost of physical activity is, of course, very variable and accounts for a variable amount of the total daily energy expenditure" [70].

1.2.2.1 Measurement of Energy Expenditure

This section aims only to **describe** methods of calorimetry and an additional method of measuring energy expenditure - doubly labelled water. Their uses will be discussed in later sections.

1.2.2.1.1 Direct Calorimetry

Direct Calorimetry is a measure of the heat output from the body [71, 72, 73, 74] using a whole-body measurement. There are various forms of direct calorimetry, including various forms of chamber calorimetry (convection, heat sink and isothermal calorimetry) and suit calorimetry.

Direct calorimetry is usually performed in a whole body sealed chamber. These chambers generally comprise a chair, table, bed, toilet and washing facilities, radio, television and sometimes exercise equipment [75, 76], and are intended to be as comfortable as possible for the subject.

Convection direct calorimetry [71, 72] is carried out in a whole body sealed chamber with thin walls. The chamber is enclosed in a water jacket in order

to prevent the outside temperature from affecting the results of the calorimetry. Sensors measure the temperature of the heat leaving the walls of the chamber, and the difference between this and the room's controlled temperature is then the energy expenditure of the subject.

In contrast, the chamber used for **heat sink** calorimetry [71] has thick, well insulated walls meaning that there is no need for a water jacket (though the temperature outside the chamber should still be monitored). Air is extracted from the chamber to a water-based heat exchanger which recirculates the air. The heat loss is calculated as the rise in the water temperature.

The walls of the chamber used for **isothermal** calorimetry are lined with an insulating layer which is in thermal equilibrium with the temperature of the chamber and the temperature of the chamber walls. Then the temperature gradient across the insulating layer is proportional to the heat loss from the subject.

In 1972, Webb, Annis and Troutman [77] described a suit calorimeter which "does not require a subject to stay quiet in order to achieve some sort of thermal steady state". The researchers report that the suit calorimeter allows subjects to "move about, exercise, eat meals, and sleep, while heat loss, heat production, and the other variables needed for energy-balance are measured continuously". The suit consists of a water-cooled garment (a network of small vinyl plastic tubing) which covers the entire body apart from the

face and the soles of the feet. This garment is worn directly on the skin and covered with several insulating layers of clothing (all closed with drawstrings or Velcro to ensure no loss of air during movement). Additionally, subjects are required to wear a facepiece to measure oxygen consumption. Water is circulated around the suit and the heat given off by the subject is measured from the increase in the temperature of the water.

The entire suit calorimeter increased a subject's weight by 20.6lb (9.4kg). Webb et al, however, found that this did not cause discomfort to subjects. The additional weight did, however, increase the oxygen cost of activity. For this reason, subjects also wore a facepiece which measured oxygen consumption.

The suit calorimeter has been developed over the years by Webb and used in various studies requiring the measurement of energy expenditure [77, 78, 75, 79]. In 1994, Hambræus et al [80] discussed the method, coming to the conclusion that it was an accurate method of measuring energy expenditure without requiring subjects to be contained in a sealed chamber.

Direct calorimetry, in all its possible forms, is an accurate and direct measure of energy expenditure [71, 72, 73]. Murgatroyd in 1993 reported that direct calorimetry is "potentially the most accurate measure of energy expenditure" [72]. However, direct calorimetry equipment is extremely expensive and requires highly trained technicians [73, 71]. Another disadvantage of direct

calorimetry is that the heat transfer from food, drink, light and so on must be accounted for [72] which can involve some lengthy measurements and calculations. Finally, direct calorimetry presents an artificial environment which is unlikely to yield the same results as free-living energy expenditure measured by an alternative method.

1.2.2.1.2 Indirect Calorimetry

Indirect calorimetry measures heat output by measuring respiratory gases [72]. The measurements of gases can then be converted to heat output. Indirect calorimetry can be conducted either in a sealed chamber or by portable methods.

In whole body indirect calorimetry, subjects are confined in a sealed chamber providing as normal an environment as possible. The chambers comprise a bed, chair, tv, shower and sometimes exercise equipment. Air flow into the chamber is continuously monitored and controlled, meaning that gas exchange from the subject can be measured. This measurement is done by extracting samples of the air from the room and measuring the gas within the samples. [72].

This method does not require sophisticated mechanical engineering and is still accurate. It is, however, very expensive and specialised, and an artificial environment for the subjects.

There are various methods of portable indirect calorimetry available today, many of which are widely used in medical practice and research. These methods measure gaseous exchange using a mask, a mouthpiece or a hood.

- Douglas Bag: The Douglas bag is a gas-impermeable bag which is carried on a subject's back during the measurement period. The subject wears a nose-clip either a mouthpiece or half mask. The bag collects the air expired by the subject, which can then be analysed to determine the gaseous exchange during the measurement time. This is then converted to heat output (i.e. energy expenditure).
- Respirometer: In respirometry, the subject blows through a valve into a machine which meters the expired air. Samples of this air are extracted for gas analysis, allowing the gasses to be measured and a conversion to heat output to be calculated.
- Ventilated hood: The subject wears a hood which allows air to be drawn over his/her head while sitting or lying. The air is controlled and monitored, allowing samples of the "used" air to be extracted and analysed to determine the gaseous exchange of the subject. Ventilated hood systems can be used in conjunction with Douglas bags to allow use outside of a clinical environment. This method is suitable for all age groups, including children. [72]

Portable methods of indirect calorimetry are more user-friendly than whole-body measurements, although nose clips and masks may be uncomfortable

for some. They are generally not useful for measurements over long periods of time. However, for short-term measurements they are generally preferred to whole-body techniques [72] and have been shown to be accurate when compared with direct calorimetry [81].

1.2.2.1.3 Doubly Labelled Water

Doubly labelled water is a method of measuring free-living total energy expenditure. A baseline urine sample is taken from the subject who then drinks a weighted dose of an oxygen and hydrogen isotope. Oxygen and carbon dioxide production can be calculated from the washout rate of both isotopes: "The slope of the washout line representing *O is steeper than the washout line representing *H, and the difference between the slopes represents CO^2 production" [30].

Measurement of energy expenditure requires at least two post-dose samples of body fluids for analysis, and can take up to 7 days for children and 14 for adults [72, 82].

The doubly labelled water method is time consuming and does involve expensive isotopes, but has been shown to be accurate [72] and suitable for people of all ages and all physical states.

1.2.2.2 Resting energy expenditure (REE)

Resting Energy Expenditure (REE) is the minimum amount of energy that the body requires when lying in physiological and mental rest. Approximately 65 - 70% of an individual's total daily energy expenditure is REE (unless that individual is extremely active, in which case this percentage is likely to be as low as 40%) [83]. REE is determined mostly by body weight and composition [84], with FFM being a much greater predictor of REE than FM. REE is known to be variable as follows [83]:

- Infants and children have a high REE for their size, due to the energy cost of growth.
- Males have a higher REE than females, because males generally have a higher percentage of FFM than females.
- Young people have a higher REE than the elderly, because of the elderly having less FFM than the young.
- REE can vary between individuals because of: medications, nutritional state, illness and temperature [85].

When strict conditions are maintained and monitored in the hours preceding the measurement of the energy expenditure, we will refer to the energy expenditure as Basal or Resting Metabolic Rate (BMR or RMR). Otherwise, we will use the term Resting Energy Expenditure (REE). REE will always be slightly higher than BMR as a result of this difference.

1.2.2.2.1 Measuring resting energy expenditure

Direct calorimetry measures energy expenditure as heat emission from the body and therefore provides a direct and accurate measurement of energy expenditure [81, 72]. In a 1995 review of measurement methods, Murgatroyd et al report that direct calorimetry is "potentially the most accurate measure of energy expenditure" [72]. Although direct calorimetry is considered to be very precise, it does have disadvantages. One is the fact that the heat transfer in food, lighting, etc must be accounted for, involving some complicated measurements and models [72, 73]. For this reason, indirect calorimetry is often preferred for measuring REE - it has been used as such from as far back as the Harris and Benedict study of 1919 [86] to studies today, such as the Avon Longitudinal Study of Parents and Children (ALSPAC) [87]. However, although indirect calorimetry is generally preferred over direct calorimetry, it remains a complicated and time consuming process, involving trained technicians and specialised equipment [88]. For this reason, it is far more convenient to be able to use prediction equations to estimate REE whenever possible. Since these equations are simply predictions rather than measurements taken under controlled conditions, they are estimates of REE rather than BMR.

A disadvantage which applies to *any* measurement recorded in a clinical setting, including direct and indirect calorimetry, is that it is an artificial environment. This means that subjects may behave differently outside of the clinical setting (for example, people with eating disorders [72]) so calorimetry measurements are unlikely to reflect real-life situations. For this reason, Seale

et al report that "calorimetry is better suited for measuring changes in EE for controlled experiments where treatment effects are being investigated." [81].

1.2.2.2.2 Models for resting energy expenditure (REE)

As discussed previously (section 1.2.2.1), it is difficult and expensive to measure REE directly so indirect methods have to be adopted. In his paper of 1915 [89], Benedict discusses various possible factors which may have an effect on resting energy expenditure. After considering the possibility of a general relationship between body weight (adults) or body surface area (adults and children) and resting energy expenditure, he reported that:

> We find here not the slightest evidence of a law governing the relationship between the total body weight and the total heat production.

> It is clear that even with normal individuals a relationship between body surface and heat production which may be expressed with any approximation to mathematical accuracy does not exist.

He did, however, accept that factors such as age and weight do have an influence on REE, but this relationship is not a simple one. Benedict concluded that "body composition ... has a great influence upon the basal metabolism", which could be said to account for apparent effects of factors such as age. In 1980, Cunningham [90] reported that "sex and age are factors influencing

the body composition of an individual, but body composition is the principal determinant of [REE]", and proceeded to suggest a model for resting energy expenditure based on lean body mass alone. Nielsen et al [91] report findings that fat mass and fat-free mass (in addition to age) were both significant and independent predictors of resting energy expenditure in both genders, from a study comprising 253 adult subjects. In contradiction to Cunningham's suggestion of a model where the only explanatory variable is fat-free mass, Nielson et al [91] reported that "expressing REE relative to FFM alone will introduce errors when lean and obese populations are compared", and suggested a model including both fat mass and fat-free mass as explanatory variables. Cunningham reports, however, that of FFM, age, height, sex and body mass, FFM is the most predictive of REE, accounting for 70% of the variability [90]. Nielsen et al consider various studies [92, 93, 90, 94] which have found no justification for including fat mass in a model for resting energy expenditure. They suggest that inaccurate measurement may be the reason for the findings, stating that "it is possible that less robust methods for assessment of body composition and/or REE have confounded the ability to find independent effects of FM on REE in some of these studies" [91].

Rather than models being based on body composition itself, prediction models for resting energy expenditure used in practice are generally dependent on other explanatory variables - which are themselves commonly considered to be predictors of body composition. Four years after Benedict published his review of factors affecting metabolism [89], he co-wrote a publication with J. Harris, introducing a set of prediction models for REE which have been

used until very recently [86]. These models are shown in equations (1.16) and (1.17) in section 1.2.2.2.2.

In the process of their research, Harris and Benedict appear to contradict Benedict's earlier findings regarding body surface area [89] by reporting that there is, in fact, a strong relationship between body surface area and metabolism. They found, in their 1919 study [86], body surface area to be a far more reliable predictor of metabolic rate than body weight. However, they did not include BSA in their final published models, stating that measurement methods must be standardised before BSA could reliably be used in BMR models. New methods for determining BSA have been developed since the DuBois method used by Harris and Benedict, but as yet BSA does not appear to be included in any resting energy expenditure prediction models.

In recent studies, many researchers have found that the Harris and Benedict (HB) models overestimate REE in today's population [93, 95, 96], with Frankenfield et al stating that the HB equations systematically overestimate BMR by $\geq 5\%$ [95]. This does appear to be quite reasonable since body composition in the population has undoubtedly changed over the last 90 years.

Various studies more recently have resulted in different sets of BMR prediction models [97, 98, 99, 100]. However, it should be noted that Frankenfield et al [95] found that the margin of error achieved with the Harris-Benedict equations has not been significantly improved in any of these later studies.

Kaplan et al in 1995 [96] compared various sets of prediction equations with REE measurements obtained from indirect calorimetry in a study consisting of 102 children. They used paired t-tests to determine whether BMR derived from the prediction equations was unbiased compared to that from the indirect calorimetry method. They considered the HB equations, as well as equations from the World Health Organisation [98] and two sets of equations from Schofield, one using weight and one using weight and height [97]. Their findings were that the Schofield equations with weight and height were the best predictors of BMR (in kilocalories per day) in children, predicting $100\% \pm 19\%$ of REE (compared to $92\% \pm 25\%$ for the HB equations (1.16) and (1.17)). Rodriguez et al [101], using the Bland-Altman method [102] to investigate the limits of agreement between several sets of prediction equations (including Harris-Benedict, WHO and Schofield) and indirect calorimetry measurements, also recommend the Schofield equations for use in a mixed population of obese and non-obese children and adolescents. These Schofield equations [101, 56, 97] are given in equations (1.18) to (1.21) in section 1.2.2.2.2.

It is quite hard to believe, however, that parameters should change so suddenly and so completely as suggested by Schofield [97]. It is possible that there is a strong linear relationship between weight and height for infants, making the estimates of parameters in these models very unstable due to a classical problem of collinearity.

The set of equations which appear to be most commonly used today were published by Henry in 2005 [100], and are known as the Oxford equations (equations (1.34) to (1.37), section 1.2.2.2.2). These equations were developed by plotting weight and BMR (by indirect calorimetry) of 10552 people, then considering the linear trendline.

While developing these equations, another new set of equations which included height as an independent variable was considered. However, the study reported "no significant advantage was afforded in predicting BMR with the inclusion of height"[100].

Ramirez-Zea [103], in another validation of predictive equations for BMR in adults, uses linear regression and concordance correlation analysis (a measure of agreement), and concludes that the Oxford equations are accurate in men, across a wide range of age and BMI. However, the study reported that none of the proposed equations were appropriate for estimating BMR in women.

In conclusion, it appears that resting energy expenditure is influenced to some extent by body composition. Therefore, it seems plausible that prediction models should include some measure of this. Of course, there is the possibility of the added complication that while factors such as height and weight can be accurately and relatively easily measured, we rely completely upon estimates of body composition. However, considering the fact that body composition can be (to some extent) predicted by anthropometric

measurements, it seems reasonable that these measurements could be used as proxies in the prediction of resting energy expenditure.

Equations (1.16) to (1.39) are some of the published linear prediction models for REE in children involving anthropometric measures, where W is weight (kg), H is height (m) and A is age (years).

Harris and Benedict (HB) [86]

$$\text{Male, 15+} \quad REE = 66.4270 + 13.7516 \times W + 500.33 \times H - 6.755 \times A \tag{1.16}$$

$$\text{Female, 15+} \quad REE = 655.0955 + 9.5634 \times W + 184.96 \times H - 4.6756 \times A \tag{1.17}$$

Schofield HW [97]

$$\text{Male, 3-10} \quad REE = 414.9 + 19.59 \times W + 130.3 \times H \tag{1.18}$$

$$\text{Male, 10-18} \quad REE = 515.5 + 16.25 \times W + 137.2 \times H \tag{1.19}$$

$$\text{Female, 3-10} \quad REE = 371.2 + 16.969 \times W + 161.8 \times H \tag{1.20}$$

$$\text{Female, 10-18} \quad REE = 200.0 + 8.365 \times W + 465.0 \times H \tag{1.21}$$

Schofield W [97]

$$\text{Male, 3-10} \quad REE = 505 + 22.7 \times W \tag{1.22}$$

$$\text{Male, 10-18} \quad REE = 693 + 13.4 \times W \tag{1.23}$$

$$\text{Female, 3-10} \quad REE = 486 + 20.3 \times W \tag{1.24}$$

$$\text{Female, 10-18} \quad REE = 659 + 17.7 \times W \tag{1.25}$$

FAO/WHO/UNU [98]

$$\text{Male, 3-10} \quad REE = 495 + 22.7 \times W \tag{1.26}$$

$$\text{Male, 10-18} \quad REE = 746 + 12.2 \times W \tag{1.27}$$

$$\text{Female, 3-10} \quad REE = 499 + 22.5 \times W \tag{1.28}$$

$$\text{Female, 10-18} \quad REE = 651 + 17.5 \times W \tag{1.29}$$

Oxford HW [100]

$$\text{Male, 3-10} \quad REE = 306 + 15.1 \times W + 74.2 \times H \tag{1.30}$$

$$\text{Male, 10-18} \quad REE = 299 + 15.6 \times W + 266 \times H \tag{1.31}$$

$$\text{Female, 3-10} \quad REE = 349 + 15.9 \times W + 210 \times H \tag{1.32}$$

$$\text{Female, 10-18} \quad REE = 462 + 9.4 \times W + 249 \times H \tag{1.33}$$

Oxford W [100]

$$\text{Male, 3-10} \quad REE = 514 + 23.3 \times W \tag{1.34}$$

$$\text{Male, 10-18} \quad REE = 581 + 18.4 \times W \tag{1.35}$$

$$\text{Female, 3-10} \quad REE = 507 + 20.1 \times W \tag{1.36}$$

$$\text{Female, 10-18} \quad REE = 761 + 11.1 \times W \tag{1.37}$$

Maffeis et al [99]

$$\text{Male, 6-10} \quad REE = 307.89 + 6.84 \times W + 564.59 \times H - 16.53 \times A \quad (1.38)$$

$$\text{Female, 6-10} \quad REE = 371.29 + 8.56 \times W + 373.21 \times H - 8.68 \times A \quad (1.39)$$

Several studies have examined the theory that resting energy expenditure (and total energy expenditure) can be expressed as some power function of body weight; i.e. $REE \propto \text{mass}^{\alpha}$ where mass is body weight in kg. Over the years, several studies have attempted to find the precise value of α. In the 1880s, Rubner [104] while studying the metabolism of dogs found this exponent to be $\frac{2}{3}$. However, Max Kleiber, in 1932 [105], suggested that $\alpha = 0.74$ and opted for the "$\frac{3}{4}$ rule" (now known as "Kleiber's Law") in a later publication [74]. It has been reported that this relationship can be applied to all mammals, from mice to elephants.

The model formulated by Kleiber is:

$$\text{REE (kcal / day)} = 70 \times \text{mass}^{0.75} \text{ [74, 106, 107]}$$

A technical report published in 1989 reported that "although infants were not included in Kleiber's investigations, this equation has been applied to infants... the exponent may be greater than 0.75 in infants, because of the additional energy requirements for growth, or because of the higher metabolic activity of adipose tissue in infants than in adults. The exponent tends to fall as growth decreases in older children" [108].

Several studies and reviews have considered this proportional relationship with varying results. Some report that the exponent of 0.75 is valid for all mammals [107, 106]. However, others have reported contradictory findings. In a re-examination of Kleiber's Law in 2001, Dodds, Rothman and Weitz concluded "we find evidence that there may not be a single scaling law for metabolic rate, and if it were to exist, we also find little compelling evidence that the exponent should be $\alpha = 3/4$" [109].

While we do know that resting energy expenditure is to some extent dependent on weight, recent research has shown that fat-free mass is more metabolically active than fat mass. It may be possible that if a scaling law does exist, it should incorporate some measure of body composition, possibly in addition to body weight.

An obvious extension of Kleiber's model is

$$REE \propto FM^a + FFM^b \tag{1.40}$$

This model will be explored in chapter 5 using nonlinear modelling techniques, as discussed in section 2.4.

1.2.2.3 Diet-induced thermogenesis (DIT)

Diet-induced thermogenesis (DIT), also known as the thermic effect of food (TEF), can be described as the amount of energy used (and therefore the amount of heat produced) over and above the basal metabolic rate during processes related to the digestion of food [110, 111]. Studies have reported a significant level of intra-individual variation in diet-induced thermogenesis [112]. DIT can be affected by many factors, including exposure to heat or cold, nicotine, caffeine and alcohol [111, 113], medication, short- or long-term illness [114, 115] and levels of physical activity [112]. Tataranni et al [116] stated that diet-induced thermogenesis is "the most difficult to measure and least reproducible component of daily energy expenditure". Reed and Hill state that it accounts for 3-10% of total daily energy expenditure [117], while Westerterp reports this to be 5-15% of TEE [113] and Granata states it to be 10% [118].

1.2.2.3.1 Measurement of diet-induced thermogenesis

There have been many studies involving measurement of diet-induced thermogenesis. In the vast majority of these studies, basal metabolic rate is measured (mostly using indirect calorimetry) and used as a baseline. After subjects have consumed a meal of controlled energy content and nutritional composition, energy expenditure is measured for a period of time. Diet-induced thermogenesis is then calculated as the difference between postprandial energy expenditure and basal metabolic rate [119, 120, 121].

While studies generally do not dispute this procedure, one thing they consistently differ on is the length of time for which the postprandial measurement should be taken in order to get an accurate DIT result. Many studies have used DIT measurement times of between 1 and 3 hours [122, 123, 124, 125]. A small number of studies take the measurement over longer periods of time. For example, Nelson [84] and Laville [121] measured over 6 hours, D'Alessio et al [126] over 8 hours and Steiniger [127] over 10 hours. A small number of studies have used 24-hour measurements, by taking the measured fasting energy expenditure (adjusted for spontaneous physical activity) as a baseline and subsequently measuring 24-hour energy expenditure under fed conditions. DIT is then calculated as the difference between the two measurements. Schutz, Bessard and Jéquier [122] in 1984 measured DIT over 24 hours and proposed that a measurement should be made over 15 hours. Twelve years later, Reed and Hill in a study comprising 131 tests, concluded that DIT "is a response lasting $\geq$6h in most people and [they] recommend that measurements last $\geq$ 5h" [117].

1.2.2.3.2 Factors affecting diet-induced thermogenesis

There are several factors which may have an effect on diet-induced thermogenesis. One such factor is the age of the subject. Morgan and York [128] and Schwartz et al [129] both reported a decrease of DIT with increasing age. Visser et al [119] reported that this effect was only found in male subjects and disappeared when body composition was taken into account, while

Tataranni et al [116] reported that diet-induced thermogenesis was inversely correlated with age for men only. Literature examining the effect of age on diet-induced thermogenesis appears to be very sparse and inconclusive.

One factor which has been discussed a great deal more in published literature is that of the energy content and the composition of meals. D'Alessio et al [126] reported no relationship between DIT and nutritional composition but a positive correlation between DIT and energy intake, a result also reported by Kinabo two years later [130]. However, several researchers have reported a clear effect of the nutritional content of meals upon diet-induced thermogenesis, with protein- and carbohydrate-rich meals resulting in a higher DIT than fat-rich meals [131, 132, 120].

Another area which seems to be highly inconclusive is whether or not body composition has an effect on DIT. Zahorska [133] in 1980 found a similar DIT between obese and non-obese subjects in a study using body surface area to determine body composition and indirect calorimetry to determine energy expenditure. In their study using densitometry and indirect calorimetry, D'Alessio et al [126] reported similar results. Maffeis et al, using skinfolds to identify obesity, also reported no effect of body composition on diet-induced thermogenesis, as did Das et al [134] in a more recent study, defining obesity by BMI tertiles. Granata and Brandon [118], in a 2002 review, concluded that there was no consensus for a link between obesity and DIT. However, in discussing this article in a letter to the editor entitled simply "The thermic effect of food is reduced in obesity", de Jonge and Bray [135] state: "In

conclusion, we agree with the authors that there is no consensus on whether TEF is reduced in obese individuals... When multiple factors are looked at simultaneously, however... the evidence for a reduction of TEF in obesity becomes significantly stronger". Many researchers have reported a similar result, including Schutz et al [122], Schutz et al [123], Armellini et al [136] and Bray et al [137]. The findings of each of these studies were that obese subjects had a lower DIT than non-obese subjects. I am unable to find any studies reporting the opposite effect of obesity on DIT. One possible reason for the differences between study results may be the differing methods used to define and determine body composition and obesity.

I have been able only to find one study resulting in a prediction equation for diet-induced thermogenesis. This study, by D'Alessio et al in 1988 [126], reported that the diet-induced thermogenesis of subjects could be predicted almost entirely ($R^2 = 0.82$) from the energy content of the meal by the formula:

$$\text{TEF} = -1.16 + 0.082 \times \text{kcal}$$

I am unable to find any subsequent references to this model in the very widespread literature on the topic. However, it seems that with the large potential that seems to have been widely demonstrated for DIT to be variable, it is unlikely that a prediction model with only one predictor would give reliable results outwith the study sample!

1.2.2.4 Physical activity (PA)

Physical activity thermogenesis can be described in two parts:

1. Exercise thermogenesis
2. Non-exercise activity thermogenesis (*NEAT*)

Exercise thermogenesis generally describes the energy cost (i.e. heat generated) during periods of physical activity, whether intentional exercise or that which occurs during day-to-day living - climbing stairs, for example. Non-exercise activity thermogenesis encompasses the energy costs of everyday living, including activities such as fidgeting, maintaining posture and muscle contraction [71].

Very often, these are considered together, and are known simply as physical activity thermogenesis. This is estimated to account for roughly 20% of daily energy expenditure in "normal" human beings. In those who are more active than "normal" (e.g. athletes), this percentage could be considerably higher [83].

1.2.2.4.1 Physical activity and body composition

It is generally accepted that the energy expended during physical activity is dependent upon the activity itself, but also upon the individual undertaking that activity [138]. The idea of a possible connection between physical activ-

ity and body composition has been around for many years. Most professionals in medicine today are likely to tell us that physical activity is important when it comes to the prevention of obesity. However, the particular effect of physical activity on this imbalance is under some debate. At the time of the 2003 Moore study [139], literature in the area was sparse. Today, although there have been many more publications in the field of physical activity and its effect on body composition, the literature available is, overall, inconclusive. Moore et al [139] tell us that at the time of their research, studies were generally cross-sectional rather than longitudinal and did not allow a significant period of follow-up on the subjects; their own study was longitudinal, extending over a period of eight years. Since this 2003 publication, there have been studies published with far longer follow up periods. For example, the 2004 paper by Hancox et al [140] described a longitudinal study with 23 years follow-up.

Conclusions from different sources are conflicting. Some studies, including Delany [141] and Goran et al [142], report no association between physical activity and body composition. In contradiction to these findings, however, many studies have reported some association between physical activity and body composition. For example, Sallis et al [143], find that obese children are less physically active than non-obese, and children who are more active appear to develop less body fat over time than less active children.

In discussing studies involving television watching (i.e. lack of physical exertion) and obesity, Anderson and Butcher [144] point out that different results

may be explained by differences in the types of study: "These mixed findings, though, tend to come from observational or prospective studies. More rigorous experimental studies consistently find that reducing children's television watching lowers their BMI."

One further criticism of studies in this field is that of small sample size. A report published in 1995 by Maffeis et al [145] detailed their study which involved, at the analysis stage, only 29 children (13 obese, 16 non-obese), while Goran et al [142] in 1998 studied 75 children and the 2003 study by Moore et al [139] involves only a slightly larger sample size of 103 - similar to that of the original Framingham Children's Study by LL Moore et al [146]. However, while these are larger than the Maffeis study, they may still be seen as being fairly small! Sallis [143] had a sample size of 286 children, a study from the National Heart, Lung and Blood Institute [147] recruited 2379, and Berkey et al [148, 149] had a sample size of 11,887.

It could of course be suggested that there is a reverse causal effect at play. For example, a paper published in 2005 by Norman et al [150] investigates the effect of excess body mass in adolescents on their ability to perform sustained exercise, concluding that overweight adolescents are "limited by the increased cardiorespiratory effort required to move their larger body mass" and "burdened by the metabolic cost of their excess mass". It is therefore clear the relationship between body composition and physical activity cannot be as straightforward as a one-directional causal relationship. It would be extremely difficult, if not impossible, to attempt to accurately investigate

and describe the complex relationships involved within any reasonable time frame.

1.2.2.4.2 Measuring activity and determining energy used

There are various methods which can be used to measure physical activity and/or determine physical activity energy expenditure. In their 2001 review, Sirard and Russell classified measures of physical activity thermogenesis into three groups: subjective measures, secondary measures and primary measures (criterion standards). [151].

Self-report is a subjective method which has been in use for a number of years. It can be immediate in the form of either an activity diary, or delayed, as a recall of physical activity over a fixed period of time. In 1986, Washburn reported that the practice of keeping an activity diary is inconvenient to the participants, so "the delayed recall technique is the most practical and commonly used approach" [152] out of the two methods. Over the past twenty years, studies have attempted to find whether or not self report is an accurate way of determining physical activity. In 1993, Sallis et al [153] using a test-retest method for analysis of accuracy, found that activity recalls were of adequate reliability and validity in children as young as age 10-11 (note, however, that the total correlation for the total group was only 0.53). However, many studies have shown that self report is actually inaccurate [152, 154], especially for children. Ekelund et al, in 2001, reported that "The self-report

methods rely on the subject's ability to recall and report physical activity and should be used with caution, especially in subjects younger than 15 yr of age" [155]. In the same year, Trost suggested that this was because "many children and adolescents have difficulty accurately recalling their past physical activity behavior" [156]. This method of determining physical activity energy expenditure relies on two calculations - the first being a calculation of the time spent on each activity, the second being a calculation of the energy cost of each activity [157].

Regardless of whether or not people have problems with memory, self-report remains a very subjective measure of physical activity. For this reason, studies have become more and more interested in validating the use of more objective measures (primary or secondary).

One such "secondary measure" is heart rate (HR) monitoring. It has been shown that, despite the inter-individual variations in heart rate, the heart rate and oxygen uptake of an individual are generally linearly related during exercise [158, 159]. Therefore, it seems logical to assume that heart rate could be useful as a proxy measure for oxygen uptake (and therefore, energy expenditure) during periods of physical activity, assuming that this linear relationship is known for the individual. This has been shown to be the case [160, 161]. It has been shown that while heart rate monitoring provides a good estimate of energy expenditure during periods of high physical activity, it is inaccurate during periods of low activity [159].

The main disadvantages to heart rate monitoring as a method of determining physical activity energy expenditure are (a) the need for the linear relationship with oxygen uptake to be determined for each individual, and (b) the delay in the response of the heart rate to activity changes [160]. The main advantage is the fact that heart rate monitoring is relatively inexpensive and noninvasive [158].

Another secondary method for determining physical activity energy expenditure is accelerometry. Accelerometers are activity sensors which can be uniaxial or triaxial. The accelerometer is worn on the body, and assesses postures and motions during the measurement period [162]. In 1999, Westerterp reported that "there is no clear difference for correspondence between indirect calorimetry and accelerometer counts" [163]. Triaxial accelerometry has been shown to be more accurate than heart rate monitoring [164] in determining the energy expended during physical activity. In a study that used multiple regression with various measures of physical activity including heart rate monitoring and accelerometry, Eston et al [160] found that triaxial accelerometry was the best single predictor ($R^2 = 0.83$ for accelerometry, 0.638 for HR) of oxygen uptake, and concluded that "a triaxial accelerometer provides the best assessment of activity". Ainslie, Reilly and Westerterp, however, report that "not all activity is reflected in acceleration or deceleration such as load carriage or on a gradient. This failure to record activity leads to large errors in predicted EE, especially participants engaged in high-intensity activity" [165].

Another type of motion sensing which has been shown to have potential in the field of physical activity energy expenditure measurement research is pedometry, which counts steps by responding to vertical acceleration [165]. Eston et al, while concluding that accelerometry was the single best assessment of physical activity, also report that "pedometry offers potential for large population studies" [160]. Ainslie et al [165] recognise the main limitation of pedometry, "they do not quantify stride length or total body displacement and are therefore of very limited utility in predicted EE", but do accept that "if overall walking activity is the outcome to be assessed, the pedometer is a useful and inexpensive instrument" [165].

While heart rate monitoring and accelerometry have been shown to be accurate measures individually, researchers find cause for combining the two methods. In the previously mentioned study of 1998, Eston et al [160] reported that the best model for predicting physical activity thermogenesis was one which included both triaxial accelerometry and HR monitoring. This model had an R^2 value of 0.85, which is only slightly larger than the model with accelerometry alone.

A "primary measure" of physical activity energy output is indirect calorimetry [151], which can be carried out either by fitting a calorimetry chamber with exercise equipment, or by using a portable method of indirect calorimetry with free-standing exercise equipment. While this has been shown to produce accurate results, it should be noted that calorimetry does not replicate real-life situations. For this reason, as well as that of the expense and

technicians required, indirect calorimetry is unlikely to be suitable for many physical activity studies. It is, however, frequently used as a validation of other methods such as heart rate monitoring and accelerometry.

Although doubly labelled water (see section 1.2.2.1.3) can not directly measure periods of activity themselves, Bar-Or et al [166] state that "If a measurement of resting metabolic rate (RMR) is also performed, then an estimate of activity energy expenditure may be obtained by difference ($TEE - [RMR + 0.1 \times TEE]$)". (Note that here, $0.1 \times TEE$ represents DIT.)

1.2.2.4.3 Activity as a level

Physical activity level (PAL) can be determined as a ratio, as shown in equation (1.41).

$$PAL = \frac{\text{total 24h energy expenditure}}{\text{Basal metabolic rate}} \tag{1.41}$$

Using this equation, people can be classified, according to 1985 FAO/WHO/UNU reference values [167] as follows in Table 1.3.

PAL value	**Description**
< 1.2	Bed rested - most likely when in care of others
$1.2 - 1.55$	Low activity level - sedentary lifestyle
$1.55 - 1.71$	Medium activity level - occasionally active, typical office work
$1.71 - 1.95$	High activity level - some manual work and/or regular exercise
> 1.95	Very high activity level - a fair amount of manual work or exercise training

Table 1.3: Reference values for PAL (FAO / WHO / UNU 1985) (adults)

1.2.2.4.4 Physical activity ratios for children

Energy expenditure for physical activity that has not been measured may, on occasion, have to be estimated. In 1990, suggestions for estimating the energy cost of activities in children were published in a report by the International Dietary Energy Consultancy Group [168]. These guidelines are as follows:

1. If the child's BMR is not known, calculate it with appropriate local formulas or with those of SCHOFIELD et al. suggested by FAO/WHO/UNU.
2. For children, 15 years or older, apply to the child's BMR the same multiple or BMR determined for equivalent activities in adults.
3. For children under 15 years of age:
   (a) For sedentary activities (with little or no movement), lying down, sitting or standing without displacement, use a factor of 1.1, 1.2 or 1.4, respectively, for all children under 15 years.
   (b) For non-walking light activities, use a factor of 2.0 or 2.2 X BMR for ages 1.5-5.9 or 6.0-14.0 years, respectively.
   (c) For walking at a normal pace on level ground and for moderate activities, use a factor of 2.2 or 2.9 X BMR for ages 1.5-5.9 or 6.0-14.9 years, respectively.

(d) For heavier activities, apply to the child's BMR the multiple of BMR determined for equivalent activities in adults, multiplied by 0.5, 0.65 or 0.8 for ages 1.5-5.9, 6.0-12.9 and 13.0-14.9 years, respectively.

These calculations will probably have a smaller error when used to estimate the energy expenditure of a group or population of children than of a single specific child.

1.2.2.5 Measuring total energy expenditure

We know that total daily energy expenditure comprises the components: resting energy expenditure, diet-induced thermogenesis and physical activity thermogenesis. However, as a result of high variability and possible inaccuracies in determination of these three components, it is likely that simply "putting together" the three measured or calculated energy expenditures will not yield an accurate calculation of total energy expenditure. Ravussin et al, in 1986, reported that "daily human energy requirements calculated from separate components of energy expenditure are inaccurate and usually in poor agreement with measured energy intake" [94]. This could be a combination of poor measurement of energy expenditure and poor measurement of energy intake. The doubly labelled water method of estimating total energy expenditure avoids most of these concerns.

Direct and indirect calorimetry can be used to calculate accurate measures of total daily energy expenditure. The subject would spend a period of 24 hours or more within a sealed calorimetry chamber as described in section 1.2.2.1. During this time, energy expenditure could be continuously measured. Portable methods of indirect calorimetry are not suitable for extended periods of time, so would not be useful in determining total daily energy expenditure.

Although whole body calorimetry does give very accurate results, it is not a natural environment and is unlikely to reflect real life. Because of its ability

to be used in free living conditions, doubly labelled water is considered the gold standard in energy expenditure measurement [166, 165]. This method of calculating total daily energy expenditure has the advantage of being suitable for all age groups and physical states of people [72]. However, it is an expensive process, and requires measurement several times over a period of up to two weeks in some cases. In 1997, Bratteby et al conducted a cohort study into daily energy expenditure measurement methods in which 374 adolescents participated [169]. The researchers report that "it was impossible to measure total energy expenditure by the DLW method in the whole study group on account of the great costs".

As with physical activity energy expenditure, total daily energy expenditure can be examined with the use of activity diaries, motion sensors or heart rate monitoring.

The use of activity diaries would require the subject to account for every minute of a 24-hour period of time. The time spent in various activities could then by multiplied by the energy cost of those activities to calculate the total energy cost of the day. Bratteby et al, in a study comprising 50 adolescents, reported in 1997 that "the activity diary... provides a close estimate of TEE in groups of adolescents, but is unsuitable for an individual estimate" [170]. Westerterp reports that "A disadvantage of questionnaires is the fact that subjects can easily overestimate or underestimate the time spent in activities, and most questionnaires are not applicable for all subject categories from children, people with and without jobs, to the elderly" [171].

While heart rate monitoring may be used for calculating total daily energy expenditure and has the advantage of being an objective measure compared with the subjectivity of activity diaries, Bratteby et al point out that "TEE estimation by HR monitoring requires, however, complex and time-consuming individual calibration procedures which limit the use of this method" [169]. It should be noted that the heart rate is affected at rest by many more factors than simply physical activity [171] though these factors do not affect the heart rate during periods of moderate to intense physical activity [72]. Therefore, heart rate at rest may not be a reliable indicator of energy expenditure.

1.2.3 Energy Intake

1.2.3.1 Total energy intake

Energy intake (EI) comes entirely from the consumption of food and drink. The energy in food can be measured using bomb calorimetry, a process which relies on the principles of direct calorimetry (see section 1.2.2.1.1). Bomb calorimetry directly measures the heat output as food is consumed completely.

When in energy-balance equilibrium ("zero energy-balance"), energy intake must equal energy output. Therefore, assuming that a subject is in equilibrium, daily energy intake can be measured by estimating daily energy expenditure. However, this method is **only** valid while in equilibrium [172].

One method of determining the energy intake of individuals is self report. This relies on subjects giving an honest account of their food consumption over a period of at least 24 hours. Naturally, this presents some issues. It could be quite likely that subjects will not remember to record everything, or will estimate portion sizes incorrectly. Bandini found that self report was unreliable for both obese and non-obese subjects when validated against TEE measured by doubly labelled water, and concluded in 1990 that "dietary intake data cannot be used to assess the role of energy intake or expenditure in the development of obesity" [172]. Using data from a Dutch health examination monitoring project comprising 2079 men and 1467 women, Braam et

al [173] find that as men and women become more overweight, they have a greater tendency to under-report energy intake. They also found that age affects self report in both sexes, and smoking habits and education level affect self report of energy intake in males only. Fisher et al [174], in a study of children, found that heavier children were more likely to under-report energy intake than 'normal-weight' children. Macdiarmid and Blundell report that "women are more likely to under-report than men, and under-reporting is more common among overweight and obese individuals". They then go on to make the logical conclusion that 'positive health image' foods are more likely to be over-reported, while 'negative health image' foods are more likely to be under-reported. In 2007, Probst and Tapsell [175] conducted a study of 147 subjects. They found that only 46.2% of the sample reported their energy intake accurately. Of the remaining subjects, a higher proportion under-reported than over-reported. The authors found, contrary to previous studies, that there was no effect of age, gender or BMI, and concluded that computerised dietary recall may yield more accurate results than face-to-face assessment. In this study, Probst and Tapsell (seemingly arbitrarily) defined under-report as $\frac{EI}{BMR} < 1.35$ and over-report as $\frac{EI}{BMR} > 2.4$

1.2.3.2 Importance of energy intake during early infancy

In 1981, a study by MS Kramer [176] recognised the potential of breastfeeding as a method of protection against obesity in later life, concluding that "breast-feeding does protect against later obesity" and attributing the conflicting results of previous studies to "insufficient attention to methodological

standards". The idea that breastfeeding, if not beneficial, is certainly not detrimental in the fight against obesity is an idea which now appears to be generally accepted throughout literature [177, 178, 179, 180] and life.

There are several possible reasons for this, as highlighted in a 2007 publication by Singhal and Lanigan [181]. The first possible explanation is that of intake regulation. When breastfeeding, infants are to a greater extent in control of how much they take than those who are formula-fed. It is not known, however, the extent to which this effect extends into adulthood. Next comes the difference in nutritional quality between breast milk and formula milk, also highlighted by Hoddinott, Tappin and Wright in 2008 [180]. One such difference is the protein content present in each type of milk. High early protein intake has been shown to be linked to later adiposity [182], and as reported by Singhal and Lanigan [181], protein intake in formula-fed infants can be up to 70% higher than that of infants who are breastfed. Hoddinott, Tapin and Wright highlight the issue of the fundamental differences between the types of feeding by stating that "formula milk is just a food, whereas breast milk is a complex living nutritional fluid that contains antibodies, enzymes, and hormones, all of which have health benefits" [180]. Finally comes the suggestion that growth acceleration in the early stages of life "may be a key programming window" for body composition in later life [181]. It has been shown that breast-fed babies have a lesser early growth acceleration than those who are formula fed [183]. Settler et al in 2005 [184] suggested that these first few weeks of post-natal life are crucial when it comes to the determination of body composition in the years to follow.

A few studies find that there is no effect of breastfeeding on protection against later-life obesity. A study in 2004 by Bogen et al [185] considered a sample of low-income black children and found no such effect. In the same year, Grummer-Strawn and Zei [186] reported a protective effect among non-Hispanic white children but a lack of such an effect among black or Hispanic children. It seems, therefore, that ethnicity may be a confounding effect. It is also known that socioeconomic factors are often confounded with breastfeeding [187]. While authors may be keen to stress that breastfeeding may indeed be beneficial in the battle against obesity, they are often just as keen to point out that more research needs to be done to determine how much of the effect is down to the breastfeeding itself, and how much is down to confounding factors [188, 179].

Overall it appears clear that, though overweight and obesity are becoming more and more apparent in today's adults and children, those first few days of life are potentially extremely important when it comes to making nutritional considerations. Furthermore, when taking such considerations into account, breastfeeding certainly does not hurt! A recent publication of reports by the Institute of Medicine [67] recommends breastfeeding for all infants. As Woodward-Lopez et al conclude: "Although the effect of breastfeeding may be small compared to other factors, the promotion and support of breastfeeding initiation and duration are a low-cost method of providing many health benefits, one of which is likely to include reducing obesity risk in children" [189].

1.2.4 Energy (im)balance and body composition

Moore et al [139] report that "obesity is an imbalance in energy intake and expenditure", a claim that is widespread in the literature [190, 191, 192, 67]. While this sounds like a relatively simple idea in theory - to maintain weight (under normal circumstances) what you do should be equal to what you eat - it is far from simple in practice. The Institute of Medicine in 2005 reported that "although 'energy intake = energy expenditure' looks like a fairly basic equation, in reality it is extraordinarily complex when considering the multitude of genetic, physiological, sociocultural, and environmental factors that affect both sides of the energy-balance equation and the interrelationships among these factors" [67].

Are increasing energy intake and decreasing energy expenditure equal contributors in the 'obesity epidemic'? Is one more important than the other when it comes to improving body composition or preventing obesity from an early age? In 2003, Labib published the 'take-home message' that "Obesity is increasing at alarming rates because of [1] a reduction in daily energy expenditure,... and [2] an increase in energy intake" [192]. While some researchers place more (or indeed all) significance on energy intake [193, 194], there are also those who place the emphasis on physical activity [195, 196, 197, 198]. A cross-sectional study published in 2000 by Atkin and Davis concluded that dietary intake did not have an effect upon body composition in childhood, but "energy expenditure, in particular physical activity, may have a greater influence on body composition in early childhood" [195].

Literature examining the link between total energy expenditure and body composition appears to be contradictory and often inconclusive. In 1998, Ravussin et al [199] concluded that "a low rate of energy expenditure may contribute to the aggregation of obesity in families", while other researchers such as Goran et al in 1998 [142] have suggested that there was in fact no significant relationship between TEE and body composition. Goran et al reported that "the rate of change in fat mass relative to fat-free mass is highly variable and is related to sex, initial fatness, and parental fatness, but is not inversely related to any of the components of daily energy expenditure". This longitudinal study involved 75 white children, and the resting energy expenditure was measured by indirect calorimetry. Several studies have shown a clear link between energy intake and body composition, but the degree of the imbalance varies from researcher to researcher.

1.2.5 Maturation: determining pubertal status

In any field of research involving young people, it is highly important to be able to consider the effects of puberty (also referred to as sexual maturation (SM)). Naturally, in order to do this, researchers must be able to determine pubertal staging. Puberty is a process that cannot be defined as clearly as, for example, age [200, 201]. In a 2006 review of assessment of pubertal status, Dorn et al state that the measurement method should be appropriate for the particular study, concluding that "in other words, we do not believe there is a one-size-fits-all answer to the frequently posed question, "What is the best way to measure puberty?" Instead, what is often required is a clarification of the research questions being addressed by a study and the subsequent components of maturation that are of greatest relevance and interest" [201].

A method of determining pubertal staging which has been widely used for more than 40 years is the Tanner method [202, 203]. This method involves comparing physical characteristics of adolescents to published photographs or line drawings depicting various stages of puberty, and thus assigning the adolescents to a pubertal stage. The Tanner stages range from 1 to 5 (prepubertal to postpubertal) and detail the growth of pubic hair in both boys and girls, genital development in boys and breast development in girls. Determining pubertal staging by this method can be done by physical examination, self report or parental report. These methods naturally differ in reliability. Various studies have shown that examination by health professionals is vital in assessing pubertal stage while self report is far less accurate [204, 205, 206].

Bonat et al, in 2002, determined that self-report of puberty was especially inappropriate for overweight adolescents [207]. While physical examination of such features may be considered embarrassing for young people, it has been reported that the key to obtaining participation in such examinations relies upon "how comfortable the investigator is in explaining the research study and the included physical exam, as well as how experienced and comfortable the health care provider is in actually performing the exam. If one or both of these aspects is missing, then the study may fail to obtain pubertal staging by physical examination" [201]. A 2000 study reported that self-report and parental report of pubertal staging are sufficient "in instances in which precise assessment of pubertal stage may not be necessary" [208].

Some researchers use different scales which were modelled on Tanner ratings. In a longitudinal study, Buckler [209] used both the Tanner scale and a scale ranging from 1 - 10. In this scale, the Tanner stages were multiplied by 2 so that, for example, Tanner stage 2 corresponded to new stage 4. However, the scale adopted by Buckler allows for the inclusion of intermediate stages which are not considered when using the Tanner system for determining pubertal staging.

Age of menarche is a potential method of determining stage of puberty in girls. It must, however, be acknowledged that this event does not occur at the start of the pubertal process and so cannot be considered representative of the onset of puberty [204]. When considering age at menarche, differing criteria can present problems for researchers (as outlined by Brooks-Gunn,

Peterson and Eichorn [210]). While one study may consider the cut-off for late maturation to be age of menarche being 14, another may consider the same cut-off to be 13.

In some longitudinal studies involving puberty, peak height velocity (PHV) is used as a marker of pubertal status. PHV for an individual adolescent can be determined from a height velocity curve. Buckler [209] found that in boys, PHV was reached at an average of 1.7 years after the first sign of pubertal onset while this figure was lower - at 1 year - in girls. It appears to be generally accepted in literature that the average difference between the age at the onset of puberty and the age of PHV is 2 years [211, 212].

Whatever method is used, it is important that precise detail is given. In the 2006 review Dorn et al discussed the importance of stating the method used: "in 79 articles it stated "pubertal development was assessed according to Tanner" but no details were provided as to whether this Tanner staging was determined by physical exam or by self-report." [201].

1.2.6 Summary

Literature on the subject of the effect of each component of the energy-imbalance is, to date, conflicting and inconclusive. Until either proven conclusively or proven wrong, we should be aware of both energy intake and physical activity in the fight against obesity!

One thing that is generally agreed in literature is that while healthy adults should have an exact energy balance in order to maintain weight, this is not in fact the case for children. The Institute of Medicine report that "growing children, even those at a healthy body weight, must be in a slightly positive energy-balance to satisfy the additional energy needs of tissue deposition for normal growth" [67]. What does not seem to be known, however, are reasonable limits within which this positive balance can be considered healthy, or how and when the positive balance of childhood should decrease to the ideal zero balance of adulthood.

It has become clear throughout the literature that the Tanner stages are considered to be extremely important in research over adolescence. While many researchers use other methods of determining pubertal staging, they are often seen being used in conjunction with the Tanner diagrams. The method appears to be highly reliable when carried out with a physical exam by a trained examiner, but care should be taken regarding self-report of pubertal staging by the Tanner method.

Two crucial questions are now:

- What are the age effects that modify the relationships between energy-imbalance and body composition?
- Is puberty a time when body composition is particularly sensitive to energy-imbalance?

1.3 Scope of research

When we set out on this piece of research in late 2006, we hoped to examine the energy-balance equation (energy-intake = energy expenditure), breaking it down into individual components (energy intake (EI), resting energy expenditure (REE), diet-induced thermogenesis (DIT) and physical activity energy expenditure (PAEE)) and modelling each of these components, and energy-balance itself, over the course of puberty.

However, it very quickly became apparent that the building blocks that would form the foundations of such research were, at best, unstable.

Take, for example, REE. We identified seven different models for estimating REE from anthropometric measurements. Using longitudinal data provided to us by ALSPAC (see section 3.2 for details), we compared these models using a representative sample of British children - and found no two in agreement. Had we gold standard REE data for these adolescents (i.e. by calorimetry, see section 1.2.2.2.1), it would have been reasonable to attempt to either validate existing models or develop our own. However, without such data, the best we could do was to show that, since no two models are in agreement, at most one of the seven is correct for the UK population of young people.

After a preliminary review of the models and data available to us, we decided on the following, revised scope of research.

- Catalogue the methods and statistical models used to quantify body composition, energy-balance and puberty, critiquing each component in terms of its accuracy and practical application on both the population and individual level.
- Utilising ALSPAC, a sample of around 14,000 British children considered to be representative of the wider population of British children, approximate the prevalence of overweight and obesity in today's society and contrast this with expected levels.
- Obtain provisional results describing the complex relationship between energy-balance and physical development during puberty and indicate what further research would be required to set this modelling on a firmer footing.

1.4 Thesis Overview

Chapter 2 is a review and, where necessary, critique of the statistical methods used throughout this project.

Chapter 3 introduces each dataset that was used in the research, starting with the Avon Longitudinal Study of Parents and Children (ALSPAC) - the main data source - and moving on to describe each data source in turn. It then concerns itself with ALSPAC and describes the subjects in this study at each wave of the study using one method of body composition assessment (Body Mass Index (BMI) standard deviation scores). Perhaps somewhat unsurprisingly, this chapter concludes that higher proportions of children in this study are overweight or obese (by this particular measurement method) than would be desirable by established standards.

Chapter 4 uses the available data to compare two widely used methods of body composition analysis - dual-energy x-ray absorptiometry (DEXA) and bioelectrical impedance analysis (BIA). This chapter then goes on to look at the possibility of developing new models for estimating body composition from BIA. Results of this modelling are presented but should be treated with extreme caution due to the lack of available data upon which to develop and test models.

Chapter 5 reviews the many models in the literature for estimating resting energy expenditure (REE) from anthropometric measurements. It opens

with an attempt to reproduce the models that were published in 1919 by Harris and Benedict [86] - which are to this day used in medical practice. This chapter then uses the ALSPAC data to estimate REE for several thousand children using seven published models - and finds a disturbing lack of agreement among models.

Chapter 6 begins by considering the ideal study in the field of body composition and energy-balance research. Longitudinal resting energy expenditure and body composition data are simulated, with a view to showing, in the absence of real data, how issues such as missing data may affect analysis.

Chapter 7 considers energy-balance as a whole and its difficulties when applied to youth, with an aim to beginning to determine roughly how much positive energy-imbalance is necessary for healthy growth. Using estimates of fat and fat-free mass gain over time, excess daily energy intake is presented for the ALSPAC subjects. Results indicate that young subjects do not require a high imbalance for healthy growth.

Chapter 8 introduces and critiques measures of pubertal status within the ALSPAC data. The consistency of self-reported data is examined and shown to be, at best, extremely tentative. Following this, potential associations between pubertal status and body composition are examined using BMI standard deviation scores and an index of fat mass (FM). When considering the results and conclusions of this chapter, we must be cautious as a result of the limitations due to the self-reported nature of the data.

Chapter 9 summarises the research as a whole, highlights the range of limitations encountered, and identifies potential areas of future modelling on this topic. This chapter finishes with a discussion of the 'ideal study' into modelling energy-imbalance over puberty.

Chapter 2

Review of statistical methods

2.1 Chapter aims

Having reviewed the clinical issues in the field of energy-imbalance in Chapter One, this chapter will introduce and discuss the statistical methods and models that are used throughout this project.

2.2 BMI for UK Children: LMS and the 1990 UK growth reference data

BMI (see equation (1.2)) as a proxy for body composition estimation in children is used throughout this project in conjunction with age- and sex-specific percentiles, standardised using the LMS method [30, 213]. Previously, smooth centile curves had been fitted to data, but with those came the problem of whether a bump or a dip was a real feature of the data or sampling error [213]. The LMS method deals with this by summarising the distribution of BMI at each age by three uncorrelated curves: median (M), coefficient of variation (S), plus a measure of skewness based on a power transformation to achieve normality (L). At any given age, this method is similar to the Box-Cox method for achieving the assumptions of linear modelling [214].

To apply the LMS method to UK data, in order to provide UK reference curves for BMI in childhood, data from eleven studies were combined [30]. These data covered the time frame 1978 to 1990 and included male and female participants from England, Scotland and Wales, aged between 0 and 23 years. Summary centiles were fitted to the data using the LMS method with parameters estimated by penalised likelihood in order to smooth the curves without the need to arbitrarily group covariates [213].

Published from this was the UK 1990 reference, comprising values of L, M and S for males and females at each age (in months) between 0 and 23 years of age. These values can then be used to compute, for any UK child, a BMI centile relative to the 1990 reference data as shown in equation (2.1), where z_α is the normal equivalent deviate for tail area α.

$$C_{100\alpha}(t) = M(t)\left[1 + L(t)S(t)z_\alpha\right]^{\frac{1}{L(t)}} \tag{2.1}$$

This can then be rearranged to give the form of the BMI Standard Deviation Scores (BMI SDS) as:

$$SDS = \frac{\left[\frac{BMI}{M(t)}\right]^{L(t)} - 1}{L(t)S(t)} \tag{2.2}$$

where L(t), M(t) and S(t) are the values of L, M and S suitable for the child's age and sex (from 1990 growth reference data [30]).

These SD scores can then be classified as:

Percentile	SDS	Category
$<$ 5th percentile	$\leq$ -1.64	Underweight
85th to $<$ 95th percentile	1.04 to $<$1.64	Overweight ("pre-obese")
$\geq$ 95th percentile	$\geq$ 1.64	Obese

Table 2.1: Criteria for overweight and obesity from gender-specific BMI-for-age standard deviation scores for children and adolescents

One important thing to remember is that the reference data use the 'what is' scenario from 1990, but we must keep in mind that 'what is' is unlikely to be 'what is ideal'. This method of standardising BMI is widely used and

adopted because it is relatively simple and intuitive, but it can be critiqued from a statistical point of view. In 1992, Cole and Green reported that "A key assumption of the LMS method is that after a suitable transformation, the data are normally distributed... the main problem with the assumption may be the presence of kurtosis, which the transformation does not adjust for, but kurtosis tends to be less important than skewness as a contributor to non-normality" [213]. Another possible criticism of the 1990 reference data is that it is based on the population as it was 22 years ago, which is unlikely to accurately represent the current UK population, far less the ideal population.

This method can be extended to find a standard deviation score for the change between two BMI SDS for an individual (BMI_i and BMI_j) using equation 2.3 [31]:

$$SDS(change) = \frac{\Delta SDS - mean(\Delta SDS)}{SD(\Delta SDS)} = \frac{\Delta SDS}{\sqrt{2(1-r)}} \qquad (2.3)$$

where $\Delta SDS = SDS_j - SDS_i$, j>i

and r = correlation between SDS_i and SDS_j for population of interest.

2.3 Linear regression

The linear regression model assumes a linear relationship between the dependent variable (y_i) and the vector of independent variables (x_i). The model takes the form:

$$y_i = x_i^T \beta + \epsilon_i, i = 1, ..., n \tag{2.4}$$

where ϵ_i is an unobserved random variable that adds noise to the linear relationship between the dependent and independent variables. The standard assumptions of the linear model require the e_i's to have zero mean and constant variance, σ^2, and to be independent. When carrying out inferences on parameters, it is often further assumed that each $e_i \sim N(0, \sigma^2)$.

In matrix notation:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon_i \tag{2.5}$$

where the standard assumption is that $\mathbf{X}$ is an $n \times p$ matrix of full rank p $(n > p)$.

Ordinary least squares (OLS) estimates β, the vector of coefficients, by miminising the sum of squared residuals. Let $\mathbf{b}$ be a candidate for β. Then the sum of squared residuals is given by

$$S(b) = \sum (y_i - x_i^T b)^2 = (\mathbf{y} - \mathbf{Xb})^T (\mathbf{y} - \mathbf{Xb}) \tag{2.6}$$

The minimum may be found by differentiating equation (2.6) with respect to $\mathbf{b}$ and setting equal to 0 [215].

With the standard assumptions, $\mathbf{X}^T\mathbf{X}$ is invertible since it is a p x p matrix with the same rank, p, as $\mathbf{X}$ itself. Therefore, the least squares estimate for $\mathbf{b}$ is given by [216]:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{2.7}$$

In this project, linear regression will be carried out in several contexts, using the `R` command `lm`, part of the `stats` package [6].

2.4 Nonlinear least-squares estimation

As a generalisation of linear regression, consider the data $\{y_i, x_{i1}, ..., x_{ip}\}_{i=1}^{n}$ and a nonlinear function $f(\cdot)$ relating the dependent variable to the independent variables [217, 218]:

$$y_i = f(\boldsymbol{\beta}, \mathbf{x_i'}) + \epsilon_i \tag{2.8}$$

where $\boldsymbol{\beta}$ is the vector of parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)$ $(n \geq p)$.

As before, we aim to estimate β by minimising the sum of squared residuals, where the sum of squared residuals [217] is

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} [y_i - f(\boldsymbol{\beta}, \mathbf{x_i'})]^2 \tag{2.9}$$

The minimum value of $S(\boldsymbol{\beta})$ occurs when the gradient, $\dfrac{\partial S(\boldsymbol{\beta})}{\partial \beta}$, is $\mathbf{0}$ [219]. Since there are p parameters, there are p gradient equations:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \beta} = -2 \sum [y_i - f(\boldsymbol{\beta}, \mathbf{x_i'})] \frac{\partial f((\boldsymbol{\beta}, \mathbf{x_i'})}{\partial \beta} \tag{2.10}$$

These gradient equations are functions of both the independent variables and the parameters, and as such, have no closed solution. Instead, initial values must be chosen for the parameters [217]. Then, the parameters are refined iteratively:

$$\beta_j \approx \beta_j^{k+1} = \beta_j^k + \Delta \beta_j \tag{2.11}$$

where k is an iteration number, and $\Delta\beta_j$ is known as the shift factor. A solution is obtained when β_j converges to some value (i.e. $\Delta\beta_j$ becomes smaller than some prescribed threshold) [220].

(Note that the procedure may not converge very well for some functions and that convergence is often greatly improved by choosing initial values close to the best-fit values [220]).

The `R` command `nls`, part of the `stats` package, performs nonlinear least-squares regression estimation using an iterative procedure to estimate $\boldsymbol{\beta}$ [6].

2.5 Linear mixed models

The assumption with standard linear models of independent error terms does not always hold. This may be because individuals are related, or measurements are taken repeatedly on the same individuals. Linear mixed models do not assume independence of errors and are therefore more suited to data with such data. These models allow variation between people in the intercept and/or slope(s) of a model.

The parameters in a linear mixed model may be classified into two types:

- fixed effects, associated with the average effect of the independent variable(s) on the dependent variable,
- variance-covariance components associated with the covariance structure of the random effects and the error term.

A general linear mixed model may be expressed as [221, 222, 223]

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \tag{2.12}$$

$$\mathbf{b}_i \sim \mathbf{N}_q(\mathbf{0}, \boldsymbol{\psi}) \tag{2.13}$$

$$\boldsymbol{\epsilon}_i \sim \mathbf{N}_{n_i}(\mathbf{0}, \boldsymbol{\sigma}^{\mathbf{2}}\boldsymbol{\Delta}_{\mathbf{i}}) \tag{2.14}$$

where

- $\mathbf{y}_i$ is the $n_i \times 1$ dependent variable vector for observations in the ith group.

- $\mathbf{X}_i$ is the $n_i \times p$ model matrix for the fixed effects for observations in group i.
- $\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed-effect coefficients.
- $\mathbf{Z}_i$ is the $n_i \times q$ model matrix for the random effects for observations in group i.
- $\mathbf{b}_i$ is the $q \times 1$ vector of random-effect coefficients for group i.
- ϵ_i is the $n_i \times 1$ vector of errors for observations in group i.
- $\boldsymbol{\psi}$ is the $q \times q$ covariance matrix for the random effects.
- $\boldsymbol{\sigma^2 \Delta_i}$ is the $n_i \times n_i$ covariance matrix for the errors in group i.

In this project, mixed models will be used to estimate parameters from the longitudinal ALSPAC data. This will be carried out using the `R` command `lmer`, from the `lme4` package [224]. These parameters will then be used to simulate longitudinal resting energy expenditure and body composition data.

2.6 Assessing agreement

2.6.1 Intraclass correlation coefficient

The intraclass correlation coefficient (ICC) assesses rating reliability or agreement by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects [225].

In this thesis, ICC will be used to assess agreement between pairs of linear prediction models for REE, using a two-way mixed model. Under this model, the subject effect is random and the 'rater' (model, here) effect is fixed. Note that inferences from a two-way mixed model are confined to the particular set of 'raters' used.

The two-way mixed effects intraclass correlation coefficient is calculated as [226]:

$$ICC = \frac{\sigma^2_{subj}}{\sigma^2_{subj} + \sigma^2_{model} + \sigma^2_{Err}} \tag{2.15}$$

where:

σ^2_{Err} is mean square (error) from analysis of variance (ANOVA)

σ^2_{subj} is $\dfrac{\text{mean square (subject)} - \text{mean square (error)}}{k}$

k is the number of models being compared

σ^2_{model} is $\dfrac{\text{mean square (model)} - \text{mean square (error)}}{n}$

n is the number of cases

The ICC can range from 0 to 1, with an ICC of exactly 1 representing perfect agreement between the models, while values close to 0 represent poor agreement. This measure of agreement gives a population measure, it does not allow us to consider agreement between models on an individual basis.

2.6.2 Bland-Altman diagrams

Bland-Altman diagrams [102, 227], also known as Tukey mean-difference plots, are used to assess agreement between two methods or models estimating the same thing.

To produce a Bland-Altman plot, the mean of the two methods is plotted against the difference between the two methods. Then the co-ordinates of a given observation C with measurements C_1 and C_2 for methods 1 and 2, respectively, are:

$$C(x, y) = \left(\frac{C_1 + C_2}{2}, (C_1 - C_2)\right) \tag{2.16}$$

95% limits of agreement, which are "only estimates of the values which apply to the whole population" [102] are calculated as:

$$\text{mean difference} \pm 1.96 \times \text{sd difference} \tag{2.17}$$

To show reasonable agreement, the points should be evenly spread between the limits of agreement across the range of the mean differences. Other things being equal, narrower limits on this plot show two models that are in closer agreement than wider limits.

2.7 Concordance analysis: Kruskal's gamma and Kendall's tau

Kruskal's gamma [228] and Kendall's tau [229] are measures of association for ordinal data. Both can vary from -1 to $+1$ and give an indication of the strength and direction of association.

When one case has a higher value than another case on both variables, the cases are said to form a "concordant" pair. When one case has a higher value than another on one variable, but a lower value on another, the cases are said to form a "discordant" pair. When the cases are equal on both variables, they form a "tied" pair.

Gamma is calculated as:

$$\gamma = \frac{N_c - N_d}{N_c + N_d} \tag{2.18}$$

Where N_c is the number of concordant pairs and N_d is the number of discordant pairs [230].

As can be seen from equation (2.18) above, Kruskal's gamma ignores all tied pairs. The reason for this is that concordance tells us about positive association, discordance about negative association, but anything else sends mixed messages or is uninformative. However, when data contain a lot of ties, this should be taken into consideration. An alternative to Kruskal's gamma, which takes ties into consideration, is Kendall's tau.

Tau-b, used for square tables, is calculated as:

$$\tau_b = \frac{N_c - N_d}{\sqrt{(N_c + N_d + T_x)(N_c + N_d + T_y)}} \tag{2.19}$$

where, N_c is the number of concordant pairs, N_d is the number of discordant pairs, T_x is the number of pairs tied on the independent variable, X and T_y is the number of pairs tied on the dependent variable, Y. [230]

Tau-c, used for non-square tables, is:

$$\tau_c = \frac{N_c - N_d}{0.5N^2 \left[(m-1)/m\right]} \tag{2.20}$$

where N is the total number of cases and m is the minimum value of the number of rows or the number of columns [230].

As mentioned above, Kruskal's gamma may overestimate the strength of the association. In a personal communication with Charles Blake of James Madison University [231] in October 2010, it was recommended that if |gamma|-|tau|>0.05, tau should be used, otherwise gamma is preferred. We have been unable to find any other criteria for choosing between the two measures.

2.8 Kaplan-Meier analysis

Kaplan-Meier analysis is used when data are 'time-to-event' (also known as 'survival' data). An individual is considered to have 'survived' to at least time t if he or she has not had the event of interest by that time. In survival data, the time of the event may not be recorded for some individuals. For those cases, the time of event is the time of their last record, and those observations are 'censored'. In this project, Kaplan-Meier will be used to examine the age at menarche. Therefore, the 'event' of interest will be reaching menarche. Data may be censored if a child drops out of the ALSPAC study before having reported an age of reaching menarche, or if the menarche age has not been reported by the age 14 Growing and Changing questionnaires.

Kaplan-Meier analysis, also known as the product-limit method [232], is a nonparametric method of estimating survival functions, used when there is the presence of censored data.

The survival function (equation (2.21), below) (also known as the reliability function) describes the proportion of individuals surviving beyond a given time [233, 232].

$$S(t) = P(T > t) \tag{2.21}$$

where t is a point in time, and T is survival time.

The Kaplan-Meier estimator, used to estimate the survival function, is defined as [233]:

$$\widehat{S(t)} = \begin{cases} 1, & \text{if } t < t_1 \\ \prod_{t_1 \leq t} \left[1 - \frac{d_i}{Y_i}\right], & \text{if } t_1 \leq t \end{cases} \tag{2.22}$$

where:

t_i is the time of the first event,

d_i is the number of events at time t_i,

Y_i is the number of individuals at risk at time t_i = the number of survivors at time t minus the number of censored cases at time t.

The quantity $\frac{d_i}{Y_i}$ estimates the conditional probability that an individual who has not experienced the event of interest just prior to time t_i experiences it at time t_i.

The Kaplan-Meier estimator is a step-down function with drops at the observed event times. The magnitude of these drops depends on both the number of events at time t_i and the pattern of censored observations prior to t_i [233].

The point estimate of the median survival time is given by t where $\widehat{S(t)} = 0.5$.

A fuller description may be found in the 1958 paper *Estimation from incomplete observations* [232].

2.9 Multiple imputation by chained equations

Missing data are a problem that can lead to inefficient analyses or biased results [234]. Multiple imputation (MI), usually restricted to data that are missing completely at random (MCAR) or missing at random (MAR) [235, 236], is one method of handling missing data. In MI, m simulated versions of the dataset are generated ($m \geq 1$) [237]. In each of the m datasets, the values which were non-missing in the original data remain the same from set-to-set, while the values that replace the missing data vary [238]. The number of simulated datasets m is generally low: Schafer reports that "Unless rates of missing information are unusually high, there tends to be little or no practical benefit to using more than five to ten imputations" [237].

This method, rather than creating a single imputed dataset, accounts for the statistical uncertainty in the imputations [239]. One such imputation method is multiple imputation by chained equations (MICE), also known as fully conditional specification (FCS) [240] or sequential regression multiple imputation (SRMI) [239], which involves specifying the imputation model separately for each variable in the data using the other variables as predictors [240].

Let Y_j, (with $j = 1, ..., p$) be one of p incomplete variables, where $Y = (Y_1, ..., Y_p)$. The observed and missing parts of Y_j are denoted by Y_j^{obs} and Y_j^{miss}, respectively. Then the observed and missing data in Y are $(Y_1^{obs}, ..., Y_p^{obs})$and $(Y_1^{miss}, ..., Y_p^{miss})$. The hth imputed data set is denoted as $Y^{(h)}$ where $h = 1, ..., m$.

Let $Y_{-j} = (Y_1, ..., Y_{j-1}, Y_{j+1}, ..., Y_p)$ denote the collection of variables in Y excluding Y_j. Let the hypothetically complete data Y be a partially observed random sample from the p-variate multivariate distribution $P(Y|\theta)$, and assume that the multivariate distribution of Y is completely specified by the vector of unknown parameters, θ.

MICE obtains the posterior distribution of θ by sampling iteratively from the conditional distributions of the form

$$\begin{aligned} &P(Y_1|Y_{-1}, \theta_1) \\ &\vdots \\ &P(Y_p|Y_{-p}, \theta_p) \end{aligned} \tag{2.23}$$

Starting from a simple draw from the observed marginal distributions, the

tth iteration of chained equations is a Gibbs sampler that successively draws

$$\begin{aligned}
\theta_1{*}^{(t)} &\sim P(\theta_1|Y_1^{obs}, Y_2^{(t-1)}, ..., Y_p^{(t-1)}) \\
Y_1{*}^{(t)} &\sim P(Y_1|Y_1^{obs}, Y_2^{(t-1)}, ..., Y_p^{(t-1)}, \theta_1{*}^{(t)}) \\
&\vdots \\
\theta_p{*}^{(t)} &\sim P(\theta_p|Y_p^{obs}, Y_1^{(t)}, ..., Y_{p-1}^{(t)}) \\
Y_p{*}^{(t)} &\sim P(Y_p|Y_p^{obs}, Y_1^{(t)}, ..., Y_{p-1}^{(t)}, \theta_p{*}^{(t)})
\end{aligned} \tag{2.24}$$

where $Y_j^{(t)} = (Y_j^{obs}, Y_j{*}^{(t)})$ is the jth imputed variable at iteration t.

This procedure repeats until convergence, with the observed data and the final set of imputed values constituting one complete dataset [239].

Analysis using standard estimating techniques is then carried out on each of the m complete datasets, and results are pooled using pooling rules known as "Rubin's rules" [236, 238].

Multiple imputation can lead to nonsensical imputed data. Specifying predictive mean matching (PMM) as the imputation model may reduce this risk. With PMM, an imputed observation takes on the value of one of a set of observed values for cases similar in terms of other variables [234].

MICE procedures are very flexible [238], can be used in a broad range of settings and, unless the fraction of missing information is unusually large, can

lead to efficient inferences even when m is small [237]. However, care must be taken when using these procedures. If there is not much information in the observed data regarding the missing values, imputations will be highly variable and likely to lead to errors in the analyses [239].

MICE is based on the assumption that the data are missing completely at random (MCAR) or missing at random (MAR), and biased estimates may be obtained if this assumption does not hold [239, 238]. In his 1999 paper, Schafer reports that MI is not necessarily the best imputation method for some problems, with alternatives such as maximum likelihood estimation being perhaps more efficient due to not involving simulation [237].

Multiple imputation will be used in the simulation section of this project, when explore the effect of missing and imputed data on analysis. This will be carried out using the `R` package `mice` [240] with predictive mean matching (PMM).

2.10 Chapter summary

This chapter has described the following methods and models that will be used throughout this project:

- LMS and the 1990 growth reference data
- Linear regression
- Nonlinear least squares estimation
- Linear mixed models
- Intraclass correlation
- Bland-Altman diagrams
- Concordance analysis in ordinal tables
- Kaplan-Meier analysis
- Multiple imputation using chained equations

Chapter 3

Datasets and investigation of body mass index in the Avon Longitudinal Survey of Parents and Children

3.1 Chapter aims

This chapter will begin by introducing the datasets that will be used throughout the rest of this project and will then explore the distribution of BMI (after accounting for age and sex) of children and adolescents in the UK using one of these datasets, considered to be representative of this population. Analysis will progress to investigate whether or not any increase seen in the weight status is a result of smaller children 'catching up' with their counterparts and whether or not 'fatter' children are likely to remain overweight or obese as they progress into the early years of their adolescent lives.

3.2 Datasets

All datasets were originally collected as part of research studies for which ethical approval was granted by the appropriate ethics committees and have been released to us for secondary use.

3.2.1 ALSPAC - main data sets

The Avon Longitudinal Study of Parents and Children (ALSPAC) [87], based in Bristol, UK, provided the main dataset for this PhD project. ALSPAC, also known as Children of the 90s, recruited more than 14,000 pregnant women in the Avon area with estimated delivery dates between early 1991 and late 1992. The women and children have been followed up since enrolment in the study, with detailed data having been collected at several points through childhood and into adolescence.

Upon commencement of this project, anthropometric data from focus groups was provided to us in the form of SPSS data files. Subject numbers and ages are as follows in Table 3.1:

Group	Sex	n	Age (years) min	median	max
Focus at 7	Male	3985	6.83	7.42	9.08
	Female	3849	7.00	7.42	9.17
Focus at 9	Male	3629	8.75	9.83	11.67
	Female	3685	8.83	9.83	11.67
Focus at 10	Male	3560	9.83	10.58	12.25
	Female	3608	9.83	10.58	12.25
Focus at 11	Male	3966	10.67	11.75	13.5
	Female	3399	10.42	11.75	13.58
Age 13 clinic	Male	2878	12.50	13.83	15.08
	Female	2960	12.58	13.83	15.17

Table 3.1: Subject numbers in ALSPAC focus data

(Note that the "Age 13 clinic" data were provided separately from the rest, along with the puberty data at the end of the project).

3.2.2 ALSPAC - puberty questionnaires

At waves 8, 9, 10, 11, 13 and 14, respondents (with or without parental help) completed a questionnaire about physical development. These data were provided to us in the late stages of this project. Subject numbers and descriptive statistics for the ages of the subjects at the time of completion are given in Table 3.2.

			Age (years)		
"Wave"	Sex	n	min	median	max
8	Male	2860	7.83	8.08	10.75
	Female	3177	8.00	8.08	11.08
9	Male	3238	9.50	9.58	11.42
	Female	3487	9.50	9.58	11.08
10	Male	3024	10.58	10.67	12.75
	Female	3318	10.58	10.67	14.67
11	Male	2866	11.42	11.67	13.33
	Female	3182	11.58	11.67	15.17
13	Male	2755	12.75	13.08	15.08
	Female	3042	13.00	13.08	14.33
14	Male	2190	14.58	14.67	15.50
	Female	2743	14.58	14.58	16.08

Table 3.2: Subject numbers in ALSPAC puberty questionnaires

3.2.3 ALSPAC - subsamples

At various points during the ALSPAC study, subsamples of 'specialised' data were collected. We have made use of the following subsamples:

			Age (years)		
Subsample	Sex	n	min	median	max
Isotopes	Male	84	11.58	11.92	12.83
	Female	92	11.50	11.83	12.75
Calorimetry	Male	660	11.90	12.40	13.12
	Female	816	11.90	12.38	13.08

Table 3.3: Subject numbers in ALSPAC subsamples

3.2.4 NHANES

The National Health and Nutrition Examination Survey (NHANES) is a nation-wide USA study run by the Center for Disease Control (CDC). Data are collected annually and provided publicly in two-yearly files on the CDC website [241]. Each

"strand" of NHANES is cross-sectional and considered to be representative of the American population. Subject numbers and ages can be seen in Table 3.4 (though it should be noted that we made use of only age 18 years and below).

			Age (years)		
Year	Sex	n	min	median	max
1999 - 2000	Male	4505	0.00	18.33	84.83
	Female	4649	0.00	19.83	84.92
2001 - 2002	Male	5232	0.00	18.67	46.33
	Female	5571	0.00	19.59	44.83
2003 - 2004	Male	4876	0.00	19.08	84.92
	Female	5023	0.00	19.83	84.92
2005 - 2006	Male	5015	0.00	18.58	84.92
	Female	5163	0.00	19.17	84.92

Table 3.4: Subject numbers in NHANES data

3.2.5 Body Composition Unit

The dataset referred to throughout this thesis as "BCU data" was provided courtesy of Professor Dympna Gallagher of the Body Composition Unit, New York Obesity Research Center, University of Colombia [242]. This dataset contained data for subjects of varying races. We are unable to assume that any modelling of body composition applies to all races. Therefore, most modelling undertaken with the BCU data has been carried out on a subset consisting of only those subjects whose race is recorded as Caucasian. Table 3.5 contains descriptive statistics for these subjects.

Data	Sex	n	Age (years) min	median	max
All races	Male	684	5.17	12.00	20.17
	Female	612	4.25	11.52	20.75
Caucasian	Male	172	6.02	11.66	20.17
	Female	145	4.25	11.71	20.56

Table 3.5: Subject numbers in BCU data

3.2.6 Harris and Benedict

In their 1919 book [86], Harris and Benedict published raw data pertaining to the resting energy expenditure of subjects, as measured by a process known as calorimetry (discussed in detail in chapter 2). These data are referred to in this thesis as Harris-Benedict data, or HB data. The subjects in this study were aged as follows:

Sex	n	Age (years) min	median	max
Male	136	16.00	24.00	63.00
Female	103	15.00	25.00	74.00

Table 3.6: Subject numbers in HB data

3.2.7 Institute of Medicine

The Institute of Medicine (IOM) database[1] is a Microsoft Excel 2003 spreadsheet that brings together several studies on energy expenditure. For the purposes of this project, we make use of the worksheets '0-2', '2-8', '9-18' and 'adults', described in Table 3.7

[1] *previously* available from http://www.iom.edu/?id=7302&redirect=0 (accessed October 2007)

Group	Male		Female	
	n	(mean age)	n	(mean age)
0 - 2	173	(0.80)	104	(0.78)
2 - 8	196	(5.53)	348	(6.63)
9 - 18	64	(11.91)	220	(11.80)
adults	276	(44.28)	262	(41.74)

Table 3.7: Subject numbers in IOM database

3.2.8 Fomon reference children

In 1982, Fomon published body composition data for the Reference Child [243], from birth to ten years of age. We have made use of the Total Body Water (TBW) data, i.e. TBW in the 'ideal child'.

3.3 BMI Standard Deviation Scores in ALSPAC

BMI as a proxy for body composition estimation in children should be used in conjunction with age- and sex-specific percentiles, standardised using Cole's 1990 LMS method [30], as described in section 2.2. Using this method, BMI SDS have been calculated for the BMI data at the following waves of ALSPAC: 7, 9, 10, 11, 13 and 14. Note that the standard deviation score for an individual child will be missing if any of age, height, weight or sex is missing for that individual. The number of BMI SDS at each wave are shown in Table 3.8:

Wave	Male	Female
7	3946	3813
9	3593	3640
10	3527	3560
11	3341	3410
13	2037	2106
14	1814	1859

Table 3.8: Number of BMI SDS in ALSPAC data

Percentile plots of these BMI SDS are shown in Figure 3.1.

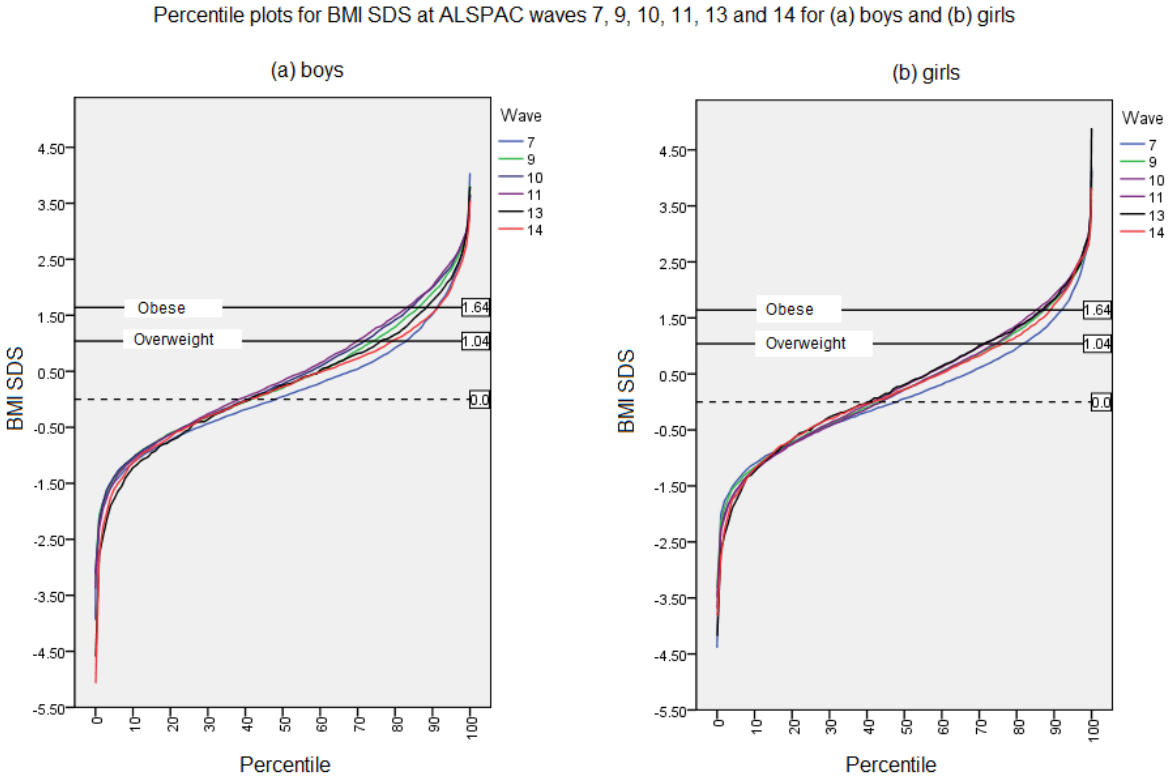

Figure 3.1: Percentile plots of BMI SDS in ALSPAC data for boys and girls at data collection waves 7 to 14

By definition, we would expect the upper ('obese') line to correspond to the 95th percentile and the lower ('overweight') to correspond to the 85th percentile. This is clearly not the case for the ALSPAC respondents. What we are seeing is that the percentile corresponding to overweight is between the 60th and 80th, depending on the wave in question - showing that the percentage of children who are overweight in this sample is very high.

Note that the data for waves 7-11 was provided separately from that for 13 and 14. As a result of this, there was a high degree of missingness between the two sets of data. Taking this into consideration, we replicated Figure 3.1 using only those

subjects for whom we had BMI SDS at EACH wave 7, 8, 9, 10, 11, 13 and 14. 281 males and 258 females met this condition. We found that the previous graph was replicated almost exactly for these 539 subjects, suggesting that the missingness did not have an effect on the overall pattern seen in the figure.

3.3.1 Examining the change in SD Scores over time (waves 7 to 11)

Due to differences in the 13-14 year data and the 7-11 focus files, the rest of this chapter will only consider wave 11 and earlier.

One area of interest in this project is examining the change in body composition over time. To consider this with BMI SDS, we can use a method published by Cole in 1997 [31] which gives a standard deviation score for the change between two BMI SDS for an individual at time i and time j (BMI_i and BMI_j) using equation 2.3 in section 2.2.

Plotting these differences for $i = 7$ and $j = 11$ gives an impression of the overall difference in BMI SD Scores over these 4 years. This is shown in Figure 3.2. What is apparent from these diagrams is the fact that although there are some SD Scores at or below the zero line (indicating no change or a decrease in BMI SDS from wave 7 to wave 11), there are considerably more individuals showing an increase in BMI SDS between these two waves. This is evident for both boys more than girls, although slightly more for boys.

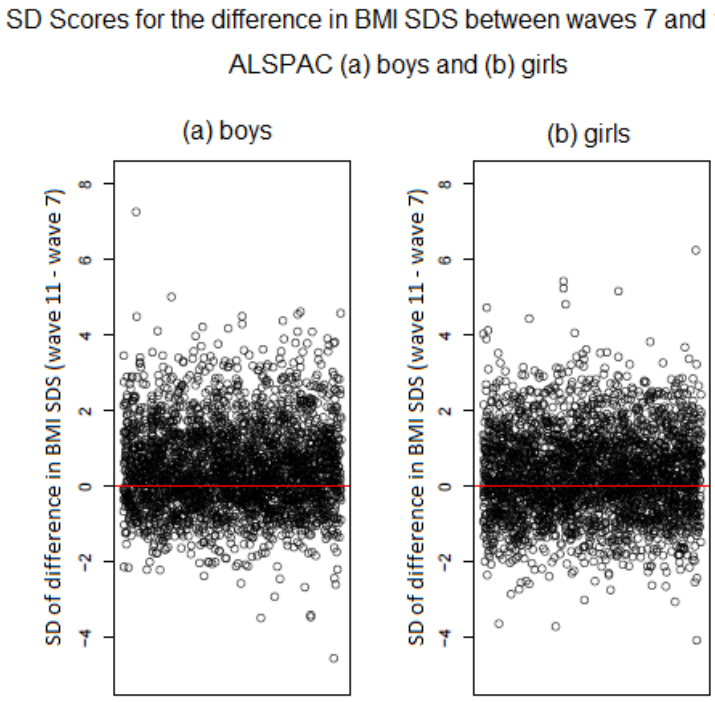

Figure 3.2: Plot of SDS for change in BMI SDS of ALSPAC participants between waves 7 and 11

Repeated measures analysis of BMI SDS over time and between sexes has been carried out using PASW v18 (previously known as SPSS, Statistical Package for Social Sciences). The results of this analysis are graphically displayed in Figure 3.3.

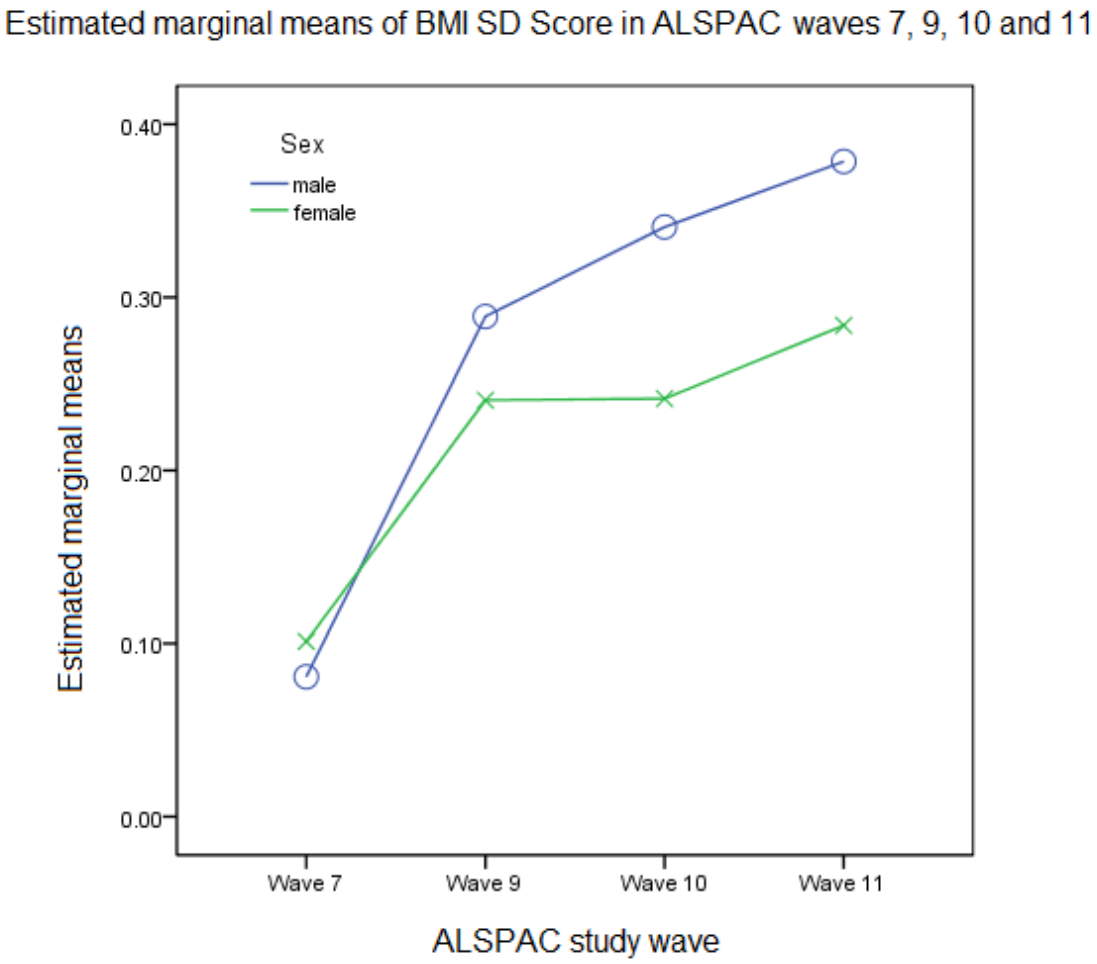

Figure 3.3: Repeated measures comparison of BMI standard deviation scores for boys and girls in ALSPAC waves 7, 9, 10 and 11.

The P-value for the interaction of sex and time was < 0.001. It is clear from Figure 3.3 that the mean BMI SDS is greater than 0 for both boys and girls at all ages. There is an increase over time for both boys and girls however it appears that the trend lines diverge with age due to the increase of excess body mass in boys.

3.3.2 Are children simply playing "catchup"?

It is possible that the increase in BMI SDS from wave 7 to 11 may be caused by children who were underweight at 7 simply "catching up" with their peers by 11. This could cause the SD for their change in BMI SDS between the two waves to be high, but would result in them being a normal weight by 11. This is examined graphically shown in Figure 3.4.

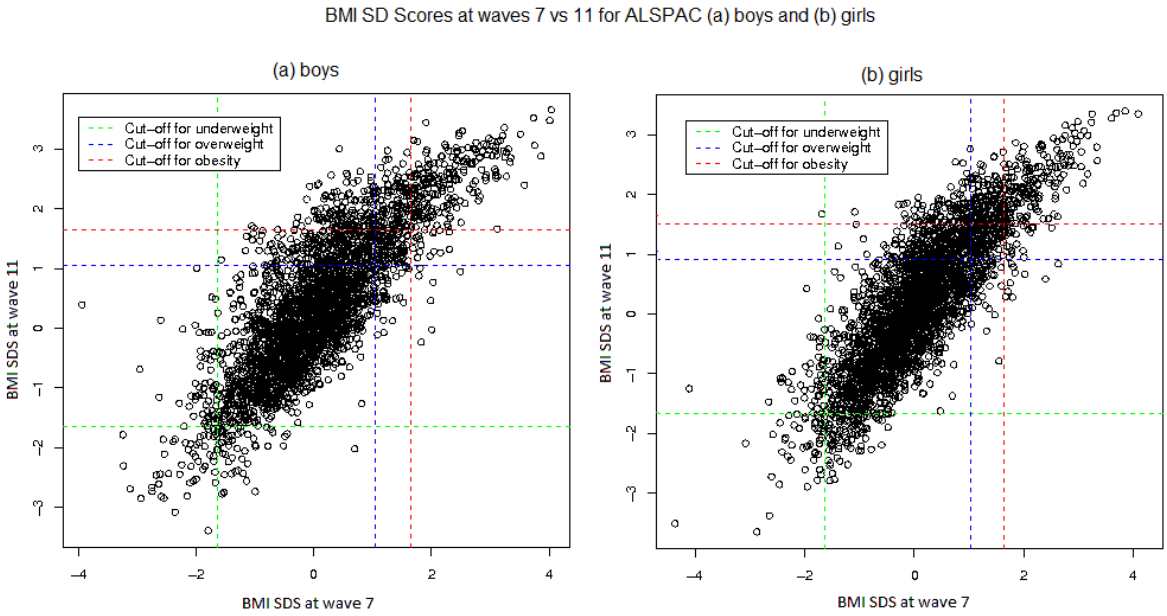

Figure 3.4: Comparison of ALSPAC BMI SDS at waves 7 and 11

If we were seeing a "catchup" effect, it would be reasonable to expect to see some data points in the middle left-hand sections of these graphs. Although there are some data points shown here, it is clear that this is not the case for the majority of the sample. What is particularly striking in Figure 3.4 is the large number of subjects who have been normal weight at wave 7 and have become overweight or obese by wave 11. Similarly, there are a number of points indicating a high rate of transition from overweight at 7 to obesity at 11.

Another idea which comes across quite clearly from these figures is that of being "stuck" in a particular state. For example, many who are overweight (or obese) at 7 are still overweight (or obese) at 11. This can be further shown with stochastic transition matrices for each step in time. These matrices are shown in table 3.3.3 on the next page.

These matrices show the probability of being in any weight classification at a point given the classification at an earlier point. The issue of being "stuck" in a particular state is obvious. For example, considering the boys 7→9, an obese boy at wave 7 has a 92% chance of being obese at wave 9. It therefore follows that if a child is overweight or obese by the age of 7, it is extremely unlikely that he or she will be at a healthy weight approximately four years later. Only 2.8% of those boys who are obese at wave 7 were not overweight at wave 11.

3.3.3 Conclusions

This section has aimed to explore the trends of overweight and obesity in the ALSPAC data. We have shown that the proportions of children who are classified as overweight or obese in the ALSPAC sample are greater than those that we might expect. We can also conclude from this analysis that children who are overweight or obese pre-puberty are at an increased risk of being overweight or obese in early adolescence.

Boys **Girls**

$7 \rightarrow 9$

		9		
		N.O.	*Ov.*	*Ob.*
7	*N.O.*	0.879	0.092	0.029
	Ov.	0.194	0.440	0.306
	Ob.	0.023	0.057	0.920

		9		
		N.O.	*Ov.*	*Ob.*
7	*N.O.*	0.886	0.093	0.021
	Ov.	0.187	0.402	0.441
	Ob.	0.004	0.152	0.884

$7 \rightarrow 10$

		10		
		N.O.	*Ov.*	*Ob.*
7	*N.O.*	0.851	0.110	0.049
	Ov.	0.162	0.412	0.427
	Ob.	0.031	0.067	0.901

		10		
		N.O.	*Ov.*	*Ob.*
7	*N.O.*	0.875	0.101	0.024
	Ov.	0.228	0.341	0.430
	Ob.	0.017	0.138	0.845

$7 \rightarrow 11$

		11		
		N.O.	*Ov.*	*Ob.*
7	*N.O.*	0.827	0.116	0.058
	Ov.	0.163	0.376	0.461
	Ob.	0.028	0.065	0.907

		11		
		N.O.	*Ov.*	*Ob.*
7	*N.O.*	0.846	0.112	0.042
	Ov.	0.212	0.338	0.450
	Ob.	0.045	0.122	0.833

$9 \rightarrow 10$

		10		
		N.O.	*Ov.*	*Ob.*
9	*N.O.*	0.941	0.054	0.005
	Ov.	0.139	0.606	0.255
	Ob.	0.007	0.088	0.905

		10		
		N.O.	*Ov.*	*Ob.*
9	*N.O.*	0.959	0.04	0.010
	Ov.	0.223	0.646	0.131
	Ob.	0.007	0.114	0.878

$9 \rightarrow 11$

		11		
		N.O.	*Ov.*	*Ob.*
9	*N.O.*	0.902	0.084	0.014
	Ov.	0.189	0.488	0.322
	Ob.	0.009	0.115	0.876

		11		
		N.O.	*Ov.*	*Ob.*
9	*N.O.*	0.921	0.069	0.010
	Ov.	0.251	0.478	0.271
	Ob.	0.015	0.167	0.818

$10 \rightarrow 11$

		11		
		N.O.	*Ov.*	*Ob.*
10	*N.O.*	0.936	0.059	0.005
	Ov.	0.195	0.559	0.206
	Ob.	0.006	0.114	0.880

		11		
		N.O.	*Ov.*	*Ob.*
10	*N.O.*	0.935	0.062	0.003
	Ov.	0.196	0.552	0.252
	Ob.	0.005	0.134	0.861

Table 3.9: Transition matrices for weight classifications between ages 7 and 11 in ALSPAC

3.4 Chapter summary

This chapter has shown that, in a sample representative of UK youth, considerably higher proportions of children than expected are overweight or obese by 1990 growth standards - i.e. children today are, on the whole, heavier (considered to be a proxy for fatness) than was considered ideal twenty years ago.

It was found that most children remain 'stuck' in a category of weight over time, indicating that an overweight child is considerably more likely to grow into an overweight adult than one of a healthy weight - a statistical finding that brings with it immense clinical and public health consequences and must not be taken lightly.

Chapter 4

Analysis of body composition models

4.1 Chapter aims

This chapter will be concerned with methods of estimating how an individual's total body mass breaks down into the two components of fat mass (FM) and fat-free mass (FFM), with specific reference to children and adolescents.

Within the datasets used in this project, there are body composition estimates from two methods: dual-xray energy absorptiometry (DEXA) and bioelectrical impedance (BIA). These methods, as discussed earlier in Chapter 2, are non-invasive and are far better tolerated by children than a method known to be more accurate, which involves complete submersion in water (hydrodensitometry).

Initial analysis will compare FM (or FFM) estimates from DEXA and BIA for the same individuals. As a consequence of the nature of our datasets, where the results do not agree, it will not be possible to determine which is most accurate.

The chapter will then be concerned specifically with BIA and will tentatively determine whether or not it is possible, with the data that we have and the statistical methods that are widely used in this field, to develop more accurate models for estimating the body composition of children using this straightforward and non-invasive procedure.

4.2 Body Composition

While BMI SDS for children and adolescents is perhaps more reliable than for adults (see section 1.2.1.2.1), it does have some potential shortcomings. Firstly, standards for overweight and obesity in childhood were established almost twenty years ago and therefore may not be applicable to today's population. Secondly, BMI does not give any approximation of fat mass or fat-free mass in the body. At best, BMI can only be considered to be a rough proxy for body composition assessment. As discussed in section 1.2.1.2.2, two methods of body composition estimation are bioelectrical impedance (BIA) and dual-energy x-ray absorptiometry (DEXA). While these methods are more accessible than the gold standard (hydrostatic weighting), they are known to currently be inaccurate.

4.2.1 DEXA and BIA: a comparison

Using the ALSPAC focus at 11 dataset, we can compare the two methods of body composition assessment. Bland-Altman [102] diagrams shown in Figure 4.1 indicate that BIA estimates of FM are higher than those from DEXA. See section 2.6.2 for a description of the Bland-Altman method. It is important, to keep in mind that neither BIA nor DEXA can be considered to be the gold standard.

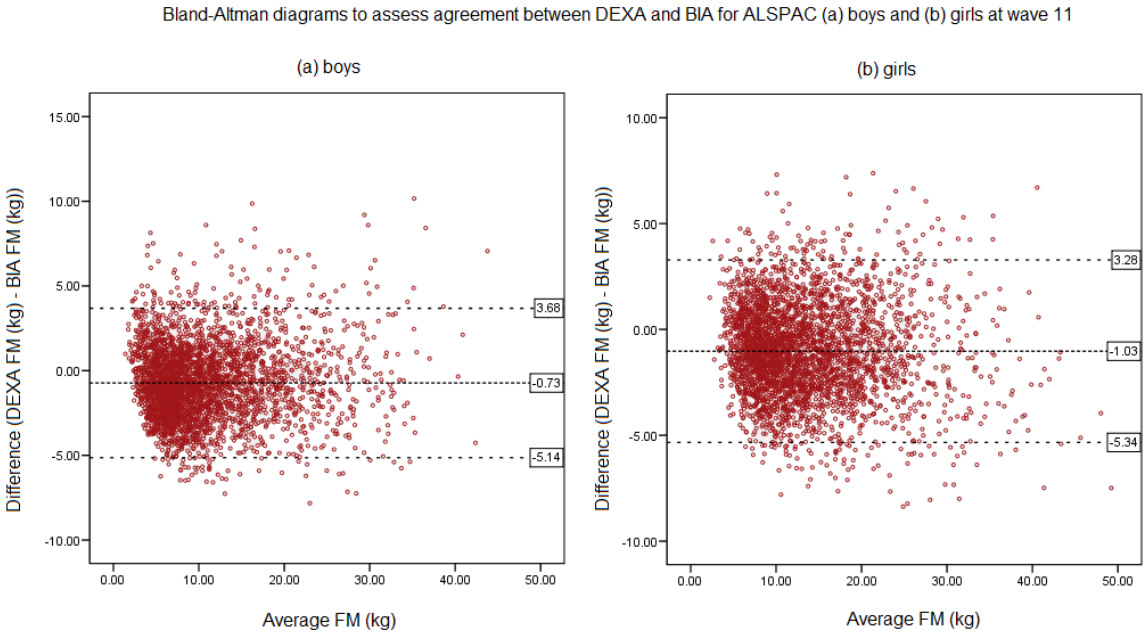

Figure 4.1: Bland-Altman diagrams for BIA and DEXA estimates of FM in the ALSPAC 11-year data

Unfortunately, as we have no gold standard measure of fat or fat-free mass, we are unable to attempt any modelling of DEXA. We can merely conclude at this stage that DEXA and BIA are not in agreement. We do, however, have total body water measured by isotope dilution, which is the gold standard, for a small subsample of ALSPAC participants. This enables us to consider modelling of the resistivity aspect of bioelectrical impedance.

4.2.2 Modelling with impedance

4.2.2.1 Resistivity

The resistivity model used for children and adolescents in this project is shown in equation (1.8) in section 1.2.1.2.2. This section aims to use the data available to us to validate the form of this model and develop it if possible.

We have gold standard TBW measured in two of the available datasets:

- ALSPAC isotope dilution subsample: n=176, age 11.5 years to 12.83 years
- New York Body Composition Unit data: n=1296, age 4.25 to 20.75 years

We begin our resistivity modelling using the BCU dataset since the larger sample size allows for considerably higher power compared to the ALSPAC subsample. Additionally, the BCU data includes race as a factor while the ALSPAC subsample is almost entirely Caucasian. Initially we will use only the Caucasian subjects (n=317) with racial effects being considered at a later stage.

As we begin our analysis, a question of interest is: does resistivity differ between the sexes? From Figure 4.2, it is unclear whether or not any difference between the slopes is statistically significant.

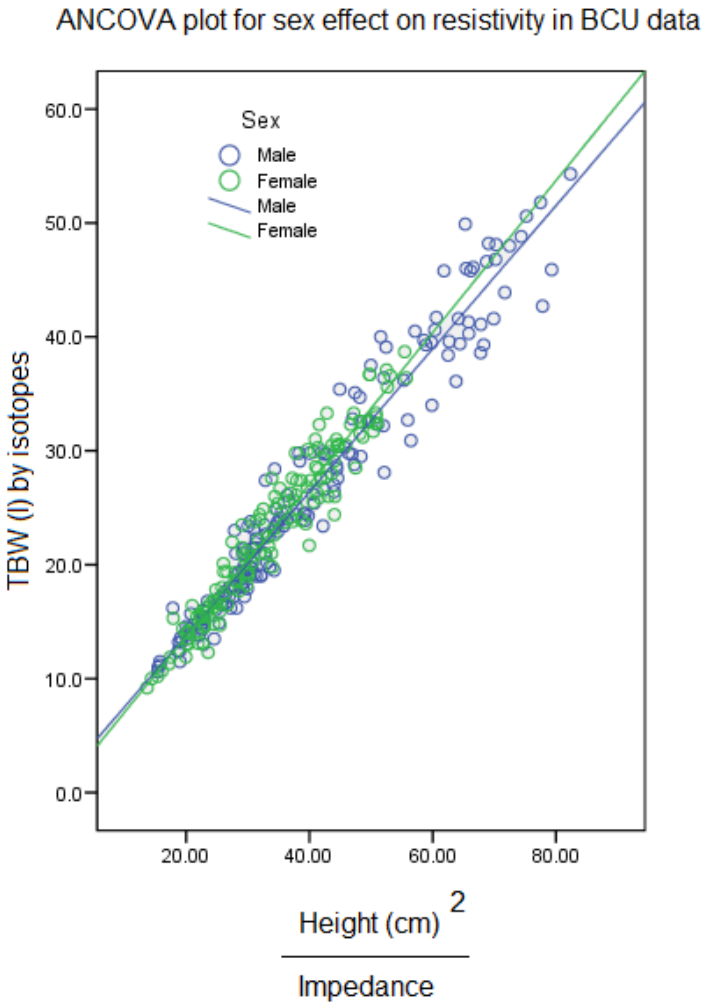

Figure 4.2: Interaction plot for sex effect on resistivity in BCU data

After removal of the non-significant interaction term, the ANCOVA P-value of 0.053 for sex is only marginally nonsignificant at the 5% level. Since inclusion of the sex term in a linear model does not increase the variation explained by that model, it is reasonable to omit the sex term and assume that resistivity does not significantly differ between the sexes.

The linear model obtained, which explains 94.4% of the variability in measured TBW is:

$$\hat{TBW} = 0.634 \times \frac{height^2}{Z} + 1.232 \tag{4.1}$$

The 95% confidence interval for the slope is (0.617, 0.651) and for the intercept is (0.552, 1.913).

The similarities between this model and the previously published model (see section 1.2.1.2.2 are clear. However, diagnostic plots (see Figure A.1 in Appendix A) for the model show clear violation of the assumption of homoscedasticity.

In an attempt to correct for this, we use the following variance-stabilising transformation (log-log model):

$$ln(TBW) = \alpha + \beta \times ln\left(\frac{height^2}{Z}\right) \tag{4.2}$$

ANCOVA for the transformed model shows, with a P-value of 0.066, no need to model the sexes separately.

The transformed model is as follows:

$$\hat{TBW} = 0.758 \times \left(\frac{height^2}{Z}\right)^{0.964} \tag{4.3}$$

Diagnostic plots (see Figure A.2) now show no violation of assumptions.

4.2.2.2 Resistivity and maturation in the BCU data

The BCU data allows us to consider the part that maturation plays in resistivity. There are two potentially useful variables: age and Tanner stage. It should be noted at this point that we have been unable to ascertain how and by whom Tanner stage was determined in the BCU data, therefore we can consider this analysis to be at best a tentative glance at pubertal effects.

Firstly, we plot residuals from the log-log model (equations (4.2) and (4.3)) against Tanner stage. While an individual effect of sex has been ruled out, it is possible that significant interactions exist. For this reason, the plots of residuals are grouped by sex. These plots are shown in Figures 4.3 and 4.4, respectively.

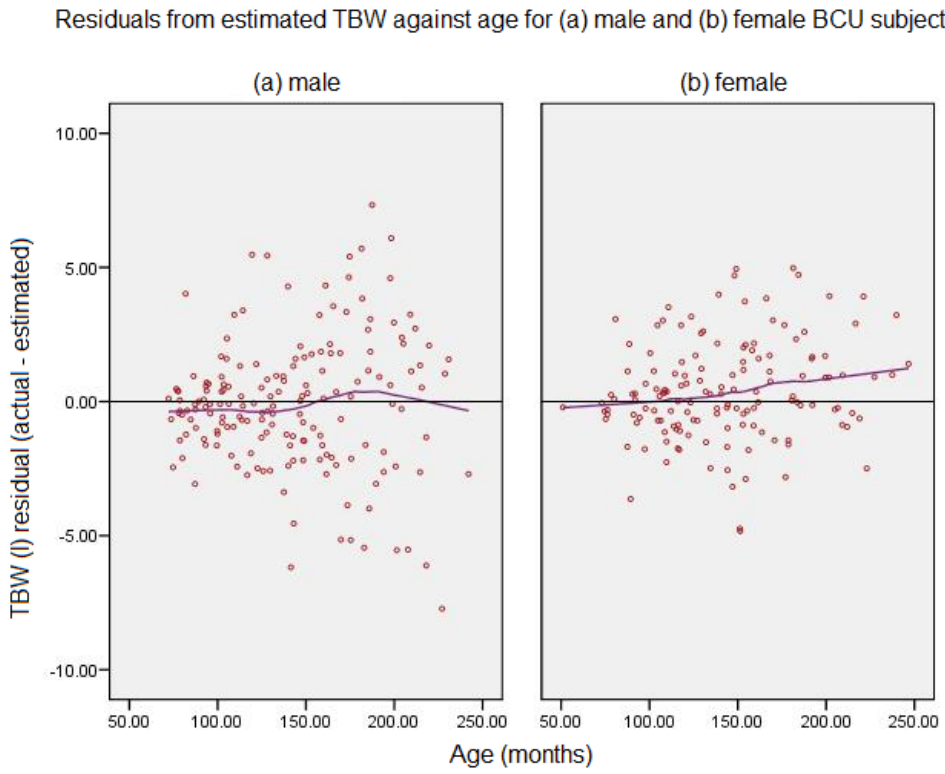

Figure 4.3: Residuals from log-log resistivity model against age of subjects

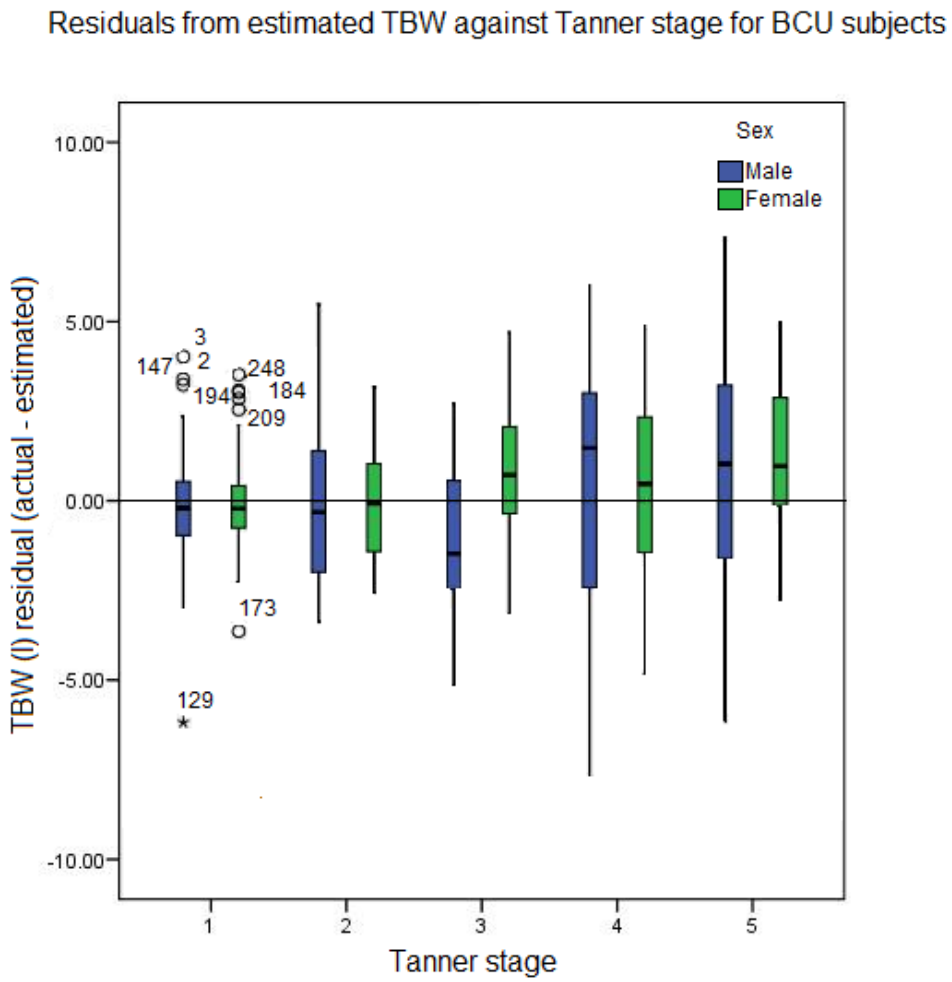

Figure 4.4: Residuals from log-log resistivity model against Tanner stage

It appears as though there does exist some effect of both age and Tanner, while any effect of sex is likely to be non-significant.

A scatterplot of residuals against age grouped by Tanner stage (see Figure 4.5) suggests that there is collinearity between the variables.

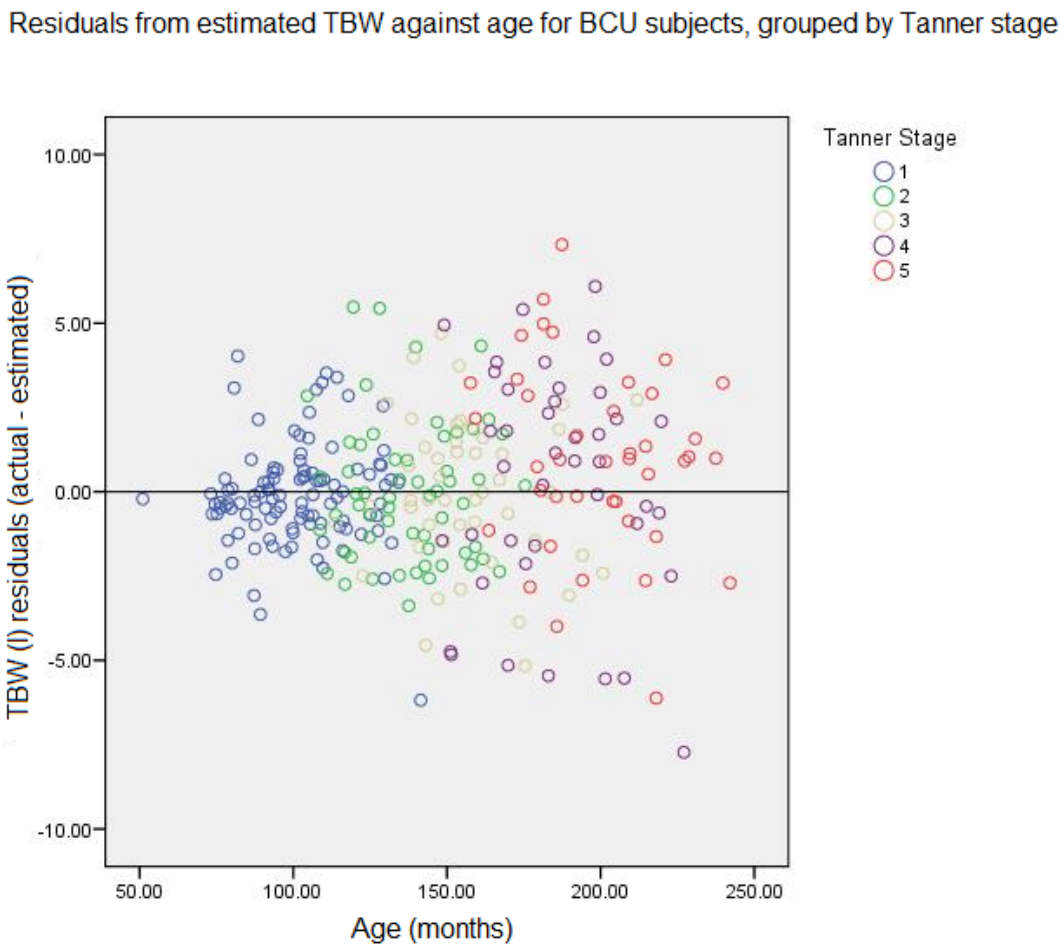

Figure 4.5: Residuals from log-log resistivity model against age grouped by Tanner stage

In order to determine which (if either) variable should be included in our model, we extend our model to include age as a covariate and Tanner stage and sex as fixed factors. We start with the most complicated model including all sensible interactions. Terms are eliminated one-by-one from the model (starting of course with the highest-order terms) according to the P-values. This process is stopped when all remaining terms have statistically significant P-values at the 5% level.

The final model from this process is:

$$ln(TBW) = -0.129 + 0.893 \times ln\left(\frac{ht^2}{Z}\right) + 0.001 \times age(m) \tag{4.4}$$

Diagnostic plots, shown in Figure A.3 on page 284 show no violation of modelling assumptions.

To compare this model with that not including an age term, we consider R^2 and s for each model. The model including age has a slightly higher R^2 (0.951 vs. 0.949) and a slightly lower s (0.084 vs. 0.085) than the simpler model. While the improvements are very slight, we are justified in continuing with the age model since age is reasonable to measure! A plot of residuals from this new model against Tanner stage in the BCU dataset (see Figure 4.6) reveals that although there might be an effect of maturation that is unexplained by age, this effect is very slight for male subjects on average, and practically non-existent on average for female subjects.

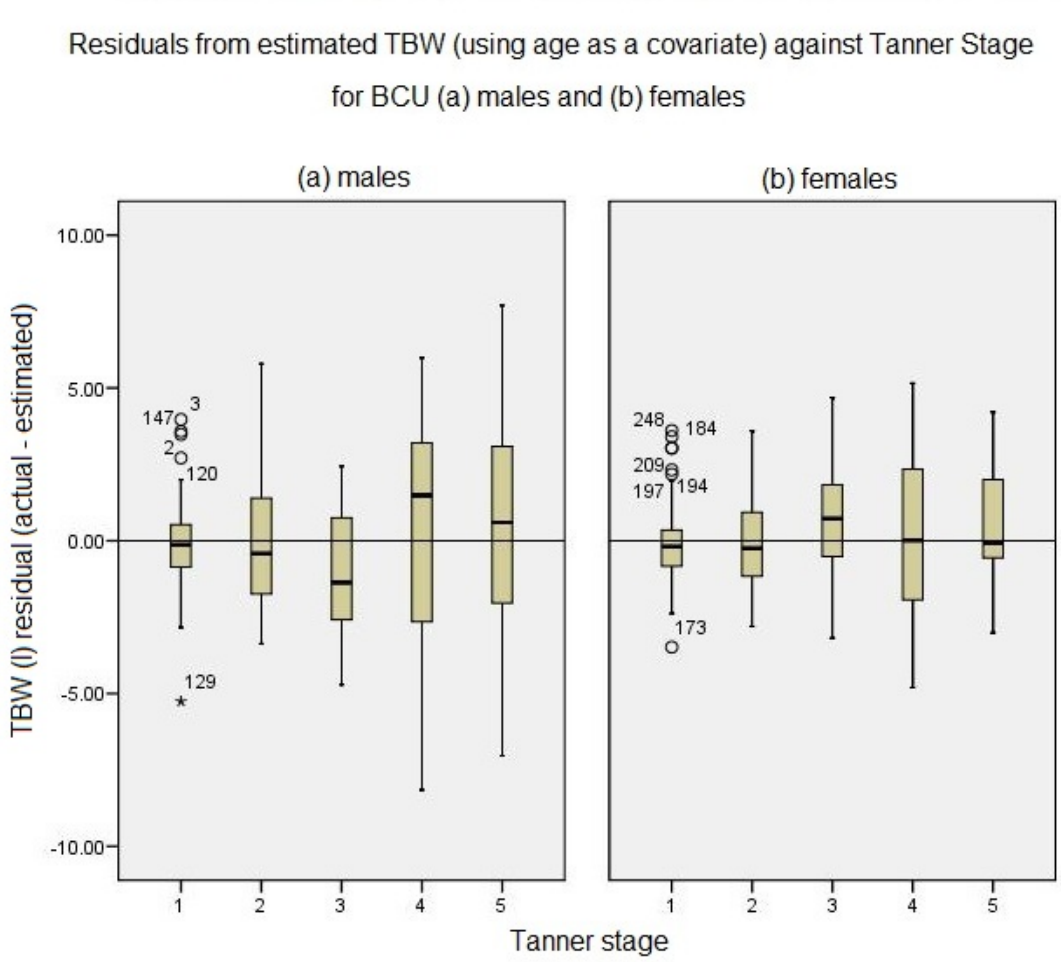

Figure 4.6: Residuals from log-log resistivity model including age as an independent variable by Tanner stage

4.2.2.3 Checking the form of the model with the ALSPAC data

While we don't have Tanner stage in the ALSPAC isotopes subsample[1], we do have age and can therefore attempt to verify the form of the model previously obtained. We begin with the following general linear model:

$$TBW_{iso} \sim sex + age(m) + ln\left(\frac{ht^2}{Z}\right) + sex \times age(m) + sex \times ln\left(\frac{ht^2}{Z}\right) \quad (4.5)$$

Modelling, we find no significant interactions. Looking at main effects, we have the following P-values:

Intercept: 0.117; $ln\left(\frac{ht^2}{Z}\right)$: < 0.001; age: 0.051

It should be noted here that although the age term is marginally nonsignificant, the ALSPAC subsample subjects were all aged roughly 11 years (mean $\pm$ sd = 11.89 $\pm$ 0.21) so it is tricky to interpret this P-value.

We can therefore assume that the ALSPAC isotopes subsample yields a model of a similar form as that developed using the BCU data. Ideally, we would have the means to show that there is no significant Tanner effect over and above the age effect, but this is unfortunately not possible using the ALSPAC data.

[1]Tanner staging data were provided near the end of this project. However, we have not used it in conjunction with this subsample because it was obtained by means of self-report (postal questionnaire) at a different time point from the clinic.

4.2.2.4 Calibrating TBW models for use with ALSPAC data

From the Bland-Altman diagram shown in Figure 4.7, our newly developed model for TBW with ht^2/Z and age over-estimates TBW in the ALSPAC subsample by approximately 2.33 litres on average.

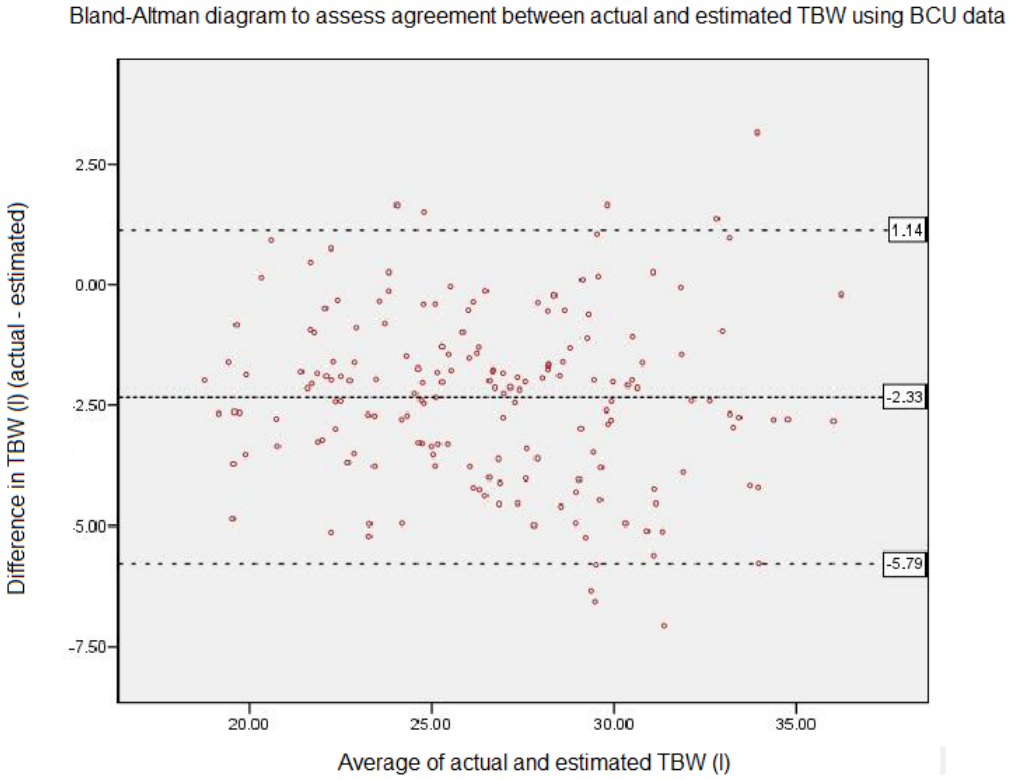

Figure 4.7: Bland-Altman diagram of actual TBW in ALSPAC subsample and TBW estimated from "age model"

It is likely that the systematic difference between actual and estimated TBW is due to differing impedance equipment between the two samples. While the equipment used in the ALSPAC clinics is known (Tanita hand-to-foot), that used in the BCU study is not. It is therefore essential, in order to be able to use the new models in conjunction with the ALSPAC data, to attempt to calibrate the TBW model

using the measured TBW in the subsample. This calibration is carried out on the log scale since this is the scale on which the model was developed. As calibration relies on the independent variable having as little error as possible, gold standard TBW by isotopes is used as the independent variable in this exercise. I.e:

$$ln(TBW_{new}) = \alpha + \beta \times ln(TBW) \tag{4.6}$$

After performing this calibration and re-arranging the model, we have the following adjustment to allow us to apply the previously developed resistivity model to the ALSPAC data:

$$ln(TBW) = -0.59 + 1.15 \times ln(TBW_{new}) \tag{4.7}$$

where TBW_{new} is given in equation (4.4). This gives:

$$ln(TBW) = -0.74 + 1.03 \times \left(\frac{ht^2}{Z}\right) + 0.001 \times age \tag{4.8}$$

4.2.2.5 Hydration

Having reached a model with which to estimate total body water, we need a model for hydration in order to get estimates of FFM in the body. We do not have any gold standard FFM measurements in any available dataset, making the task of accurately modelling hydration impossible. The current models in use for subjects younger than 18 years old are shown in equations (1.11) and (1.12) in section 1.2.1.2.2. However, those models are not equal to the "adult" hydration value of 0.732 until ages 31 and 38, for males and females, respectively.

In a 1982 paper, Fomon et al [243] published hydration values of the "reference child" from birth to 10 years of age. Figure 4.8 shows these reported values.

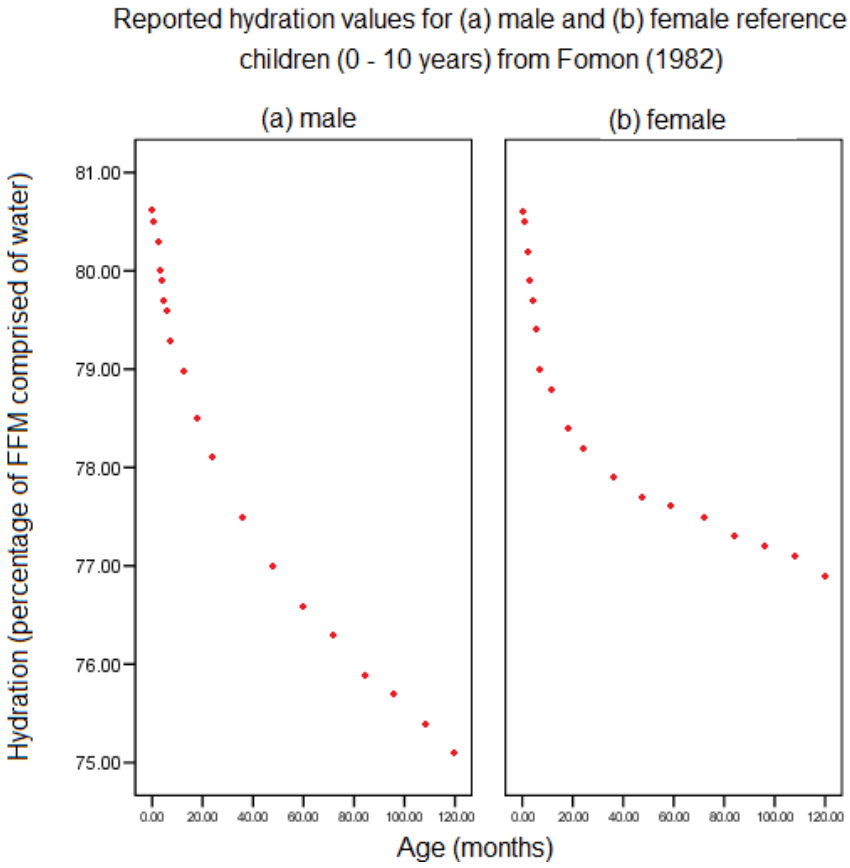

Figure 4.8: Hydration values of the 'ideal' male and female children

This graph raises a number of points:

- It is clear that the reference child hydration does not follow linear models
- It is possible that the values are already beginning to converge to a limit (possible adult value) far earlier than age 18
- This limit may be considerably higher than 0.732, and could differ between the sexes
- Girls seem to be maturing, in terms of hydration, earlier than boys

It could be said that infants, who are known to have much higher hydration than their older counterparts, would heavily influence any modelling carried out on these data.

Using only those "subjects" of age 24 months and above, we attempt to fit a model to the reference child hydration using curve estimation. Note that age is in **months**. We obtain the following models (equations (4.9) and (4.10) shown graphically in Figure 4.9):

BOYS:

$$h = 0.7969 - (0.0007910 \times age) + \left(0.00000580 \times age^2\right) - \left(0.00000002 \times age^3\right) \tag{4.9}$$

GIRLS:

$$h = 0.7896 - (0.0004104 \times age) + \left(0.00000398 \times age^2\right) - \left(0.00000002 \times age^3\right) \tag{4.10}$$

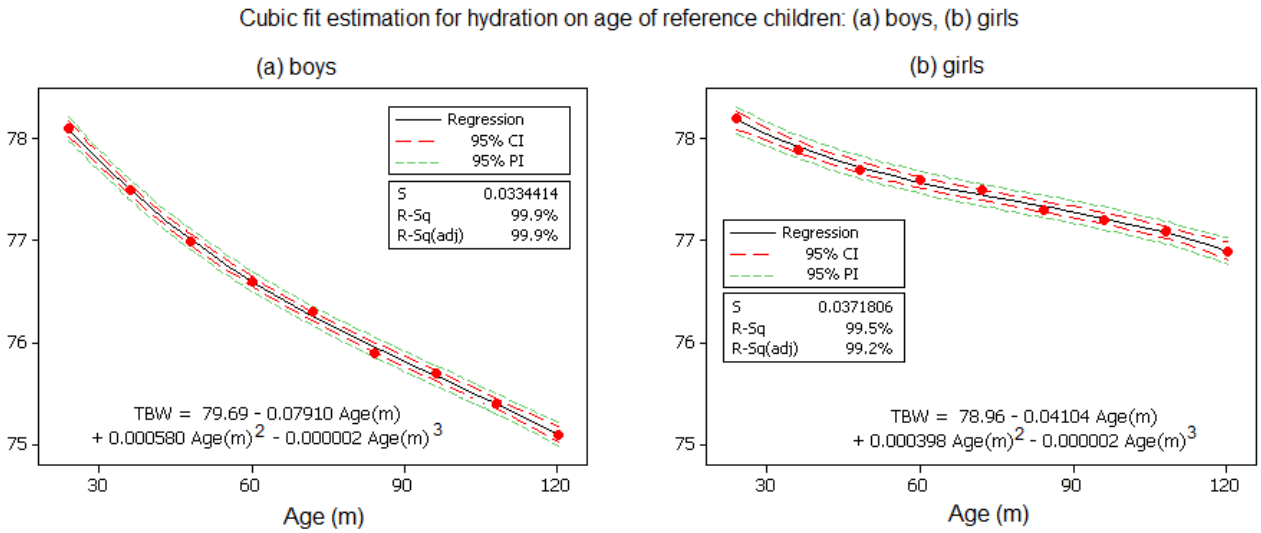

Figure 4.9: Cubic models for hydration on age: male and female reference children aged 24 months and above

In conjunction with the resistivity model (equation (4.8)), these new hydration models could potentially be used to estimate body composition in the larger ALSPAC datasets.

4.2.2.6 Applying new models to ALSPAC data

Figures 4.11 and 4.10 for boys and girls, respectively, show the difference (published - new) in FM from the two sets of impedance models applied to the ALSPAC data. It is clear that for both sexes, particularly for younger children, estimations from the two models are similar, but difference increases on the whole as age increases.

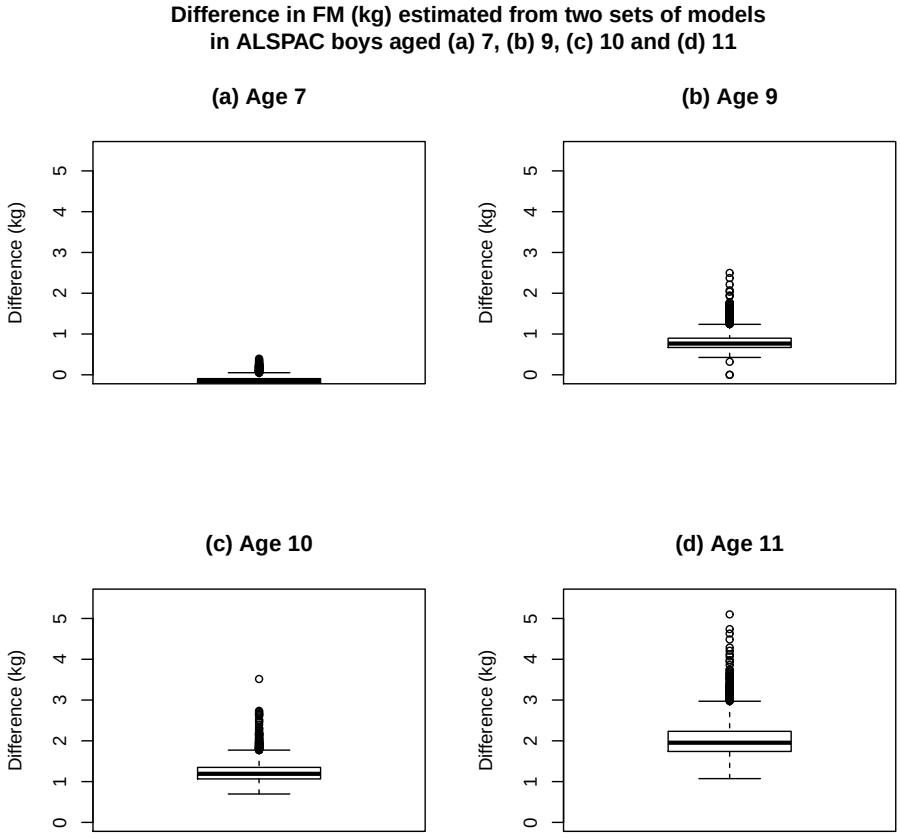

Figure 4.10: Difference in FM (kg) estimated from published and new BIA models for ALSPAC boys aged (a) 7, (b) 9, (c) 10 and (d) 11

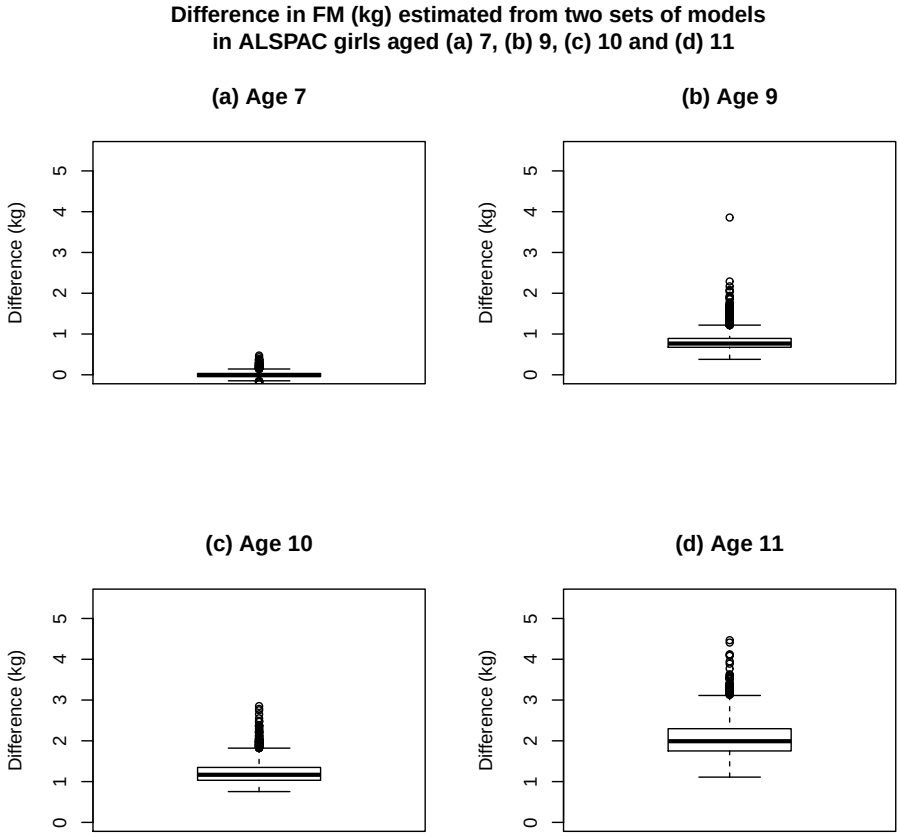

Figure 4.11: Difference in FM (kg) estimated from published and new BIA models for ALSPAC girls aged (a) 7, (b) 9, (c) 10 and (d) 11

4.2.3 Over-estimation of FM

A known problem with BIA is that, using current models, FFM is likely to be over-estimated for those subjects with impedance values at the low end of the scale. Since FM is the residual from weight and FFM, this problem results for some individuals in inaccurate FM estimations, including negative estimations. The number of negative FM estimations in the ALSPAC data is small, but must be noted regardless as it makes it clear that these models cannot be considered to be accurate on the individual level. Table 4.1 shows the number and percentage of such estimations in the ALSPAC data, using both the published BIA models and those developed in this chapter.

Sex	Wave	n (%)	n (%)
M	7	11 (0.28%)	9 (0.23%)
	9	11 (0.31%)	29 (0.81%)
	10	7 (0.20%)	7 (0.20%)
	11	7 (0.21%)	66 (1.98%)
F	7	0 (0.00%)	0 (0.00%)
	9	1 (0.03%)	1 (0.03%)
	10	2 (0.06%)	4 (0.11%)
	11	2 (0.06%)	5 (0.15%)

Table 4.1: Occurrence (n (%)) of negative FM results using two sets of BIA models with the ALSPAC data

While the new models over-estimate FFM for more subjects than the Glasgow models, this is not to say that they are, on a population level, less accurate. It must be noted that these new models suffer due to having been developed without:

- knowledge of impedance equipment in the BCU data
- gold standard body composition data (hydrodensitometry)
- accurate hydration data - it is unreasonable to assume that children today

are physiologically similar to Fomon's Reference Children of 1982. This requires significant further research as hydration currently involves several as-yet unproven assumptions

This chapter has highlighted data collection issues that must be addressed, carried out and brought together in order to replicate this analysis in future with results that will be, hopefully, meaningful, accurate and applicable to the current population.

4.3 Chapter summary

The opening analysis of this chapter has uncovered a startling and highly important fact: two widely used methods of estimating body composition do not give equivalent results. Unfortunately, with no gold standard data to work with, we were unable to validate either method, with its current underlying models - we can simply conclude that at most one method has been accurate in estimating the FM and FFM of children on the population level when applied to the ALSPAC sample.

The rest of this chapter isolated BIA with an aim to critiquing and potentially developing the models used to determine FM and FFM from BIA data. The first conclusion reached was that BIA cannot be said to be accurate on the individual level - which, given the desire in this project to investigate energy balance on a subject-by-subject basis, is disappointing. De-constructing the current models it was found that hydration values for children (dependent on age) do not meet the constant adult value until the mid 30s. It is also apparent that while hydration is considered to be equal for male and female adults, these values may actually differ, with female FFM comprising more water than that of males. These conclusions are extremely tentative and cannot be furthered until more suitable data are collected for the purpose.

While attempts to further the modelling in this chapter were, on the whole, unsuccessful, they did serve to identify areas in which current research is insufficient, with the potential to guide further research in the future.

Chapter 5

Analysis of resting energy expenditure models

5.1 Chapter aims

Chapter 5 centres around resting energy expenditure (REE). When reviewing literature (Chapter One), several sets of models for estimating REE from anthropometric measurements in children were identified. What was apparent from the literature was that no one model has yet been identified as clearly superior and several are in current use in research and clinical practice. One of these sets of models was developed and published, with the supporting data, in 1919 by Harris and Benedict and the initial aim of this chapter will be to replicate their analysis with modern day software (such as the statistical software packages R and SPSS) with a view to determining whether or not the form of their models was the most ideal given the data they had.

The analysis will then move from this to applying each REE model in turn to the ALSPAC data in order to determine if they are in agreement with one another. Unfortunately, as with Chapter Four, it will not be possible at this stage to validate any model or models absolutely, since no 'gold standard' results are available.

Following on from this, modelling of REE using different variables and statistical techniques will be attempted - with a view not to developing new models for use but rather to exploring different possibilities for how research into REE in childhood and adolescents might proceed.

5.2 REE in Harris-Benedict Study

The Harris and Benedict models (equations (1.16) and (1.17)) [86] for REE are frequently used today in both research and clinical practice, but were published over 90 years ago. In addition to the concern that what applied to the population in the early 20^{th} century may not apply to the population today, there is also the fact and therefore potential issue of methodology having changed to some extent since 1919. The aim of this section is to attempt to validate the models shown above, making use of modern methods and statistical computing facilities. Here, we apply the original and any further models only to the original 1919 data. This analysis is not intended to validate these models with respect to today's population.

The publication by Harris and Benedict gives data on 136 males and 103 females. These data have been input to SPSS v 15.0 and R v 2.6.0 for analysis. Due to the significance of sex as an interaction term, we will treat males and females separately.

As a subjective initial analysis of these data, we considered correlation matrices and scatterplot matrices of all variables, for both males and females. We found that potential all independent variables have some relationship with the dependent variable (REE) for both males and females, though relationships for females are generally quite weak. There may be some issue of multicollinearity, particularly between body surface area (BSA) and weight.

Upon analysing these data using multiple linear regression, we obtain the following prediction equations. Note that this is simply an attempt to reproduce the analysis by Harris and Benedict.

$$REE(male) = 67.307 + 13.750 \times W(kg) + 499.816 \times H(m) - 6.748 \times age(y) \quad (5.1)$$

$$REE(female) = 659.7920 + 9.6873 \times W(kg) + 176.3872 \times H(m) - 4.6278 \times age(y) \quad (5.2)$$

These parameter estimates are essentially the same as those obtained by Harris and Benedict. The small differences are likely due to modern computational accuracy. These models have adjusted R^2 values of 0.7478 and 0.5121 respectively, and s values of 103.153 and 108.923 respectively. Checking diagnostics, as shown in Figure A.4 reveals that the assumption of normality of residuals may be questionable.

Considering the coefficients of the parameters obtained in this analysis, we see that the parameter Height is not significant for females (Table 5.1) and the intercept is not significant for either sex.

	Male		Female	
	Coefficient	P-value	Coefficient	P-value
Intercept	67.3070	0.763	659.7320	0.056
Weight	13.7500	< 0.001	9.6873	< 0.001
Height	199.8160	< 0.001	176.8372	0.422
Age	-6.7480	< 0.001	-4.6278	< 0.001

Table 5.1: Parameters and associated P-values from linear regressions of REE on age, height and weight using the HB data

Re-fitting the models without the intercept gives the following predictive equations for REE, with standard errors of the estimates for males and females of 102.800 and 110.410 respectively:

$$REE(male) = 13.638 \times W(kg) + 541.590 \times H(m) - 6.669 \times age(y) \tag{5.3}$$

$$REE(female) = 9.359 \times W(kg) + 596.204 \times H(m) - 4.678 \times age(y) \tag{5.4}$$

Note that there appears to be little change in the coefficients for weight or age, while the lack of significance previously seen for height in females is now resolved. Why, though, height should be so sensitive to the presence or absence of the intercept, is unknown. All parameters are now significant.

Since we know that body composition is an important factor in determining resting energy expenditure, it seems logical that there should be some proxy measure of body composition included in a prediction equation. However, while considering only the data from the Harris-Benedict publication [86], we have limited access to such measures. we have therefore performed Stepwise regression (in an attempt to combat multicollinearity) including the variables BSA and BMI (calculated from height and weight), both potential proxies for body composition. The result of this modelling (steps not shown) was the following models for REE:

$$REE(male) = 421.992 - 446.243 \times H(m) - 6.679 \times age(y) + 1226.603 \times BSA(m^2) \tag{5.5}$$

$$REE(female) = -6.524 \times age(y) + 1025.764 \times BSA(m^2) \tag{5.6}$$

This suggests that when body surface area is accounted for in a model, there is no need to include a representation of body mass index. This could be due to multicollinearity, with the correlations between BMI and BSA in the 1919 data being 0.623 (males) and 0.845 (females). Note that the variability explained by these models (0.7478 and 0.5121 respectively) is roughly equal to that explained by models (5.1) and (5.2).

5.2.1 Conclusions

We should note that the models for males appear to consistently explain around 23% more of the variation in REE than those for females. This is, perhaps, to be expected because a key element of REE is fat-free mass - which generally occurs in higher proportions in males than in females, and Harris and Benedict had no direct measure of body composition.

During this analysis, it became clear that the most stable independent variable was age - with parameters remaining relatively unchanged and always highly significant. This suggests that age should always be included somehow in **any** resting energy expenditure model for adults. However, it seems unlikely that this should be done by developing different models for different age groups (as has been done in many of the models listed in section 1.2.2.2.2), since the other parameters are unlikely to change so dramatically at a specific age (unless age to be considered to be a marker for body composition - in which case, the relationship between age and body composition must be far better understood than at present)!

The aim of this analysis was to be able to validate the models originally developed by Harris and Benedict in 1919, using more modern approaches than would have been available to the authors at the time of the original study. The aim appears to be fulfilled for the data given in the original publication, although it is unlikely that these models will apply to today's population.

5.3 REE in ALSPAC data

5.3.1 Introduction

We will now attempt to apply the linear models for REE identified from the literature (section 1.2.2.2.2) to the ALSPAC "Children of the 90s" data - not with an aim to deciding on the *correct* model, but with the intention of comparing each model to each of the others. Although the Harris-Benedict models were developed on a sample all aged 15 years and over, we include for completeness these models in our analysis of children.

Note that The Oxford HW equation for boys aged 3-10 (equation 1.30) as given by Henry [100] may not be accurate in the original paper. This is based on the fact that, when applied to the data, the values for this model are considerably lower than all other models, while Henry stated that there was no significant difference between this and the Oxford H model. To date, attempts to establish contact with the author have been unsuccessful.

5.3.2 Applying models to ALSPAC data

Each REE model has been applied to the focus groups 7 and 11, categorised by sex, with descriptive statistics given in Tables 5.2 and 5.3. There are clear discrepancies between models. For example, there is a mean difference of almost 200 kcal/d between the WHO and Maffeis models for girls in age 7 group. There are differences between the sexes: the mean resting energy expenditure given by the Schofield HW model is higher for boys than for girls in the focus at 11 group by over 100 kcal / d. This is perhaps to be expected, since males are believed

to have a higher REE than females. However females at 11 have a considerably higher mean REE than boys in two of the three models that are dependent only upon weight.

Sex	Model	n	min	max	mean	sd
Male	HB	3946	782	1395	1000	80.8
	Schofield HW	3946	864	1576	1082	90.1
	Schofield W	3947	859	1640	1088	98.5
	WHO	3947	849	1630	1078	98.5
	Oxford HW	3946	623	1165	787	68.4
	Oxford W	3947	877	1679	1112	101
	Maffeis	3946	899	1304	1071	55.1
Female	HB	3813	963	1517	1098	52.7
	Schofield HW	3813	791	1736	1011	86.8
	Schofield W	3813	770	1850	1009	96.4
	WHO	3813	814	2011	1079	106.9
	Oxford HW	3813	808	1709	1021	83.7
	Oxford W	3813	788	1858	1025	95.5
	Maffeis	3813	848	1396	994	56.2

Table 5.2: Descriptive statistics of REE (kcal/day) from 7 published linear prediction models when applied to the ALSPAC Focus at 7 data

Sex	Model	n	min	max	mean	sd
Male	HB	3341	944	2203	1325	160.3
	Schofield HW	3341	1068	2391	1415	165.8
	Schofield W	3345	1004	2054	1265	131.2
	WHO	3345	1029	1986	1267	119.5
	Oxford HW	3341	1001	2320	1364	166.0
	Oxford W	3345	1008	2450	1367	180.2
	Maffeis	3341	987	1756	1253	97.9
Female	HB	3410	1054	1921	1307	108.4
	Schofield HW	3410	984	1869	1278	111.9
	Schofield W	3413	1059	2525	1450	183.9
	WHO	3413	1047	2496	1433	181.8
	Oxford HW	3410	993	1874	1259	110.3
	Oxford W	3413	1102	1931	1257	115.3
	Maffeis	3410	942	1803	1217	108.3

Table 5.3: Descriptive statistics of REE (kcal/day) from 7 published linear prediction models when applied to the ALSPAC Focus at 11 data

Another point which may be seen from this table is the discrepancy between the Harris-Benedict model and the other models at wave 7. The Harris-Benedict model gives a considerably lower REE for boys and a considerably higher REE for girls. We should note, again, that this model was not developed on children - suggesting that we do indeed require separate models for different periods of life and growth!

5.3.3 Testing for differences between models

Paired t-tests have been used to formally consider any significant mean differences between predictions obtained from different models (Tables 5.4 and 5.5), where all confidence intervals are given in kcal/d. All P-values for the correlation coefficients and the t-tests were < 0.001 for both focus groups. Note that - due to an overlap in samples - the standard deviation for the difference between Schofield W and WHO was 0.000 for the boys in the focus at 7 group, so no t-test could be performed. The performed t-tests clearly show that no two models are statistically similar at either age or for either sex. While some methods differ by a great deal and others considerably less, all differences were statistically significant at the 5% level.

Pair	Model 1	Model 2	Boys			Girls		
			Mean	95% Confidence Interval		Mean	95% Confidence Interval	
				Lower	Upper		Lower	Upper
1	HB	Schofield HW	-82.6	-83.2	-82.1	87.1	86.0	88.2
2		Schofield W	-88.4	-89.2	-87.5	89.0	87.6	90.4
3		WHO	-78.4	-79.2	-77.5	19.3	17.6	21.1
4		Oxford HW	212.3	211.7	212.9	76.5	75.5	77.5
5		Oxford W	-112.8	-113.7	-111.9	73.2	71.8	74.6
6		Maffeis	-71.8	-72.7	-71.0	104.0	103.7	104.2
7	Schofield HW	Schofield W	-5.7	-6.1	-5.4	1.9	1.5	2.3
8		WHO	4.3	3.9	4.6	-67.8	-68.5	-67.1
9		Oxford HW	294.9	294.2	295.6	-10.6	-10.7	-10.5
10		Oxford W	-30.2	-30.5	-29.8	-14.0	-14.3	-13.6
11		Maffeis	10.8	9.5	12.1	16.9	15.8	17.9
12	Schofield W	WHO	-	-	-	-69.7	-70.0	-69.3
13		Oxford HW	300.7	299.7	301.6	-10.6	-13.0	-12.0
14		Oxford W	-24.4	-24.5	-24.3	-14.0	-15.9	-15.8
15		Maffeis	16.5	14.9	18.1	16.9	13.6	16.4
16	WHO	Oxford HW	290.7	289.7	291.6	57.2	56.4	58.0
17		Oxford W	-34.3	-34.5	-34.3	53.8	53.5	54.2
18		Maffeis	6.5	4.9	8.1	84.6	82.9	86.4
19	Oxford HW	Oxford W	-325.1	-326.1	-324.0	-3.4	-3.8	-2.9
20		Maffeis	-284.1	-284.9	-283.3	27.4	26.5	28.3
21	Oxford W	Maffeis	40.9	39.2	42.6	-69.7	29.4	32.2

Table 5.4: Mean differences in REE (kcal/day) between models and 95% confidence intervals (model 1 minus model 2): Focus at 7 data

Pair	Model 1	Model 2	Boys			Girls		
			Mean	95% Confidence Interval		Mean	95% Confidence Interval	
				Lower	Upper		Lower	Upper
1	HB	Schofield HW	-90.2	-90.9	-89.5	29.8	29.3	30.3
2		Schofield W	59.9	58.6	61.2	-142.4	-144.9	-139.8
3		WHO	58.1	56.6	59.7	-125.4	-127.9	-122.9
4		Oxford HW	-39.2	-39.7	-38.7	48.5	48.4	48.7
5		Oxford W	-41.6	-42.8	-40.4	50.5	50.1	50.9
6		Maffeis	71.9	69.7	74.2	90.6	90.2	90.9
7	Schofield HW	Schofield W	150.1	148.9	151.3	-172.2	-174.9	-169.5
8		WHO	148.3	146.7	149.9	-155.2	-157.9	-152.6
8		Oxford HW	51.0	50.7	51.2	18.7	18.3	19.1
10		Oxford W	48.6	48.0	49.1	20.7	19.8	21.6
11		Maffeis	162.1	159.5	164.7	60.8	60.5	61.0
12	Schofield W	WHO	-1.8	-2.2	-1.4	16.9	16.9	17.0
13		Oxford HW	-99.1	-100.4	-97.8	190.9	188.4	193.5
14		Oxford W	-101.5	-103.1	-99.8	192.9	190.6	195.2
15		Maffeis	12.0	10.4	13.7	232.9	230.2	235.6
16	WHO	Oxford HW	-97.3	-99.0	-95.7	174.0	171.5	176.5
17		Oxford W	-99.7	-101.8	-97.7	175.9	173.7	178.2
18		Maffeis	13.8	12.4	12.2	216.0	213.4	218.6
19	Oxford HW	Oxford W	-2.4	-3.1	-1.7	2.0	1.5	2.5
20		Maffeis	111.1	108.6	113.6	42.0	41.8	42.3
21	Oxford W	Maffeis	113.6	110.4	116.7	40.1	39.3	40.8

Table 5.5: Mean differences in REE (kcal/day) between models and 95% confidence intervals (model 1 minus model 2): Focus at 11 data

5.3.4 Checking the agreement between models

We consider pairwise intraclass correlations (ICC), with a two-way mixed model, checking for absolute agreement between models (for the groups focus at 7 and focus at 11) as shown in Tables 5.6 and 5.7.

		HB	Sch HW	Sch W	WHO	Ox HW	Ox W
Boys	Schofield HW	0.668					
	Schofield W	0.645	0.993				
	WHO	0.693	0.994	0.995			
	Oxford HW	0.193	0.124	0.128	0.136		
	Oxford W	0.540	0.945	0.971	0.944	0.115	
	Maffeis	0.597	0.832	0.772	0.786	0.080	0.690
		HB	Sch HW	Sch W	WHO	Ox HW	Ox W
Girls	Schofield HW	0.510					
	Schofield W	0.503	0.992				
	WHO	0.765	0.786	0.806			
	Oxford HW	0.563	0.991	0.976	0.821		
	Oxford W	0.578	0.981	0.986	0.871	0.986	
	Maffeis	0.351	0.837	0.826	0.534	0.850	0.784

Table 5.6: Intraclass correlations between pairs of estimated REE (kcal/day) from different models in the ALSPAC Focus at 7 data

		HB	Sch HW	Sch W	WHO	Ox HW	Ox W
Boys	Schofield HW	0.860					
	Schofield W	0.892	0.647				
	WHO	0.871	0.621	0.996			
	Oxford HW	0.968	0.954	0.795	0.770		
	Oxford W	0.951	0.958	0.788	0.760	0.993	
	Maffeis	0.768	0.494	0.908	0.926	0.640	0.611
		HB	Sch HW	Sch W	WHO	Ox HW	Ox W
Girls	Schofield HW	0.955					
	Schofield W	0.603	0.527				
	WHO	0.648	0.568	0.996			
	Oxford HW	0.910	0.980	0.488	0.527		
	Oxford W	0.902	0.956	0.503	0.542	0.991	
	Maffeis	0.738	0.866	0.392	0.423	0.929	0.920

Table 5.7: Intraclass correlations between pairs of estimated REE (kcal/day) from different models in the ALSPAC Focus at 11 data

For a description of ICC, see section 2.6.1. These correlations give a general measure of agreement between models. We can see that, at age 7, the Oxford HW model is not in agreement with any other model for the boys. This may be because of the possible error in the formula that was identified earlier in this thesis. What is clear here is that the HB models are in poor agreement with the other models at age 7, and better at age 11. This may be because the HB models were not developed on children. Among the other models, for both boys and girls at both ages, the agreement seems, in general, to be reasonable.

Pairwise agreement between models can also be examined using Bland-Altman diagrams [102] (See section 2.6.2 for a description of this method). These have been produced for both boys and girls at each of the focus groups 7 and 11 and appear to show a consistent lack of agreement, shown by an increasing or decreasing pattern, between the models. A selection of these Bland-Altman diagrams (for boys at wave 7) are shown in Figure 5.1. From these plots, we can see the individual discrepancies between the models that are not shown in the intraclass correlations. In general, the higher the average REE, the higher the discrepancy between the models. Here, the Bland-Altman diagrams are far more informative in terms of assessing agreement between models than the intraclass correlations.

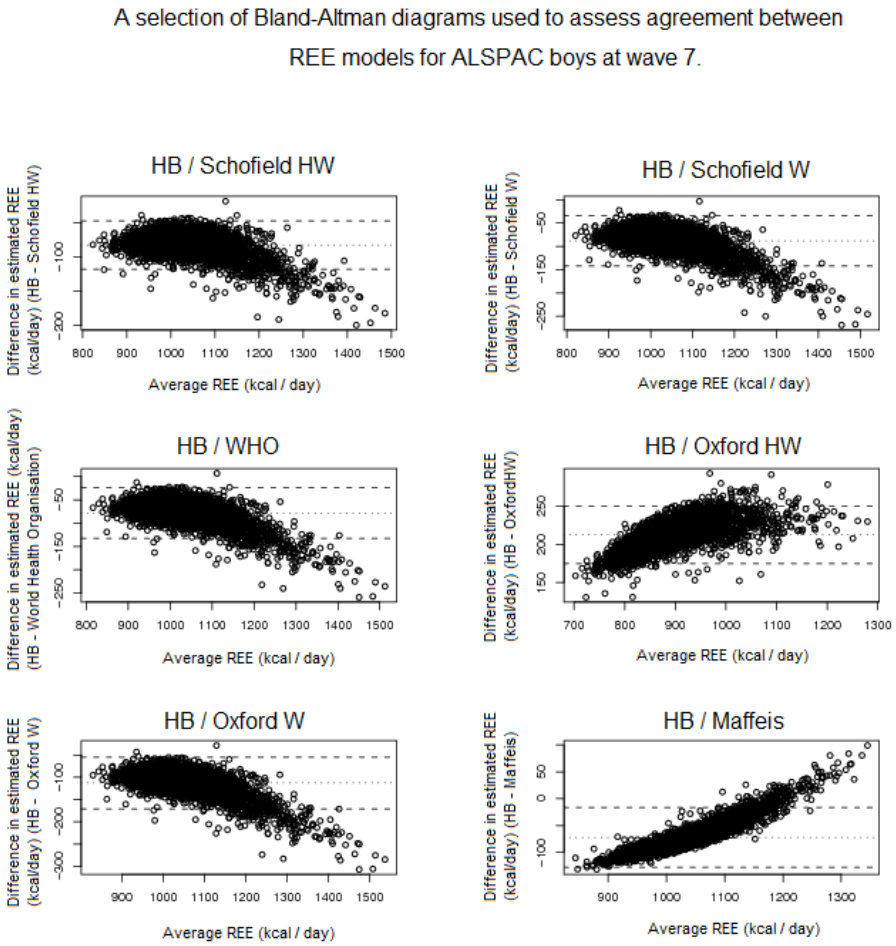

Figure 5.1: A selection of Bland-Altman diagrams assessing agreement between REE (kcal/day) estimated from different published models applied to the ALSPAC data

5.3.5 Conclusions

This analysis has attempted neither to separate "correct" models from "incorrect" models, nor to mark any one set of models as a gold standard for use today. Rather, the aim was to determine how closely each published model agreed with each of the others when applied to data outwith that of the original studies. From the results presented throughout this section, it is abundantly clear that no two models are in agreement.

We can conclude, therefore, that *at most* one model can be correct. This is not, however, to say that there is one correct model!

5.4 Looking for a scaling relationship

As discussed in the review of relevant literature, the model $REE \propto M^{0.75}$ has been suggested [74] and shown to be relatively accurate for all adult mammals.

5.4.1 Harris - Benedict data

Re-analysing the Harris and Benedict data [86], we investigate whether there is an appropriate scaling relationship for this specific sample.

We find the following relationships (equations (5.7) and (5.8)) when performing simple linear regressions with the natural logs of REE and weight:

$$Male \;\; REE = 117.9 \times Weight^{0.631} \tag{5.7}$$

$$Female \;\; REE = 292.9 \times Weight^{0.379} \tag{5.8}$$

These models had R^2 values of 64.0 and 38.7 for males and females, respectively. Of course, residual plots for these models must be considered. As well as plotting residuals against fitted values, we will also examine plots of residuals against age in order to determine whether or not age may be useful as a proxy measure of body composition where more accurate measures are not available. Diagnostic plots (not shown) for these models show that the models fit the data very well but as shown in Figure 5.2, plots of residuals against age, but there is a clear effect of age that we have not accounted for - and that may be useful in representing body composition in models.

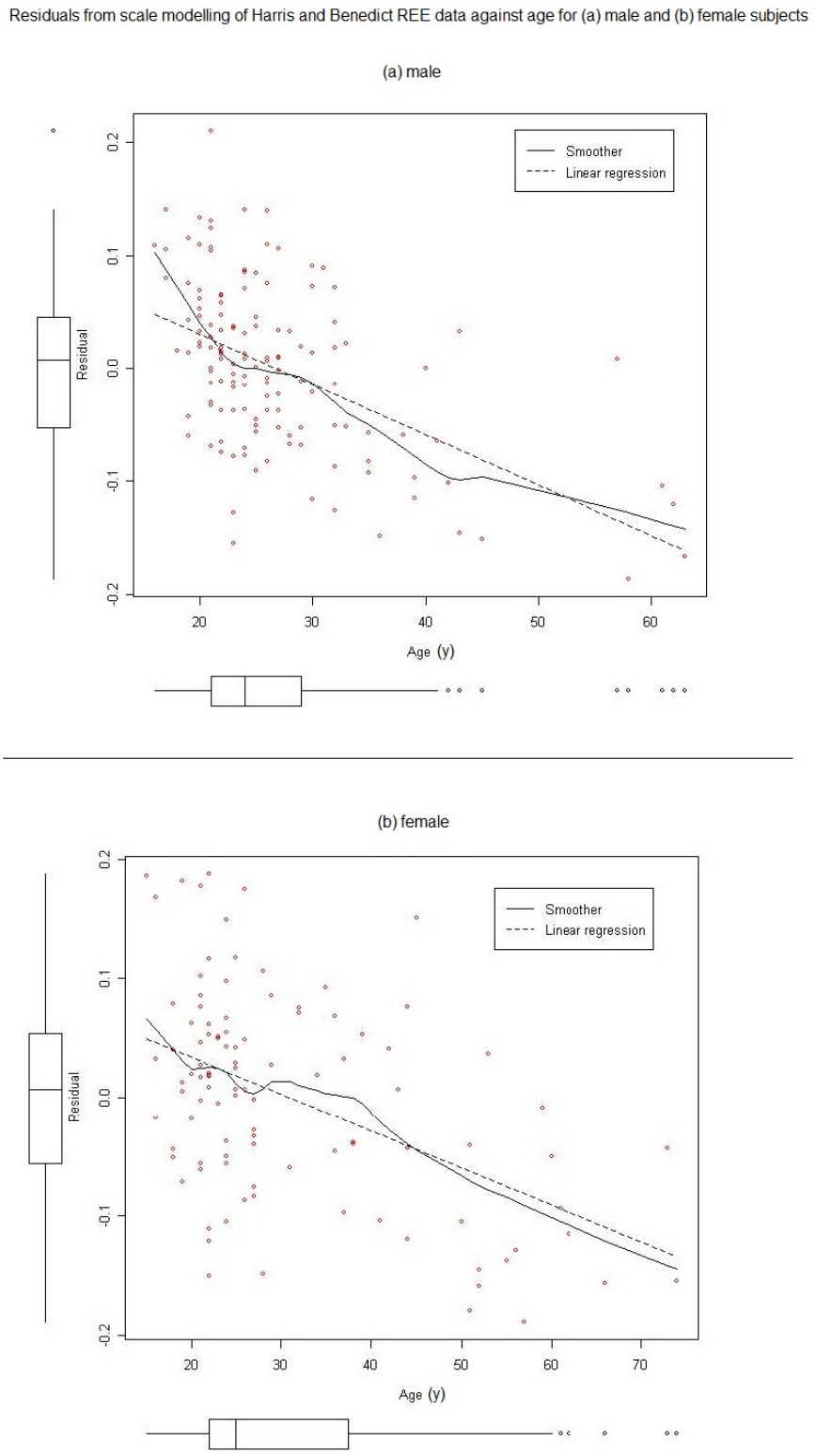

Figure 5.2: Residuals from REE scale models (equations 5.7 and 5.8) applied to the HB data

5.4.2 Institute of Medicine Database

The database from the Institute of Medicine brings together data from several reputable published sources into one dataset with anthropometric and energy expenditure data for several age groups of both males and females. Unfortunately, the database does not include detailed data on body composition. However, with the data available, it is possible to investigate whether or not there is a scaling relationship between REE and weight for each age group within each sex. Additionally, we consider the data for each sex combining all ages (excluding 0-2 years). The models found by the regression on the natural logs are summarised in Table 5.8.

Sex	Age group	Model (REE (kcal/d)=)	R^2
Male	0 - 2	$127.7 \times Weight^{0.567}$	33.4
	2 - 8	$200.3 \times Weight^{0.539}$	33.8
	9 - 18	$109.9 \times Weight^{0.684}$	77.6
	Adult	$79.0 \times Weight^{0.711}$	27.3
	2+	$395.4 \times Weight^{0.296}$	63.4
Female	0 - 2	$101.5 \times Weight^{0.681}$	33.4
	2 - 8	$278.7 \times Weight^{0.410}$	51.0
	9 - 18	$368.7 \times Weight^{0.326}$	44.4
	Adult	$108.9 \times Weight^{0.607}$	27.6
	2+	$314.2 \times Weight^{0.390}$	81.9

Table 5.8: Models produced for expressing REE (kcal/day) proportional to a power of body mass using the IOM data

The differences between the constants and exponents can be shown graphically as follows in Figure 5.3.

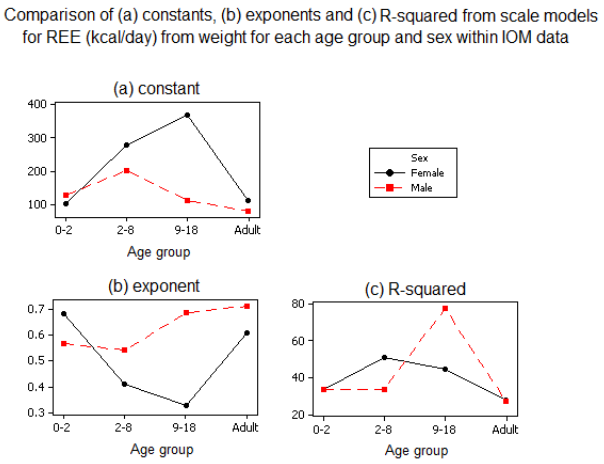

Figure 5.3: Comparisons of constants, exponents and R-squared for scale models within IOM data

Residuals vs. fitted values plots (not shown) revealed no violation of modelling assumptions for this data.

5.5 Modelling REE based on FM and FFM

5.5.1 Linear models

For a subsample of 246 ALSPAC participants (110 male, 136 female) with a mean age of 12.41, we have REE provided in two measures: oxygen uptake, and a conversion from this to energy expenditure calculated using the Weir formula [244] which is considered to be the gold standard conversion equation and is widely used in current research. For each subject, we have measurements taken at various different levels of activity and we isolate the data for the "lying". Since we have no information regarding the procedures preceding measurement, we attempt to model resting energy expenditure rather than basal metabolic rate. We also have DEXA measurements of fat mass (FM), lean mass (LM) and bone mass (BM) previously recorded at the ALSPAC Focus at 11 group. FFM is considered to be equal to BM+LM. In order to make some of the REE data , we impute body composition measurements for the subsample by calculating the proportion of body weight given by both FM and FFM at the time of the DEXA measurements for each subject, and applying these proportions to the corresponding body weights in the later subsample.

5.5.1.1 Exploratory analysis

Plotting measured REE against age (for boys and girls separately) as shown in Figure 5.4, we see that there are some considerably high REE values for both sexes. We also see a general increase in REE over time, which appears to be stronger for the girls than for the boys.

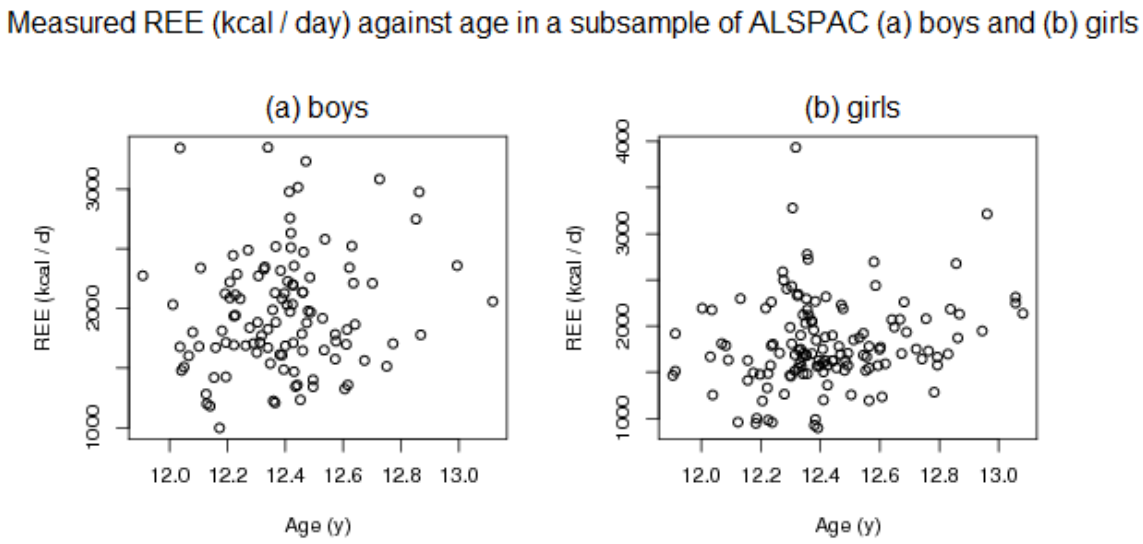

Figure 5.4: Scatterplot of age (years) and REE (kcal/d) in a subsample of ALSPAC participants

Comparing the weights, fat and fat-free masses and BMI SDS for the two samples (using imputed FM and FFM for the later sample), we found that each variable increased for all subjects between the Focus at 11 group and the subsample measurements approximately one year later.

5.5.1.2 Formal analysis

Butte et al [245] give equations for calculating REE for boys and girls based on their Tanner stage. We assume the Tanner stage of the subsample respondents to be, on average, 2 (given the ages). Therefore, the Butte equations for these data would be:

BOYS:

$$REE(kcal/d) = (37.9 \times FFM) + (6.45 \times FM) \tag{5.9}$$

GIRLS:

$$REE(kcal/d) = (40.2 \times FFM) + (6.45 \times FM) \tag{5.10}$$

We attempted to generate our own models using the method of linear least squares regression, but found clear violations of the modelling assumptions for both sexes. The models were:

BOYS:

$$REE(kcal/d) = (55.79 \times FFM) + (15.10 \times FM) \tag{5.11}$$

GIRLS:

$$REE(kcal/d) = (56.323 \times FFM) + (6.398 \times FM) \tag{5.12}$$

Using the Butte equations ((5.9) and (5.10)), we can estimate the REE for the subsample. This can then be compared to the estimated REE from our own models (despite violation of assumptions), and the actual measured values from the subsample. Means and standard deviations are shown in Table 5.9, and boxplots in Figures 5.5 and 5.6.

Sex	n	Result from	Mean REE	sd
Boys	110	Measured	1963.6	490.1
		Ours	1946.4	316.7
		Butte	1280.5	194.5
Girls	136	Measured	1807.0	473.0
		Ours	1805.9	266.1
		Butte	1312.5	197.0

Table 5.9: Descriptive statistics for REE (kcal/day) produced from 3 linear models using FM and FFM in a subsample of ALSPAC participants

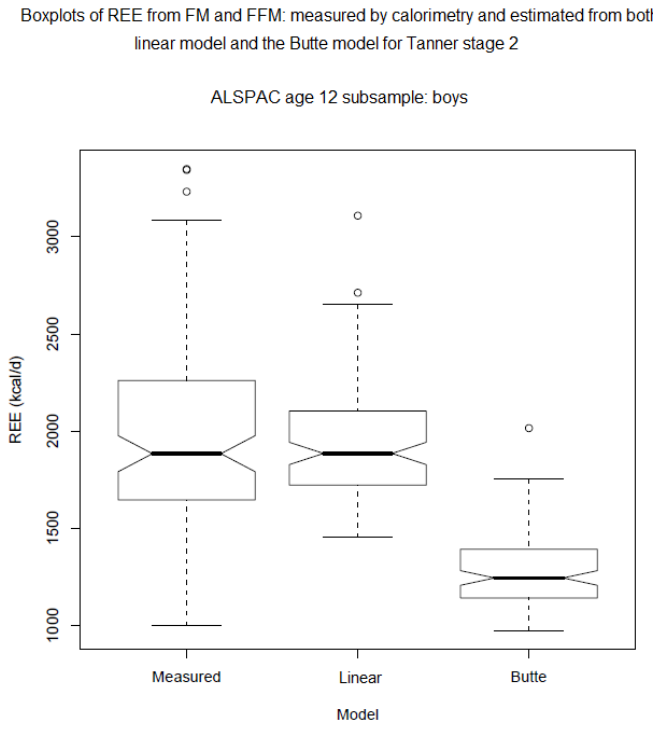

Figure 5.5: Boxplots of REE (kcal/day) produced from 3 linear models using FM and FFM in a subsample of ALSPAC boys

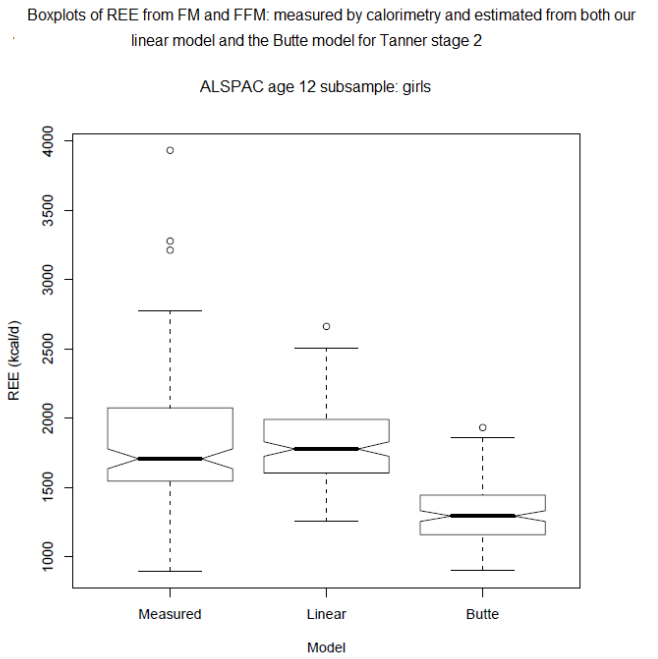

Figure 5.6: Boxplots of REE (kcal/day) produced from 3 linear models using FM and FFM in a subsample of ALSPAC girls

It seems from the statistics and boxplots that there are significant differences between our model and the Butte model, and between the actual measurements and the Butte model. It should also be noted that in both our model and the measured REE values, there are some considerably high results. Realistically, would a child of age 12 have a resting energy expenditure as high as 2000 to 3000 kcal per day?

5.5.2 Nonlinear models

Using R's nonlinear least squares function [6] function, as described in section 2.4, we can fit a nonlinear model to our data. The model we are aiming to fit is:

$$REE = \beta_0 FM^{\beta_1} + \beta_2 FFM^{\beta_3} \tag{5.13}$$

The arbitrarily chosen starting values for this model are a=300, b=1, c=300, d=1. The number of iterations to convergence was 9 for the boys and 26 for the girls.

We get the following equations for boys and girls separately (again, there was an issue with modelling assumptions):

BOYS:

$$REE(kcal/d) = (360.1741 \times FM^{0.3436}) + (199.5909 \times FFM^{0.5142}) \tag{5.14}$$

GIRLS:

$$REE(kcal/d) = (169.1893 \times FM^{0.2798}) + (32.2767 \times FFM^{1.1159}) \tag{5.15}$$

Table 5.10 and Figures 5.7 and 5.8 summarise the estimated REE values for each sex using these models as well as previous models.

Sex	n	Result from	Mean REE	sd
Boys	110	Measured	1963.6	490.1
		Ours	1946.4	316.7
		Ours (Nonlinear)	**1963.6**	**231.3**
		Butte	1280.5	194.5
Girls	136	Measured	1807.0	473.0
		Ours	1805.9	266.1
		Ours (nonlinear)	**1807.2**	**258.8**
		Butte	1312.5	197.0

Table 5.10: Descriptive statistics for REE (kcal/day) produced from linear and nonlinear models using FM and FFM in a subsample of ALSPAC participants

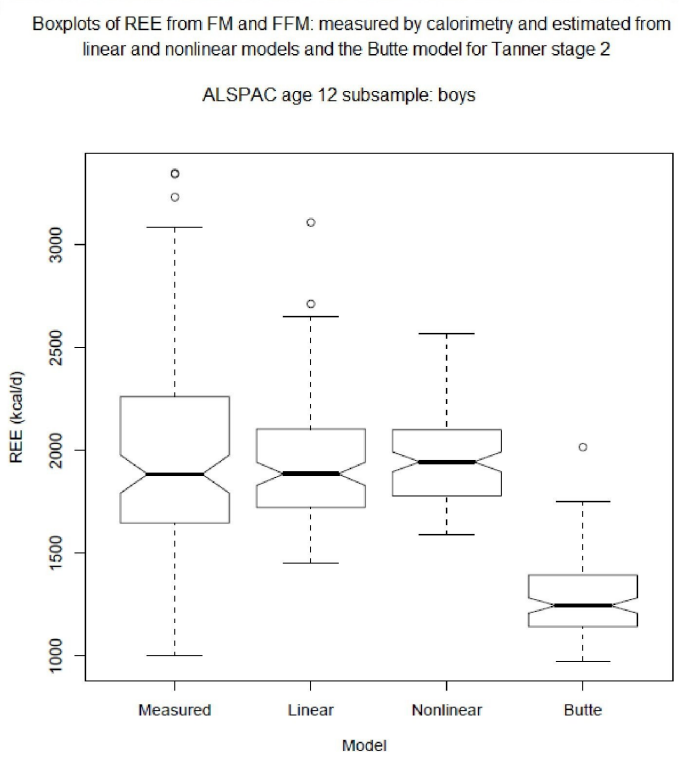

Figure 5.7: Boxplots of REE (kcal/day) produced from linear and nonlinear models using FM and FFM in a subsample of ALSPAC boys

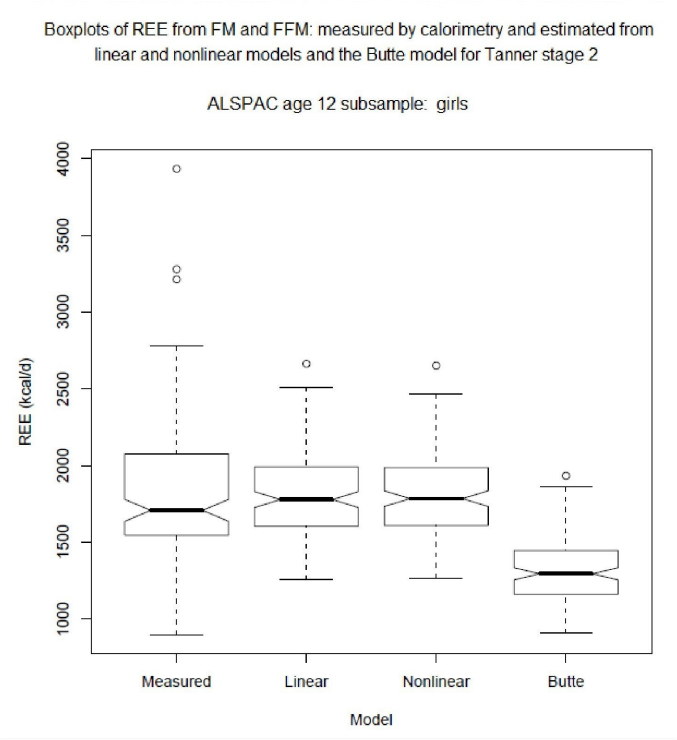

Figure 5.8: Boxplots of REE (kcal/day) produced from linear and nonlinear models using FM and FFM in a subsample of ALSPAC girls

Leaving aside the issue of modelling assumptions for now, we can compare the models obtained in this section (summarised on page 163) using Bland-Altman diagrams [102] to assess agreement, or lack thereof, among models. These are shown in Figures 5.9 and 5.10 for boys and girls, respectively.

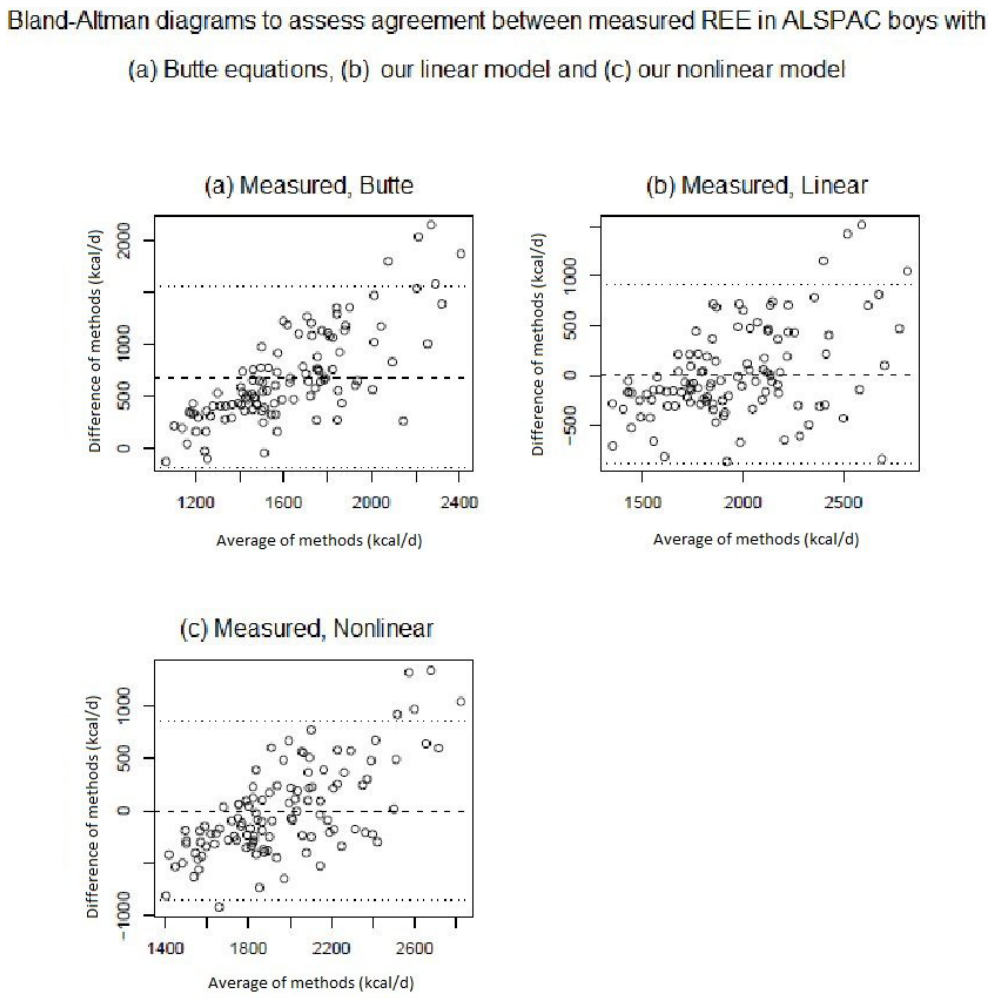

Figure 5.9: Bland-Altman plots for assessing the agreement between REE as measured by ALSPAC and REE as estimated from our models - ALSPAC subsample (boys)

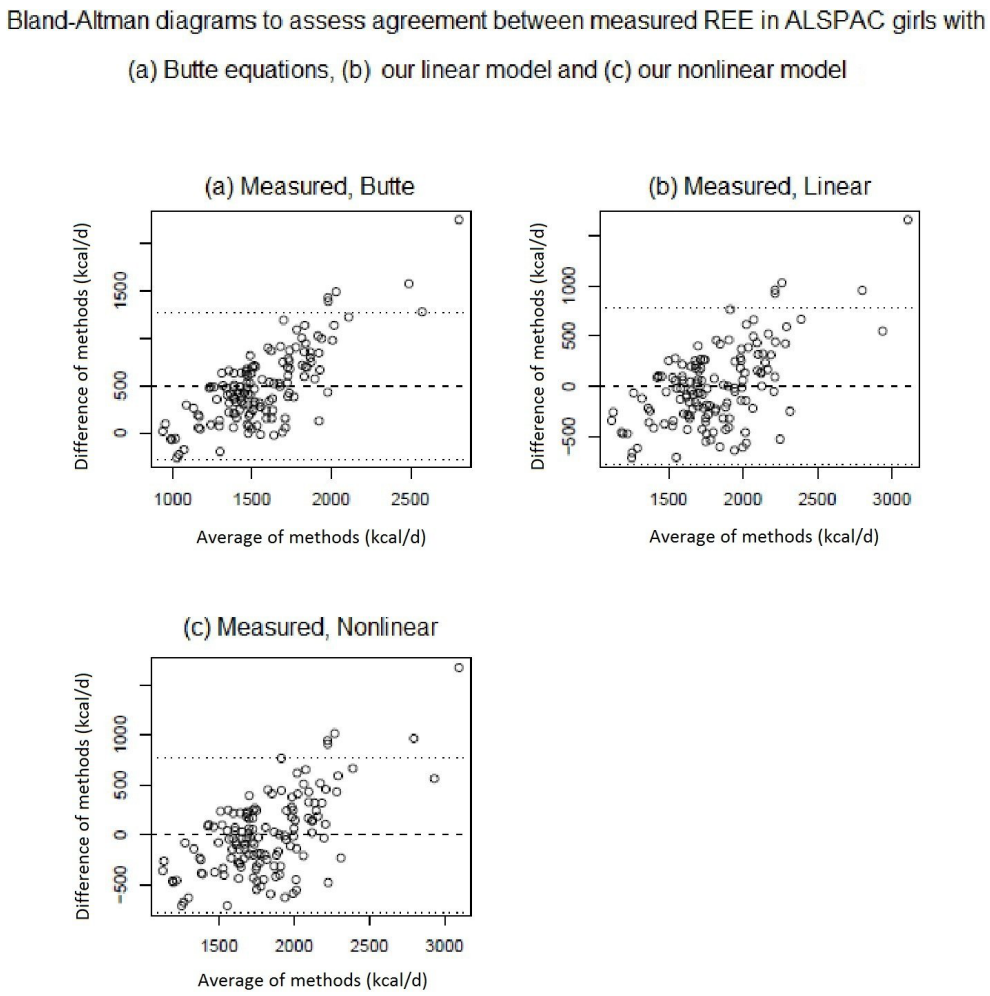

Figure 5.10: Bland-Altman plots for assessing the agreement between REE as measured by ALSPAC and REE as estimated from our models - ALSPAC subsample (girls)

5.5.3 Modelling REE with another dataset

The dataset used in this analysis was provided by Professor Jonathan Wells of the Institute of Child Health in London, in the spring of 2008. Resting energy expenditure measured by calorimetry (kcal/d) and body composition (fat mass and fat-free mass as determined by DEXA) are included for 16 boys and 14 girls of ages 8.1 - 12.1 years (mean = 9.7, sd = 1.3) and 8.2 - 12.4 years (mean = 10.1, sd = 1.4), respectively.

5.5.4 Linear models

We began by investigating separate linear models of REE on FM and FFM for each sex, which gave the following:

$$REE(boys) = 55.378 \times FFM - 13.845 \times FM \quad (5.16)$$

$$REE(girls) = 46.098 \times FFM + 6.028 \times FM \quad (5.17)$$

However, it should be noted that sample size, and therefore statistical power, is extremely low, rendering these models tentative at best. In order to both increase power and investigate whether or not sex is a statistically significant covariate when it comes to these pre- or early-pubescent children (no data are available on pubertal staging), we combine the data for both sexes to consider ANCOVA models, treating sex as a covariate and FM and FFM as independent variables. As there is no biological realism in investigating an interaction between FM and FFM, the three way interaction will not be considered. Therefore, the first ANCOVA

model considered will be:

$$REE \sim sex + FM + FFM + sex * FM + sex * FFM \tag{5.18}$$

This model returns a F-statistic of 15.05 on 5 and 21 degrees of freedom, with P-values for the interactions of 0.98180 ($sex * FM$) and 0.52376 ($sex * FFM$). Since the two-way interactions are non-significant, we remove the one with the least significance ($sex * FM$) and re-fit the ANCOVA model as:

$$REE \sim sex + FM + FFM + sex * FFM \tag{5.19}$$

This model returns an F-statistic of 19.23 on 4 and 22 degrees of freedom, and a P-value for the interaction of 0.534, showing no statistical significance at any reasonable level. As a result, we are able to remove this term and fit a model with just the main effects:

$$REE = sex + FM + FFM \tag{5.20}$$

This procedure returns a marginally nonsignificant P-value for sex (0.52). However, removing the term sex from the model lowers the adjusted R^2 from 0.744 to 0.712, suggesting that since sex isn't difficult or expensive to determine, it should perhaps remain in the model. Diagnostics for this model, as shown in Figure A.5 on page 285, show that there may be problems with the assumptions of modelling - particularly normality of residuals.

5.5.5 Nonlinear models

We went on to investigate nonlinear modelling of the BMR data provided by Professor Wells. We came across some initial problems with running the iterative model. To overcome these difficulties, it was necessary to increase the number of iterations used in the procedure from 50 to 150.

The model we are aiming to fit with this procedure is:

$$REE = \beta_0 FM^{\beta_1} + \beta_2 FFM^{\beta_3} \tag{5.21}$$

This results in the following equations for boys and girls:

BOYS:

$$REE(kcal/d) = (99.6 \times FM^{0.34}) + (141.5 \times FFM^{0.66}) \tag{5.22}$$

GIRLS:

$$REE(kcal/d) = (1034 \times FM^{0.005}) + (0.176 \times FFM^{2.774}) \tag{5.23}$$

Considering the exponents in these models, we notice that the girls' exponents are not remotely similar to anything we have already seen or would in fact expect from this modelling.

We investigate the variance covariance matrices of these models, finding the following eigenvalues:

BOYS: 6496528.00, 5856.39, 0.06, 0.00005

GIRLS: 34276.68, 0.93, 0.0003, 0.000003

From these eigenvalues, we can assess the stability of the parameter estimates. If the condition number (the ratio of the largest to the smallest eigenvalue) is small, we may be confident in the stability of the estimates. A large condition number implies that any small change in the data could drastically alter parameter estimates. For these models, it is clear that the condition numbers for both boys and girls are extremely large, showing numerically very unstable models. This may be down to the small sample size. We can therefore not continue with this modelling.

5.6 Chapter summary

This chapter opened by exploring the data published by Harris and Benedict in 1919, which they used to develop models for estimating resting energy expenditure (REE) from anthropometric measurements.

The findings of this section were that despite the lack of techniques and equipment that would have been available many years ago, the best models that can be derived today from the published data are in-fact the same models that were originally published. It must be kept in mind, however, that these models were developed almost a century ago and therefore on a different population, yet are still in use to this day. It must also be kept in mind that while these models were developed using adult subjects, they are currently used to estimate REE in younger subjects.

Following on from this, seven sets of models for estimating REE from anthropometric measurements were applied to the ALSPAC datasets, showing that no two models are in agreement on the individual level. It was not possible to progress to marking any model as correct, however, due to the lack of gold standard data collected by ALSPAC.

The final part of this chapter considered alternative possibilities for REE modelling, including modelling based on body composition.. However, until the issues raised in Chapter 4 have been addressed - that is, until we have accurate methods of estimating body composition - it will not be possible to develop reliable models using such variables as predictors.

Chapter 6

The ideal study and a simulation of resting energy expenditure

6.1 Chapter aims

Chapter 6 will begin by describing an 'ideal study' in body composition and resting energy expenditure research. Following on from this, longitudinal resting energy expenditure and body composition data will be simulated, in a manner which generates reasonably realistic data, for several thousand children between the ages of 7 and 10 years. The chapter will focus mainly on potential issues with real-life data, such as sample size and missing data.

6.2 The ideal study

In attempting to design the ideal study with which to model energy-imbalance over puberty, it quickly becomes clear that such a study, despite the potential of unlimited resources, is virtually an impossibility.

The guiding principle of such a study in this field would be that every aspect of the energy-balance equation (equation (1.1)) is measured directly and professionally rather than having to rely on the indirect estimates based on the statistically dubious models that are often used in practice. This, however, is rendered impossible by ethics and by the very nature of energy measurements - it is near impossible and highly unethical for anybody, particularly a research group, to have complete control over, or accurate knowledge of, the energy intake of children for any useful length of time. Unable to control accurately measure individual aspects of the energy-balance equation (equation (1.1) in section 1.1), we must settle for indirect estimates, particularly in terms of physical activity energy expenditure and energy intake. While it may be possible to improve upon these estimates with statistical modelling, significant residual error will unfortunately remain a feature of energy-balance research. Knowing, therefore, that the ideal study is an impossibility, we must then consider how to limit the error and yield reasonably accurate results, i.e. by considering the design and implementation of the best possible study.

One of the first things that should be considered when designing this study is who will be studied, for how long and by whom. We know that energy-balance differs over time and between the sexes, therefore the best study should be longitudinal in design with a male and female sample. In order to allow researchers insight into

pre-, peri- and post-pubertal processes, it is suggested that the participants are followed up from pregnancy until early to mid 20s.

It must be understood that a study of such a nature brings with it very large potential for dropout over time, resulting in missing data. While there do exist imputation methods for dealing with missing data, such as multiple imputation (see section 2.9), these methods rely on missing data being missing at random - which may not be the case for such data. The sample must therefore be large enough to ensure that statistical analyses retain adequate power after such dropout, though of course it should be noted that no-one can ever be sure that the complete cases are representative of the target population.

The aim of this study is to examine energy-balance in healthy UK adolescents. Therefore, any child with health concerns that may influence energy-balance should be excluded. In order to ensure a sample that is representative of the UK population today, participants should be of an ethnic and socio-economic mix. This raises further problems in that certain groups may be under-represented. It may therefore be necessary to over-sample from specific areas to ensure adequate representation of each sub-group from the population. It is likely that certain sub-groups of the population, possibly those from lower socioeconomic classes, will drop out more quickly or in larger numbers, which is another reason for over-sampling from such groups in the first place. Some adjustment of the final analysis, on a post-hoc stratification basis, might then be necessary.

While, as previously discussed, residual error in this field of research is inevitable, this error can be limited by ensuring that measurements are as accurate as possible.

As little as possible should be self-reported, as it has been shown that self-reported data are not consistent [172]. All variables should therefore be determined or measured according to agreed protocol by staff trained to a consistently high standard. All sites involved in any form of data collection must have identical equipment and reporting procedures must be standardised across such sites.

Anthropometric measures should be taken by trained staff at every occasion when the children are seen. These measures should include height, weight and possibly circumferences (waist / hip). To reduce measurement error, if possible, the same member of staff should perform all these measurements on a given child over the course of the study. These anthropometric measures would allow BMI standard deviation scores to be calculated in order to study the distribution of childhood body composition. BMI SDS should be calculated according to the 1990 growth reference data[30], in the absence of a more recent reference, using the LMS method as described in section 2.2.

To develop accurate models of body composition, an accurate method of determining the proportions of FM and FFM in the body would be needed, along with data from those methods for which we would like to develop models. While it is known that hydrodensitometry, a very accurate method of determining body composition, is perhaps not well tolerated by children, it has been suggested (see section 1.2.1.2.2) that air-displacement plethysmography is both accurate and safe for use with children. It would therefore be desirable to have the 'best possible' estimates of FM and FFM, determined using air-displacement plethysmography, taken as close as possible to carrying out BIA and DEXA scans. This would allow models for determining body composition from either BIA or DEXA, or both, to

be developed in conjunction with the best possible estimates of FM and FFM. It is important that these measurements are taken at regular intervals as children grow, to allow the effect of age to be modelled. Ideally, the data should be split into two datasets: one on which to develop models, and another on which to test such models.

In order to develop models for REE, as close as possible to anthropometric measurements and AP estimates of body composition being recorded, REE should be measured using calorimetry (as described in section 1.2.2.2.1). Physical activity energy expenditure cannot be directly measured - the best method of obtaining information on physical activity would be to have children wear accelerometers. This could be done for a period of time as close as possible to the REE and body composition measurements being taken. Penpraze et al. in 2006 recommend the use of accelerometry for 7 days, and for 10 hours on each day [246]. Energy intake, too, cannot be directly measured. Intake diaries could be used - however, it must be kept in mind that self-reported energy intake data are subject to error for reasons described in section 1.2.3.1.

This project was concerned with the study of energy-balance and body composition over puberty. Therefore, the ideal study in this field should include estimates of pubertal staging. For both sexes, Tanner stage (see section 1.2.5) should be recorded by clinicians. It has been shown that if clinicians are well trained and able to put the child at ease, the child is more likely to be comfortable with the process [201]. Age of menarche should be recorded for the girls. It is not possible to record this with certainty, as it must be self-reported. One possibility for limiting inaccuracies could be that at each clinic, girls could be asked if they have

had their first period yet. This would give interval censored estimates which may be more accurate than asking girls to recall their age of menarche at a much later point.

Along the course of the study, it may be of interest to take other measures, including measures from parents. These may include height and weight in order to assess the weight status of the parents and study relationships with the body composition of offspring. Unless height and weight are recorded in medical records before pregnancy, pre-pregnancy measurements would be self-reported and subject to recollection bias, and therefore not recommended. Once parents are recruited into the study, these measurements should be periodically taken by trained clinicians. It would not be necessary to take parental measures as often as child measures. Socio-economic status should also be recorded, as it may be of interest to study links between this and childhood body composition.

6.3 Simulation

A fundamental problem with research into energy-balance and body composition is the lack of real data, for example, the absence of gold standard body composition data. In the absence of real data, synthetic data may be used to illustrate how real data could be analysed. This would allow the effect of real-data issues, such as missing data, to be explored. The conclusions from such a simulation study may then influence practices in future research.

6.3.1 Generation of cross-sectional data

Using ALSPAC data at each of waves 7, 9 and 10, models for height based on age, and weight based on height and age were produced. These independent variables were chosen because they were measured in the ALSPAC data.

In the first simulation, cross-sectional data were generated for 10,000 male and 10,000 female subjects of each age 7, 8, 9 and 10. For each subject, height was generated as

$$height_i \overset{iid}{\sim} N(\mu, \sigma^2) \tag{6.1}$$

For ages 7, 9 and 10, the parameter values used in this simulation are based on those from the ALSPAC models at the same ages. For age 8, parameters were chosen to produce data that lay between age 7 data and age 9 data. Different parameters were used to generate heights for boys and girls, as shown in Table 6.1. Weight was then generated from the formula

$$weight_i = \beta_0 + \beta_1 \times height_i + height_i^2 \times e_i \tag{6.2}$$

with $e_i \overset{iid}{\sim} N(0, 1.5^2)$ and parameters (again, derived from measured data) as shown in Table 6.2. Note that height is in metres and weight is in kilograms.

	Age	μ	σ^2
Boys	7	1.280	0.054
	8	1.344	0.057
	9	1.393	0.061
	10	1.457	0.064
Girls	7	1.253	0.054
	8	1.315	0.060
	9	1.386	0.065
	10	1.435	0.070

Table 6.1: Parameter values used in the simulation of height data

	Age	β_0	β_1
Boys	7	-45.552	58.083
	8	-62.000	70.000
	9	-75.420	78.552
	10	-82.589	83.557
Girls	7	-49.925	60.467
	8	-65.000	72.000
	9	-79.478	82.220
	10	-58.101	85.827

Table 6.2: Parameter values used in the simulation of weight data

From these synthetic data, FFM was generated as:

$$FFM_i = \gamma_0 + \gamma_1 \times height_i \tag{6.3}$$

with Gaussian noise added.

The parameters used to generate FFM were found from ALSPAC data along with models for determining body composition from bioelectrical impedance (see section 1.2.1.2.2). These parameters are shown in Table 6.3. FM can then be considered to be the difference between weight and FFM.

	Age	γ_0	γ_1	σ^2
Boys	7	-37.95	46.84	2.00
	8	-44.51	49.96	2.00
	9	-57.49	59.41	2.00
	10	-64.85	63.03	2.00
Girls	7	-29.58	38.36	2.00
	8	-37.72	43.98	2.00
	9	-49.42	52.80	2.00
	10	-58.78	58.95	2.00

Table 6.3: Intercept (γ_0) and slope (γ_1) parameters used in the simulation of FFM data

As identified in section 4.2.3, the BIA models can result in negative values of FM. It could therefore be expected that any subsequent modelling with data generated from these models may have a similar problem. In order to limit this risk, ALSPAC cases with negative FM values were excluded when equation (6.3) was modelled. Investigation of the data simulated from this model revealed 47 (0.12%) negative values of FM out of 40,000 (10,000 at each age) for boys and 16 (<0.01%) out of 40,000 for girls.

Resting energy expenditure was simulated based on the Harris and Benedict models (see equations (1.16) and (1.17) in section 1.2.2.2.2) with Gaussian noise added as follows:

Boys

$$ree_i \sim N((66.4270 + 13.7516 \times weight_i + 500.33 \times height_i - 6.755 \times age_i), \sigma_B^2) \quad (6.4)$$

Girls

$$ree_i \sim N((655.0955 + 9.5634 \times weight_i + 184.96 \times height_i - 4.6756 \times age_i), \sigma_G^2) \quad (6.5)$$

where the values for σ_B^2 and σ_G^2 are shown in Table 6.4. These values were chosen by examining plots of REE as calculated from the simulated data using equations (6.4) and (6.5) with different values of σ_B^2 and σ_G^2, and selecting values that resulted in 'plausible' data.

Age	σ_B^2	σ_G^2
7	30^2	25^2
8	40^2	25^2
9	40^2	25^2
10	40^2	25^2

Table 6.4: Variances used to add Gaussian noise to simulated REE by gender and age

From equations (6.4) and (6.5), along with the parameters given in Table 6.1 and Table 6.2, it is possible to calculate the resulting expected values of REE for each sex and age. These are shown in Table 6.5.

Age	Boys	Girls
7	1055.531	1101.241
8	1125.982	1144.755
9	1170.186	1200.077
10	1266.282	1237.747

Table 6.5: Expected values of REE (kcal/day) by age and gender

These expected values are reasonably similar to the mean values of REE predicted for children of these ages, using the Harris and Benedict models, in the ALSPAC data (see Table 5.3 in section 5.3.2).

The purpose of the cross-sectional simulation was to be confident that the models used to generate data would yield plausible results, before simulating longitudinal data. Plots of the ALSPAC data and the simulated data reveal that the simulated data are certainly not implausible. For one example of such a plot, see Figure 6.1.

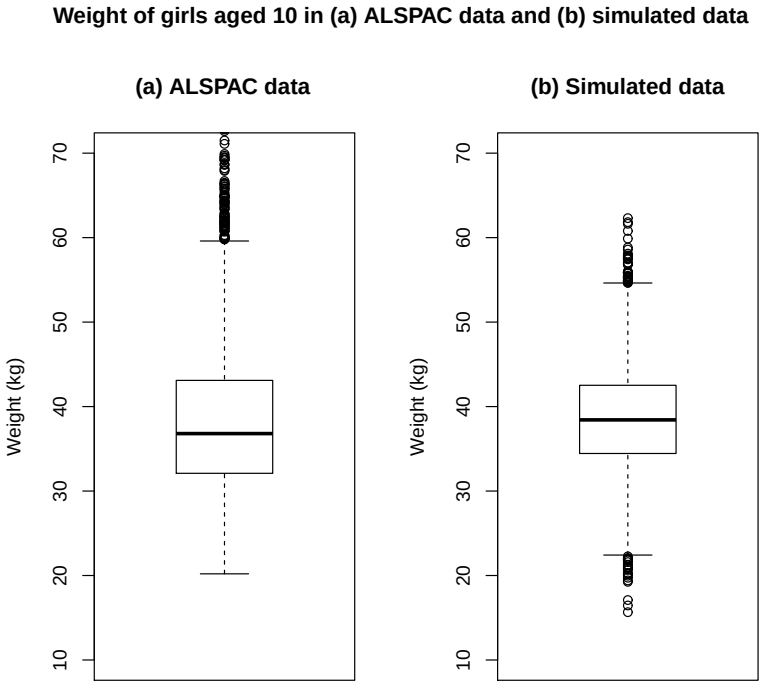

Figure 6.1: Measured height (cm) from ALSPAC data and simulated height (cm) for girls aged 10

The models discussed in this section will now be used to generate longitudinal data.

6.3.2 Generation of longitudinal data

In a second simulation study, longitudinal height data were created for both boys and girls from a linear mixed effects model produced from the ALSPAC data with height as the dependent variable and age as the independent variable. These data were created for children of ages 7, 8, 9 and 10. Heights were generated based on the linear mixed effects (random coefficient) regression model:

$$Height_{ij} = B_0 + B_1 \times Age_{ij} + b_{0ij} + b_{1ij} \times Age_{ij} + e_i \tag{6.6}$$

for the ith child at the jth age. The coefficient and error terms are independent for both boys and girls, and $e_i \sim N(0, \sigma^2)$.

The data were created using the following parameters:

Boys

$$\mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 15.09 & -0.71 \\ -0.71 & 0.23 \end{pmatrix} \right)$$

Girls

$$\mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 26.10 & -0.52 \\ -0.52 & 0.51 \end{pmatrix} \right)$$

where

- n is number of subjects,
- p is number of time points,
- e_i (1.08 for boys, 1.26 for girls) is the variance of the error,
- B_0 (93.7 for boys, 90.2 for girls) is the fixed intercept effect (average group

intercept),

- B_1 (5.7 for boys, 6.0 for girls) is the fixed slope effect (average group slope),
- $\text{Var}(b_0)$ (15.09 for boys, 26.10 for girls) is the variance of individual intercepts,
- $\text{Var}(b_1)$ (0.23 for boys, 0.51 for girls) is the variance of individual slopes,

Upon examining the simulated data, it was found that, for some children, height decreased over time. Rejection sampling was therefore used to ensure monotonicity. 10000 cases for girls and 10000 cases for boys were generated.

From these heights along with age (where age is 7, 8, 9 or 10), weight was created using the models described in Table 6.2. From these variables, FM, FFM and REE were simulated as described in section 6.3.1. The resulting REE is shown in Figure 6.2.

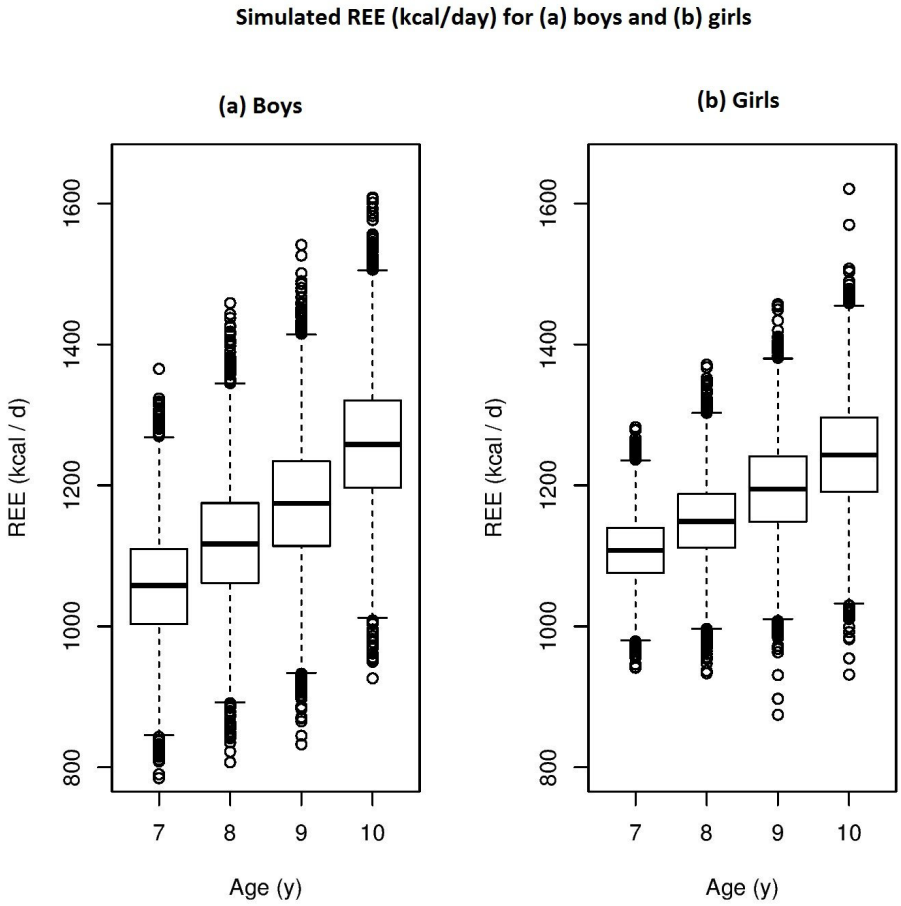

Figure 6.2: Simulated REE (kcal / day) for ages 7, 8, 9 and 10 for (a) boys and (b) girls

Linear mixed models were fitted to the simulated data with REE as the dependent variable, FFM and age as fixed effects and subject as a random effect. A random intercepts model was used. Resulting models were as follows:

Boys

$$REE_i = 345.570 + 25.315 \times FFM_i + 22.996 \times age_i \tag{6.7}$$

with the variance component estimate of the intercept being 4307.1.

Girls

$$REE_i = 776.893 + 17.897 \times FM_i - 0.047 \times age_i \tag{6.8}$$

with the variance component estimate of the intercept being 2016.2.

Fitted values from these models are compared to the simulated REE in Figure 6.3. This figure shows that the fitted model for REE is in reasonable agreement with the (simulated) REE.

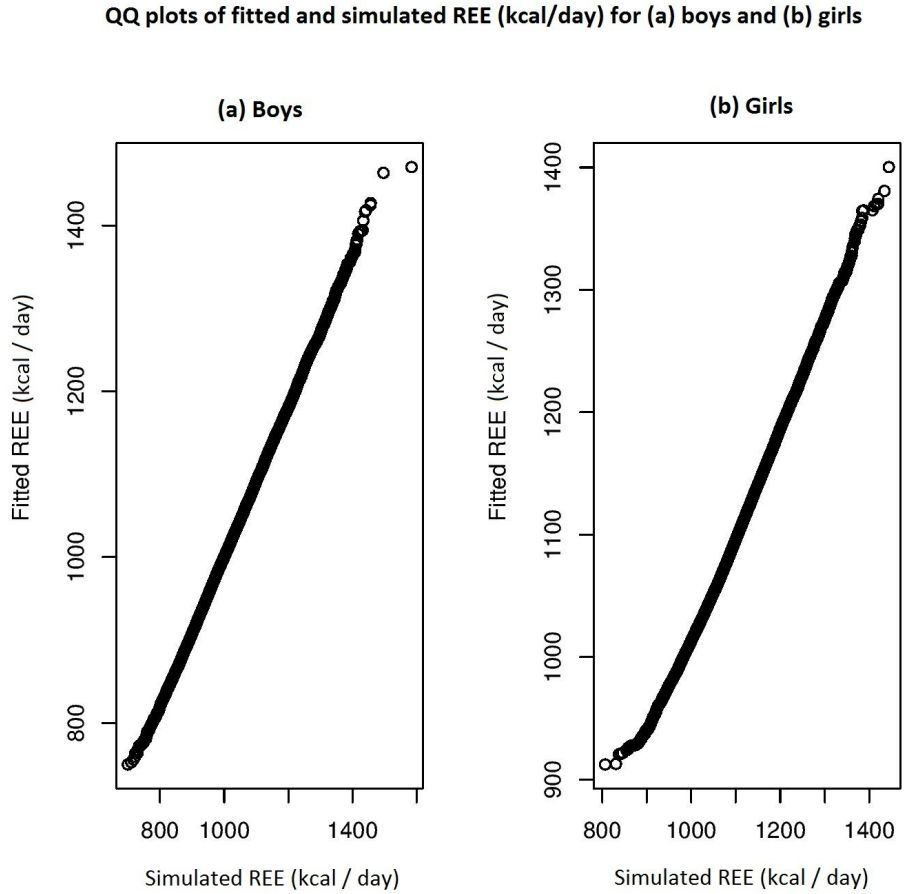

Figure 6.3: Plot of simulated and fitted values of REE (kcal/day) for (a) boys and (b) girls

6.3.3 The effect of real-life data problems

6.3.3.1 Sample size

100 samples of data were simulated for both boys and girls using the method described in section 6.3.2. For each sample, at each age 7, 8, 9 and 10, the following linear model was fitted:

$$REE_i = \alpha + \beta \times FFM_i \tag{6.9}$$

Let the mean REE estimated for age j ($j = 7, ..., 10$) from simulated sample k ($k = 1, ..., 100$) be $\overline{REE}_{kj}$. The mean square error can be calculated from the expected values of REE given in table 6.5 as:

$$MSE(\overline{REE}_j) = \frac{\sum_{k=1}^{100}(\overline{REE}_{kj} - E(REE_j))^2}{100} \tag{6.10}$$

This was calculated for 100 samples of each size: 50, 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 and 10000, for both boys and girls. Results are given in Figure 6.4 and Figure 6.5 for boys and girls, respectively. The sample size shown on these graphs has been restricted to [0,3000] to allow us to focus the smaller sample sizes.

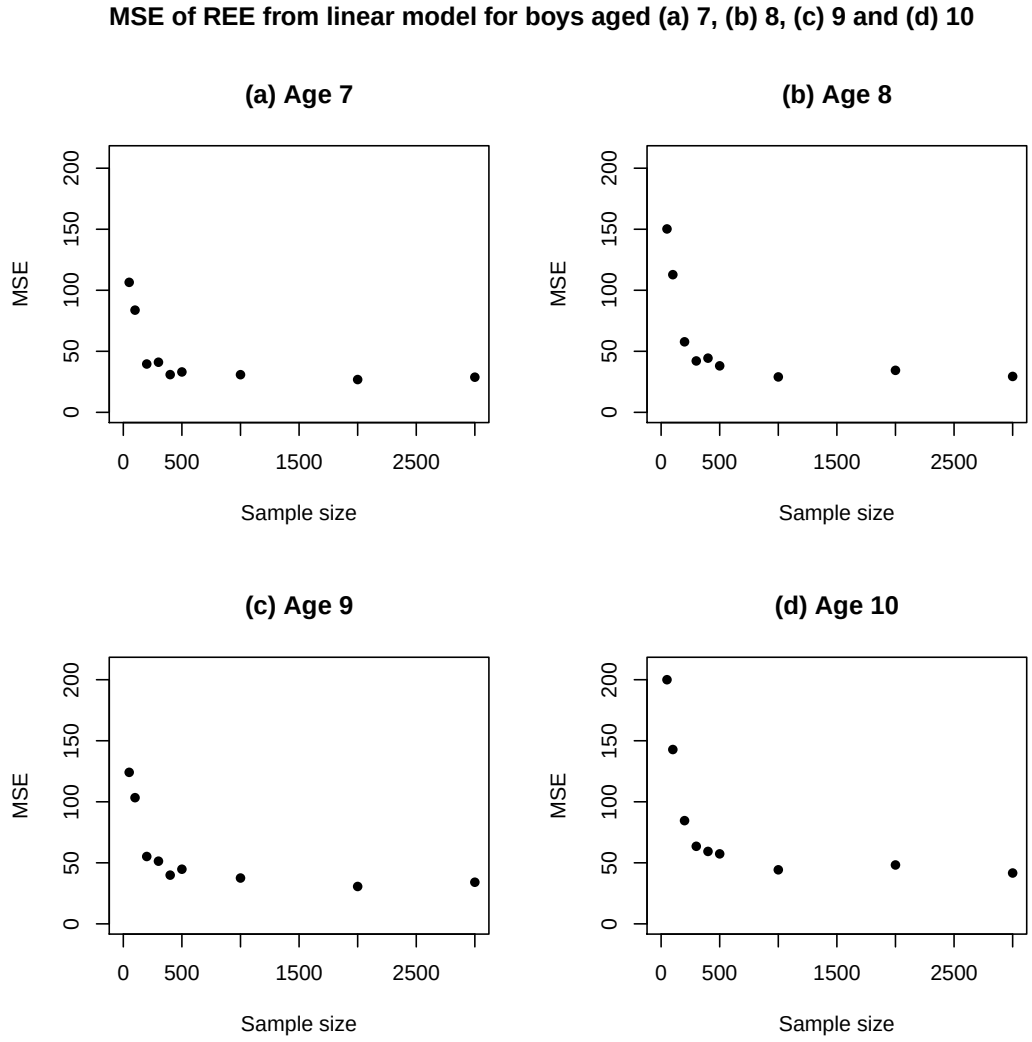

Figure 6.4: MSE of $\overline{REE}_j$ $(j = 7, ..., 10)$ from linear model for boys aged (a) 7, (b) 8, (c) 9 and (d) 10

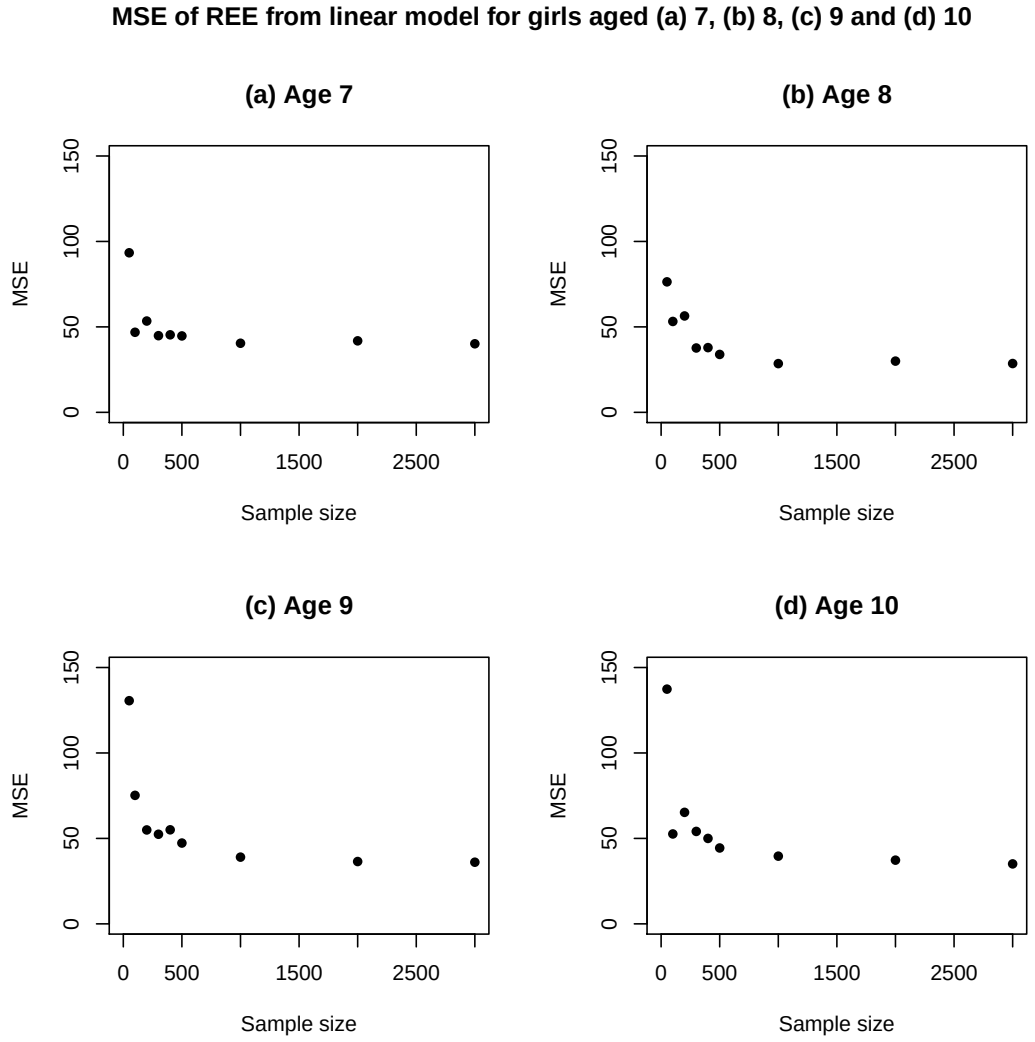

Figure 6.5: MSE of $\overline{REE}_j$ $(j = 7, ..., 10)$ from linear model for girls aged (a) 7, (b) 8, (c) 9 and (d) 10

These plots appear to suggest that, with complete data, a sample size of around $n = 1000$ is required to reduce the MSE in such modelling to an acceptable level.

From Tables 6.1 and 6.2, along with the linear functions used to simulate FFM (equation 6.3) and REE (equations 6.4 and 6.5), the true values of α and β for the relationship shown in equation (6.9) were calculated. These are shown in Table 6.6.

Age	Boys		Girls	
	α	β	α	β
7	445.237	27.734	733.452	19.897
8	463.373	29.823	745.189	19.861
9	497.952	26.604	762.023	18.395
10	560.146	26.168	797.342	17.061

Table 6.6: True values of α and β from equation (6.9) by age and gender

It is of interest to determine whether or not sample size has a significant effect on results. With the sample size varied between 50 and 10000 cases as before, we can determine what percentage (of 100 samples of each sample size) of linear model 95% confidence intervals contain the true values shown in Table 6.6. Results are shown in Figure 6.6 for boys and Figure 6.7 for girls.

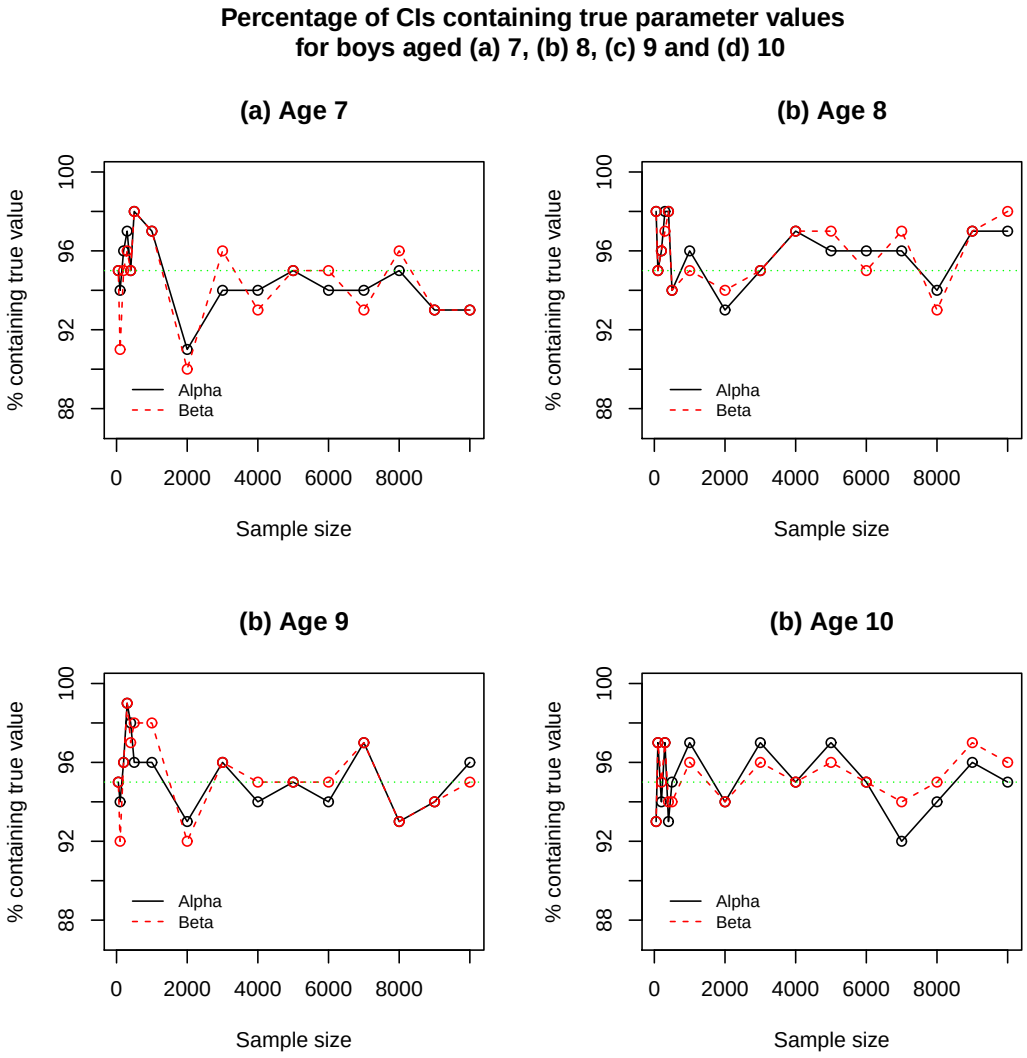

Figure 6.6: Percentage of 95% confidence intervals containing true values of α and β for boys aged (a) 7, (b) 8, (c) 9 and (d) 10

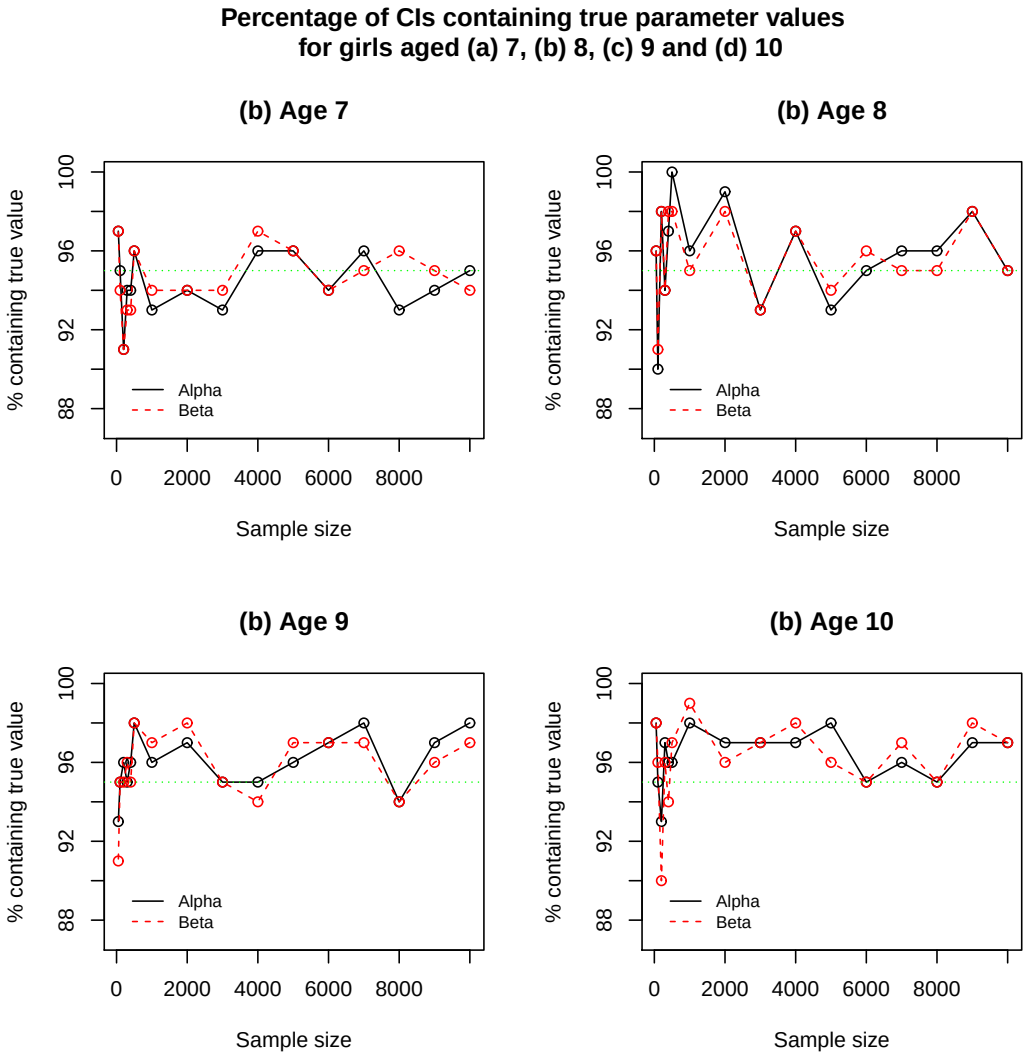

Figure 6.7: Percentage of 95% confidence intervals containing true values of α and β for girls aged (a) 7, (b) 8, (c) 9 and (d) 10

By the very definition of 95% confidence intervals, regardless of sample size, on average 95% of the confidence intervals should contain the true value. Indeed this is what we appear to see: for some of these samples, coverage probability is lower than 95%, for others higher - even at small sample sizes. It is possible that any pattern here is obscured by the relatively small number of simulations (100).

Another method of assessing the accuracy of models is the MSE. The MSEs of the parameters in equation (6.9) are calculated as:

$$MSE(\hat{\alpha}_j) = \frac{\sum_{k=1}^{100}(\hat{\alpha}_{kj} - \alpha_j)^2}{100} \tag{6.11}$$

$$MSE(\hat{\beta}_j) = \frac{\sum_{k=1}^{100}(\hat{\beta}_{kj} - \beta_j)^2}{100} \tag{6.12}$$

where each $\hat{\alpha}_{kj}$ and $\hat{\beta}_{kj}$ are parameter estimates for the model shown in equation (6.9) when fitted to the kth simulated sample ($k = 1, ..., 100$) at age j (j=7,...,10), and the true values of α_j and β_j are as shown in Table 6.6. These MSEs have been calculated for each sample size of 50, 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 and 10000. Figures 6.8 and 6.9 show the mean square error of $\hat{\alpha}_j$ and $\hat{\beta}_j$, respectively, for boys as the sample size changes. As before, the sample size shown on these graphs has been restricted to [0,3000] to allow us to better consider the smaller sample sizes.

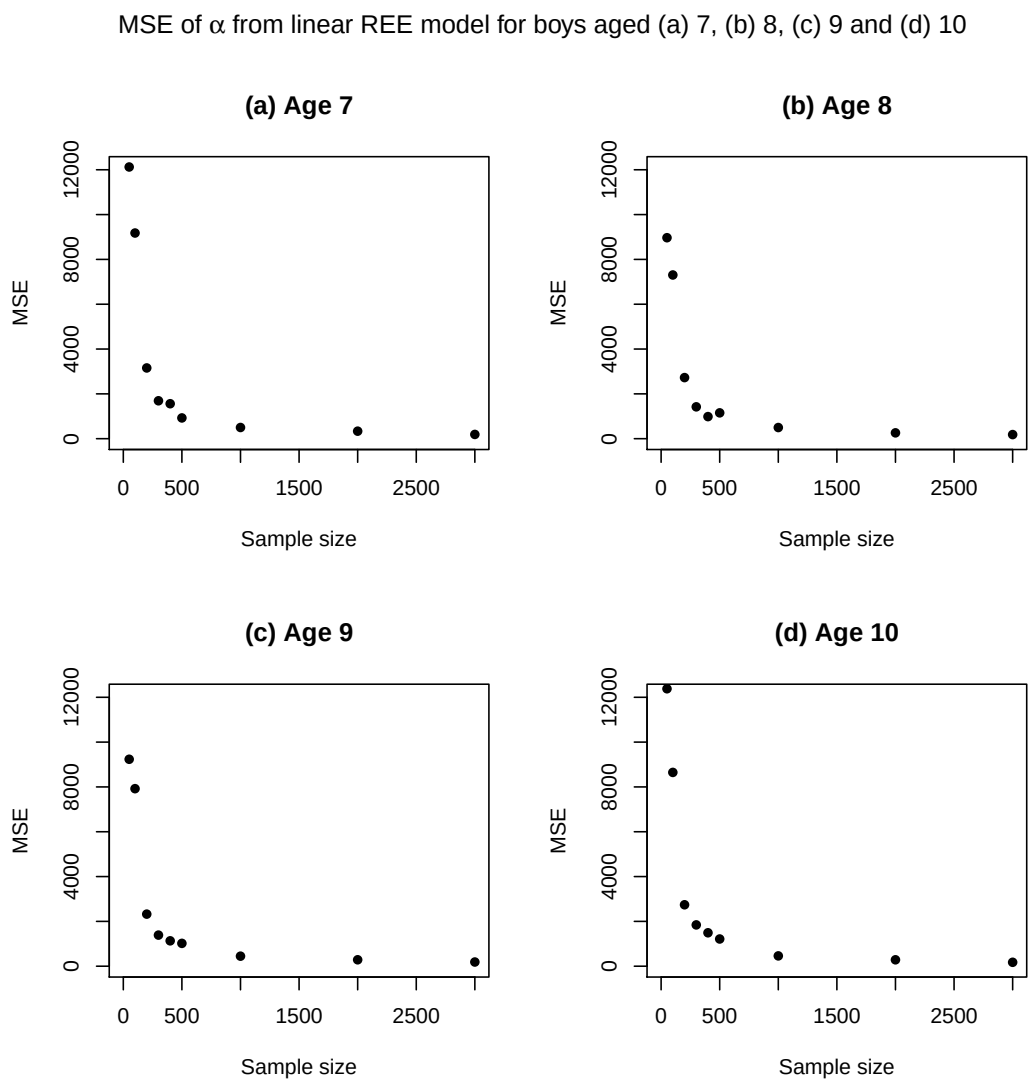

Figure 6.8: MSE of $\hat{\alpha}_j$ in linear REE model for various sample sizes for boys aged (a) 7, (b) 8, (b) 9 and (d) 10

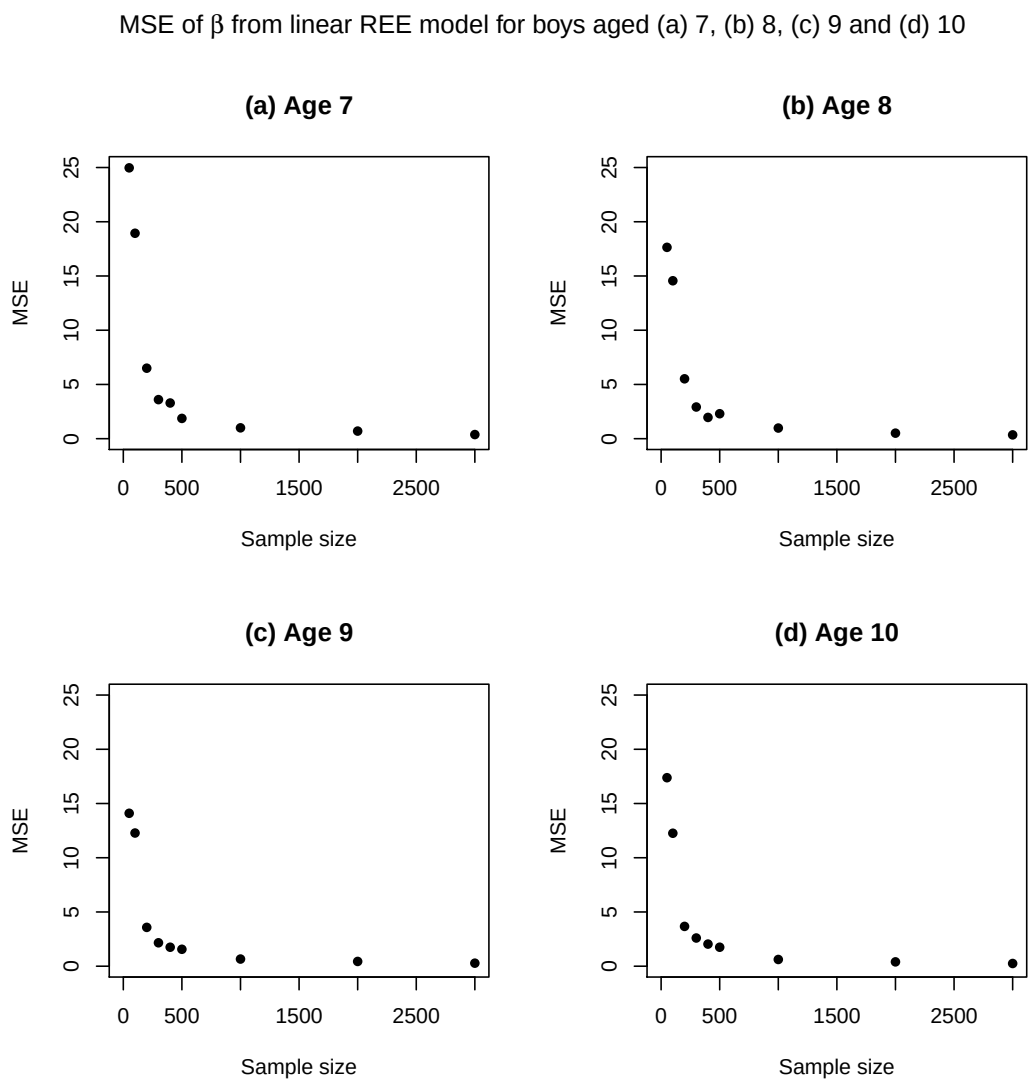

Figure 6.9: MSE of $\hat{\beta}_j$ in linear REE model for various sample sizes for boys aged (a) 7, (b) 8, (c) 9 and (d) 10

A very similar picture was seen for the MSEs of both parameters for girls (not shown). These plots indicate that acceptable MSE can be achieved with quite a small sample size (n around 1000), as long as each observation is complete. However, the complete data generated here are not representative of real-world data where, of course, observations may be missing values for individual variables.

6.3.3.2 Missing data

In order to investigate the effect of missing data points on the analysis, samples were simulated, using the method described in section 6.3.2, with between 10% and 90% of observations randomly removed from FFM and REE. For each number of points removed, 100 samples were simulated.

It may be possible that, in large studies, complete data may be obtained for a much smaller sample of cases. This data could then be used to develop models to predict variables, and these models used to impute the missing values in the larger sample. To investigate the effect of this type of imputation, the following model was developed on two complete samples of size $n = 50$ (one for boys, one for girls):

$$FFM_i = \alpha + \beta_1 \times height_i + \beta_2 \times weight_i \tag{6.13}$$

These models were then applied to the cases in the full samples ($n = 10000$) where FFM was missing. This method will be referred to throughout this chapter as 'OLS imputation'.

As discussed in section 2.9, multiple imputation by chained equations (MICE) is another method of imputing missing data, assuming the data are missing completely at random (MCAR) or missing at random (MAR).

The assumption of random missingness can be subjectively checked using margin plots, a feature of R's "VIM" package [247]. Details can be found in the description of MICE in the Journal of Statistical Software [240]. It should be noted that this is a subjective check, rather than a definitive result. One example of such a plot for this simulation, for age 7 data, is shown in Figure 6.10. For each variable, the outer boxplot represents the marginal distribution of the variable for the complete cases. The inner boxplot represents the distribution of the variable for cases who have a missing value in the other variable. If these distributions are similar, as seen in Figure 6.10, it is likely (though not guaranteed!) that the assumption holds.

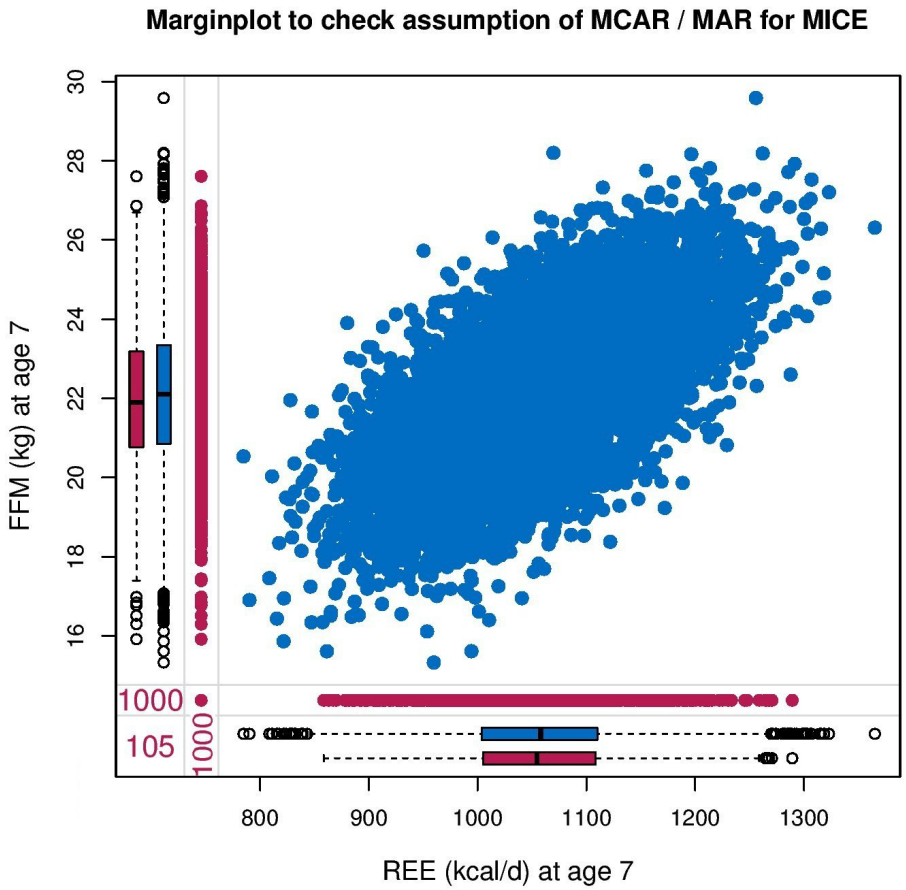

Figure 6.10: Margin plot to check the assumption of random missingness in the age 7 simulated data

MICE has been implemented for the simulated data with varying amounts of missing data, using 5 iterations and 5 imputed datasets per imputation. The imputation model used was predictive mean matching (PMM). Only missing values in the independent variable (FFM) were imputed, missing values in the dependent variable (REE) were not imputed at any point.

Mean square error, as shown in equations (6.11) and (6.12), was calculated from 100 samples of each amount of missing data - first with a complete case analysis, then with data imputed using each of OLS imputation and MICE. These results, presented only for boys, are shown in Figures 6.11 and 6.12 for $\hat{\alpha}_j$ and $\hat{\beta}_j$, respectively. Similar results were found for girls.

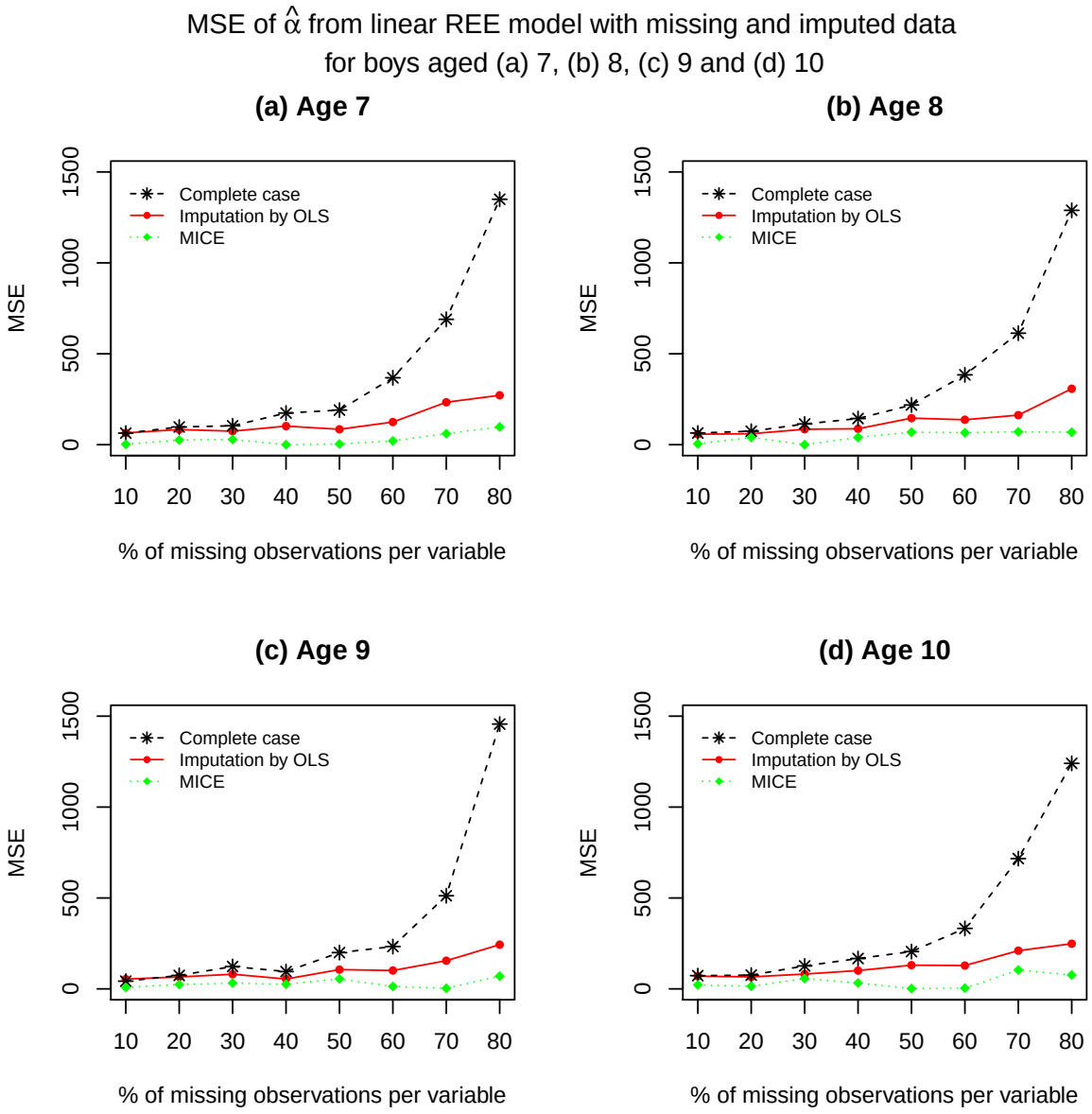

Figure 6.11: MSE of $\hat{\alpha}_j$ $(j = 7, ..., 10)$ in linear REE model for various amounts of missing data for boys aged (a) 7, (b) 8, (c) 9 and (d) 10 (with missing and imputed data)

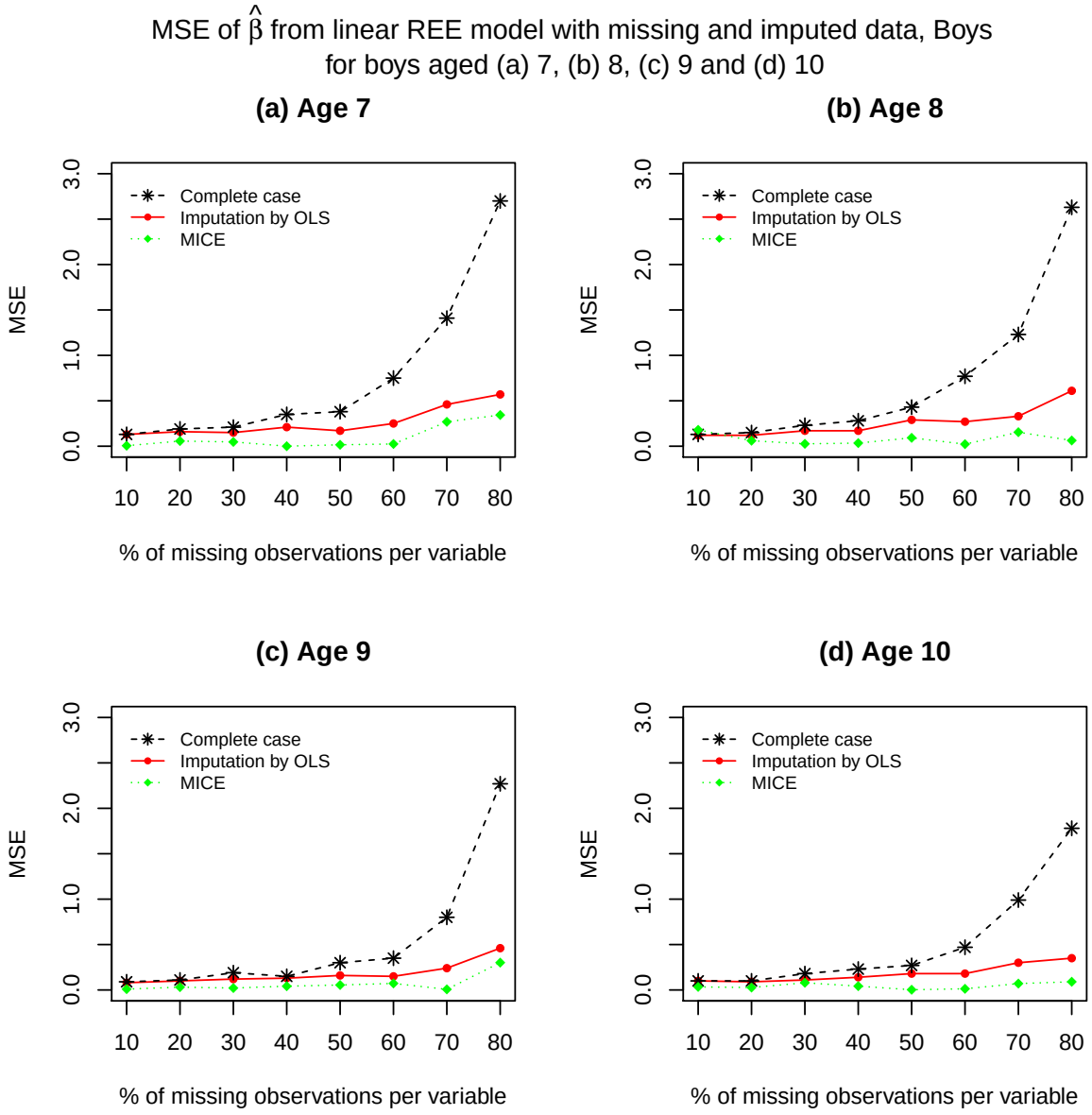

Figure 6.12: MSE of $\hat{\beta}_j$ $(j = 7, ..., 10)$ in linear REE model for various amounts of missing data for boys aged (a) 7, (b) 8, (c) 9 and (d) 10 (with missing and imputed data)

From these plots, it is clear that, with a complete-case analysis, MSE increases as the percentage of missing data increases. Both methods of imputation resulted in greatly improved parameter estimates (compared to complete-case analysis) for

large amounts of missing data, while the improvement for a small fraction of missing data was slight (as MSE for the complete-case analysis was low with these datasets).

Each method of imputation has advantages and disadvantages - the OLS method takes far less computational time than MICE, particularly as the fraction of missing data increases, but results in a lesser improvement of estimates. It must also be kept in mind that if a smaller, complete-case sample of data does exist, it is likely to be a biased sample. On the other hand, MICE is considerably slower as it involves MCMC simulation (between several hours and several days, depending on the sample size and fraction of missing data, compared to mere seconds for the OLS method) but results in improved parameter estimates and is widely regarded as one of the best tools for dealing with missing data [239, 248, 249, 237]. Each of these things, along with careful consideration of the methods' assumptions (and, in the case of the OLS method, the availability or lack thereof of a smaller, complete sample), should play a part in deciding how to deal with missing data.

6.4 Chapter summary

This chapter began by exploring, using parameters estimated from the ALSPAC data, relationships between age, height, weight, REE, FM and FFM in children aged 7 - 10, with a view to generating longitudinal data that are reasonably realistic. Firstly, cross-sectional data were generated, in order to check the reasonability of the data generated by these models.

Longitudinal height data were then created from a linear mixed effects model for 10,000 boys and 10,000 girls at each age 7, 8, 9 and 10, and the models for REE and body composition were applied to these data (ensuring as much as possible that, at each stage, the data generated were reasonable). These data were simulated 100 times as complete datasets.

To allow consideration of real-world data problems, the sample size was reduced by varying amounts, and 100 datasets of each sample size were generated for both boys and girls. Finally in the data creation process, varying amounts of data were randomly removed from the dependent and independent variables of a linear model, with a view to exploring techniques for dealing with missing data.

This chapter has shown that it is possible to generate reasonably realistic longitudinal data in the field of resting energy expenditure and body composition, using a combination of published models and models developed from measured data.

The effect of sample size was considered, and it was shown that reasonably low modelling error could be achieved with a sample size of around $n = 1000$ (of each sex), assuming data are complete.

In considering real-data issues, it was shown that with complete data, as sample size increases, error in parameter estimates is greatly reduced. Sample sizes up to $n = 10,000$ were considered, but there was little improvement in MSE observed with $n = 10,000$ compared to that observed when $n = 2,000$.

Finally, looking at missing data, this chapter has demonstrated that as the fraction of missing data increases, so does error in parameter estimates. Two methods of dealing with missing data were considered:

- imputing the independent variable(s) using a linear regression model developed on a much smaller, complete dataset
- imputing the independent variable(s) using multiple imputation by chained equations, assuming that the data are missing at random.

Both methods resulted in a considerable improvement in parameter estimates, particularly when the fraction of missing data was large. Of course, these methods are not the only methods of imputation available to researchers. When deciding how to deal with missing data, consideration must be given to the complexity and computational time of each possible method, along with the underlying assumptions.

Chapter 7

Quantifying energy-imbalance

7.1 Chapter aims

One of the original aims of this project was to model energy-imbalance over time and subsequently determine whether ALSPAC participants could be categorised by a small or large degree of energy-imbalance over the transition into puberty.

Without direct measurements or reasonable estimates of the components of the energy-balance equation, it will not be possible to directly determine this with accuracy. However, as a consequence of positive energy-imbalance resulting in weight gain, when we have estimates of FM and FFM gain over time, as well as of the energy cost of depositing each type of mass, it is possible to retrospectively estimate the degree of energy-imbalance that has occurred over the time period.

While BIA, with its current underlying models (as detailed in section 1.2.1.2.2), was critiqued in Chapter 4 as being inaccurate on the individual level, it should be noted that these models are the best available published models at this time. As such, for the purposes of this chapter, these models and the arising body composition estimates will be considered to be reasonable on the population level.

This chapter will therefore use these estimates of FM and FFM derived from BIA for the ALSPAC participants at focus groups 9 and 11, with an aim to determining the magnitude of energy-imbalance over this period and estimating the average daily energy-imbalance required between ages 9 and 11 for children to get to (or remain at) a healthy weight at age 11 according to BMI SDS.

7.2 Quantifying energy-imbalance retrospectively

During infancy, childhood and adolescence, people must gain some weight - this is a natural and normal part of maturation towards adulthood. However, with more and more children becoming overweight or obese, it is clear that many of today's children are gaining more weight than is necessary for normal growth. It is therefore of interest to develop ways of being able to determine what degree of energy-imbalance is 'just right' for healthy growth. In keeping with the rest of this project, we will consider boys and girls separately.

In order to directly measure the degree of energy-imbalance over a time period, we would require data on almost every component of energy-balance to have been collected very routinely throughout that period, namely: resting energy expenditure, physical activity energy expenditure and energy intake. This has not been done and is unlikely to be done - it is almost impossible and highly unethical to have complete control over a child's energy intake for a sustained period of time, for example.

In the absence of this data, we must work retrospectively. From impedance data, we are able to calculate FM and FFM estimates for each child. This chapter will consider waves 9 and 11. Therefore, we can estimate the gain over the roughly two-year period for each child in terms of both FM and FFM. In 1995, Westerterp published the energy costs of storing each type of mass [1] [250], allowing us, from the FM and FFM estimates, to estimate roughly how much energy-imbalance oc-

[1] 9076 kcal / kg for FM and 1433 kcal / kg for FFM

curred over the time period for each child. Additionally, since we can calculate the number of days between focus visits 9 and 11 for each child, we can estimate the average imbalance per day over this period. With the ability to calculate age- and sex-specific BMI SDS for the participants at both time periods, we can estimate the energy-imbalance required for transition from one degree of fatness to another - for example, the amount of calories needed for a child who is overweight at 9 to be at a healthy weight at 11.

In this analysis, we consider only those subjects for whom we have impedance data at both waves 9 and 11. Further to this, we exclude any subject for whom the current BIA models give a negative value of FM (as discussed in section 4.2.3). Therefore, subject numbers for this section are: 2979 boys and 3072 girls.

We first explore this data in order to determine how the 'fatness' of these children changed over time with BMI SDS as an estimation of fatness. Recall the cut-off points for 'fatness' classification as shown in Table 2.1 in section 2.2. Table 7.1 indicates that children remain, for the most part, 'stuck' in one category of fatness. Certainly, very few subjects transition by two categories - that is, for example, very few normal weight children at 9 have become obese by 11 and very few who were obese at 9 have become normal weight by 11.

Sex	Status at 11	Status at 9			
		Underweight	Normal weight	Overweight	Obese
Boys	Underweight	70	34	0	0
	Normal weight	24	1897	61	4
	Overweight	0	190	160	45
	Obese	0	32	106	356
Girls	Underweight	79	75	0	0
	Normal weight	31	1937	97	5
	Overweight	0	158	186	63
	Obese	0	23	107	311

Table 7.1: Transition of 'fatness' in ALSPAC ages 9 to 11 (only those subjects for whom we have impedance at both 9 and 11)

This message is not a new one to this thesis (it was uncovered earlier with the full sample, see section 3.3.3) but is nonetheless important enough to be re-stated. It would appear that the key to young people progressing through adolescence at a healthy weight is for them to be at a healthy weight prior to their adolescent years.

Incorporating estimated daily energy-imbalance, as described above, with this change in BMI SDS over time tells a very interesting tale indeed. It is common to hear parents say 'he / she is a growing boy / girl' while handing out calorie-laden snacks, justifying the unhealthy treats with the fact that the child needs extra energy to grow. Without in-depth research and public education, this misconception may not be the fault of the parents but, as Figure 7.1 and 7.2 show, is in fact incorrect and therefore potentially damaging to the long-term health of the children.

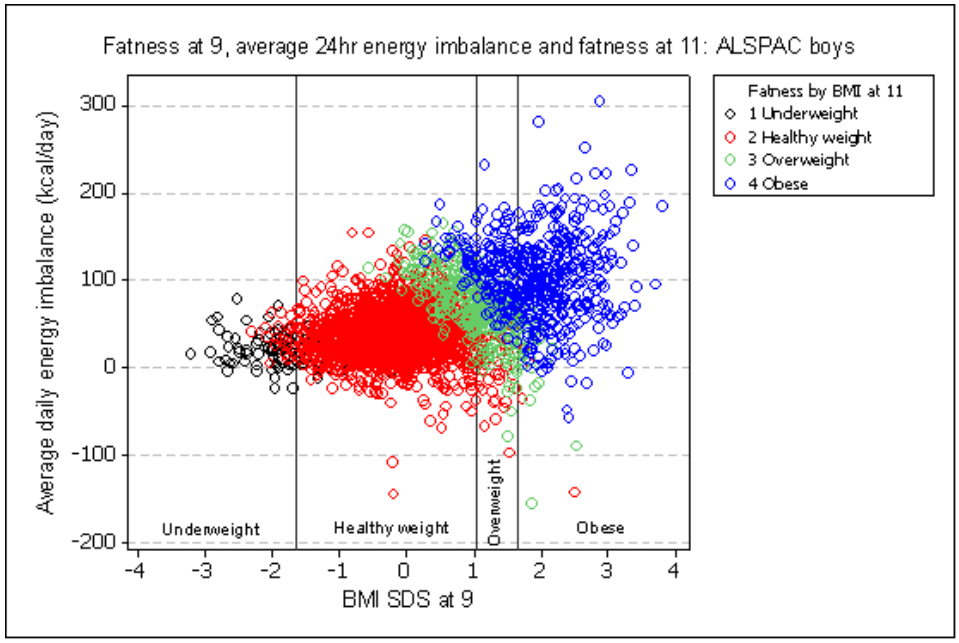

Figure 7.1: Average daily energy imbalance (kcal/day) between ages 9 and 11 against BMI SDS at 9, grouped by BMI SDS at 11, ALSPAC boys

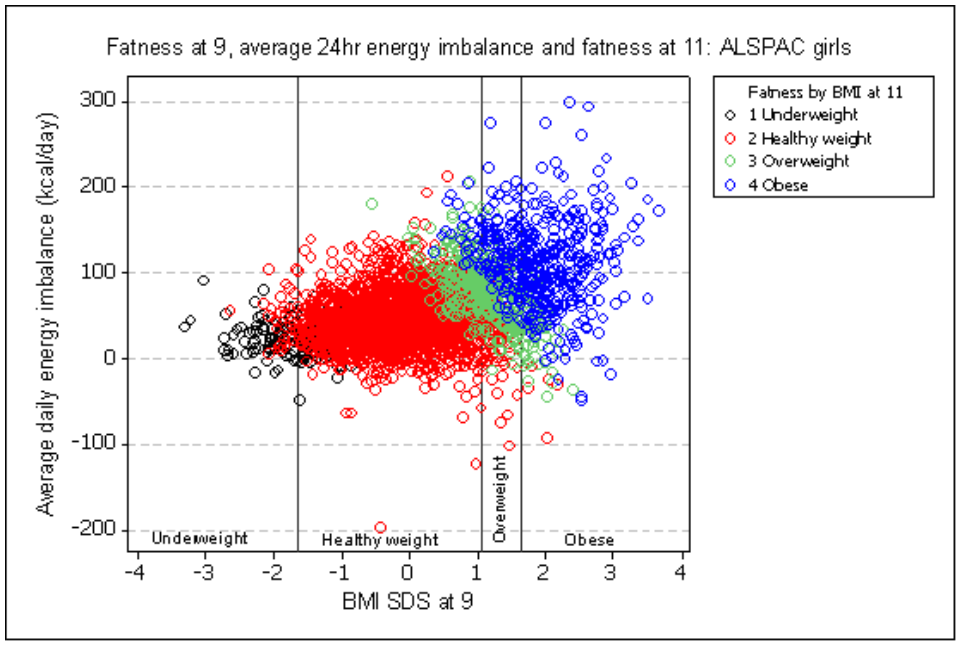

Figure 7.2: Average daily energy imbalance (kcal/day) between ages 9 and 11 against BMI SDS at 9, grouped by BMI SDS at 11, ALSPAC girls

Figures 7.1 and 7.2 raise some very interesting points:

- Looking at those who are 'above average' but not overweight at 9 (i.e. SDS between 0 and 1.04), it is clear that as BMI SDS tends towards overweight, the degree of average daily energy-imbalance resulting in remaining healthy at 11 decreases considerably. A boy with a standardised BMI just above average would appear to remain healthy with an imbalance of up to 150 kcal/day on average over this period (though presumably those with an intake at the high end would be closer to overweight than average at 11) while those boys with an SDS of 1 at 9 - still in the healthy bracket - are likely to become overweight by 11 with an average daily imbalance of just 50 kcal/day.
- Some children, both boys and girls, with the highest BMI SDS at 9 and the highest average daily energy-imbalance appear to be overweight rather than obese at age 11. This is a surprising result and could possibly be explained by the additional energy cost of daily life when obese. It is intriguing though why this should be the case when others, with a lower energy-imbalance, remain obese.

7.2.1 How much imbalance is healthy?

A key question of this research is: how much energy-imbalance is required in order for children to remain in - or get to - a healthy weight as they enter puberty? Table 7.2 shows the average daily imbalance for those children at each weight status at 9 and at a healthy weight at 11. Note that due to small sample sizes in some combinations (obese at 9 and normal weight at 11, for example), the point estimates and confidence intervals have resulted from nonparametric one-sample tests and

are therefore estimates of the medians. With only 9 subjects in total having been obese at 9 and healthy weight at 11, the estimates for these groups are extremely tentative but are included for completeness.

Imbalance required for healthy weight at 11				
Status at 9	Sex	n	Median	95% CI
Underweight	Boys	24	34.9	(27.4, 40.7)
	Girls	31	45.7	(35.8, 55.9)
Healthy weight	Boys	1897	38.3	(37.1, 39.6)
	Girls	1937	45.2	(43.9, 46.5)
Overweight	Boys	61	11.9	(3.3, 20.0)
	Girls	97	27.2	(21.0, 33.5)
Obese	Boys	4	-15.2	(-142.5, 21.1)
	Girls	5	-31.4	(-93.2, -9.9)

Table 7.2: Median energy-imbalance (kcal/day) required to transition from each 'fatness' stage at 9 to healthy weight at 11.

Table 7.2 is, in conjunction with Figures 7.1 and 7.2, is very eye-opening. We can see, for example, that the estimated median energy-imbalance for maintaining a healthy weight in the early stages of puberty is 37 to 40 kcal per day on average for boys over the two year period and 44 to 47 kcal per day on average for girls. This is a tiny amount, especially when we consider the energy content of snacks aimed at children.

7.2.2 How much is too much?

In addition to knowing how much imbalance leads 9 year olds to a healthy weight at 11, it is also of interest to know how much imbalance is associated with being normal weight at 9 years old and overweight at 11 years old. This is shown, by means of nonparametric one-sample tests for consistency with the previous section, in Table 7.3.

Imbalance resulting in overweight at 11			
Sex	n	Median	95% CI
Boys	190	93.0	(88.5, 97.4)
Girls	158	96.3	(91.4, 101.6)

Table 7.3: Median energy-imbalance (kcal/day) resulting in transition from healthy weight at 9 to overweight at 11.

Similarly to Table 7.2, Table 7.3 is perhaps very surprising, suggesting that normal-weight 9-year olds who eat around 90 - 100 kcal more than they expend per day, on average, are likely to become overweight by age 11.

7.3 An important take-home message

In an age when children are constantly faced with high-energy and easily available foods, often coupled with the fallacy that growing children require a very high degree of energy-imbalance for growth it is very important to bring the message uncovered in this chapter to the attention of the public:

According to this research, healthy pre- or early-pubertal children, in order to remain healthy, require less than 50 kcal per day, on average, of 'excess' energy. An excess of 100 kcal per day- less than they would get from, for example, a small packet of crisps or similar snack - is likely to lead to that child becoming overweight before their teenage years and at significantly increased risk of adult obesity and co-morbidities such as diabetes and cardiovascular disease.

7.4 Chapter summary

This chapter has retrospectively explored energy-imbalance, as indicated by weight gain over time, and its relationship with fatness, as indicated by age- and sex-standardised BMI, in ALSPAC participants over the ages 9 to 11.

The overwhelming result from this chapter has been that while most children do require a positive energy imbalance over this time period, the reality seems to be that the required imbalance is relatively small.

It is estimated in this chapter that a 9 year boy of healthy weight is likely to be of a healthy weight at age 11 with an average daily energy-imbalance of just 37 - 40 kcals/day over those two years (for girls the estimation is 44 - 47 kcals/day), while a child who is overweight at 9 requires an imbalance of 30 or fewer kcals/day to be at a healthy weight by age 11. Conversely, a healthy weight 9 year old who consumes 100 kcals/day (or fewer as BMI SDS tends towards overweight) more than he or she expends is at risk of being overweight by 11.

In an age and culture where energy-dense foods are very readily available, sedentary activities replace exercise and a common opinion is that children require a moderate to high energy-imbalance to grow, the results presented in this chapter are both startling and highly important. Given that entering puberty overweight or obese is known to be potentially detrimental to long-term physical and mental health, people must become more aware that pre- or early- pubescent children do not, unless very active, require a high energy intake for normal, healthy growth.

Chapter 8

Modelling during puberty

8.1 Chapter aims

In the late stages of this project, data from ALSPAC's 'Growing and Changing' questionnaires was made available. This chapter begins by exploring the distribution of the pubertal status measures (Tanner stage and age of first menstrual period) at each questionnaire 'wave' (corresponding roughly to ages 9, 10, 11, 13 and 14). In addition to the distribution of the pubertal staging, this chapter will consider the consistency of these measurements over time, since all of the data in these questionnaires was self-reported.

Following on from this, the main focus of this chapter will be to investigate any associations arising between self-reported pubertal stage and body composition. Without gold-standard measures, the estimates of body composition used in this chapter will be:

- BMI standard deviation scores,
- Fat mass index, calculated from FM and height.

8.2 Measures of pubertal stage in ALSPAC

At each wave (9, 10, 11, 13 and 14), ALSPAC respondents were asked, separately from 'focus' visits, to complete and return a postal questionnaire entitled 'Growing and Changing' [251]. We were granted access to some of these data in the late stages of this project. It should be noted that, as a consequence of how it was collected, ALL variables were self-reported - a feature to be critiqued later in this chapter (from p 224). In particular, we are interested in working with two variables from these questionnaires: Tanner stage and age of menarche.

8.2.1 Tanner stage

One measure of pubertal status is self- (or parentally-) reported Tanner stage, as discussed briefly in section 1.2.5.

The Tanner stages (as described in the ALSPAC SPSS labels) are as follows:

Boys

- Stage 1: About the same as when younger
- Stage 2: Penis + testes bit bigger, scrotum dropped + changed
- Stage 3: Penis longer, testes grown + dropped lower
- Stage 4: Penis longer + wider + bigger head, scrotum darker + bigger
- Stage 5: Size and shape of mans

Girls

- Stage 1 Nipple is raised a little, rest of breast still flat
- Stage 2 Breast bud stage - small breast mound, larger areola
- Stage 3 Larger areola and breast. Areola not sticking out
- Stage 4 Areola and nipple form mound above breast
- Stage 5 Mature adult stage - only nipple sticks out

It should be noted at this stage than Tanner staging may be misleading. For example, Tanner staging for females considers the size and shape of the breasts. It is known that this is highly variable in fully mature adults, so of course must be variable in adolescents. This is not reflected in Tanner staging. Therefore, it is useful to consider other measures of pubertal staging in conjunction with Tanner, when available (such as age of menarche). Indeed, as discussed in Chapter One, Tanner stage is considered by some to be a poor method of 'measuring' maturity, which is, of course a latent variable. It is understood that other measures of pubertal staging were recorded in ALSPAC, but we were not granted access to these data.

We first look at the number of respondents reporting to be in each Tanner stage at each wave of questionnaires, as shown in Tables 8.1 and 8.2.

BOYS			Wave		
Tanner	**9**	**10**	**11**	**13**	**14**
1	783 (24.9%)	584 (20.2%)	262 (09.9%)	72 (03.1%)	19 (01.0%)
2	1162 (36.9%)	1034 (35.8%)	753 (28.5%)	304 (13.2%)	86 (04.5%)
3	846 (26.9%)	874 (30.2%)	980 (37.1%)	700 (30.4%)	335 (17.6%)
4	215 (06.8%)	262 (09.1%)	463 (17.5%)	784 (34.1%)	965 (50.8%)
5	8 (00.3%)	12 (00.4%)	43 (01.6%)	193 (08.4%)	494 (26.0%)
Not sure	133 (04.2%)	124 (04.3%)	142 (05.3%)	248 (10.8%)	n/a
n	*3147*	*2890*	*2643*	*2301*	*1899*
Missing	4172	4429	4676	5018	5420
Total	7319	7319	7319	7319	7319

Table 8.1: Tanner stage by data collection wave, ALSPAC boys *n* (%)

GIRLS			Wave		
Tanner	**9**	**10**	**11**	**13**	**14**
1	2134 (61.8%)	1320 (40.2%)	398 (12.7%)	50 (01.7%)	3 (00.1%)
2	1097 (31.8%)	1188 (36.2%)	1048 (33.4%)	294 (09.9%)	39 (01.5%)
3	182 (05.3%)	680 (19.2%)	1131 (36.1%)	1010 (34.0%)	457 (17.1%)
4	24 (00.7%)	132 (04.0%)	469 (15.0%)	1145 (38.6%)	1475 (55.3%)
5	2 (00.1%)	7 (00.2%)	65 (02.1%)	387 (13.0%)	695 (26.0%)
Not sure	13 (00.4%)	6 (00.2%)	26 (00.8%)	83 (02.8%)	n/a
n	*3452*	*3283*	*3137*	*2969*	*2669*
Missing	3375	3544	3690	3858	4158
Total	6827	6827	6827	6827	6827

Table 8.2: Tanner stage by data collection wave, ALSPAC girls *n* (%)

It should be noted that there is a high level of drop-out, among both sexes, as can be seen from the previous tables. It seems as though, while more boys than girls are reporting higher Tanner stages at an early age, girls tend to reach full maturity earlier than boys. By wave 14, boys appear to be "catching up" with girls, though a slightly higher percentage of boys at wave 14 still report themselves to be at Tanner stage 1 than girls at the same wave.

Due to the small numbers of respondents in some combinations of wave and Tanner stage, we have grouped Tanner stage for most of the formal analysis in this chapter as: '1 or 2', '3' and '4 or 5'.

8.2.2 Age of menarche

Each year, girls (and / or their parents) were asked whether or not the respondent had had her menstrual period yet. Results were as shown in Table 8.3.

	Period yet?			
Wave	Yes	No	*n*	Missing
9	17 (00.5%)	3440 (99.5%)	*3457*	3370
10	76 (02.3%)	3220 (97.7%)	*3296*	3531
11	507 (16.1%)	2649 (83.9%)	*3156*	3671
13	1878 (62.0%)	1151 (38.0%)	*3029*	3798
14	2595 (95.0%)	136 (05.0%)	*2731*	4096

Table 8.3: Occurrence of menstrual period at each ALSPAC questionnaire wave, *n* (%)

The first feature to be noted is that less respondents answered this question than the ones about Tanner stage at each wave. Secondly, at wave 14, 4.9% (136) of the respondents claim to have not yet started menstrual periods. It may be interesting to look at the distribution of Tanner stages (where reported) among only those subjects. This is shown in Table 8.4.

Tanner	n (%)
1	1 (00.7%)
2	20 (14.7%)
3	62 (45.6%)
4	47 (34.6%)
5	3 (02.2%)
Not stated	3 (02.2%)

Table 8.4: Distribution of Tanner stage among those ALSPAC participants who had not reached menarche by wave 14 (n=136)

We can see that, despite not having started menstrual periods, 50% of these subjects claim to be in the later stages of physical puberty by wave 14. This suggests that when considering maturity, we should perhaps use a measure of "menarche age" in addition to Tanner staging for girls.

It should be noted that the menarche data were censored - not all girls had reached menarche by the last time they submitted a questionnaire response (whether at wave 14 or earlier). In order to get a measure of how censored the menarche data were, we look at a Kaplan-Meier survival curve. Where respondents have not reached menarche by wave 14, the age variable is the last recorded age and data are censored, otherwise age is the reported age of menarche. There was age at menarche at at least one wave for 3010 of the 4440 respondents for whom age was available. The survival plot is shown in Figure 8.1, where 'survival' is NOT having reached menarche - and is consistent with the data. For a description of the Kaplan-Meier method, see section 2.8.

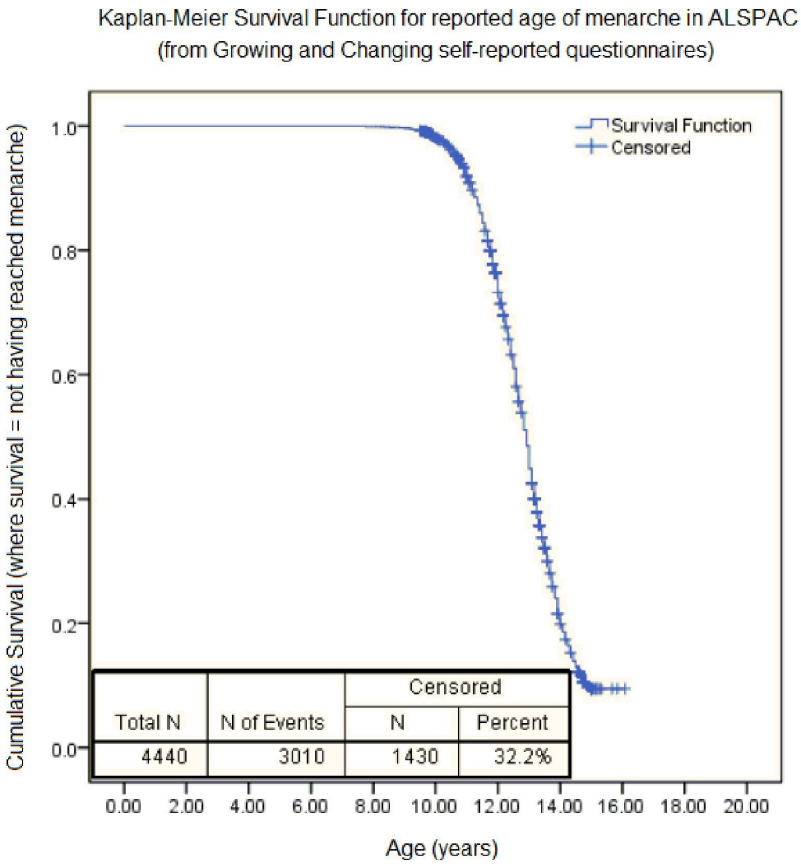

Figure 8.1: Kaplan-Meier plot of self-reported age of menarche in the ALSPAC Growing and Changing questionnaires

From Figure 8.1 we can see that 32.2% of observations were censored. The median age of menarche given by the Kaplan Meier estimator is 12.92 years (95% CI (12.88, 12.96)).

8.3 Consistency of self-reported data

Both literature and common sense tell us that postal questionnaires, i.e. self-reported data, may not result in entirely accurate data. This could present some problems in interpreting results - there is a significant possibility, for example, that 'plumper' respondents think they are at a more advanced Tanner stage than they actually are, while under-reporting weight! In this section, we explore the consistency of self-reported data in ALSPAC.

8.3.1 Who completed questionnaire? Association with Tanner stage response

It is important to consider who reported these Tanner stages. Up to and including wave 13, the question "Questionnaire completed by...?" was given, and possible answers have (for the purposes of this research) been grouped into:

- 'Child not involved' (parent / other / parent and other)
- 'Child and other' (parent and child / parent, child and other / child and other)
- 'Child only'

Results are summarised in Tables 8.5 and 8.6.

BOYS	Wave			
	9	**10**	**11**	**13**
Child not involved	2317 (74.5%)	2040 (70.4%)	1434 (52.4%)	1194 (45.2%)
Child and other	680 (21.9%)	713 (24.6%)	779 (28.5%)	695 (26.3%)
Child only	113 (03.6%)	146 (05.0%)	552 (19.1%)	750 (28.4%)
n	*3110*	*2899*	*2735*	*2639*
Missing	4209	4420	4584	4860
Total	7319	7319	7319	7319

Table 8.5: Who completed the ALSPAC Growing and Changing questionnaire (boys)

GIRLS	Wave			
	9	**10**	**11**	**13**
Child not involved	2638 (76.2%)	2423 (73.6%)	1803 (57.1%)	1616 (53.8%)
Child and other	747 (21.6%)	794 (24.1%)	969 (30.7%)	765 (25.5%)
Child only	78 (02.3%)	75 (02.3%)	388 (12.3%)	624 (20.8%)
n	*3463*	*3292*	*3160*	*3005*
Missing	3364	3535	3667	3822
Total	6827	6827	6827	6827

Table 8.6: Who completed the ALSPAC Growing and Changing questionnaire (girls)

It is clear that, as a child ages, responsibility for assessing pubertal stage moves away from the parent and towards the child. Between waves 9 and 13, though, the majority of the respondents answered the questionnaire along with a parent. At wave 14, the respondents were simply asked whether they had had help to complete the questionnaire, to which 11.1% of boys and 14.6% of girls answered "yes".

Having found these differences, concordance was tested for who completed the questionnaire and grouped Tanner stage using Goodman and Kruskall's gamma or Kendall's tau. Both statistics were computed for each ordinal table. Where Kendall's tau is reported, we use tau-b for waves 8-13 and tau-c for wave 14. For

a description of these methods, see section 2.7. Using a condition from Charles Blake of James Madison University [231] (discussed in personal correspondence in October 2010): if |gamma|-|tau|>0.05, tau is preferred (as gamma is therefore assumed to have overestimated the strength of the association), otherwise gamma is preferred. Both statistics are reported in Table 8.7, with 'preferred' statistics highlighted in bold type.

Sex	Wave	Gamma	Tau	P-value
Boys	9	0.199	**0.106**	<0.001
	10	0.235	**0.133**	<0.001
	11	0.185	**0.124**	<0.001
	13	**-0.073**	-0.051	0.007
	14	**-0.041**	-0.010	0.529
Girls	9	0.252	**0.117**	<0.001
	10	0.194	**0.102**	<0.001
	11	**0.083**	0.053	<0.001
	13	**0.064**	0.041	0.008
	14	**0.051**	0.015	0.293

Table 8.7: Concordance analysis of the association in ALSPAC Growing and Changing questionnaires between reported Tanner stage and who completed the questionnaire

When initially analysing these data, we also looked at χ^2 tests of association between Tanner stage and who completed the questionnaire. However, we have reported only the concordance analysis because Gamma and tau are potentially more conservative than χ^2, and report both strength and direction of association (as opposed to simply significance). Additionally, Gamma and tau are more powerful than χ^2 in this instance because they utilise the fact that the data are ordinal, while χ^2 doesn't. From the table of statistics, we see that the where the association is significant, tables are always concordant (Tanner stage higher as children have more input), the association is not significant for either sex at wave 14, and where it is significant, it is usually very weak.

8.3.2 Tracking reported Tanner stage over time

One potential concern with self-reported pubertal staging is that an individual may report him/herself at one Tanner stage at wave i, and at a lower stage in wave j, for $i < j$.

Exploring this concern with the reported Tanner stages in ALSPAC, we have focused on differences between consecutive waves. The numbers of respondents whose Tanner stage was reported as lower than the previous wave are shown in Table 8.8. For most of these subjects, the decrease was by only one stage. However, there were a small number with a larger decrease - for example, males reported as being at Tanner stage 4 at wave 13 and Tanner stage 1 at wave 14.

Wave i	Wave j	Boys	Girls
9	10	291	66
10	11	218	37
11	13	101	29
13	14	139	135

Table 8.8: Count of ALSPAC respondents with a lower reported Tanner stage at questionnaire wave j than at wave i ($i < j$)

We have looked, for each of the respondents in the above table, at who completed the questionnaire at each of the to ages (parent and other, child and other, child alone). We found that, in most cases, the questionnaire was completed by the same person or people in both years. This raises a serious issue of lack of consistency of self-report over time.

8.3.3 Consistency of reported age of menarche at different ages

We suspect that as someone gets further in age from her first menstrual period, her recollection of when it occurred is likely to become more inaccurate. For example, the youngest age of menarche stated was 7 years and 3 months, this response was given (without parental help) at the 'age 14' wave of questionnaires. It should be noted, however, that this respondent did not give a response at waves 9 - 11, and had given the response (with parental help) to be 12 years and 2 months when asked at wave 13.

Paired-samples t-tests were used to check the consistency of the reported age of menarche over time reported by the same individual at waves 9, 10, 11, 13 and 14. These tests are shown in Table 8.9, with statistically significantly different pairs in bold type. Bland-Altman [102] limits of agreement for these significantly different pairs of menarche responses have been calculated as: *mean* difference $\pm 1.96 \times$ sd difference

		Paired Differences				P-value (2-tailed)
		Mean	Std. Deviation	Intervals 95% CI	Intervals Agreement	
Pair 1 (N=7)	age_first_menarche_9_y - age_first_menarche_10_y	-0.18	1.03	(-1.13, 0.77)	(-2.19, 1.84)	0.66
Pair 2 (N=2)	age_first_menarche_9_y - age_first_menarche_11_y	0.33	0.94	(-8.13, 8.80)	(-1.51, 2.18)	0.70
Pair 3 (N=2)	age_first_menarche_9_y - age_first_menarche_13_y	-0.83	2.71	(-25.19, 23.52)	(-6.15, 4.48)	0.74
Pair 4 (N=5)	**age_first_menarche_9_y - age_first_menarche_14_y**	**-3.20**	**1.38**	**(-4.92, -1.48)**	**(-5.91, -0.49)**	**0.01**
Pair 5 (N=48)	**age_first_menarche_10_y - age_first_menarche_11_y**	**-0.12**	**0.38**	**(-0.23, -0.01)**	**(-0.23, 0.09)**	**0.03**
Pair 6 (N=38)	age_first_menarche_10_y - age_first_menarche_13_y	0.12	0.60	(-0.08, 0.31)	(-1.05, 1.28)	0.23
Pair 7 (N=40)	age_first_menarche_10_y - age_first_menarche_14_y	-0.32	1.25	(-0.71, 0.08)	(-2.76, 2.12)	0.12
Pair 8 (N=315)	**age_first_menarche_11_y - age_first_menarche_13_y**	**0.15**	**0.63**	**(0.08, 0.22)**	**(-1.09, 1.39)**	**0.00**
Pair 9 (N=283)	**age_first_menarche_11_y - age_first_menarche_14_y**	**0.13**	**0.78**	**(0.04, 0.22)**	**(-1.40, 1.65)**	**0.01**
Pair 10 (N=1089)	age_first_menarche_13_y - age_first_menarche_14_y	0.02	0.72	(-0.02, 0.07)	(-1.38, 1.43)	0.25

Table 8.9: Paired t-tests for the differences in age menarche as reported in the ALSPAC Growing and Changing questionnaire at multiple time points

Although the mean differences from the paired t-tests are small, the limits of agreement are wide. This suggests that while the mean difference in reported age of menarche at different times is small on a population basis, there is potential for large differences on an individual level. In further analyses with the menarche data, since we suspect that a self-report is more likely to be accurate closer to the occurrence of the first menstrual period than later, we will use the first reported age of menarche for each respondent.

We therefore have "age of first menstrual period" for 3010 respondents, with descriptive statistics as shown in Table 8.10 (note that the mean in this table was obtained using uncensored data only). It is important to note that these data are censored - not all respondents had had their first period by wave 14.

N	Min	Median	Max	Mean	SD
3010	7.58	12.92	15.00	13.00	1.08

Table 8.10: Descriptive statistics of age of menarche in the ALSPAC Growing and Changing questionnaires

8.3.4 Association between time after menarche and Tanner stage

Time after menarche is calculated for each girl as reported menarche age subtracted from age when completing the postal questionnaire, for waves 9, 10, 11, 13 and 14, summarised in Table 8.11. While time after menarche clearly cannot be used for all girls in clinical practice, it can, as a retrospective measure, be used to explore relationships between pubertal staging techniques, and (later), to study the development of body composition over maturation.

Wave	N	Min	Median	Max	Mean	SD
9	2422	-5.17	-2.92	2.08	-2.86	1.08
10	2376	-4.08	-1.83	3.08	-1.79	1.07
11	2376	-3.17	-0.83	3.33	-0.79	1.07
13	2490	-1.67	0.58	5.50	0.62	1.04
14	2431	0.00	2.00	7.08	2.07	1.08

Table 8.11: Descriptive statistics for the years between reported age of menarche and completing the ALSPAC Growing and Changing questionnaire

Figure 8.2 shows time after menarche plotted for each Tanner stage (ungrouped) at waves 9, 10, 11 and 13.

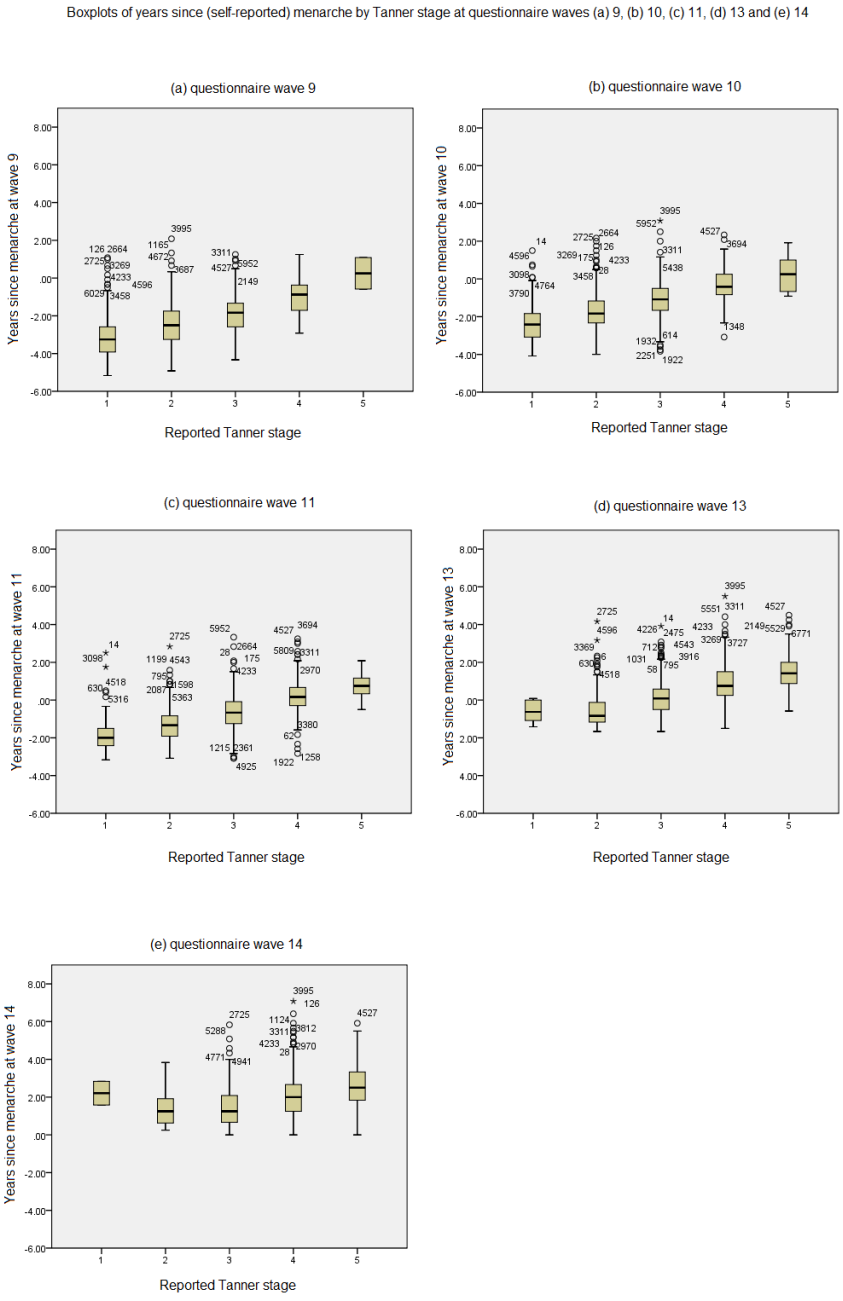

Figure 8.2: Difference in years between reported age of menarche and completion of each ALSPAC Growing and Changing questionnaire, grouped by Tanner stage

It is obvious from this graph that there is a positive (non-causal) relationship between Tanner stage and time after menarche. It seems possible, therefore, that

these methods of assessing maturation are in reasonable agreement on a population basis. However, we see a very different story on the individual level. Individual 7779, for example, reported Tanner stage 2 at wave 13 despite her first menstrual period reportedly having been more than 4 years previous.

8.3.5 Conclusions about self-reported data

The previous subsections clearly highlight an important issue - one which is widely discussed in literature (see section 1.2.5) - the potential of inconsistency with self-reported data. We have found that Tanner staging is not consistent when reported by the same person at different times, and there is a significant association between who reported the Tanner stage and the Tanner stage itself. We have also found that reported age of menarche changes considerably as girls age.

Therefore, while we will continue to use the self-reported data in the absence of an alternative, we must treat any results and conclusions from the remainder of this chapter with extreme caution.

8.4 Investigating the association between pubertal status and body composition

8.4.1 Tanner stage and BMI standard deviation scores

Using the pubertal staging data along with the BMI SDS calculated using the LMS method[1] [30], we were able to perform one-way ANOVAs to determine whether there appeared to be any association between pubertal status at a given wave and BMI standard deviation score (at the same wave). Again, we do need to remember that these data are self-reported.

Figure 8.3 displays BMI SDS against Tanner stage for boys and girls, respectively, with P-values from a standard one-way ANOVA added.

[1]Please note that the BMI SDS used in this section differ slightly from those used earlier in the project. This is because those used in this section have been calculated from the BMI as stated in the puberty dataset, while the others were calculated from the BMI in the focus files.

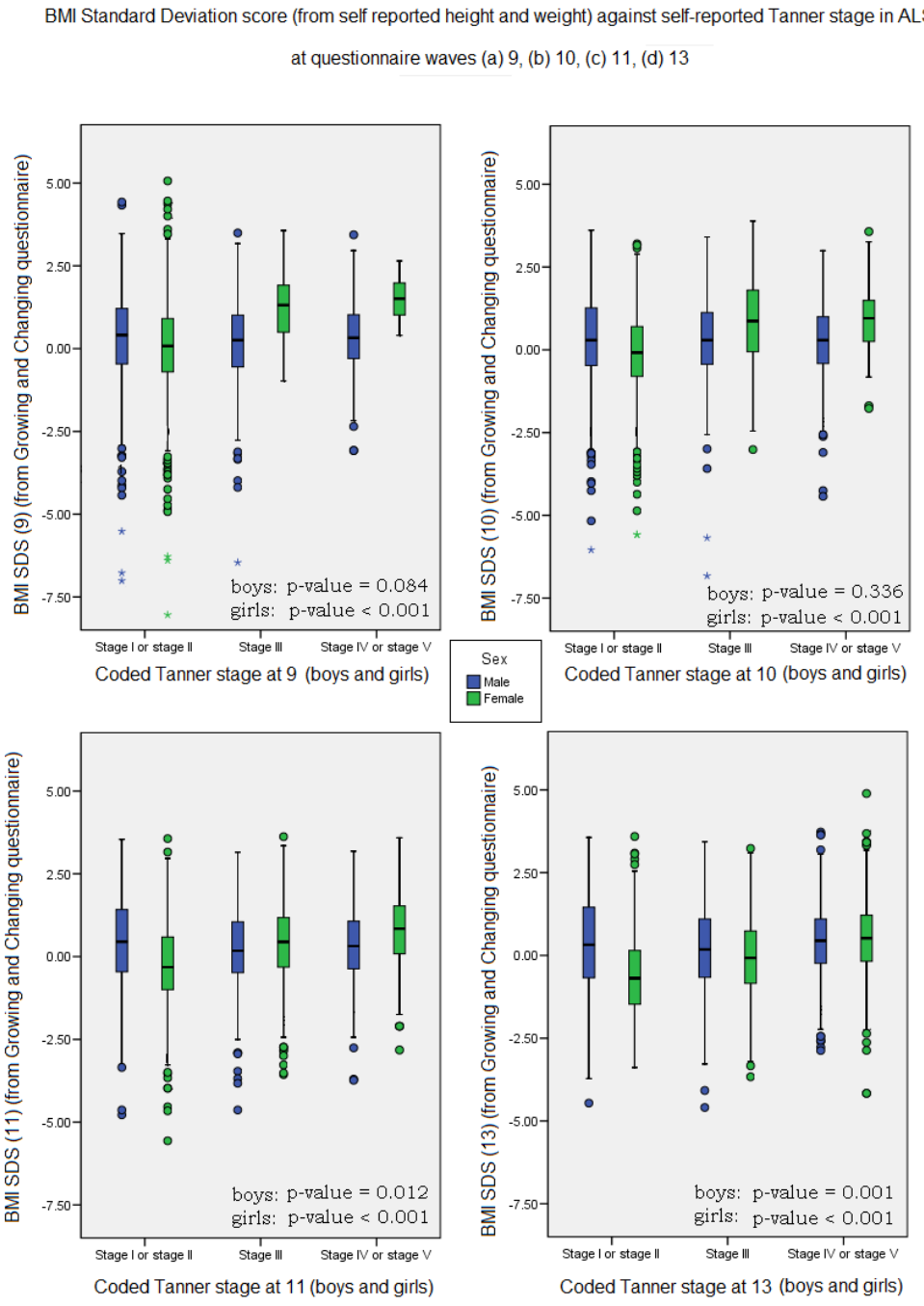

Figure 8.3: Coded Tanner stage vs BMI SDS in ALSPAC waves 9 to 13

From Figure 8.3, there is evidence of a significant association between Tanner stage and BMI Standard Deviation Score, the direction of the association is positive. It is important to note that we are not implying causation.

8.4.2 Tanner stage and fat mass (fat index) from bioelectrical impedance

Fat mass (FM) was calculated at each wave from bioelectrical impedance using the published model which was described in section 1.2.1.2.2. Percentage FM is calculated from the FM obtained from this model along with the recorded weight. We have used this model, while it is not completely accurate, as it is the best available published model for BIA.

As before, applying these models to the ALSPAC data resulted in a small number of negative values for FM. For the purposes of this analysis, any subject with a negative FM result at any age has been excluded.

Descriptive statistics for %FM are shown in Table 8.12.

Sex	Wave	N	Min	Median	Max	Mean	SD
Boys	9	2396	1.20	24.78	58.47	25.90	9.44
	10	2326	0.80	26.78	56.83	27.26	9.40
	11	2216	0.66	26.95	58.60	27.61	9.66
	13	1899	0.41	20.72	52.48	21.34	10.08
Girls	9	2237	0.12	26.63	53.05	26.05	9.68
	10	2191	2.39	27.53	58.02	27.68	9.40
	11	2113	0.72	27.23	59.21	27.65	10.09
	13	1966	4.21	34.19	58.68	33.74	8.79

Table 8.12: Descriptive statistics for percentage fat mass from BIA calculated for both boys and girls from each wave of ALSPAC puberty data

One very interesting point to note is that the percentage FM rises between waves 9 and 10, remains relatively unchanged at wave 11, then rises wave 13 for girls and decreases for boys.

JCK Wells [252] recommends adjusting mass for height as:

$$\mathrm{FM}_{\mathrm{index}} = \frac{\mathrm{FM}}{\mathrm{height}^2} \tag{8.1}$$

Figure 8.4 displays $\mathrm{FM}_{\mathrm{index}}$ against Tanner stage for boys and girls, respectively. One-way ANOVAs have been computed to assess association between $\mathrm{FM}_{\mathrm{index}}$ and Tanner stage. P-values for these tests have been added to the graphs.

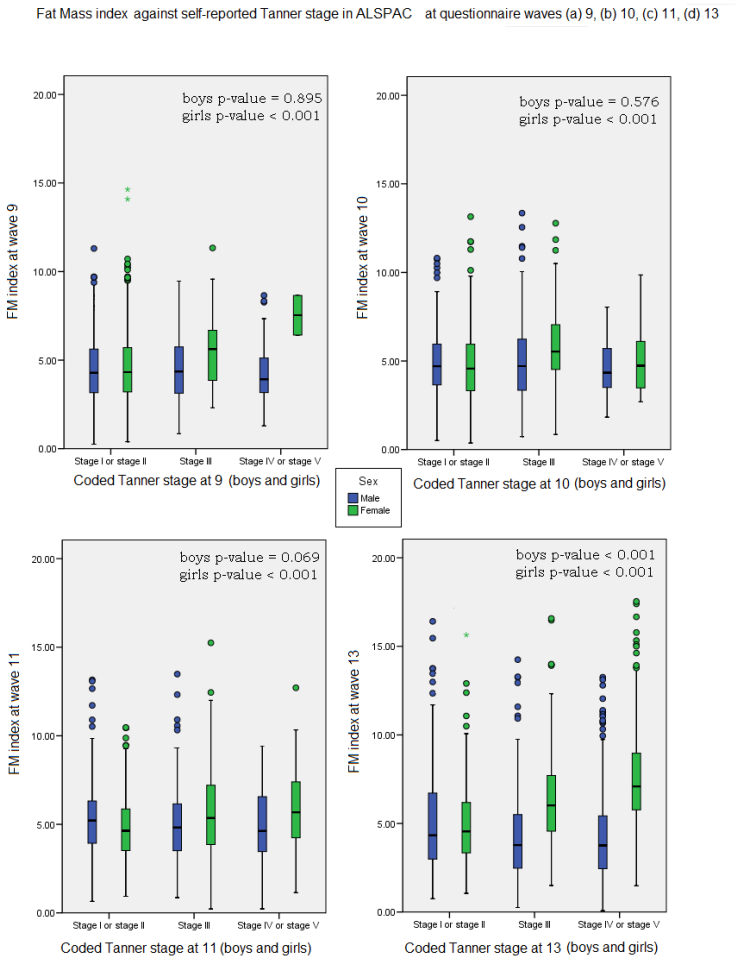

Figure 8.4: Coded Tanner stage vs FM_{index} in ALSPAC waves 9 to 13, with P-values

There is not significant evidence of association between %FM (from BIA) and Tanner stage for boys at waves 7 to 11, while the association is significant and positive at all waves for girls. At wave 13, the significant relationships for males and females are in opposing directions. Again, we do not imply causation.

Comparing this graph with that for BMI SDS and Tanner stage (Figure 8.3), the relationships between puberty and fat mass are much more obvious at wave 13 than any relationship between puberty and BMI SDS!

We cannot test the association after wave 13, due to impedance data not being available.

8.4.3 Time after menarche and BMI SDS

Having already considered the potential association between Tanner stage and BMI SDS, it is of interest to consider whether there is a similar relationship between time after menarche and BMI SDS. This has been assessed using least squares linear regression (Figure 8.5)

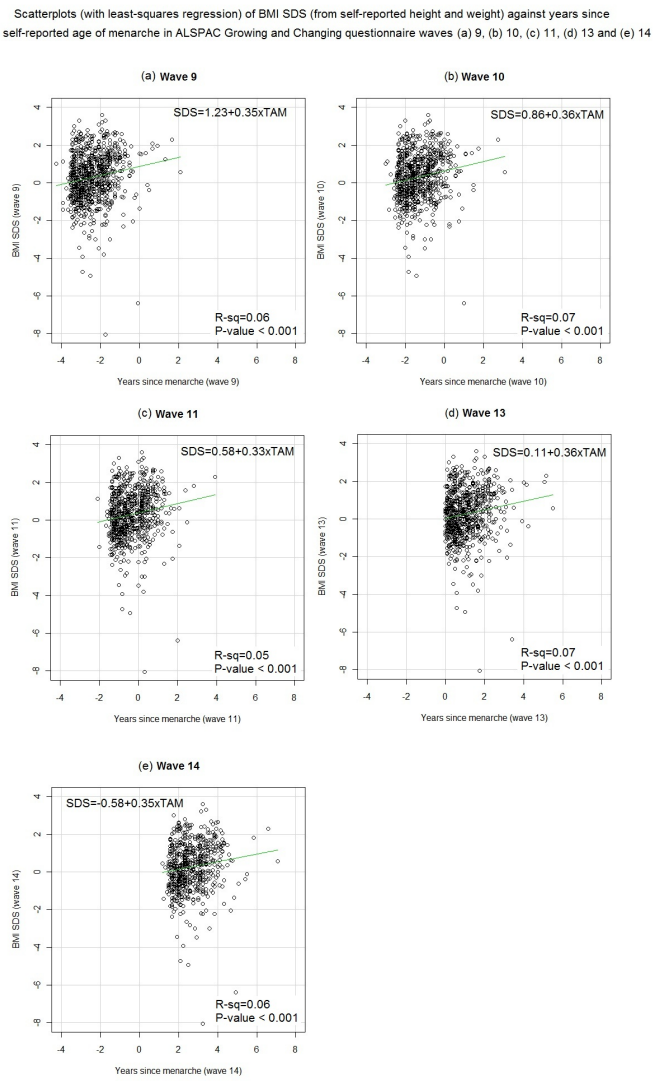

Figure 8.5: BMI SDS by age after menarche for ALSPAC questionnaire waves 9 to 14 (least-squares regression line superimposed)

It is clear that there is a non-causal positive association between BMI SDS and time after menarche at every wave. The slopes are positive and significantly different from zero. Additionally, the intercept decreases with time. This shows that the earlier menarche occurs, the higher BMI is on average.

This will be considered using the wave 14 data, where the average age of the subjects is 14.68.

We know that a BMI SDS of 1.04 is the cutoff for overweight, and 1.64 for obesity. The linear prediction model for age 14 is:

$$SDS = -0.59 + 0.35 \times \tau_j \tag{8.2}$$

where τ_j is time after menarche corresponding to a BMI SDS of j.

Substituting SDS of 1.04 and 1.64 into this equation gives $1.04 = -0.59 + 0.35 \times \tau_{1.04}$ and $1.64 = -0.59 + 0.35 \times \tau_{1.64}$, resulting in $\tau_{1.04} = 4.616$ and $\tau_{1.64} = 6.321$. This suggests that, at the wave 14 questionnaire, reaching menarche 4.6 years previous is likely to be associated with overweight, and 6.3 years previous is likely to be associated with obesity. Again, we are not implying causation.

However, while the relationships are significant, they are extremely weak (explaining no more than 7% of the variance in BMI SDS) - suggesting that there may be issues with interpreting these relationships on the individual level.

8.4.4 Time after menarche and fat mass index

We now consider the potential association between time after menarche and FM index (as defined in equation (8.1)) where FM has been estimated from bioelectrical impedance. It is important to note that while the impedance measurements were taken at a clinic, the puberty questionnaire was issued at a different point in the year. As with BMI SDS, association has been assessed by way of linear regression, shown in Figure 8.6.

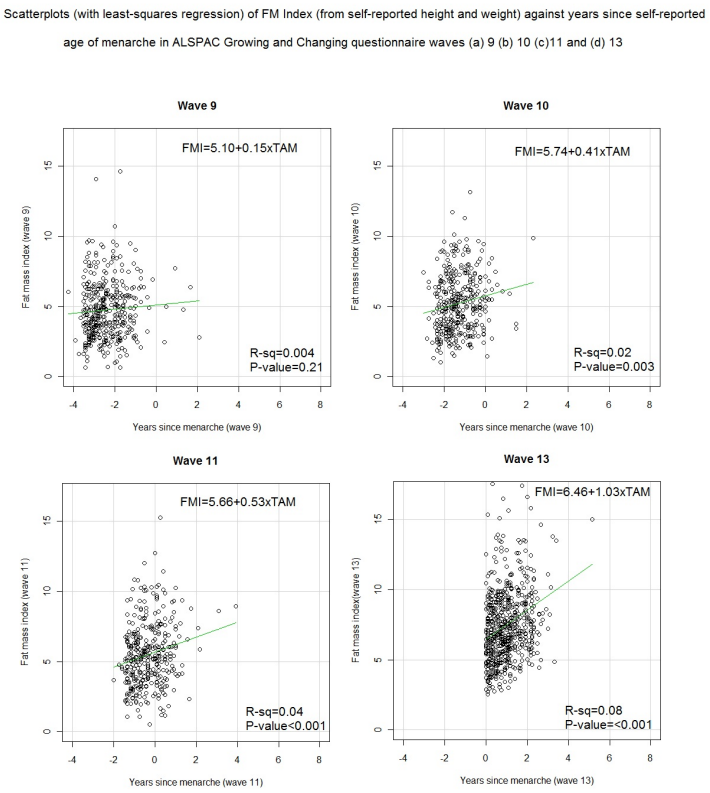

Figure 8.6: FM_{index} by age after menarche for waves ALSPAC questionnaire 9 to 13 (least-squares regression line superimposed)

It is clear from these graphs and hypothesis tests that there is an association between FM index and time after menarche at waves 10-13, but not at age 9. As before, we are not claiming a causal relationship, merely association. For each

wave 10, 11 and 13, the (non-causal) association is positive - as age after menarche increases, FM index also increases. Again, while significant, relationships are weak - therefore, care should be taken when interpreting such relationships on the individual level at this stage.

Also obvious is the increase in the slopes over time. All slopes other than FM index at wave 9 are statistically significant at the 5% level.

At this point, it is important to again acknowledge the censored nature of the data, as the censoring of Time After Menarche must affect the analysis described in this section. It should also be noted that we have not attempted to give a full description of missingness in the dataset (which might be a MNAR mechanism). As a result, conclusions are tentative.

8.5 Chapter summary

This chapter began by exploring the distributions of measures of pubertal status in ALSPAC, collected by means of annual self-reported questionnaires between the ages of 7 and 14. As we would expect, this showed that most children were at the early stages of puberty at age 7 and the later stages by age 14. This analysis raised the potential problem that self-reported data may not be consistent, and this was explored using concordance analysis. Indeed, it was found that there were marked discrepancies in the self-reported data from year to year, for example, some children reported one pubertal stage one year and a lesser stage in subsequent years.

Chapter Seven then explored relationships between pubertal status and body composition in the ALSPAC data. In addition to being cautious as a result of the aforementioned inconsistencies, we must keep in mind that we do not have any gold-standard measures of body-composition in this dataset. As proxies for body-composition, we used both BMI SDS (see section 2.2) and fat mass index, as shown in equation 8.1, where FM(kg) is estimated from the Glasgow BIA models (see section 1.2.1.2.2). We found a positive relationship between pubertal stage as determined by self-reported Tanner stage and BMI SDS. In addition to Tanner stage, we were able, for the girls, to look at the association between age of menarche and BMI SDS. As mentioned, there were inaccuracies in the consistency of the self-reported data: girls often reported different ages of menarche at each wave of questionnaires. In a naive attempt to deal with this, we used each girl's *earliest* reported age of menarche. From this, we were able to calculate the time between each wave of questionnaires and the time of menarche. This difference was plotted against BMI SDS and fat mass index, and statistically significant positive relationships were found for both.

While we must be very careful with the results from this chapter, these analyses suggest that there are relationships between puberty and body composition. It would appear that, as children progress into adolescence, both BMI SDS and FM increase significantly.

Chapter 9

Review and discussion

9.1 Summary

This thesis has described an attempt to track energy-imbalance and body composition through puberty. The main tools have been:

- an extensive search of literature, identifying and considering in turn the main components of body composition and the energy-balance equation,
- careful analysis of the (admittedly limited) data to investigate the issues identified in the literature search,
- a simulation study, used to consider the effect of real-life data issues on analysis.

In reviewing the literature on body composition, we found that Tim Cole, in 1990, published a method of calculating age- and sex-specific body mass index (BMI) standard deviation scores (SDS) for children. We identified that, despite the concerns about the use of BMI as a marker for fatness in adults, this method (the LMS method) is a relatively straightforward and accurate representation of body size in childhood. In terms of exploring energy-balance and weight gain, however, BMI SDS falls short. While it gives a useable indication of an individual child's size relative to other children, it gives no information about the body in terms of absolute values or proportions of fat mass (FM) and fat-free mass (FFM). For this, far more complicated methods are required. We researched the methods that are currently used for determining fat-free mass and fat mass in both adults and children, and how effective and practical each method was considered to be by leading researchers in the field. We found a clear trade-off in this methodology: the more accurate the method, the more time-, cost- and personnel- intensive the procedure. An accurate two-component method of determining body composition in live subjects is hydrodensitometry, a procedure that is unsuitable for many indi-

viduals - particularly children - as it involves complete submersion in water. Dual energy x-ray absorptiometry (DEXA) and bioelectrical impedance (BIA) are two methods of achieving estimates of FM and FFM in the body. They are both more practical for routine use - but less accurate with current models - than hydrodensitometry.

We were able to explore these methods with the Avon Longitudinal Study of Parents and Children (ALSPAC) data. Firstly, we classified children as not overweight, overweight or obese according to their BMI standard deviation score. In an ideal population, we could perhaps hope to find around 85% of the population were not overweight (i.e. underweight or normal weight), 10% overweight and 5% obese. From public knowledge and the results of the literature search, we expected to find that the numbers of overweight and obese children in the sample would be higher than the ideal - which was precisely what we did find. At the 'age 11' wave of questionnaires and clinics, for example, we found that almost 20% of both sexes were classified by their BMI SDS as obese.

Using the ALSPAC datasets as well as other data that we were able to acquire, we showed that DEXA and BIA gave body composition estimates that were not in agreement with one another - which, having carried out the literature review, was what we expected to find. Unfortunately, due to a lack of appropriate data, we were unable to continue with further modelling of DEXA. Continuing with BIA, we explored recently developed models of body composition and also proceeded to develop our own models. Unfortunately, we did not have data from gold standard methods on which to test such models. Therefore, while we are not yet in a position to suggest models for use in clinical practice, we have succeeded in showing

that the models which are currently used are not accurate on the individual level for this study, representative of the UK population. This was shown by the occurrence of very low, sometimes negative, FM estimates for many individuals.

Moving on to the energy-balance equation (equation 1.1 in section 1.1), we identified the main elements of energy expenditure (EE) to be resting energy expenditure (REE) (or basal metabolic rate (BMR)), diet-induced thermogenesis (DIT) and physical activity energy expenditure (PAEE); and energy intake (EI) is on the other side of the equation. While there do exist accurate methods for measuring energy expenditure (direct and indirect calorimetry), the literature review identified the complexity of their use which means that they are not suitable for routine use with many people - including children and adolescents. PAEE is often estimated from self-report of activity - a method that is known to be highly unreliable. There have been several studies of physical activity which have involved the use of heart-rate monitoring and / or accelerometry. While these methods are more objective than self-report, they are not without fault. They still require models to get from a measurement of heart rate to a measure of PAEE - and as with models for many things in this area, there is no one model that is considered to be the correct one. PAEE can be measured by exercising inside a calorimeter, but this does not replicate real-life situations and is not suitable for the young. A similar issue arises with energy intake, which is usually estimated by self-report through the use of food diaries.

The lack of availability of data limited what we could do with each aspect of the energy-balance equation. With anthropometric data available at several stages in ALSPAC, we were able to look at some of the many models for REE that have

appeared in the literature. Firstly, we replicated and attempted to extend the models published by Harris and Benedict in 1919, using data that was published in their book [86]. We concluded that the published models were in fact the best models available for the original data. However, we must keep in mind that what was considered representative of the population a century ago is unlikely to be representative of the population today!

Following on from this, we collated several recently published models for REE and applied them to the ALSPAC datasets. In this section, we could not attempt to identify models as correct or incorrect - since we did not have results from a gold-standard method for comparison - we merely intended to show agreement, or lack thereof, among the models. Using descriptive statistics, t-tests, Bland-Altman diagrams and intraclass correlations, it was shown that no two models are in agreement. While many of these models are currently used in clinical practice, their lack of agreement shows that, at most, only one of them can be correct. We also considered nonlinear modelling but, due to insufficient data, were unable to draw concrete conclusions. It is clear that a single, accurate model for REE must be developed and validated for use, allowing for potential differences due to factors such as gender, age, race and anthropometry.

In Chapter Six, we began by considering the perfect study in this area. It quickly became clear that this study is an impossibility with current equipment and methodology - the best we can hope for is to reduce error. We simulated longitudinal resting energy expenditure and body composition data, with a view to considering how data issues such as sample size and missing observations affect analysis. We found that error was reduced to an acceptable level with sample sizes

of around $n = 1500$ (with no missing data - a very unrealistic scenario!). Looking at the effect of missing data, we considered complete-case analysis and two imputation methods: (1) developing models for dependent variables on smaller, complete sub-samples and using these models to impute missing observations, and (2) multiple imputation using chained equations (MICE). Both imputation methods resulted in improved parameter estimates over the complete-case analysis. Estimates from data imputed using MICE were lower in error than those using the other method, but it must be noted that MICE is costly in terms of time and computational resources, and assumes that the data are missing at random (MAR) or missing completely at random (MCAR).

Data on other aspects of energy-(im)balance was not available to us for the ALSPAC study - physical activity data were extremely limited and energy intake data were simply not gathered. For this reason, we looked at quantifying energy-imbalance over time using changes in body composition. The main aim of this section was to determine how much energy-imbalance over a short period of time resulted in maintaining or changing body composition according to BMI standard deviation scores. We know that young people usually require a positive energy-imbalance to maintain healthy growth - but it is unknown how much is healthy and how much is too much.

We were able to calculate, from BIA-derived estimates of fat mass and fat-free mass gained over time, the estimated energy-imbalance for each individual over the time period (wave 9 to wave 11). Overall, we found that healthy growth was achieved with a very small degree of energy-imbalance - often less than 100 kcal per day. There were other, surprising results - for example, those with a high BMI

SDS at 9 and a high degree of energy-imbalance were actually at a lower BMI SDS by 11. With no scientific answer to this, we must, at this point, simply attribute it to the fact that 'heavier' people require more energy than their counterparts for everyday living.

The study concluded by looking at puberty and the determination of pubertal status. We identified the Tanner scale to be the most widely used method of determining the pubertal stage of an individual. We did, however, discover that there are flaws in this method. The main flaw is that Tanner stage is often reported by adolescents themselves, or by their parents, rather than by trained medical professionals. This brings the potential problem of inaccurate responses - whether intentional or otherwise. For example, it is possible that a 'plump' adolescent girl would consider herself to be further along the pubertal process than would be determined by a trained professional. We considered links identified in the literature between pubertal status and body composition, finding that this is a subject much more widely written about for girls than for boys. What literature was available for boys suggested that those with a higher BMI at a young age seemed to experience puberty later than their lower-BMI counterparts. The consensus for girls was the opposite - greater fatness in youth was reported to lead to earlier sexual maturation. It must be noted, however, that whatever relationships exist between puberty and body composition are intensely complex and certainly cannot be considered to be causal at this stage. As an alternative, or addition, to Tanner stage for assessing pubertal development, we identified age of menarche for girls and peak height velocity (PHV) for boys. While PHV can be determined given frequent clinic visits, age of menarche must be self-reported and is therefore susceptible to inaccuracy.

The ALSPAC puberty data were made available to us in the late stages of this project, and allowed us to consider a number of the issues that were raised in the literature review. These data were collected by means of a postal questionnaire completed by the children themselves, parents or teachers. We described the ALSPAC children in terms of their reported pubertal status, and it quickly became obvious that there we were encountering difficulties with the accuracy of the reported data. For example, there were several hundred children for whom reported Tanner stage decreased over time, and there was a clear association (among all children) between the reported Tanner stage and the person who completed the questionnaire. We also considered the accuracy of the reported age of menarche, finding that often, girls who reported their first age of menarche in more than one questionnaire were not consistent. Using these self-reported data in conjunction with the (at best tentative) body composition estimates discussed earlier (BMI SDS and fat mass by BIA), we explored the relationships between puberty and body composition. While we found a positive association between BMI SDS and Tanner stage for girls, we found a lack of evidence for such associations among boys. We were therefore unable to draw conclusions similar to those found in the literature regarding BMI and age of sexual maturation. Instead, we were simply able to conclude that there appears to be a stronger positive relationship between body composition (by proxy) and pubertal stage for girls (by both Tanner stage and age of menarche) than for boys (by Tanner stage alone, in the absence of PHV data).

9.2 Limitations of research

As with the majority of research projects, this PhD was not without a number of limitations - some we were able to overcome in some way, others we were not.

One such limitation was uncovered very early on in the project, during the initial search of the literature, and concerned the original scope of research. As mentioned earlier, we had originally hoped to model energy-imbalance. While researching energy-imbalance, the sheer breadth of the topic became clear and it was obvious that it would be necessary to narrow our aims somewhat. As we began examining the available data, this necessity became clear for another reason - there was a distinct lack of data available on several aspects of energy-imbalance. Often this was because data were not collected, perhaps because it was not possible or practical to do so. For example, one of the original aims was to use the ALSPAC data to validate DEXA as a means of determining body composition. However, this would have required hydrodensitometry data, currently the only body composition determination procedure deemed to be a gold-standard. Hydrodensitometry, as described in section 1.2.1.2.2, involves complete submersion in water - a procedure generally considered unsuitable for young people - and was therefore not carried out by ALSPAC. Without this, the best we were able to offer regarding DEXA was a comparison with BIA, itself not a gold-standard measurement.

As we progressed with this research, data - whether the lack availability or the timing of our access - became a major stumbling block. We had intended to investigate methods and models over puberty, but were nearing the end of the project before we received puberty data from ALSPAC. Further to this, once received, we

found the puberty data to be flawed in several ways. All puberty data were right-censored: the last wave of the questionnaire was wave 14 and not all respondents had (according to self-report) reached puberty by this wave (using certain agreed definitions of puberty). Some respondents of both sexes reported early Tanner stages in the final wave, and some girls stated that they had not reached menarche by this time. The collected data were not ideal - for example, while we had a measure of menarche in addition to Tanner stage for girls, there was no alternative provided for boys. Had we had it, we could have made use of PHV data - often used as an indicator of puberty for boys. This would have required measurements to have been made regularly - at the very least quarterly - throughout the study duration. The biggest issue was that the growth data were collected by means of an annual postal questionnaire, resulting in entirely self-reported data, a method known to be highly unreliable. Additionally, these questionnaires were administered at different time points from the practitioner-led clinics, making it difficult to draw conclusions when analysis involved both sets of data - for example, examining body composition by BIA (from clinic data) over puberty (from questionnaires).

On top of data issues, we came across problems with previous modelling of several aspects of body composition and energy-imbalance. While much has been written on most individual areas, the literature is often conflicting. For example, for resting energy expenditure (REE), we were able to easily identify seven sets of models for estimation based on anthropometric measurements, none of them gold-standard and no two in agreement.

Very often, several of these limitations co-existed. For example, we had hoped to be able to determine the degree of energy-imbalance that was necessary for healthy

growth. Without measurements of the energy-balance components, we attempted to overcome this by using a proxy - estimates of fat-mass gain and fat-free-mass gain over time multiplied by the energy required to store this mass. Without gold-standard body composition data, we were forced to use FM and FFM estimates from BIA - by a model we have shown to be inaccurate on the individual level. In addition to this, the estimates of the energy cost of storing such mass were simply that - estimates, reported by one researcher.

This project has succeeded in highlighting various problems with energy-balance research as it stands today. As a result of the limitations discussed, conclusions from modelling should be treated with great care - while many pieces of modelling can be applied with reasonable confidence to the population, it is impossible to apply them with the required level of accuracy to the individual, which would ultimately be the aim with such a research topic. Much more work must be done in the areas identified in this project before models can be considered 'gold-standard' and applied with confidence to the individual adolescent.

9.3 Contributions of research

Until now, research into body composition and energy-balance has been largely focused on individual aspects thereof. One of the main contributions of this research is the integration of the literature on these individual aspects into a comprehensive review.

This project has shown that, in a sample representative of UK children, higher proportions of children are overweight or obese than would be expected, and prepubertal children who are overweight or obese are likely to remain as such as they enter puberty.

In the chapter on body composition models, we confirmed for this dataset that BIA and DEXA are not in agreement with one another (a discrepancy that had already been identified in other studies). Furthermore, it was pointed out that hydration equations commonly used for children are not equal to the adult value at age 18, suggesting that further work needs done to develop models for body composition from impedance. Further illustrating this point, it was found that current BIA models can result in negative FM values for some individuals.

From the review of literature on energy expenditure, there were several models for REE identified which are currently used. This project has shown that, of these REE equations, no two are in agreement.

This thesis further contributes to research by, through the use of reasonably realistic simulated REE and body composition data, examining the effect that real

-life data issues might have on analysis. It has shown that multiple imputation using chained equations considerably reduces the mean square error (MSE) of parameter estimates even when very large amounts (up to 80%) of data are missing from the independent variable, even when similar amounts are missing from the (non-imputed) dependent variable.

An important contribution to the research occurred when considering the degree of energy-imbalance required for healthy growth in childhood and adolescence. This project has shown that such growth is achieved with a very small degree of positive imbalance. It appears, therefore, that children do not need an energy intake that is much larger than their energy expenditure in order to grow healthily.

The final chapter of analysis showed that self-reported pubertal stage data over time is likely to be inconsistent for some individuals. This is an important contribution to research as it suggests that, for accurate and reliable results, data should not be self-reported if possible.

The project then went on to show that BMI SDS increases with pubertal staging, and importantly, there may be an association between obesity and early puberty for girls.

9.4 Future work

At this stage in the research, further, more complicated modelling of current data is not what is required. While it is infuriating, as a statistician, to see consistently poor practice (lack of willingness to use or build on models developed by other researchers, for example), the fundamental problems do not lie with the statistical modelling. What is required are foundational changes in how data are collected and used, i.e. models for individual components of energy-imbalance must be able to be validated against gold-standards.

However, far more important than this, it is crucial that there is much more of an awareness that everything in this area of research is interrelated. Without gold-standard measures of body composition, for example, we cannot hope to develop accurate models for resting energy expenditure. However, we must keep in mind that designing and implementing such a study would be very costly indeed.

Having said that, the most fundamental need at this stage is to be able to obtain, on a routine basis, accurate estimates of fat-mass and fat-free mass. Without this, as mentioned above, modelling of other components becomes obsolete and we therefore cannot hope to make any useful progress in tracking energy-imbalance over time.

In terms of modelling and data issues, further consideration should be given to imputation of missing data, given that this type of data are particularly susceptible to missingness. This thesis explored two methods of imputation with simulated data. It would be worthwhile in future investigating other imputation methods

and models, if possible, on real data. There is a huge literature on missing values and imputation. Perhaps all that is really required is for researchers in this area to apply existing research. Of course, there will always remain the problem that data are missing not at random (MNAR). Those children who are unusually light or heavy are likely to be differentially removed from studies of growth and body composition by their parents. That is a non-statistical problem to be overcome!

Bibliography

[1] Stamatakis E. Anthropometric measurements, overweight, and obesity. In Health Survey for England 2002. Volume 1: The health of children and young people. Sproston K, Primatesta P, editors. London: The Stationery Office; 2003.

[2] Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of Overweight and Obesity in the United States, 1999-2004. J Am Med Assoc. 2006;295(13):1549 – 1555.

[3] Rennie K, Jebb S. Prevalence of obesity in Great Britain. Obes Res. 2005;6:11 – 12.

[4] SPSS Inc. Released 2007. SPSS for Windows, Version 15.0. Chicago, SPSS Inc.;.

[5] SPSS Inc. Released 2009. PASW Statistics for Windows, Version 18.0. Chicago: SPSS Inc.;.

[6] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2011. ISBN 3-900051-07-0. Available from: `http://www.R-project.org/`.

[7] Reily JJ. Obesity in childhood and adolescence: evidence based clinical and public health perspectives. Postgrad Med J. 2006;82:429 – 437.

[8] Fry J, Finley W. The prevalence and costs of obesity in the EU. P Nutr Soc. 2005;64:359 – 362.

[9] Etisham S, Barrett T, Shaw N. Type 2 diabetes mellitus in UK children - an emerging problem. Diabetic Med. 2008;17:867 – 871.

[10] Eisenberg M, Neumark-Sztainer D, Story M. Associations of weight-based teasing and emotional well-being among adolescents. Arch Pediatr Adolesc Med. 2003;157:733 – 738.

[11] National Institutes of Health. Energy - Information About Energy Balance. http://science.education.nih.gov/supplements/nih4/energy/guide/info-energy-balance.htm; Updated 2005; Accessed January 2008.

[12] Williams J, Wake M, Hesketh K, Maher E, Waters E. Health-Related Quality of Life of Overweight and Obese Children. J Am Med Assoc. 2005;293(1):70 – 76.

[13] Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, et al. Prevalence of Obesity, Diabetes, and Obesity-Related Health Risk Factors, 2001. J Am Med Assoc. 2003;289(1):76 – 78.

[14] Campbell University. Body Composition Measurements; 2004. Accessed January 2008. Electronic resource, `http://www.drwoolard.com/fitness_testing/body_comps.htm`.

[15] Mackie A, Hannan W, Tothill P. An introducion to body composition models used in nutritional studies. Clin Physiol Meas. 1982;10(4):297 – 310.

[16] Fields DA, Goran MI. Body composition techniques and the four-compartment model in children. J Appl Physiol. 2000;89:613 – 620.

[17] Withers RT, LaForgia J, Pilland RK, Shipp NJ, Catherton BE, Shultz CG, et al. Comparisons of two-, three-, and four-compartment models of body composition in men and women. J Appl Physiol. 1998;85(1):238 – 245.

[18] Wells JCK, Fuller NJ, Dewit O, Fewtrell MS, Elia M, Cole TJ. Four-component model of body composition in children: density and hydration of fat-free mass and comparison with simpler models. Am J Clin Nutr. 1999;69:904 – 912.

[19] Withers RT, Laforgia J, Heymsfield SB. Critical appraisal of the estimation of body composition via two-, three-, and four-compartment models. Am J Hum Biol. 1999;11(2):175 – 185.

[20] Baumgartner RN, Heymsfield SB, Lichtman S. Body composition in elderly people: the effect of criterion estimates on predictive equations. Am J Clin Nutr. 1991;53:1345 – 1353.

[21] Huang B, Rodriguez B, Burchfiel C, Po-H Chyou, Curb J, Sharp D. Associations of adiposity with prevalent coronary heart disease among elderly men: the honolulu heart program. Int J Obes. 1997;21:340 – 348.

[22] Rexrode K, Carey V, Hennekens C, Walters E, Colditz A, Stampfer M, et al. Does abdominal adiposity affect coronary heart disease in women? JAMA. 1998;280:1483 – 1484.

[23] World Health Organisation. Weight for Height; 2007. Accessed 05/04/07. Electronic resource, http://www.who.int/childgrowth/standards/weight_for_height/en/index.html.

[24] Poustie V, Walting R, Ashby D, Smith R. Reliability of percentage ideal weight for height. Arch Dis Child. 2000;83:183 – 184.

[25] Mei Z, Grummer-Strawn LM, Pietrobelli A, Goulding A, Goran MI, Dietz WH. Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. Am J Clin Nutr. 2002;75:978 – 985.

[26] Daniels S, Khoury P, Morrison J. The utility of body mass index as a measure of body fatness in children and adolescents: differences by race and gender. Pediatrics. 1997;99:804 – 807.

[27] Gallagher D, Visser M, Sepulveda D, Pierson R, Harris T, Heymsfield S. How useful is body mass index for comparison of body fatness across age, sex and ethenic groups? Am J Epidemiol. 1996;143(3):228 – 239.

[28] World Health Organisation. BMI-for-age; 2007. Accessed 10/04/07. Electronic resource, http://www.who.int/childgrowth/standards/bmi_for_age/en/index.html.

[29] Choo V. WHO reassesses appropriate body-mass index for Asian populations. The Lancet. 2002;360:235.

[30] Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990. Arch Dis Child. 1990;73:25 – 29.

[31] Cole TJ. Growth monitoring with the British 1990 growth reference. Arch Dis Child. 1997;76:47 – 49.

[32] Centers for Disease Control and Prevention. About BMI for Children and Teens; 2006. Accessed 04/04/07. Electronic resource, http://www.cdc.gov/nccdphp/dnpa/bmi/childrens_BMI/about_childrens_BMI.htm.

[33] Dietz W, Bellizzi M. Introduction: the use of body mass index to assess obesity in children. Am J Clin Nutr. 1999;70(suppl):123s – 125s.

[34] Chan D, Watts G, Barrett P, Burke V. Waist circumference, waist-to-hip ratio and body mass index as predictors of adipose tissue compartments in men. QJM. 2003;96:441 – 447.

[35] Taylor R, Jones I, Williams S, Goulding A. Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk mass, as measured by dual-energy X-ray absorptiometry, in children 3-9 y. Am J Clin Nutr. 2000;72:490 – 495.

[36] Florida's Health. Floridy Obesity Prevention Program; 2000 - 2007. Accessed 05/04/07. Electronic resource, `http://www.doh.state.fl.us/family/obesity/whatis.html`.

[37] Taylor R, Keil D, Gold E, Williams S, Goulding A. Body mass index, waist girth, and waist-to-hip ratio as indexes of total and regional adiposity in women: evaluation using receiver operating characteristic curves. Am J Clin Nutr. 1998;67:44 – 49.

[38] Dalton M, Cameron A, Zimmet P, Shaw J, Jolley D, Dunstan D, et al. Waist circumference, waist-hip ratio and body mass index and their correlation with cardiovascular disease risk factors in Australian adults. J Intern Med. 2003;254:555 – 563. On behalf of the Ausdiab Steering Committee.

[39] Fernández J, Redden D, Pietrobelli A, Allison D. Waist Circumference Percentiles in Nationally Representative Samples of African-American, European-American, and Mexican-American Children and Adolescents. J Pediatr. 2004;145:439 – 444.

[40] Ashwell M. A new shape chart for assessing the risks of obesity. Proc Nutr Soc. 1995;54:86A.

[41] Ashwell M, LeJeune S, McPherson K. Ratio of waist circumference to height may be better indicator of need for weight management. BMJ. 1996;312:377.

[42] Ashwell M. Waist to height ratio and the Ashwell shape chart could predict the health risks of obesity in adults and children in all ethnic groups. Nutr Food Sci. 1995;35(5):359 – 364.

[43] Gustat J, Elkasabany A, Srinivasan S, Berenson GB. Relation of abdominal height to cardiovascular risk factors in young adults. Am J Epidemiol. 1999;151(9):885 – 891.

[44] Ward A, Pollock M, Jackson A, Ayres J, Pape G. A comparison of body fat determined by underwater weighing and volume displacement. Am J Phisiol. 1978 January;234(1):E94 – E96.

[45] Kravitz L, Heyward V. Getting a grip on body composition. IDEA today. 1992;10(4):34 – 39.

[46] Cyrino E, Okano A, Glaner M, Romanzini M, Gobbo L, Makoski A, et al. Impact of the use of different skinfold calipers for the analysis of the body composition. Rev Bras Med Esporte. 2003;9(3):150 – 153.

[47] Georgia State University. Body Composition; 1998. Accessed 05/04/07. Electronic resource, `http://www2.gsu.edu/~wwwfit/bodycomp.html`.

[48] Chan Y, Leung S, Lam W, Peng X, Metreweli C. Body fat estimation in children by magnetic resonance imaging, bioelectrical impedance, skinfold and body mass index: A pilot study. J Paediatr Child Health. 1998;34(1):22 – 28.

[49] Demerath E, Guo S, Chumlea W, Towne B, Roche A, Siervogel R. Comparison of percent body fat estimates using air displacement plethysmography and hydrodensitometry in adults and children. Int J Obes. 2002;26:389 – 397.

[50] Fields D, Goran MI, McCrory M. Body-composition assessment via air-displacement plethysmography in adults and children: a review. Am J Clin Nutr. 2002;75:453 – 467.

[51] BlueCross BlueShield of Tennessee Medical Policy Manual. Air Displacement Plethysmography for the Measurement of Body Composition; 2006. Accessed 04/06/2007. Electronic resource, `http://www.bcbst.com/MPManual/Air_Displacement_Plethysmography_for_the_Measurement_of_Body_Composition.htm`.

[52] Litaker M, Barbeau P, Humphries M, Gutin B. Comparison of Hologic QDR-1000/W and 4500W DXA Scanners in 13- to 18-Year Olds. Obes Res. 2003;11:1545 – 1552.

[53] University of Alberta: Human Nutrition Research Unit. Dual Energy X-Ray Absorptiometry (DXA); 2002 - present. Accessed 11/04/07. Electronic resource, http://www.hnru.afns.ualberta.ca/index.asp?page=/Services/DXA.

[54] Van Loan M, Mayclin P. Body composition assessment: dual-energy X-ray absorptiometry (DEXA) compared to reference methods. Eur J Clin Nutr. 1992;46(2):125 – 30.

[55] de Mello M, Damaso A, Antunes H, Siqueira K, Castro M, Bertolino S, et al. Body composition evaluation in obese adolescents: the use of two different methods. Rev Bras Med Epsorte. 2005;11(5):251e – 254e.

[56] Wong W, Butte N, Hergenroeder A, Hill R, Stuff J, O'Brian Smith E. Are basal metabolic rate prediction equations appropriate for female children and adolescents? J Appl Physiol. 1995;81(6):2407 – 2414.

[57] Williams JE, Wells JCK, Wilson CM, Haroun D, Lucas A, Fewtrell MS. Evaluation of Lunar prodigy DEXA for assessing body composition in healthy persons and patients by comparison with the criterion 4 component model. Am J Clin Nutr. 2006;83:1047 – 1054.

[58] Heymsfield S, Wang Z, Baumgartner R, Ross R. Human body composition: Advances in models and methods. Annu Rev Nutr. 1997;17:527 – 558.

[59] Wright CM, Sherriff A, Ward S, McColl JH, Reilly JJ, Ness AR. Development of bioelectrical impedance - derived indices of fat and fat-free mass for assessment of nutritional status in childhood. Eur J Clin Nutr. 2007;p. 1 – 8.

[60] Sherriff A, Wright CM, Reilly JJ, McColl J, Ness A, Emmett P. Age- and sex-standardised lean and fat indices derived from bioelectrical impedance analysis for ages 7 - 11 years: functional associations with cardio-respiratory fitness and grip strength. Brit J Nutr. 2009;101:1753 – 1760.

[61] Villoria M. Statistical modelling of body composition in children and their parents. University of Glasgow. Glasgow; 2008.

[62] Kronmal R. Spurious correlation and the fallacy of the ratio standard revisited. Journal of the Royal Statistical Society Series A (Statistics in Society). 1993;156(3):379 – 392.

[63] Kyle U, Bosaeus I, de Lorenzo A, Deurenberg P, Elia M, Gomez J, et al. Bioelectrical impedance analysis - part 1: review of principles and methods. Clin Nutr. 2004;23:1226 – 1243.

[64] Houtkooper L, Going S, Lohman T, Roche A, Van Loan M. Bioelectrical impedance estimation of fat-free body mass in children and youth: a cross-validation study. J Appl Physiol. 1992;72(1):366 – 373.

[65] Jebb S, Cole TJ, Doman D, Murgatroyd P, Prentice A. Evaluation of the novel Tanita body-fat analyser to measure body composition by comparison with a four-compartment model. Brit J Nutr. 2000;83:115 – 122.

[66] Sardinha L, Going S, Teixeira P, Lohman T. Receiver operating characteristic analysis of body mass index, triceps skinfold thickness and arm girth for obesity screening in children and adolescents. Am J Clin Nutr. 1999;70:1090 – 1095.

[67] Institute of Medicine, ed Koplan J and Liverman C and Kraak V. Preventing Childhood Obesity. Health in the Balance. Washington DC: The National Academic Press; 2005.

[68] Sing A, Mulder C, Twisk J, van Mechelen W, Chinapaw M. Tracking of childhood overweight into adulthood: a systematic review of the literature. Obes Res. 2008;9(5):474 – 88.

[69] Haslam D. Obesity: a medical history. Obes Res. 2007;8(suppl 1):31 – 36.

[70] Ravussin E, Bogardus C. A brief overview of human energy metabolism and its relationship to essential obesity. Am J Clin Nutr. 1992;55:242S – 245S.

[71] Levine J. Measurement of human energy expenditure. Pub Health Nutr. 2005;8(7A):1123 – 1132.

[72] Murgatroyd P, Shetty P, Prentice A. Techniques for the measurment of human energy expenditure: a practical guide. Int J Obes. 1993;17:549 – 568.

[73] Battley E. The advantages and disadvantages of direct and indirect calorimetry. Thermochimica Acta. 1995;250:337 – 352.

[74] Kleiber M. The Fire of Life. Robert R Krieger Publishing Company; 1975.

[75] Webb P. The measurement of energy expenditure. J Nutr. 1991;121:1897 – 1901.

[76] Jebb S, Murgatroyd P, Goldberg G, Prentice A, Coward W. In vivo measurement of changes in body composition: description of methods and their validation against 12-d continuous whole-body calorimetry. Am J Clin Nutr. 1993;58:455 – 462.

[77] Webb P, Annis J, Troutman, JR S. Human calorimetry with a water-cooled garment. J Appl Physiol. 1972;32(3):412 – 418.

[78] Webb P. 24-hour energy expenditure and the menstrual cycle. Am J Clin Nutr. 1986;44:614 – 619.

[79] Webb P, Sangal S. Sedentary daily expenditure: a base for estimating individual energy requirements. Am J Clin Nutr. 1991;53:606 – 611.

[80] Hambræus L, Sjödin A, Webb P, Forslund A, Hambræus K, Hambræus T. A suit calorimeter for energy-balance studies on humans during heavy exercise. Eur J Appl Physiol. 1994;68:68 – 73.

[81] Seale J, Rumpler W, Conway J, Miles C. Comparison of doubly labeled water, intake-balance, and direct- and indirect-calorimetry methods for measuring energy expenditure in adult men. Am J Clin Nutr. 1990;52:66 – 71.

[82] Cole TJ, Bellizzi M, Flegal K, Dietz W. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ. 1990;320:1240 – 1245.

[83] Rowett Research Institute. Fact Sheet 1: Energy expenditure;. Accessed May 2007. Electronic resource, `http://www.rowett.ac.uk/edu_web/sec_pup/energy_expenditure.pdf`.

[84] Nelson K, Weinsier R, Long C, Schutz Y. Prediction of resting energy expenditure from fat-free mass and fat mass. Am J Clin Nutr. 1992;56:848 – 856.

[85] Fujii T, Phillips B. Quick review: The metabolic cart. Internet J Intern Med. 2003;3(2).

[86] Harris J, Benedict F. A Biometric Study of the Basal Metabolism in Man. Washington, DC: Carnegie Institute of Washington;. Publication No 279.

[87] University of Bristol. Avon Longitudinal Study of Parents and Children. http://www.alspac.bris.ac.uk/welcome/index.shtml; 2006. Accessed 10 Nov 2006.

[88] Flancbaum L, Choban P, Sambucco S, Verducci J, Burge J. Comparison of indirect calorimetry, the Fick method, and prediction equations in estimating the energy requirements of critically ill patients. Am J Clin Nutr. 1999;69:461 – 469.

[89] Benedict F. Factors affecting basal metabolism. J Biol Chem. 1915;20:263 – 299.

[90] Cunningham J. A reanalysis of the factors influencing basal metabolic rate in normal adults. Am J Clin Nutr. 1980;33:2372 – 2374.

[91] Nielsen S, Hensrud D, Romanski S, Levine J, Burguera B, Jensen M. Body composition and resting energy expenditure in humans: role of fat, fat-free mass and extracellular fluid. Int J Obes. 2000;24:1153 – 1157.

[92] Cunningham J. Body composition as a determinant of energy expenditure: a synthetic review and a proposed general prediction equation. Am J Clin Nutr. 1991;54:963 – 969.

[93] Mifflin M, St Jeor S, Hill L, Scott B, Daugherty S, Koh Y. A new predictive equation for resting energy expenditure in healthy individuals. Am J Clin Nutr. 1990;51:241 – 247.

[94] Ravussin E, Lillioja S, Anderson T, Christin L, Bogardus C. Determinants of 24-hour energy expenditure in man. Methods and results using a respiratory chamber. J Clin Invest. 1986;78:1568 – 1578.

[95] Frankenfield D, Muth E, Rowe W. The Harris-Benedict studies of human basal metabolism: History and limitations. J Am Diet Assoc. 1998;98:439 – 445.

[96] Kaplan A, Zemel B, Neiswender K, Stallings V. Resting energy expenditure in clinical pediatrics: Measured versus prediction equations. J Pediatr. 1995;127:200 – 205.

[97] Schofield W. Predicting basal metabolic rate, new standards and review of previous work. Clin Nutr. 1985;39c(1s):5 – 42.

[98] World Health Organisation. Energy and protein requirements. Report of a joint FAO/WHO/UNU expert consultation. 1985;WHO Technical Report Series(No 724). Geneva: World Health Ogranisation.

[99] Maffeis C, Schutz Y, Micciolo R, Zoccante L, Pinelli L. Resting metabolic rate in six- to ten-year-old obese and non-obese children. J Pediatr. 1993;122:556 – 562.

[100] Henry CJK. Basal metabolic rate studies in humans: measurement and development of new equations. Pub Health Nutr. 2005;8(7A):1133 – 1152.

[101] Rodriguez G, Moreno L, Sarria A, Fleta J, Bueno M. Resting energy expenditure in children and adolescents: agreement between calorimetry and prediction equations. Clin Nutr. 2000;21(3):255 – 260.

[102] Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;i:307 – 310.

[103] Ramirez-Zea M. Validation of three predictive equations for basal metabolic rate in adults. Pub Health Nutr. 2005;8(7A):1213 – 1228.

[104] Rubner M. Über den einfluss der körpergrösse auf stoff- und kraftwechsel. Zeitschrift für Biologie. 1883;19:536 – 562.

[105] Kleiber M. Body size and metabolism. Hilgardia. 1932;6:315 – 353.

[106] Wang Z, O'Connor T, Heshka S, Heymsfield S. The reconstruction of Kleiber's law at the organ tissue level. J Nutr. 2001;131:2867 – 2970.

[107] Smil V. Laying down the law. Every living thing obeys the rules of scaling discovered by Max Kleiber. Nature. 2000;403:597.

[108] Activity, Energy Expenditure and Energy Requirements of Infants and Children. International Dietary Energy Consultancy Group; 1989.

[109] Dodds P, Rothman D, Weitz J. Re-examination of the "3/4-law" of metabolism. J Theor Biol. 2001;209:9 – 27.

[110] Encyclopedia Britannica. Human Nutrition. Encyclopedia Britannica Online. 2007;Accessed 20/08/2007 from http://www.britannica.com/eb/article-250632.

[111] G G. Nutrition Bulletin. In: Calorimetry 2000: Symposium on Energy Regulation Research. vol. 26. University of Maastricht. British Nutrition Foundation; 2001. p. 95 – 97.

[112] Weststrate J. Resting metabolic rate and diet-induced thermogenesis: a methodological reappraisal. Am J Clin Nutr. 1993;58:592 – 601.

[113] Westerterp KR. Diet induced thermogenesis. Nutr Metab. 2004;1:5.

[114] Gougeon R. Effect of insulin and energy restriction on the thermic effect of protein in type 2 diabetes mellitus. Obes Res. 2001;9(4):241 – 250.

[115] Poizot-Martinand K Benourine I, Philibert P, Manelli J, Gastaut J. Diet-induced thermogenesis is increased in HIV infected patients. vol. 9 of Int Conf AIDS. France; 1993. p. 530. Abstract no PO-B36-2367.

[116] Tataranni P, Larson D, Snitket S, Ravussin E. Thermic effect of food in humans: methods and results from use of a respiratory chamber. Am J Clin Nutr. 1995;61:1013 – 1019.

[117] Reed G, Hill J. Measuring the thermic effect of food. Am J Clin Nutr. 1996;63:154 – 169.

[118] Granata G, Brandon L. The thermic effect of food and obesity: Discrepant results and methodological variations. Nutr Rev. 2002;60(8):223 – 233.

[119] Visser M, Duerenberg P, van Staveren W, Hautvast J. Resting metabolic rate and diet-induced thermogenesis in young and elderly subjects: relationship with body composition, fat distribution, and physical activity level. Am J Clin Nutr. 1995;61:772 – 778.

[120] Maffeis C, Schutz Y, Grezzani A, Provera S, Piacentini G, Tatò L. Meal-induced thermogenesis and obesity: Is a fat meal a risk factor for fat gain in children? J Clin Endocr Metab. 2001;86:214 – 219.

[121] Laville M, Cornu C, Normand S, Mithieux G, Beylot M, , et al. Decreased glucose-induced thermogenesis at the onset of obesity. Am J Clin Nutr. 1993;57:851 – 856.

[122] Schutz Y, Bessard T, Jéquier E. Diet-induced thermogenesis measured over a whole day in obese and nonobese women. Am J Clin Nutr. 1984;40:542 – 552.

[123] Segal K, Lacayanga I, Dunaif A, Gutin B, Pi-Sunyer F. Impact of body fat mass and percent fat on metabolic rate and thermogenesis in men. Am J Physiol. 1989;256:E573 – E579.

[124] Sharief N, Macdonald I. Differences in dietary-induced thermogenesis with various carbohydrates in normal and overweight men. Am J Clin Nutr. 1982;35:267 – 272.

[125] Swaminathan R, King R, Holmfield J, Siwek R, Baker M, Wales J. Thermic effect of feeding carbohydrate, fat, protein and mixed meal in lean and obese subjects. Am J Clin Nutr. 1985;p. 177 – 181.

[126] D'Alessio D, Kavle E, Mozzoli M, Smalley K, Polansky M, Kendrick Z, et al. Thermic effect of food in lean and obese men. J Clin Invest. 1988;81:1781 – 1789.

[127] Steinger J, Karst H, Noack R, Steglich R. Diet-induced thermogenesis in man: thermic effects of single protein and carbohydrate test meals in lean and obese subjects. Ann Nutr Metab. 1987;31:117 – 125.

[128] Morgan J, York D. Thermic effect of feeding in relation to energy-balance in elderly men. Ann Nutr Metab. 1983;27:71 – 77.

[129] Schwartz R, Jaeger L, Veith R. The thermic effect of feeding in older men: The importance of the sympathetic system. Metabolism. 1990;39:733 – 737.

[130] Kinabo J, Durnin J. Thermic effect of food in man: effect of meal composition and energy content. Brit J Nutr. 1990;64:37 – 44.

[131] Westerterp KR, Wilson S, Rolland V. Diet induced thermogenesis measured over 24h in a respiration chamber: effect of diet composition. Int J Obes. 1999;23:287 – 292.

[132] Schwartz R, Ravussin E, Massari M, O'Connell M, Robbins D. The thermic effect of carbohydrate versus fat feeding in man. Metabolism. 1985;34(4):285 – 293.

[133] Zahorska-Markiewicz1 B. Thermic effect of food and exercise in obesity. Eur J Appl Physiol. 1980;44(3):231 – 235.

[134] Das S, Saltzman E, McCrory M, Hsu L, Shikora S, Dolnikowski G, et al. Energy expenditure is very high in extremely obese women. J Nutr. 2004;134:1412 – 1416.

[135] de Jonge L, Bray G. The thermic effect of food is reduced in obesity. Nutr Rev. 2002;60(9):295 – 297.

[136] Armellini F, Zamboni M, Mino A, Bissoli L, Micciolo R, Bosello O. Postabsorptive resting metabolic rate and thermic effect of food in relation to body composition and adipose tissue distribution. Metabolism. 2000;49(1):6 – 10.

[137] Bray G. Epidemiology, risks and pathogenesis of obesity. Meat Science. 2005;71(1):2 – 7.

[138] Wrigglesworth J. Energy and Life. Modules in Life Sciences. Taylor & Francis Ltd; 1997.

[139] Moore LL, Gao D, Bradlee ML, Couples LA, Sundarajan-Ramamurti A, Proctor MH, et al. Does early physical activity predict body fat change throughout childhood? Prev Med. 2003;37:10 – 17.

[140] Hancox R, Milne B, Poulton R. Association between child and adolescent television viewing and adult health: a longitudinal birth cohort study. Lancet. 2004;364:257 – 262.

[141] Delany JP. Role of energy expenditure in the development of pediatric obesity. Am J Clin Nutr. 1998;68(suppl):950s – 955s.

[142] Goran MI, Shewchuk R, Gower B, Nagy T, Carpenter W, Johnson R. Longitudinal changes in fatness in white children: no effect of childhood energy expenditure. Am J Clin Nutr. 1998;67:309–316.

[143] Sallis JF, Patterson TL, Buono MJ, Nader PR. Relation of cardiovascular fitness and physical activity to cardiovascular disease risk factors in children and adults. Am J Epidemiol. 1988;127(5):933 – 41.

[144] Anderson P, Butcher K. Childhood obesity: Trends and potential causes. Future Child. 2006;16(1):19 – 45.

[145] Maffeis C, Zaffanello M, Pinelli L, Schutz Y. Total energy expenditure and patterns of activity in 8 - 10 year old obese and nonobese children [original article]. J Pediatr Gastr Nutr. 1995;23(3):256 – 261.

[146] Moore L, Nguyen U, Rothman K, Cupples L, Ellison R. Preschool physical activity level and change in body fatness in young children: The Framingham Children's Study. Am J Epidemiol. 1995;142(9):982 – 988.

[147] Obarzanek E, Schreiber G, Crawford P, Goldman S, Barrier P, Frederick M, et al. Energy intake and physical activity in relation to indexes of body fat: the National Heart, Lung, and Blood Institute Growth and Health Study. Am J Clin Nutr. 1994;60:15 – 22.

[148] Berkey C, Rockett H, Field A, Gillman M, Frazier A, Jr CC, et al. Activity, dietary intake, and weight changes in a longitudinal study of preadolescent and adolescent boys and girls. Pediatrics. 2000;104(4):56 – 65.

[149] Berkey C, Rockett H, Gillman M, Colditz G. One-year changes in activity and in inactivity among 10-15 year old boys and girls: Relationship to change in body mass index. Pediatrics. 2003;111(4):836 – 843.

[150] Norman A, Drinkard D, McDuffie J, Ghorbani S, Tanoff L, Yanovski J. Influence of excess adiposity on exercise fitness and performance in overweight children and adolescents. Pediatrics. 2005;115:e690 – e696.

[151] Sirard J, Russell R. Physical activity assessment in children and adolescents. Sports Med. 2001;31(6):439 – 454.

[152] Washburn R, Montoye H. The assessment of physical activity by questionnaire. Am J Epidemiol. 1986;123(4):563 – 576.

[153] Sallis J, Buono M, Roby J, Micale F, Nelson J. Seven-day recall and other physical activity self reports in children and adolescents. Med Sci Sports Ex. 1993;25(1):99 – 108.

[154] Bar-Or O. A commentary to children and fitness: a public health perspective. Res Quart Ex Sport. 1987;58:304 – 307.

[155] Ekelund U, Sjöström M, Yngve A, Poorvliet E, Nilsson A, Froberg K, et al. Physical activity assessed by activity monitor and doubly labeled water in children. Med Sci Sports Ex. 2001;33(2):275 – 281.

[156] Trost S. Objective measurement of physical activity in youth: current issues, future directions. Ex Sport Sci Rev. 2001;29(1):32 – 36.

[157] Durnin J, Passmore R. Energy, Work and Leisure. Heinemann Educational Books Ltd; 1967.

[158] Schutz Y, Weinsier R, Hunter G. Assessment of free-living physical activity in humans: An overview of currenty available and proposed new measures. Obes Res. 2001;9(6):368 – 379.

[159] Livingstone M. Heart-rate monitoring: the answer for assessing energy expenditure and physical activity in population studies? Brit J Nutr. 1997;78(6):869 – 871.

[160] Eston R, Rowlands A, Ingledew D. Validity of heart rate, pedometry, and accelerometry for predicting the energy cose of children's activities. J Appl Physiol. 1998;84(1):362 – 371.

[161] Wareham N, Hennings S, Prentice A, Day N. Feasability of heart-rate monitoring to estimate total level and pattern of energy expenditure in a population-based epidemiological study: the Ely young cohort feasibility study 1994 - 5. Brit J Nutr. 1997;78:889 – 900.

[162] Bussman J, Martens W, Tulen J, Schasfoort F, van den Berg-Emons H, Stam H. Measuring daily behaviour using ambulatory accelerometry: The activity monitor. Behav Res Meth Ins C. 2001;33(3):349 – 356.

[163] Westerterp KR. Physical activity assessment with accelerometers. Int J Obes. 1999;23(Suppl 3):S45 – S49.

[164] Meijer G, Westerterp KR, Koper H, ten Hoor F. Assessment of energy expenditure by recording heart rate and body acceleration. Med Sci Sports Ex. 1989;221:343 – 347.

[165] Ainslie P, Reilly T, Westerterp KR. Estimating human energy expenditure: a review of techniques with particular reference to doubly labelled water. Sports Med. 2003;33(9):683 – 698.

[166] Bar-Or O, Foreyt J, Bouchard C, Bromwell K, Dietz W, Ravussin E, et al. Physical activity, genetic, and nutritional considerations in childhood weight management. Med Sci Sports Ex. 1998;30(1):2 – 10.

[167] International Dietary Consultancy Group. Activity, Energy Expenditure and Energy Requirements of Infants and Children. Proceedings of an I/D/E/C/G Workshop held in Cambridge, USA; 1985.

[168] Torun B. Energy cost of various physical activities in healthy children. In: Schurch B, Scrimshaw N, editors. Activity, energy expenditure and energy requirements of infants and children in Cambridge, Massachusetts, USA. Massacheusetts, USA: International Dietary Energy Consultancy Group; 1990. .

[169] Bratteby L, Sandhagen B, Lötborn M, Samuelson G. Daily energy expenditure and physical activity assessed by an activity diary in 374 randomly selected 15-year-old adolescents. Eur J Clin Nutr. 1997;51:592 – 600.

[170] Beateby L, Sandhagen B, Fan H, Samuelson G. A 7-day activity diary for assessment of daily energy expenditure validated by the doubly labelled water method in adolescents. Eur J Clin Nutr. 1997;51:585 – 591.

[171] Westerterp KR. Assessment of physical activity level in relation to obesity: current evidence and research issues. Med Sci Sports Ex. 1999;31(11, supplement 1):S522 – S525.

[172] Bandini L. Validity of reported energy intake in obese and nonobese adolescents. Am J Clin Nutr. 1990;52(3):421 – 425.

[173] Braam L, Ocké M, Bueno-de-Mesquita H, Seidell J. Determinants of obestity-related under-reporting of energy intake. Am J Epidemiol. 1998;147(11):1081 – 1086.

[174] Fisher J, Johnson R, Lindquist C, Birch L, Goran MI. Influence of body composition on the accuracy of reported energy intake in children. Obes Res. 2000;8(8):597 – 603.

[175] Probst Y, Tapsell L. Over- and under-reporting of energy intake by patients with metabolic syndrome using an automated dietary assessment website. Nutr Diet. 2007;64:280 – 284.

[176] Kramer M. Do breast-feeding and delayed introduction of solid foods protect against subsequent obesity? J Pediatr. 1981;98:883 – 887.

[177] Armstrong J, Reilly JJ, the Child Health Information Team. Breastfeeding and lowering the risk of childhood obesity. The Lancet. 2002;359:2003 – 2004.

[178] Dewey K. Is breastfeeding protective against child obesity? J Hum Lact. 2003;19(1):9 – 18.

[179] Owen G, Martin R, Whincup P, Smith G, Cook D. Effect of infant feeding on the risk of obesity across the life course: a quantative review of published evidence. Pediatrics. 2005;115:1367 – 1377.

[180] Hoddinott P, Tappin D, Wright C. Breast feeding. BMJ. 2008;336:881 – 887.

[181] Singhal A, Lanigan J. Breastfeeding, early growth and later obesity. Obes Res. 2007;8(Suppl 1):51 – 54.

[182] Günther A, Remer T, Kroke A, Buyken A. Early protein intake and later obesity risk: which protein sources at which time points throughout infancy and childhood are important for body mass index and body fat percentage at 7 y of age? Am J Clin Nutr. 2007;86:1765 – 1772.

[183] Ong K, Preece M, Emmett P, Ahmed M, Dunger D. Size at birth and early childhood growth in relation to maternal smoking, parity and infant breastfeeding: longitudinal birth cohort study and analysis. Pediatr Res. 2002;52:863 – 867.

[184] Settler N, Stallings V, Troxel A, Zhao J, Schinnar R, Nelson S, et al. Weight gain in the first week of life and overweight in adulthood: a cohort study of European American subjects fed infant formula. Circulation. 2005;111:1897 – 1903.

[185] Bogen D, Hanusa B, Whitaker R. The effect of breast-feeding with and without formula use on the risk of obesity at 4 years of age. Obes Res. 2004;12:1528 – 1536.

[186] Grummer-Strawn L, Mei Z. Does breastfeeding protect against pediatric overweight? Analysis of longitudinal data from the centers for disease control and prevention of pediatric nutrition surveillance system. Pediatrics. 2005;113:e81 – e86.

[187] Woo J, Dolan L, Morrow A, Geraghty S, Godman E. Breastfeeding helps explain racial and socioeconomic status disparities in adolescent adiposity. Pediatrics. 2008;121(3):e458 – e465.

[188] Li L, Parsons T, Power C. Breast feeding and obesity in childhood: cross sectional study. BMJ. 2003;327:304 – 305.

[189] Woodward-Lopez G, Davis Ritchie L, Gerstein D, Crawford P. Obesity. Dietary and Developmental Influences. CRC Press; 2006.

[190] Hohepa M, Schofield G, Kolt G. Adolescent obesity and physical inactivity. NZ Med J. 2004;117(1207).

[191] Jago R, Baranowski T, Baranowski JS, Thompson D, Greaves KA. BMI from 3-6y of age is predicted by TV viewing and physical activity, not diet. Int J Obes. 2005;29:557 – 564.

[192] Labib M. The investigation and management of obesity. J Clin Pathol. 2003;56:17 – 25.

[193] Weinsier R, Nelson K, Hensrud D, Darnell B, Hunter G, Shutz Y. Metabolic predictors of obesity. J Clin Invest. 1995;95:980 – 985.

[194] Stunkard A, Berkowitz R, Stallings V, Schoeller D. Energy intake, not energy output, is a determinant of body size in infants. Am J Clin Nutr. 1999;69:524 – 530.

[195] Atkin L, Davies P. Diet composition and body composition in preschool children. Am J Clin Nutr. 2000;72:15 – 21.

[196] Deheeger M, Rolland-Cachera M, Fontvielle A. Physical activity and body composition in 10 year old French children: linkages with nutritional intake? Int J Obes. 1986;21:372 – 379.

[197] Stubbs C, Lee A. The obesity epidemic: both energy intake and physical activity contribute. Med J Aust. 2004;181(9):489 – 491.

[198] Prentice A, Jebb S. Obesity in Britain: gluttony or sloth? BMJ. 1995;11:437 – 439.

[199] E ER, Lillioja S, Knowler W, Christin L, Freymond D, Abbott W, et al. Reduced rate of energy expenditure as a risk factor for body-weight gain. New England J Med. 1988;318(8):467 – 472.

[200] Dorn L. Measuring puberty. J Adolescent Health. 2006;39:625 – 626.

[201] Dorn L, Dahl R, Woodward R, Biro F. Defining the boundaries of early adolescence: a user's guide to assessing pubertal status and pubertal timing in research with adolescents. Appl Devel Sci. 2006;10(1):30 – 56.

[202] Marshall W, Tanner J. Variations in pattern of pubertal changes in boys. Arch Dis Child. 1970;45:13 – 23.

[203] Marshall W, Tanner J. Variations in pattern of pubertal changes in girls. Arch Dis Child. 1969;44:291 – 303.

[204] Coleman L, Coleman J. The measurement of puberty: A review. J Adolesc. 2002;25:535 – 550.

[205] Duke P, Litt I, Gross R. Adolescents' self-assessment of sexual maturation. Pediatrics. 1980;66:918 – 920.

[206] Brooks-Gunn J, Warren M, Rosso J, Gargiulo J. Validity of self-report measures of girls' pubertal status. Child Devel. 1987;58:829 – 841.

[207] Bonat S, Pathomvanich A, Keil M, Field A, Yanovski J. Self-assessment of pubertal stage in overweight children. Pediatrics. 2002;110:743 – 747.

[208] Dorn L, Susman E, Nottleman E, Innof-Germain G, Chrousos G. Perceptions of puberty: Adolescent, parent and health care personnel. Devel Psych. 1990;26(2):322 – 329.

[209] Buckler J. A Longitudinal Study of Adolescent Growth. London: Springer-Verlag; 1990.

[210] Brooks-Gunn J, Peterson A, Eichorn D. The study of maturational timing effects in adolescence. J Youth Adolesc. 1985;14(3):149 – 161.

[211] Karlberg J, Kwan C, Gelander L, Albertsson-Wikland K. Pubertal growth assessment. Hormone Research. 2003;60(suppl 1):27 – 35.

[212] Liu Y, Albertsson-Wikland K, Karlberg J. New reference for the age at childhood onset of growth and secular trend in the timing of puberty in Swedish. Acta Pædiatrica. 2000;89:637 – 643.

[213] Cole T, Green P. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med. 1992;11:1305 – 1319.

[214] Box G, Cox D. An analysis of transformations. J Royal Stat Soc, Series B. 1964;26(2):211 – 252.

[215] Ravishanker N, Dey D. A First Course in Linear Model Fitting. Texts in Statistical Science. Chapman and Hall / CRC; 2002.

[216] Jobson J. Applied Multivariate Data Analysis. vol. 1: Regression and Experimental Design. Springer-Verlag New York, Inc; 1991.

[217] Fox J. Nonlinear regression and nonlinear least squares. Appendix to An R and S-PLUS Companion to Applied Regression; 2002.

[218] Gallant A. Nonlinear regression. Am Stat. 1975;29(2):73 – 81.

[219] Kelly C. Iterative Methods for Optimization. Philadelphia: Society for Industrial and Applied Mathematics; 1999.

[220] Weisstein E. Nonlinear Least Squares Fitting; (Accessed 27/07/2012. From MathWorld–A Wolfram Web Resource.

[221] Jiang J. Linear and generalized linear mixed models and their applications. Springer series in statistics. Springer; 2007.

[222] Fox J. Linear mixed models. Appendix to An R and S-PLUS Companion to Applied Regression; 2002.

[223] Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. Springer Series in Statistics. New York: Springer; 2000.

[224] Bates D, Maechler M, Bolker B. lme4: Linear mixed-effects models using S4 classes; 2012. R package version 0.999999-0. Available from: http://CRAN.R-project.org/package=lme4.

[225] Uebersax J. Intraclass Correlation and Related Methods; 2006. Accessed May 2012. Webpage.

[226] de Vet H, Terwee C, Mokkink L, Knol D. Measurement in Medicine. Practical Guides to Biostatistics and Epidemiology. Cambridge University Press; 2011.

[227] Bland J, Altman D. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8:135 – 160.

[228] Goodman L, Kruskal W. Measures of Association for Cross Classifications. Springer Series in Statistics. Springer-Verlag; 1979.

[229] Kendall M. Rank Correlation Methods. Hafner Publishing Co., New York; 1955.

[230] Healey J, Prus S. Statistics: A Tool for Social Research, 1st edition. Nelson Education Ltd.; 2010.

[231] Blake C. Cross national research methods. James Madison University, Harrisonburg, VA, Class INTA-295; 2004.

[232] Kaplan E, Meier P. Estimation from incomplete observations. J Am Stat Assoc. 1958;53(282):457 – 481.

[233] Klein J, Moeschberger M. Survival Analysis: Techniques for Censored and Truncated Data. Statistics for Biology and Health. Springer-Verlag New York, Inc.; 1997.

[234] Horton N, Kleinman K. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. Am Stat. 2007;61(1):79 – 90.

[235] Little R, Rubin D. Statistical analysis with missing data. Wiley-Blackwell; 2002.

[236] Rubin D. Multiple Imputation for Nonresponse in Surveys. Wiley-Blackwell; 1987.

[237] Schafer J. Multiple imputation: a primer. Stat Methods Med Res. 1999;8(1):3 – 15.

[238] van Buuren S. Flexible imputation of missing data. Interdisciplinary statistics. Chapman and Hall/CRC; 2012.

[239] Azur M, Stuart E, Frangakis C, Leaf P. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011;20(1):40 – 49.

[240] van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45(3):1–67. Available from: http://www.jstatsoft.org/v45/i03/.

[241] Centers for Disease Control and Prevention (CDC) National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. http://www.cdc.gov/nchs/nhanes.htm: Hyattsville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention; Updated September 2008; Accessed September 2008.

[242] Columbia University. Obesity Research Center. http://www.nyorc.org/index.html; 2009. Accessed March 2009.

[243] Fomon S, Haschke F, Ziegler E, Nelson S. Body composition of reference children from birth to age 10 years. Am J Clin Nutr. 1982;35:1169 – 1175.

[244] Weir J. New methods for calculating metabolic rate with special reference to protein metabolism. J Physiol. 1949;109:1 – 9.

[245] Butte N, Christiansen E, Sørensen T. Energy-imbalance underlying the development of childhood obesity. Obesity. 2007;15(12):3056 – 3066.

[246] Penpraze V, Reilly JJ, MacLean C, Montgomery C, Kelly L, Paton J, et al. Monitoring of physical activity in young children: how much is enough? Pediatr Ex Sci. 2006;18:483 – 491.

[247] Templ M, Alfons A, Kowarik A, Prantner B. VIM: Visualization and Imputation of Missing Values; 2013. R package version 3.0.3 http://CRAN.R-project.org/package=VIM.

[248] ;.

[249] Rubin D. Multiple imputation after 18+ years (with discussion). J Am Stat Assoc. 1996;91(434):473 – 489.

[250] Westerterp KR, Donkers JHHL, Fredrix EWHM, Boekhoudt P. Energy intake, physical activity and body weight: A simulation model. Brit J Nutr. 1995;73:337 – 347.

[251] ALSPAC. Bristol University , Avon Longitudinal Study of Parents and Children, Questionnaires; Updated January 2010; Accessed September 2010. Electronic resource, `http://www.bristol.ac.uk/alspac/sci-com/quests/`.

[252] Wells JCK. A critique of the expression of paediatric body composition data. Arch Dis Child. 2001;85:67 – 72.

Appendix A

Diagnostic plots

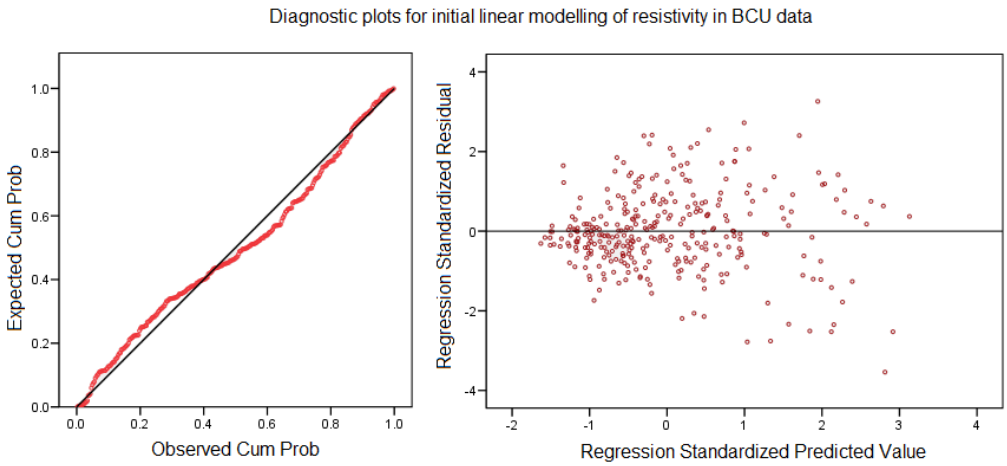

Figure A.1: Diagnostic plots for initial linear modelling of resistivity in BCU data (page 120)

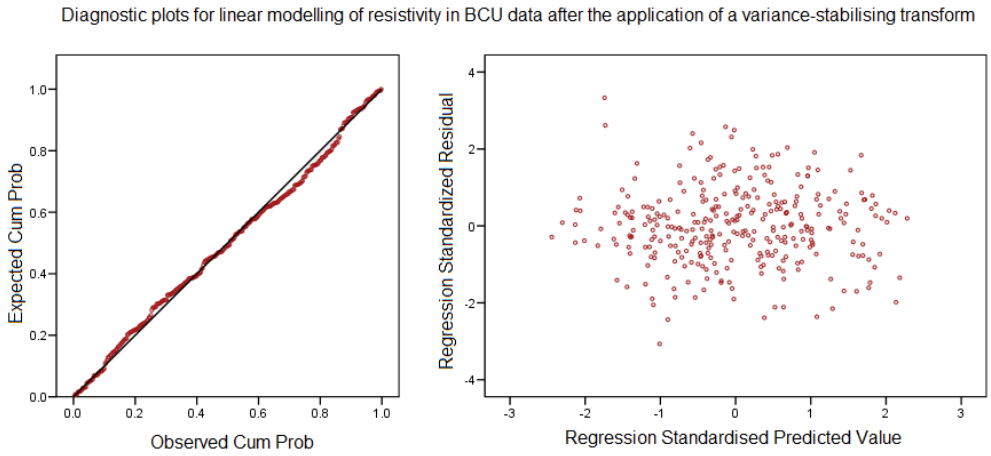

Figure A.2: Diagnostic plots for transformed model of resistivity in BCU data (page 120)

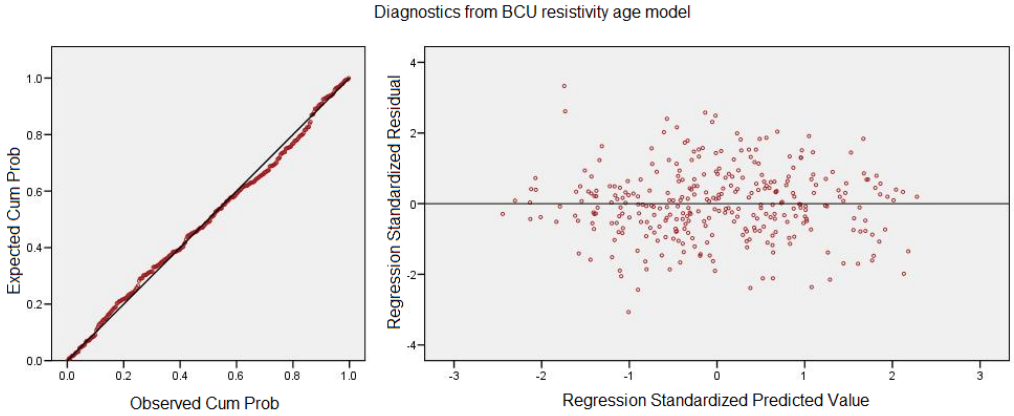

Figure A.3: Diagnostics from resistivity model with age (page 124)

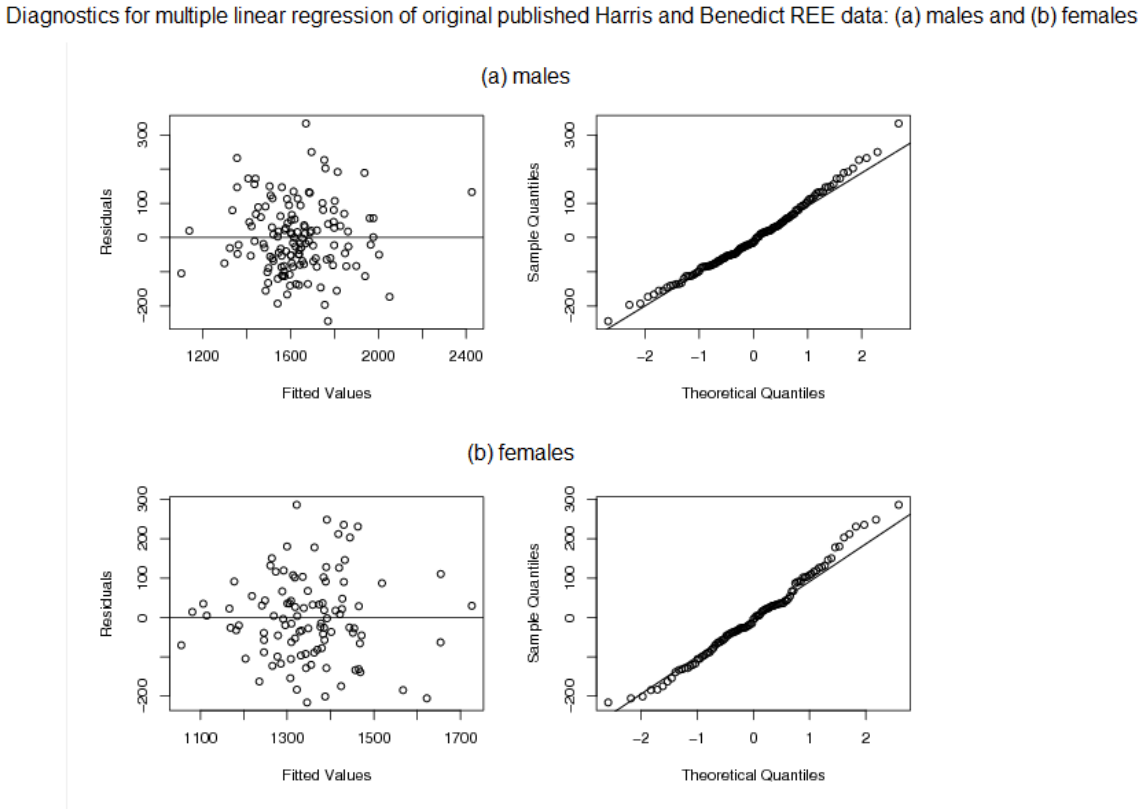

Figure A.4: Diagnostics for multiple linear regression: HB energy expenditure data (page 140)

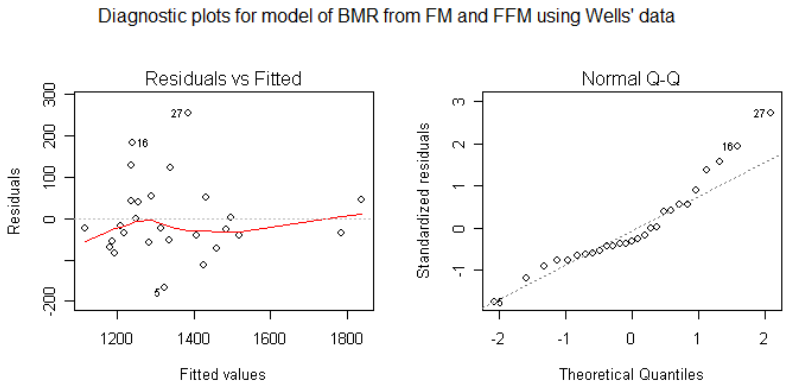

Figure A.5: Diagnostics for model of BMR from FM and FFM (excluding sex) using Wells' data (page 168)