# Statistical Analysis of Child Growth Data

A thesis presented for the degree

of Doctor of Philosophy at the

University of Durham

Jennifer Argyle

*Department of Mathematical Sciences,*

*University of Durham,*

*Durham, DH1 3LE.*

May 2002



# Statistical Analysis of Child Growth Data

*Submitted for degree of Doctor of Philosophy by Jennifer Argyle*

The study of child growth is complex. There are many clinical questions to answer but not necessarily the statistical methodology to deal with these questions. Human growth begins at conception and continues into adult life.

In chapter 1 we discuss the characteristics of the growth process from conception to maturity and the purpose of growth monitoring. In chapter 2 we summarise the mathematical approaches to growth data. In chapter 3 we summarise the approaches that have been used to detect growth faltering. In this chapter we introduce the conditional gain Z-score.

The data set analysed within this thesis is from the Newcastle growth and development study. In infancy we have routine weights of 3415 term infants. A sub-sample of these infants were followed-up at 7-9 years as part of a research study. These children belonged to three subgroups: cases were children that were defined as failing to thrive in infancy, controls were matched to cases and a 20% systematic sample. The school entry data of the sub-sample followed at 7-9 years were retrieved from school health records.

In chapter 4 we carry out a preliminary analysis of the routine infancy weight Z-scores. The infancy data provided the opportunity to generate the correlation structure of routine weight Z-scores in infancy. In chapter 5 we develop a model for this correlation structure. In chapter 7 we explore patterns in the conditional weight gain Z-scores and also suggest some alternative criteria for identifying growth faltering in infancy.

In chapters 6, 8 and 9 we analyse the anthropometric data obtained at follow-up and school entry. In childhood, the conditional gain Z-score is used to contrast height with mid-parental height and height at follow-up with height at school entry. The anthropometric data of the case and control children will be compared.

*To my Nanna, Agnes Argyle (1918-1996). She is sadly missed but her strength of character will be with me forever.*

# Contents

# List of Tables

# List of Figures

# Declaration

This thesis is the result of research carried out between October 1995 and August 2000, and between January 2002 and April 2002 in the department of mathematical sciences at the University of Durham. I declare that the material presented in this thesis has not been submitted previously for any degree in either this or any other University. Chapters 1 to 3 form a literature review and as such no claim for originality is made. The rest of the work for this thesis is entirely my own, unless stated otherwise. The statistical software used to carry out the analysis in this thesis is the statistical programming language R discussed in Ihaka and Gentleman (1996).

# Acknowledgements

I wish to thank both of my supervisors David Wooff and Allan Seheult for their patience, guidance and support in both an academic and personal context. Many thanks to Charlotte Wright (University of Glasgow) and Sally Corbett (University of Durham) for allowing me to use their data and answering my numerous queries. I offer my gratitude to both Charlotte Wright and Robert Drewett (University of Durham) for shaping my understanding of the condition of failure-to-thrive and their lively discussions on the food versus love debate.

I wish to thank the University of Durham for financial support whilst I was a teaching assistant in the department of mathematical sciences. I express my gratitude to the department of mathematical sciences for the studentship I received for my last six months of supervised study.

We are all a product of our own histories. I wish to acknowledge my numerous school teachers for their encouragement and never doubting once my capabilities. In particular, Mr Henry for sparking my interest in statistics at the age of 16 years. Thank you to the staff and my fellow M.Sc. students within the probability and statistics department at the University of Sheffield for providing me with both the knowledge and skills for analysing data. I wish to thank my colleagues in the probability and statistics section for their support. In particular, the late Dyfrig Williams for his companionship and patience when I was learning the S-PLUS language.

Finally, I wish to thank my family and friends for their support and encouragement over the years. Many thanks to my Mum and Dad for supporting my ambition to go to University. To my sisters, Nicola and Claire for just being themselves! A big thank you to my friends: Sharon John, Susan and Ian Stewart, Tammy Long, Ailsa Carrick, Mark Pilling, Sonya Warne, Allison Cook, Miranda Morgan and my numerous friends from St. Mary's College.

# Chapter 1

# Clinical issues in the study of child growth

The study of child growth is complex. There are many clinical questions to answer but not necessarily the statistical methodology to deal with these questions. Human growth begins at conception and continues into adult life. Is this process regular and predictable? What influences a child's path to final adult size? From a clinicians point of view, the main question is: given this child's weight or height at this age, is this child 'normal'. Does their pattern of weight or height gain cause concern? Many of the approaches in use only consider one or two weight (or height) measures. Growth is not a process that occurs only between two points, so the primary issue is how we interpret weight or height measures as they evolve.

The interest in child growth spans many disciplines: statistics, medicine, nutrition, psychology, education and anthropology. In statistics the study of 'growth models' initiated by Potthoff and Roy (1964) has become a field in its own right. However the majority of work developed in this area is only applicable to individuals with complete data and with all measurements taken at the same time-points. Even in a research study setting it is unlikely to arrive at data of this quality.

The monitoring of the growth process raises statistical complexities in itself. Consider weight and height, which are usually monitored at school entry. The actual height or weight measures are subject to error, either by the process of measuring or biological variation[1]. In a research study setting there are issues of missing data, weight and height measures that are correlated, interpretation of the longitudinal

[1]For example: humans are taller in the morning than in the evening. This is termed diurnal variation.

element and expected variability of measures themselves. In a routine setting, there are many more issues: irregularly spaced data, height and weight measures are themselves subject to more error, 'missing' data may be absent for a reason, variability in the number of weight measures and potential sources of bias because light infants may be weighed more frequently, short or heavy children monitored more closely.

In section 1.1 we look at the general pattern of growth for the typical child. The growth process from birth to adulthood is briefly described and the observed differences in males and females highlighted. Although the discussion in this section covers all phases of the growth process, the majority of the work within this thesis focuses on the infancy period from birth to about 18 months. Section 1.2 discusses the need for a growth reference and introduces the revised UK 1990 reference (Cole et al. 1998). Within this same section we also stress the advantages of working on a Z-score scale in contrast to the other alternatives. In section 1.3 we discuss the influence of hereditary and environment on growth. We also discuss the main disorders of growth, placing emphasis on growth faltering in infancy, termed failure-to-thrive. Finally, in section 1.4, we conclude with a summary of the current guidelines with regards to growth monitoring.

## 1.1 Physical growth from conception to maturity

### 1.1.1 Characteristics of the human growth curve

In general when we talk about the curve of human growth we are referring to the growth in height. However the process of a child's growth is intricate; it doesn't only involve an increase in linear dimensions. For example, over the course of a child's life (and even into adulthood), a child's weight will, in general, increase. This in itself reflects changes in body composition[2], lengthening of bones (and thus increase in bone weight) and biological variation[3]. Nevertheless, most body measurements as good as follow the growth curves described below for height (Tanner 1989).

> Physical growth is a complex, stepwise process that requires the orchestration of hormonal regulatory systems, adequate energy and other nutritional resources, and the necessary antecedent bi-

[2]In nutrition literature, body composition refers to relative proportions of fat, lean body mass and water that contribute to overall weight.

[3]Biological variation refers to fluctuations in body weight observed due to foods and liquids consumed, as well as material excreted.

> ological events. ... Ongoing growth is not only dependent upon current and future physiological and nutritional resources but also upon past physiological experiences (Woolston 1991, pp11).

Definition of height velocity

$$V = \frac{H_2 - H_1}{\Delta t} \tag{1.1}$$

where $H_1$ and $H_2$ are the two height measurements, and $\Delta t$ is the time interval between measurements.

In the growth literature, the human growth curve is often considered to be made up of three sequential phases:

- **Infancy**: This term is often used to refer to the age period from birth until about 2 years. Infancy is characterised by a period of rapid growth, that rapidly decelerates.
- **Childhood**: This term is used to describe the period after infancy, from 2 years until the onset of puberty. This period is characterised by a steady and slowly declining growth. In childhood, growth occurs at about 5-6 cm per year.
- **Adolescence/Puberty**: This term usually refers to the time period between the start of the pubertal spurt and age when final adult height is reached. Growth in adolescence is characterised by a rapid growth spurt at puberty followed by a rapid deceleration in growth. At the peak of the growth spurt, height velocity is about 8-12 cm per year depending on the gender. The age of peak velocity for males is about 2 years later than for females and the peak velocity reached is higher.

Further descriptions of growth during these three phases are outlined below, along with prenatal and post-adolescent growth.

### 1.1.2 Prenatal growth

Prenatal growth refers to the time period between conception and birth. Growth in the prenatal period is fundamental to the child's future well-being (Tanner 1989). Traditionally the gestational age of the foetus is calculated from the first day of the last menstrual period. On average this is about two weeks before ovulation, but

this can vary by up to six days in either direction. Therefore 40 weeks is the most frequent gestational age at birth, but in fact only represents 38 weeks of true foetal age (Tanner 1989). Lengths of gestation from 37 to 42 weeks are regarded as normal. Babies born earlier than 37 weeks are considered to be 'pre-term', whereas babies born within the normal age range are said to be 'term' deliveries (Tanner 1989).

In the embryonic period, the first two months, the velocity is low (Tanner 1989). Following this period the foetus experiences a high rate of growth until around 20 weeks, the age at which peak velocity is reached. However, this peak velocity can occur up to six weeks later (Tanner 1989). The peak in velocity, during gestation, corresponds to the highest peak in the velocity growth curve from conception to maturity (Woolston 1991). Growth in the weight of the foetus follows a similar pattern. However, the peak velocity is reached later, at about 34 weeks (Tanner 1989). During the last ten weeks of gestation the foetus stores very considerable amounts of energy in the form of fat and this is viewed to serve as a protective mechanism for the critical period after birth (Tanner 1989). After the peak velocity in length (weight) is reached it is then followed by a period of deceleration in length (weight) gain until birth. Tanner (1989) states that there is substantial evidence that, beginning at 34 to 36 weeks, the growth of the foetus slows down due to the influence of the uterus, whose available space is by then becoming fully occupied.

Birth weight is influenced by many factors, many of them maternal. Tanner (1989) states that poor environmental circumstances, in particular nutritional status of the mother, can result in lowered birth weights and this seems to be largely due to a reduced rate of growth in the last 2 to 4 weeks. Roddam (1998) used a graphical modelling approach to explore relationships between maternal factors and a child's height or weight from birth until 5 years for 1163 infants born in two small towns in south Wales in 1972-73. Roddam (1998) found birth weight to be associated with birth length, child's sex, maternal height, maternal weight, mothers weight gain between 20 and 36 weeks, smoking status at 36 weeks, maternal age and parity, and oedema and albumin status at 20 weeks. Birth weight was not found to be associated with maternal blood pressure or social class. The latter finding contradicts the opinion held by some authors. Sinclair (1985) states that mothers of lower socio-economic class have smaller babies.

### 1.1.3 Growth in infancy

In infancy, babies have their birth weights routinely monitored. Up until the age of 2 years, the supine length of an infant is measured: this is on average about 1 cm more

than the measurement of standing height taken on the same child (Tanner 1989). In infancy, weight is easier to measure than length. Infancy is an important phase of a child's growth, the nutritional demands are high for this period of rapid growth. Thus poor nutrition or illness can greatly influence a child's growth progress. Skuse et al. (1994) suggested that the first few months of life may constitute a 'sensitive period' in terms of mental development.

In the first few days after birth the infant experiences a transient loss in weight, usually about 5% of birth weight (Sinclair 1985). Birth weight is regained in about 10 days (Sinclair 1985). This weight loss is due to a diminished intake of fluid (Sinclair 1985) and represents a period of adjustment to breast or bottle feeding.

> In the first year of life, linear growth and weight gain continue at a rate that is still remarkable, although less spectacular than during fetal life. By 1 year, the normally developing infant has accumulated generous stores of adipose tissue, has tripled its birth weight, and has grown 25 additional centimetres, an increase of 50% of birth length. In the second year, there is continued deceleration of linear growth rates; by the age of 2 years, linear growth has stabilised at a rate that is characteristic of the childhood years. Growth velocity for weight gain follows similar patterns. (Woolston 1991, pp13)

After this initial weight loss, the first twelve weeks after birth sees a rapid increase in height velocity (Tanner 1989). Tanner (1989) hypothesises that this rapid growth after birth represents a period of catch-up growth for those new-borns that have been delayed the most in the uterus. On average, the smaller the baby, the more it grows at this time (Tanner 1989). However Garn and LaVelle (1984) hold conflicting views, namely that there is no strong inverse relationship between birth weight and subsequent weight gain, but the latter may be associated with maternal weight. After this peak in height or weight velocity, the velocity of height or weight growth decelerates rapidly until about 2 years. At around 1.75 years in girls and 2 years in boys respectively, a child will be about half as tall as they will be as adults (Tanner 1989).

Another measure often monitored in infancy is head circumference, because this is associated with brain size. At birth, a baby has a proportionally large head. The peak in velocity of head circumference occurs before birth, at a gestational age of about 15 to 17 weeks (Tanner 1989). This relatively high velocity continues until a

gestational age of about 32-34 weeks (Tanner 1989). After this age the deceleration in head circumference growth is rapid (Tanner 1989). The head circumference of boys tends to be larger than in girls (Roche et al. 1986). After infancy, head circumference is rarely monitored but continues to increase until 18 years and also has a pubertal spurt (Roche et al. 1986). In the revised UK 1990 reference, growth in head circumference is essentially complete by the age of 5 years (Cole et al. 1998).

Other measures that are sometimes monitored are skinfold thickness in certain sites of the body. These tend to be prone to more error in the untrained hand. In third world countries arm and abdominal circumferences are sometimes also monitored.

### 1.1.4 Childhood growth

Childhood represents the period of growth from the age of 2 years until the onset of puberty. It is usually characterised by relatively stable rates of gain in height (5 to 7.5 cm/year) and weight (2 to 2.5 kg/year) (Woolston 1991). During childhood there is a slight deceleration in linear growth rate but a slight acceleration in weight gain. Resulting in a gain in adipose tissue in the years prior to puberty (Woolston 1991). However this is not the full story. Butler et al. (1990) examined longitudinal height data from 135 children (80 boys and 55 girls) in the Edinburgh growth study. These children were measured 6 monthly between the age of 3 years and the onset of adolescence. A spurt was defined as an height velocity increase that was more than twice the measurement error of velocity ($\pm$ 1.8 mm) (Butler et al. 1989). Butler et al. (1990) identified the usual mid-growth spurt[4] at an age of about 7 and 6.7 years in boys and girls, respectively. In addition Butler et al. (1990) also identified two other spurts, they termed these the 'pre-school spurt' (at 4.8 and 4.6 years in boys and girls, respectively) and 'late-childhood spurt' (at 9.2 and 8.6 years in boys and girls, respectively). Some children were also found to have a further spurt just before the onset of puberty, Butler et al. (1990) observed this spurt in children with average-to-late pubertal onset.

[4]The mid-growth spurt has been identified by several authors but isn't always a feature of a child's growth profile.

### 1.1.5 Growth at adolescence: gender differences in the pattern of growth and differences in tempo

At all ages until adolescence the typical boy tends to be taller than the typical girl (Tanner 1989). The difference in tempo starts about halfway through the fetal period. At birth the typical boy grows slightly faster than the typical girl but these velocities converge at around 7 months (Tanner 1989). From this point until the age of 4 years, the velocity of height growth for girls is higher than boys (Tanner 1989). After 4 years, the velocity in both sexes is then about the same until adolescence (Tanner 1989).

On entering adolescence, starting at around 10.5 years and reaching peak velocity at about 12 years, the typical girl overtakes the typical boy in attained height (Tanner 1989). She surpasses the typical boy at about 11 years, but is then overtaken again by the typical boy at around 14 years (Tanner 1989). The pubertal spurt for boys is higher than that for girls, with the typical boy starting and reaching his peak some 2 years later than the typical girl (Tanner 1989). It is this delay in onset of the pubertal spurt and its higher peak that is thought to contribute to the observed difference in height between males and females (Tanner 1989).

The pattern observed for weight is similar to that of height. The typical girl weighs a little less than the typical boy at birth, their weights then converge at about 8 years, the typical girl then becomes heavier from around 9 to 10 years until 14.5 years (Tanner 1989). Sinclair (1985) states that the spurt in weight lags behind the peak velocity for height by about 3 months. The timings quoted here were published just over 10 years ago. At that time Tanner (1989) stated that children in the USA reached their peak at around 6 months earlier than in the UK. This difference arises due to the 'secular trend'[5] with the US being in advance of the UK. Thus the actual timings for the typical boy and girl may be slightly earlier than quoted here.

The timing of the pubertal spurt differs from one individual to the next, even in the same sex. However the sequence of events is the same. The onset of hormonal changes associated with puberty can be as young as 8 in girls or 9-10 in boys (Bee 1995). The different rate of maturation isn't a phenomenon of adolescence, but does become evident then (Tanner 1989). 'Tempo of growth' is the term used to describe the tendency for development to be rapid or slow (Tanner 1989). The possible

[5]The 'secular trend' refers to tendency over the last hundred years for children within industrialised countries, to be larger and grow to maturity more rapidly (Tanner 1989). There is some evidence to suggest that in the USA that this may have levelled off.

influences on tempo are discussed in section 1.3.1. This difference in tempo leads to problems in creating growth charts in adolescence, because the pubertal spurt will be averaged over early and late maturers leading to a flattening in the curve (Tanner 1989).

### 1.1.6 Post-adolescent growth

Growth of the skeleton does not entirely cease at the end of adolescence (Tanner 1989). Although growth in length of limbs ceases; the vertebral column continues to grow until about 30 years. Thus leading to a very small increase in height after adolescence, on average 3 to 5 millimetres (Tanner 1989). In the age period 30 to 45 years attained height remains static. After the mid-forties height then begins to decline. For some individuals, a greater loss in height is experienced because of curvature in the spine due to osteoporosis.

Growth in stature is defined to have virtually ceased, when thereafter only some 2% of height is added (Tanner 1989). Using this definition, in western countries, the average boy and girl stop growing, at 17.5 years and 15.5 years, respectively (Tanner 1989). There is a normal range of variation of 2 years, in either direction, about these values (Tanner 1989).

The behaviour of weight post adolescence is less straight forward. Some individuals experience fluctuations in their attained weight whereas others maintain the same weight. In general there tends to an increase in weight in middle age.

This subsection and the last were included here to provide a 'full-picture' of the growth process. The maximum measurement age is around 9 years for the growth data discussed within this thesis, thus the issues generated by the timing of the pubertal growth spurt will not be considered further here.

## 1.2 The use of Growth charts

### 1.2.1 Purpose of growth reference

The normal growth process described in the last section describes a child's journey to adult height if the child is adequately fed, is not deprived, grows up in a nurturing environment and experiences no illness. Tanner (1989) views the child's curve of growth to be a 'target seeking' function, where 'target' refers to the child's genetic

potential. The term 'canalization' is used to describe this power to stabilise and return to a predetermined growth curve after being pushed off course (Tanner 1989). Life events such as illness and starvation are situations in which a child's growth could be affected. The rapid growth after a period of growth restriction is termed 'catch-up' growth (Tanner 1989; Prader et al. 1963). Tanner (1989) stated that there were two ways in which catch-up could occur and that it was likely to be a combination of these responses:

1. an increase in velocity to such an extent that the original curve is attained and then growth proceeds normally
2. a delay in maturity, so that growth is resumed at the correct velocity for bone age[6]

> One thing seems clear about canalization. Regulation is better in females than in males. ... Similarly, girls recover from growth arrest more quickly than boys. The physiological reason for the greater stability is not known. (Tanner 1989, pp171)

If circumstances don't improve, then 'catch-up' may not occur, resulting in a final height that is lower than expected. The term used for this is 'stunted', a child is said to be stunted if they don't reach a final adult height within the region of their genetic potential. However, parents themselves may have been subjected to poverty and poor nutrition and thus may not provide the best indicator of genetic potential. This issue will be expanded on in Chapter 2. The term 'stunting' is used to describe the process of slowing up in height growth, the deficit in attained length or height compared to a reference (defined below) (Waterlow 1988).

The motivation for the study of child growth differs between developed and underdeveloped countries. In underdeveloped countries, there are often seasonal fluctuations in weight gain or even weight loss and observed stunting in height growth is the norm. Therefore in underdeveloped countries, children's growth is viewed to be a sensitive index of the health and nutrition of the population (Tanner et al. 1966). In developed countries, where infectious diseases are well controlled and children are in general adequately fed, it is desirable to detect growth disorders early

[6]Bone age is a measure of developmental age and is based on successive stages of development of the skeleton. It is usually assessed by scoring a radiograph of the wrist region, but such study has gone out of vogue because of the risks from exposure to radiation. This is viewed to be more representative than chronological age in assessing how far an individual has progressed along his or her road to maturity (Tanner 1989).

so that catch-up growth is likely to occur (Tanner et al. 1966). In other words, the purpose of growth monitoring in the developed world is to detect organic conditions, poor nutrition and growth disorders such as Growth Hormone deficiency and Turners syndrome; whereas in poorer countries the main interest is in malnutrition and the possible consequences. A growth reference can be used to monitor a child's growth in order to identify individuals at risk, they may be used to monitor the response to treatment or, as described above, as an indicator of the general health/nutrition of a population (Tanner 1989).

Currently there is much interest in references for the body mass index (BMI), defined by equation (1.2), because this is viewed to be an indirect method of assessing obesity and malnutrition. The use of BMI does have drawbacks: it may be dependent on stature, affected by relative leg length or sitting height and may reflect both lean and fat tissue (Garn et al. 1986). There is a worrying trend of increasing obesity that has generated much public health interest, in terms of future health outcomes for the obese individual. Power et al. (1997) suggest that there are elevated risks of adult obesity for overweight children, but that the prediction of adult obesity from child adiposity measures was only moderate. However, research from the thousand families cohort study in Newcastle suggests that the tracking from childhood overweight to adult obesity was poor (Wright et al. 2001).

$$\text{BMI} = \frac{\text{weight (Kg)}}{[\text{height (m)}]^2} \tag{1.2}$$

### 1.2.2 The growth reference

In order to say anything about a child's height or weight on a one off occasion we need to compare it to a reference. Growth references are used to compare an individual relative to other children of the same age and sex. The quality of this comparison depends on how accurately the sample population represents the whole population (Tanner 1989).

> A growth standard or growth reference is a dataset representing the distribution of a given anthropometric measurement as it changes with some covariate - usually age - in the two sexes, based on a specified reference sample of children. (Cole 1993, pp21)

Growth standard and growth reference are often used interchangeably in the growth literature. A growth standard is taken to represent an attained height or

weight to aspire to; in other words 'good' or optimal growth (Cole 1993). The implication is that a standard is prescriptive, with no need to for continual revision (Voss and Mulligan 1999a). Tanner et al. (1966) used the term standard for the previous British growth charts. A growth reference provides a reference (Cole 1993), in the sense that it describes the status-quo and requires updating (Voss and Mulligan 1999a). The current UK growth charts are termed growth references (Freeman et al. 1995).

Growth references are derived from cross-sectional data and require large sample sizes, Eveleth and Tanner (1990) state that a sample of about 1000 of each sex are required for each age group. Furthermore, the number of children measured in each group should be approximately proportional to the rate of growth (Eveleth and Tanner 1990). The quality of the height and weight data is important, in particular the measurement technique used. Thus routinely collected data would be unsuitable in the construction of a growth reference; because of clothing, poor installation of measuring equipment and multiple observers.

The distribution of a growth reference is usually summarised by selected centiles which, in the case of height, are symmetric about the median (50th centile). The extreme centiles, such as the third centile, are subject to greater sampling error than the more central centiles (Eveleth and Tanner 1990). Centile charts are easy to use fairly accurately if the height or weight falls within the more central centiles (Cole 1993). However, in the tails the distribution of the centiles are very widely spaced (Cole 1993) and determining the location of children above or below the extreme centiles is problematic. This is likely to be the case, for children that have growth disorders. One advantage of using centiles is that in order to derive them we do not need to know the form of the underlying distribution (Cole 1993).

One alternative to centile specification, used frequently in the nutrition literature, is percentage of median:

$$\frac{\text{weight}}{\text{median weight for age and sex}} \times 100 \tag{1.3}$$

This can be used similarly for height. One advantage of this approach is that it is still interpretable in the extremes of the distribution for height or weight (Cole 1993). Stunting is usually defined as less than 90% height-for-age and wasting is defined as less than 80% weight-for-height. The downside to using the percentage-of-median approach is that the interpretation of the observed deficit depends both on the measurement and measurement age (Cole 1993).

An alternative, advocated by Waterlow et al. (1977), is to use the Z-score. This is the approach used in this thesis. Z-scores are by definition normally distributed, with mean 0 and a standard deviation of 1. There are many advantages to using Z-scores: they are interchangeable with centiles if the underlying distribution is known, they can be as large or as small as necessary so heights and weights below the third or above the ninety seventh can be represented (Cole 1993). In addition, if the growth reference adequately adjusts for age, children of slightly different ages can be easily compared.

The distribution of height is close to normal for the majority of childhood (Cole 1993). As height appears to be normal distributed the Z-scores for height are defined in the usual way.

$$Z(\text{height}) = \frac{\text{height - median (mean) height for age and sex}}{\text{SD of height for age and sex}} \tag{1.4}$$

However, weight is not normally distributed, with more of the population having high weights rather than low weights. The distribution of weights has a long right tail in comparison to the left tail. This results in centiles that are more widely spaced above the median, than below (Cole 1993). The same is true for BMI and skinfold data. Cole (1988) devised the LMS method, based on the Box-Cox transformation (Box and Cox 1964). This works by removing the skewness in the non-normal data to give a symmetric distribution. The LMS method and other alternatives will be discussed in more detail in Chapter 2. However, the basics of the LMS method are outlined below. Using the LMS method a Z-score is defined by equation (1.5).

$$\text{Z-score} = \frac{[\text{Measurement}/M(t)]^{L(t)} - 1}{S(t)L(t)} \tag{1.5}$$

where measurement refers to child's measurement (e.g. height, weight, BMI, etc.), $t$ is age, $L(t)$ is the power transformation, $M(t)$ is the median and $S(t)$ is the coefficient of variation (ratio of standard deviation to median). In the case of height, which is already normally distributed, the power transformation ($L(t)$) is 1 at all ages. The centiles can then be obtained from $L(t)$, $M(t)$ and $S(t)$ using equation (1.6).

$$C_{100\alpha}(t) = M(t)(1 + L(t)S(t)z_\alpha)^{1/L(t)} \tag{1.6}$$

where $C_{100\alpha}(t)$ is the centile curve plotted against $t$ and $z_\alpha$ is the normal equivalent deviate for that centile.

Growth references are based on cross-sectional data; thus they are only useful for one-off measures. Tanner et al. (1966) refers to this type of reference as a 'distance' standard. However a child's size at one given age represents growth since conception,

it doesn't tell you anything about the process of how the child reached this height or weight. If we have two weight or two height measures then we are able to say something about a child's growth between these two points. Tanner et al. (1966) gave this type of reference the term 'velocity' standard. These will be discussed in more detail in Chapter 3.

Growth references are created and used for the monitoring of height, weight, BMI, weight-for-height, head circumference and skinfolds. In some poorer countries there are local references for arm and abdominal circumferences.

### 1.2.3 Tanner-Whitehouse charts

Tanner et al. (1966) developed the first growth chart for British children. The Tanner-Whitehouse growth charts (Tanner et al. 1966) were derived from London children measured in 1959 but adjusted slightly to be appropriate for 1965. The major height and weight centiles on the Tanner-Whitehouse chart were 3rd, 10th, 25th, 50th, 75th, 90th and 97th (Tanner et al. 1966). The height and weight data that contributed to the standard were derived from three sources: Supine lengths, heights and weights from birth to 5.5 years were from a longitudinal sample of around 80 children from central London measured at the London Child study centre (Moore et al. 1954); heights and weights from 5.5 to 15.5 years were a cross-sectional sample (approximately 1000 boys and 1000 girls for each year of age) taken from the London County Council survey of 1959 (Scott 1961); and heights and weights from 16.5 to 20 years were a longitudinal sample of 30 children from the Harpenden Growth study (Tanner et al. 1966).

In infancy, the sample size on which the Tanner-Whitehouse charts is based is very small, in view of infancy being such a critical period of a child's growth. However Tanner et al. (1966) does suggest that these measures are in reasonable agreement with those taken on 250 children (of each sex) in the Oxford Child health survey (Tanner (1958) in Tanner et al. (1966)). In addition, growth of children in London may not represent the full picture of growth experienced in other areas of the UK.

Obviously, at the time of creation of these charts the computing technology was far behind what is possible today, so much of the smoothing of centile curves was done by eye (Tanner et al. 1966). In adolescence, children mature at different rates. Tanner et al. (1966) gives this the term 'phase-difference'. The shape of the distance curve at adolescence was derived from 49 boys and 41 girls from the

Harpenden Growth study that had sufficient data over the adolescent time period. There is a break in the Tanner-Whitehouse height charts at 2 years, the time at which length measurement switches to height measurement. Tanner et al. (1966) also presented the first British height and weight velocity references; discussion of these will be delayed until Chapter 3.

### 1.2.4 National Centre for Health Statistics (NCHS) growth reference and its revision

On average, children in North America have a slightly faster tempo of growth and grow to a slightly greater adult height than British children (Tanner 1989). The original NCHS growth reference (Hamill et al. 1977) was derived from three sources: Fels longitudinal study (1929-1975)[7] from birth to 3 years and nationally representative data from the National Health Examination Surveys (NHES II and III: 1963-70) and the first Health and Nutrition Examination Survey (HANES I: 1971-74) from 2 to 18 years.

The NCHS created references for two age groups: birth to 36 months and 2 to 18 years. This resulted in a discontinuity in the growth reference, with some discrepancy in the age range 2-3 years. In the age range, birth to 3 years, charts were created for weight, length, weight-for-length and head circumference. From 2 to 18 years; charts were created for height and weight. Weight-for-height charts were only created for prepubescent boys and girls (Hamill et al. 1979). Major centiles on these growth charts were the 5th, 10th, 25th, 50th, 75th, 90th and 95th. These were smoothed using least-squares-cubic-splines (Hamill et al. 1979). More extreme centiles were not obtained as the numbers in each yearly age group varied between 300 and 600 (Tanner 1989).

The NCHS growth reference was adopted in the late 1970's by the World Health Organisation (WHO) to provide an international growth reference (Waterlow et al. 1977). However, work by WHO has illustrated that this reference is sufficiently flawed to interfere with the health and nutritional management of infants and young children (de Onis and Blössner 1997). The main problem is that in infancy, the Fels infants are primarily formula fed, introduced to solids very early[8] and homogeneous

[7]The growth data from the Fels longitudinal study is unique in terms of following participants into adulthood and representing 3 or 4 generations of the same family (Roche 1992).

[8]Early introduction of solids can be detrimental in underdeveloped countries because of the risks from contaminated food

in terms of genetic, geographic, socio-economic backgrounds (de Onis and Garza 1997). The sample sizes that contributed to this reference under the age of 5 were also very small (Guo et al. 2000). The childhood weight-for-height reference is also criticised for assuming age-independence (Cole 1985).

After many calls for updating this reference, the NCHS/WHO reference has undergone revision and is now known as the Centers for Disease Control and Prevention (CDC) 2000 growth reference (Kuczmarski et al. 2000; Ogden et al. 2002). Major changes to the NCHS reference are the inclusion of the third and ninety-seventh centiles, Fels longitudinal infancy data replaced by nationally representative data, extension of charts to 20 years and elimination of disjunction between curves for infants and older children (Kuczmarski et al. 2000). The CDC 2000 growth charts include a new BMI-for-age reference covering the age range: 2 to 20 years (Ogden et al. 2002). The BMI growth chart also includes the 85th percentile because this is the recommended threshold for identifying overweight children (Kuczmarski et al. 2000). The CDC 2000 growth reference was constructed from five national health examination surveys collected from 1963 to 1994 (NHES II and III, and NHANES I, II and III) and five supplementary data sources (Kuczmarski et al. 2000). In order to avoid the influence of increased body weight and BMI in the most recent national survey, data from NHANES III above the age of 6 years was excluded. Kuczmarski et al. (2000) suggest that this national data provide a better representation of the racial or ethnic diversity and the size and growth patterns of combined breast- and formula-fed infants in the USA.

The centiles of the CDC 2000 growth reference were derived using a two stage process. Initially the empirical centile curves were smoothed using either parametric or non-parametric approaches depending on the growth variables considered (Kuczmarski et al. 2000). In infancy the Guo et al. (1988) model was applied to length, weight and head circumference data. Whereas in childhood a non-linear model was applied to stature data, a polynomial regression model was applied to weight-for-stature data, and a locally weighted regression procedure was applied to weight and BMI data (Kuczmarski et al. 2000). In the final stage smoothed centile curves were estimated using a modified LMS estimation procedure (Kuczmarski et al. 2000; Ogden et al. 2002).

### 1.2.5 The UK 1990 reference and its revision

Tanner et al. (1966) stated at the time of publication of their standard that the growth reference should be updated every 10 to 15 years. Many authors (Wright

et al. 1993; Wright et al. 1992; Voss et al. 1987; Whitehead et al. 1989b) raised their concerns that the Tanner-Whitehouse reference was out of date and in need of revision. The main concerns were that the growth data that formed the reference was from the South East of England (mainly London) and the secular trend to earlier maturity and greater adult height (Freeman et al. 1995). The Tanner-Whitehouse references were based primarily on 'bottle-fed' children, whereas present day feeding practices promote breast-feeding. In infancy use of the Tanner-Whitehouse reference for Cambridge infants that were breast-fed lead to the impression that the child's growth was faltering from 3-4 months after having an initial advantage (Whitehead et al. 1989b). In Newcastle (the data set studied within this thesis), where the proportion of breast-fed and bottle-fed children is unknown, a similar pattern was observed for both the NCHS and Tanner-Whitehouse references (Wright et al. 1993).

Conventionally, growth charts had always been characterised by the 3rd, 10th and 25th centiles below the median and 75th, 90th and 97th centiles above. The World Health organisation uses cut-offs based on Z-scores (Cole 1994b). Cole (1994b) proposed that the format of a growth chart should be revised from a 7 to 9 centile chart, with each centile spaced two-thirds of a Z-score apart. Thus making the interpretation of Z-scores and centiles compatible. In the production of the UK 1990 reference this proposal was put into action. The distribution of UK 1990 reference is summarised by the 0.4th, 2nd, 9th, 25th, 50th, 75th, 91st, 98th and 99.6th centiles. Using the 0.4th centile as a cut-off results in only one normal child in 260 that lie below this cut-off (Cole 1994b).

Growth data from seven sources were used to create the UK 1990 growth reference for height and weight (Freeman et al. 1995). The reference sample totalled over 25000 individuals from growth surveys between 1978 and 1990 (Freeman et al. 1995). The National Study of Health and Growth (NSHG) (Rona and Chinn 1986; Rona and Chinn 1987) height and weight measurements were used as the reference data set as these were the most recent (Freeman et al. 1995). The other data frames were then adjusted accordingly. Ethnic minorities were excluded because these populations are known to exhibit different growth patterns (see section 1.3.1).

Cole et al. (1995) published the first UK reference for the body mass index. This was derived from the same data sources as the original UK 1990 reference (Freeman et al. 1995). The body mass index of children changes substantially with age (Cole et al. 1995). In infancy it rises steeply to a peak at about 8 months, it then falls in the preschool years and flattens out around 5.5 years (often termed the 'adiposity rebound'[9]) and finally rises into adulthood (Cole et al. 1995). There is also a greater

[9]The timing of the 'adiposity rebound' is often thought to be indicative of later obesity,

degree of skewness in the distribution of body mass index than for weight (Cole et al. 1995).

The original UK 1990 reference was shown to have a sex bias for weights in infancy (Wright et al. 1996). There were two and half times more girls than boys with weights below the third centile during the first year. The UK 1990 reference was then revised (Preece et al. 1996) and according to Cole et al. (1998) there is no longer a sex bias in the current reference.

In the construction of the revised UK 1990 Growth reference data for weight, height, body mass index and head circumference were constructed from 12 sources (17 distinct surveys, all cited from Cole et al. (1998)):

1. Human Measurements Anthropometry and Growth (HUMAG) infants, toddlers, boys, and girls (British Standards Institution 1990)
2. HUMAG men and women (Jones et al. 1993)
3. National Study of Health and Growth (NSHG) (Rona and Chinn 1986; Rona and Chinn 1987)
4. Tayside growth study (no ref.)
5. Cambridge infant growth study (Whitehead et al. 1989b)
6. Whittington birth data study (Colley et al. 1991)
7. National Diet and Nutrition Survey (NDNS) (Gregory et al. 1995)
8. MRC Dunn Nutrition Unit Premature Baby Study (Lucas et al. 1984)
9. Cambridge Rosie Premature Neonates 1985-94 (no ref.)
10. Cambridge Rosie Neonates 1992-93 (no ref.)
11. Northern Region Premature Neonates 1991-92 (no ref.)
12. Edinburgh Growth Study (Ratcliffe et al. 1994)

The data from these surveys were analysed by maximum penalised likelihood using the LMS method (Cole and Green 1992). It assumes that for independent positive data $y_i$ at ages $t_i$ $(i = 1, \ldots, n)$ an age-specific Box-Cox transformation

there is an increased risk of adult obesity the earlier the adiposity rebound (Cole et al. 1995)

(Box and Cox 1964) can be applied to the data to make them normally distributed. In general the $t_i$ are replicates from a smaller set of distinct ordered ages $T_j$ ($j = 1, \ldots, m$). The distribution of $y_i$ at $T_j$ (where $t_i = T_j$) is summarised by the median $M(T_j)$, coefficient of variation $S(T_j)$, and Box-Cox power $L(T_j)$. The formula below converts the measurement $y_i$ to its Normal equivalent deviate $z_i$:

$$z_i = \begin{cases} \frac{(\frac{y_i}{M(T_j)})^{L(T_j)}-1}{L(T_j)S(T_j)} & (L(T_j) \neq 0) \\ \\ \frac{\ln(\frac{y}{M(T_j)})}{S(T_j)} & (L(T_j) = 0) \end{cases} \tag{1.7}$$

The quantities L, M and S are natural cubic splines with knots at each $T_j$, and are estimated by maximum penalised likelihood. The required centile curves were as follows:

$$C_{100\alpha}(T_j) = \begin{cases} M(T_j)(1 + L(T_j)S(T_j)Z_\alpha)^{1/L(T_j)} & (L(T_j) \neq 0) \\ M(T_j)\exp[S(T_j)Z_\alpha] & (L(T_j) = 0) \end{cases} \tag{1.8}$$

where $Z_\alpha$ is the normal equivalent deviate of size $\alpha$.

The LMS coefficients are available in Microsoft Excel format or as text files from the Child Growth Foundation (Child Growth Foundation 1996b). LMS values are available from 23 weeks gestation to 23 years for weight, 33 weeks gestation to 23 years for height, 33 weeks gestation to 23 years for BMI and 23 weeks gestation to 17 (boys) or 18 (girls) years for head circumference. In the postnatal period, LMS values are given every calendar month (twelfths of a year). Linear interpolation between these values can then be used to obtain the values of L, M and S for the child's decimal age. The UK 1990 centile chart for boys (pre-term to 52 weeks) can be found in figure 1.1.

Recently concern has been expressed that there are several growth references in use in clinics and hospitals in the UK (Wright et al. 2002; Cameron 2002). Wright et al. (2002) considered the following references: Tanner-Whitehouse (Tanner et al. 1966), Gairdner-Pearson (Gairdner and Pearson 1971; Gairdner and Pearson 1985), Buckler-Tanner (Tanner and Whitehouse 1976; Buckler and Tanner 1997) and the UK 1990 growth reference (Freeman et al. 1995; Cole et al. 1998). These four growth references are widely used at present or in the past (Wright et al. 2002). The overall consensus was that the Tanner-Whitehouse and Gairdner-Pearson references were obsolete and that for clinical purposes the use of the revised UK 1990 reference is advocated (Wright et al. 2002). However, there are no suitable references for head circumference beyond infancy (Wright et al. 2002). The Buckler-Tanner chart is an update of the clinical longitudinal chart of Tanner-Whitehouse (Tanner and

Whitehouse 1976) formed using the amalgamated data set from Freeman et al. (1995) and longitudinal data from 198 Sheffield adolescents (Cameron 2002). It is suggested that the Buckler-Tanner charts are suitable for assessing height after the age of 2 (Wright et al. 2002) and its use is advantageous in the monitoring of individual children in adolescence (Cameron 2002). However concern has been expressed that the longitudinal source sample in the Buckler-Tanner charts is too small and may only be valid for a subgroup of adolescents (Preece (1998) in Cameron (2002) and Wright et al. (2002)). However, Cameron (2002) concluded that the Buckler-Tanner reference should be used wisely in adolescence.

## 1.3 Growth disorders

### 1.3.1 Interaction of heredity and environment

In this section we consider the factors that influence a child's path to adult size. Tanner (1989) viewed child growth to be a product of the continuous and complex interaction of heredity and environment.

Heredity refers to the genetic make-up of a child. Children of tall parents tend to be taller than children of short parents providing environmental circumstances are adequate. There is also thought to be a genetic influence on the tempo of growth, thus a late maturing girl may have a late-maturing father or a late-maturing mother (Tanner 1989). However, Tanner (1989) held the opinion that the genetic control of tempo was largely independent of genetic control of final adult height.

Environmental influences include nutritional, ethnic, seasonal, illness, psychosocial stress, urbanisation, socio-economic status and family size (Tanner 1989).

1. **Nutrition:** Malnutrition delays growth. Children subjected to an episode of acute starvation recover more or less completely by the 'catch-up' process, described above, provided the adverse conditions are not too severe and do not last too long (Tanner 1989). Children that experience chronic undernourishment during the majority of their childhood end up with lower final adult heights (Tanner 1989). Infancy is the period when the child is most at risk from malnutrition (Eveleth and Tanner 1990). In Western countries, today, there is more of a problem with over-nutrition. Research by Reilly and Dorosty (1999) suggests that in children, with ages from 6 to 15 years, there may be

an increasing trend of obesity with age.

2. **Ethnic:** Populations of different ethnic origin differ in their average adult size, tempo of growth and final shape (Tanner 1989). Asian and Japanese children are shorter (Tanner 1989). American children of African descent have almost identical curves to the Americans (Tanner 1989). Europeans of African descent have a higher tempo growth than Europeans at all ages (Tanner 1989). An additional complication is that children of ethnic origin grow differently in other countries than in their country of origin, largely due to other environmental influences. Furthermore, mixed marriages also complicate the issue. Thus, in general, ethnic minorities are excluded in the creation of growth references. It has been suggested that the centiles of the revised UK 1990 reference could be shifted upwards or downwards for different ethnic minorities (Chinn et al. (1996) in Cameron (2002)).

3. **Season:** In European countries, children aged 7-10 years grow faster in height in spring and summer than in autumn and winter (Marshall 1971). In tropical countries, seasonal variations arise because of rainy and dry periods, which in turn influence food availability and infection (Tanner 1989).

4. **Illness:** Obviously illness has a major impact in poorer countries due to lack of medical resources. However, in well-nourished children in richer countries the effects on growth of minor illness are minimal (Tanner 1989). Complete catch-up growth after illness may not occur in children with poor diets (Tanner 1989). For instance in children that fail-to-thrive, poor nutrition leads to a higher frequency of minor ailments which in turn influences food intake (Woolston 1991).

5. **Psychosocial stress:** Growth faltering can occur in children that experience psychosocial stress. Two mechanisms for this delay are proposed (Tanner 1989): alterations in appetite and inhibition of growth hormone secretion. The views of Tanner (1989) feed into the 'food versus love' debate, in the origin of failure-to-thrive; see Chapter 3. Taitz and King (1988) illustrated that catch up could occur in abused children placed in foster homes.

6. **Urbanisation:** In industrialised countries children in urban areas are usually larger and have a more rapid tempo of growth than children

from rural areas (Tanner 1989). Tanner (1989) attributes this to rural children eating less and expending more energy, whereas in urban areas there is access to: food, health services, education and welfare facilities (Tanner 1989).

7. **Socioeconomic status:** In most societies children from different social classes differ in size and in tempo of growth (Tanner 1989). Upper class children tend to be taller and reach maturity earlier (Tanner 1989). Only the weight-height relationship is affected differently, with lower classes having higher levels of obesity (Tanner 1989).

8. **Family size:** Children from large families tend to grow slower (Tanner 1989). This is thought to be due to having more mouths to feed and increased incidence of infections (Tanner 1989). During childhood, first-born children tend to be taller than later-born children with the same number of siblings, because the first-born child has had a period of being an only child (Tanner 1989).

### 1.3.2 Disorders of child growth

Table 1.1 summarises the main conditions affecting a child's growth (adapted from Hall (2000)). Jellinek and Hall (1994) found that in a survey of parents that had children with growth disorders, that the majority (46%) of growth disorders were first identified because of parental concern whereas only 8% of children's growth disorders were identified using routine height monitoring.

In childhood, research focuses on children with short stature. However, in infancy, the emphasis is on growth faltering, usually weight. This is given the term 'failure-to-thrive', which is categorised as non-organic and organic failure-to-thrive. The latter is growth faltering due to some medical cause, such as Down's syndrome. Failure-to-thrive and growth faltering will be discussed further in Chapter 3. It is generally accepted that 'failure-to-thrive' in infancy can lead to reduced attained height in childhood or shorter than expected final adult height.

> Epidemiological surveys using longitudinal data have shown many short children at school-entry failed to thrive during infancy (Skuse 1998, pp342)

Thus, there is some overlap in the populations that are studied for 'short stature' and for 'failure-to-thrive' in childhood.

Table 1.1: A summary of the main conditions affecting growth (adapted from Hall (2000))

| Short Stature or growth failure | Tall stature or accelerated growth |
|---|---|
| Isolated growth hormone deficiency | Marfan's Syndrome |
| Multiple pituitary hormone deficiency | Klinefelter's syndrome (XXY) |
| Turner's syndrome (girls) | XYY syndrome |
| Psychosocial deprivation | Soto's Syndrome |
| Silver-Russell syndrome | Thyrotoxicosis |
| Skeletal dysplasias and bone disorders | Congenital adrenal hyperplasia |
| Noonan's syndrome | Premature sexual maturation |
| Neurofibromatosis | Pituitary gigantism |
| Hypothyroidism | |
| Inflammatory bowel disease | |
| Coeliac disease | |
| Chronic renal disease | |

Short stature can be due to genetic shortness, deprivation, malnutrition or of organic origin (or a combination of these). The majority of research focuses on the psychological consequences of short stature (Skuse 1987; Siegel et al. 1991) and influence on mental ability. Lacey and Parkin (1974) carried out a study of children of short stature in Newcastle at the age of 10 years. Lacey and Parkin (1974) found that there was a tendency for children from poor homes to be shorter, have less subcutaneous fat and to have greater delay in skeletal maturation. All these short children scored poorly in tests of mental ability and attainment (Lacey and Parkin 1974). This work has been replicated recently by Voss et al. (1989). They identified 180 children whose height at school entry lay below the third percentile using the Tanner-Whitehouse standards (Tanner et al. 1966) from 14,346 children in 2 health districts entering school during two consecutive years (1985-7). One hundred and forty of these short normal children (remaining 32 excluded for pathology, 5 from ethnic minorities and 3 non-compliant) were matched with 140 controls of average height. The controls were matched with cases for age, sex and school class. However, the children in the Wessex growth study were selected on the basis of height alone; the socio-economic status (SES) of the short participants is lower than that of the control group (Stratford et al. 1999). This cohort of children have been followed over the last 12 years (Voss 1999) at 6 monthly intervals for height, with two psychometric follow ups in 1989/91 and 1994/96. Downie et al. (1997) concluded that until 11 to 13 years that social class had more influence than height on a child's psychological

development.

## 1.4 The Coventry Consensus - current guidelines for growth monitoring

### 1.4.1 Health for all children

Hall (1996) reviewed the role of growth monitoring. In relation to weight monitoring in infancy, Hall (1996) stated that there was little justification for regular weighing once the parent and primary care team were satisfied that the baby is feeding normally and gaining weight. Hall (1996) proposed that length should be measured at 6-8 weeks and that height should be checked at 1.5-2 years, 3.5 years, 5 years and proposed a further check at 7. The following guidelines were proposed for identifying slow-growing children (Hall 1996):

1. In pre-school children (less than 5 years) referral was recommended if height crosses two channel widths between any pair of measurements.[10]

2. In school age children (5-9 years) referral was recommended if a child's height crosses one or more channels between any pair of measurements.

As pointed out by Mulligan et al. (1998) these guidelines are ambiguous, because growth rate requires both the change in height and the time interval between measurements to be considered. Mulligan et al. (1998) considered the change in height Z-score between 5 and 8 years for children measured in a research study and community setting. Considering a change in height Z-score greater than 0.67 to indicate 'slow' or 'fast' growing, they found this would not result in an excess number of inappropriate referrals. Although more girls than boys were identified as 'slow-growing' (Mulligan et al. 1998).

The above guidelines have been revised as a result of the 'Coventry Consensus'. This was a meeting of a multidisciplinary team of health professionals held in Coventry (July 1998). The issues raised at this meeting are summarised by Hall (2000) and Wright (2000) for children over and under 2 years old, respectively. As a result of the consensus it was

[10] channel width refers to the distance between major centiles in a growth reference

agreed that the potential benefits of growth monitoring include: identification of chronic disorders; provision of reassurance to parents; monitoring the health of the nation's children; and supporting future research. (Hall 2000, pp10)

### 1.4.2 Growth monitoring in infancy

In infancy, as discussed above, the primary role of weight monitoring is to identify children at risk of failure-to-thrive (Wright 2000). In clinical practice, the full picture is considered, not just the child's weight chart. The overall conclusion from the Coventry consensus was that infants should be weighed less often but with more attention paid to the recorded weights (Wright 2000). The recommendations from the Coventry consensus with regards to growth monitoring in children under 2 years were:

- Birth weight (but not length), related to gestational age, is essential both for growth monitoring and as an important epidemiological marker.
- No justification has been found for the routine monitoring of length before the age of 2 years, except where there is concern.
- Length should be measured only where there is concern about a child's growth or weight gain.
- Babies who are growing normally should only be weighed at immunisation and surveillance contacts, and should not be weighed more than once every two weeks under the age of 6 months and once a month thereafter.
- A child should be weighed whenever there is clinical concern.
- Average children (initially $>$ 9th and $<$ 91st centile) often cross through one centile space during the 1st year. A sustained fall through two weight centile spaces justifies a more detailed assessment, which should initially be primary care based. (Wright 2000, pp7, Table 1)

### 1.4.3 Growth monitoring in childhood

In childhood, the primary justification for growth monitoring is for the identification of short stature due to growth hormone deficiency or Turners syndrome (Hall 2000). These are the only disorders of growth from table 1.1 with few other clues except short stature (Hall 2000). The main conclusions from the Coventry consensus with regards to height monitoring, of interest here, were:

- Single height measurements with a cut-off using the 0.4 centile on the 1990 charts, come closest to satisfying the criteria for screening.
- School entry offers a good opportunity to screen the whole population. The theoretical advantages are low marginal cost when combined with other school entry screening procedures, potentially high coverage, an acceptable yield of new cases of isolated Growth hormone deficiency and Turner's Syndrome, secondary benefits in case-finding for other disorders and (when combined with weight) a contribution to a core data set for child public health.
- Correction for parental height should not at present be undertaken as part of screening.
- Because the school entry measurement provides the best opportunity to identify growth disorders, the measurement must be done to a high standard, so reliable equipment must be supplied and correctly assembled or installed, and staff training is essential.

...

...

- Children whose height is above the 99.6 centile need be referred only if there are other unexplained symptoms or signs.
- Height measurement at other ages, using the 0.4 centile to trigger action, is good clinical practice. It should be undertaken on an opportunistic basis when a child is seen for other reasons, whether in primary or secondary care, but should not be regarded as a total population screening programme.

- Routine growth monitoring to detect centile crossing has too low a sensitivity and specificity to be regarded as screening. (Hall 2000, pp13)

The situation with regards to weight monitoring after infancy is still less clear. It is usual to judge a child's weight with reference to their height, this can be achieved by using the UK 1990 reference for BMI (Cole 1995). Voss and Mulligan (1999b) compared the body mass indices of 120 children of average height at school entry to the UK 1990 BMI reference. Before puberty the mean BMI was on the 45th centile, whereas at 16-17 years the mean BMI was on the 65th centile (Voss and Mulligan 1999b). This would seem to suggest a trend to increasing obesity over a very short time scale (Voss and Mulligan 1999b). However, the policy with regards to the current UK 1990 BMI reference is to be 'freeze' it at its present state, in order to compare prevalence of overweight with reference to situation in 1990 (Cole et al. 1999). Cole and Roede (1999) have also suggested using BMI centiles derived from Dutch children in 1980 as a baseline to assess obesity. The working party for obesity (Dietz and Robinson 1998) have suggested identifying centiles corresponding to adult cut-offs of 25 and 30 $Kg/m^2$ and extrapolating back to childhood. Cole et al. (2000) have put this suggestion into action, producing an international reference with these suggested cut-offs. However the underlying assumption in the creation of this reference is that prevalence of overweight or obesity is the same throughout childhood (Mulligan 2000). The 'Coventry Consensus' concluded that 'screening' for obesity would not fulfil accepted criteria (Hall 2000), but that recording height and weight together would have greater clinical and public health value than height alone. The consensus concluded that:

> monitoring the changing weights and BMI's of the nation's children is important in view of the high and increasing prevalence of obesity, and could be facilitated by a policy of universal measurement when children start school. (Hall 2000, pp13)

There are no formal referral guidelines for weight, but community guidelines suggest the following rule-of-thumb: a child's weight centile that deviates from their height centile by over 2 centile bands (Schilg and Hulse 1997). This is equivalent to a difference in weight and height Z-scores of more than four thirds.

## 1.5 Summary of thesis content

In this chapter we have discussed the physical features of the growth process for children growing 'normally'. We have highlighted the importance of a growth reference in terms of assessing a child's attained weight or height status. In this thesis, the child's anthropometric measures will be converted to Z-scores using the revised UK 1990 reference. Some of the possible disorders of growth were introduced. The focus in this thesis will be primarily on children whose weight gain falters in infancy. We concluded by presenting current guidelines for growth monitoring.

A statistical literature review of the various aspects of human growth can be found in chapter 2. The approaches used to derive centiles for growth references will be discussed, along with the techniques that have been used to model, primarily, the human height growth curve. We conclude chapter 2 by introducing the concept of mid-parental height.

Chapter 3 focuses on the approaches that can be used to detect growth faltering. In infancy the focus is on weight faltering whereas in childhood the focus is on height faltering. In this chapter we introduce the conditional gain Z-score. A conditional weight gain Z-score provides a way of assessing a child's weight gain between two occasions, provided the correlation between weight Z-scores at these two time-points is known. Within this chapter we also discuss the numerous approaches used to detect failure-to-thrive in infancy. Currently there isn't an agreed research definition of failure-to-thrive, so published results on outcome after growth faltering in infancy are not comparable.

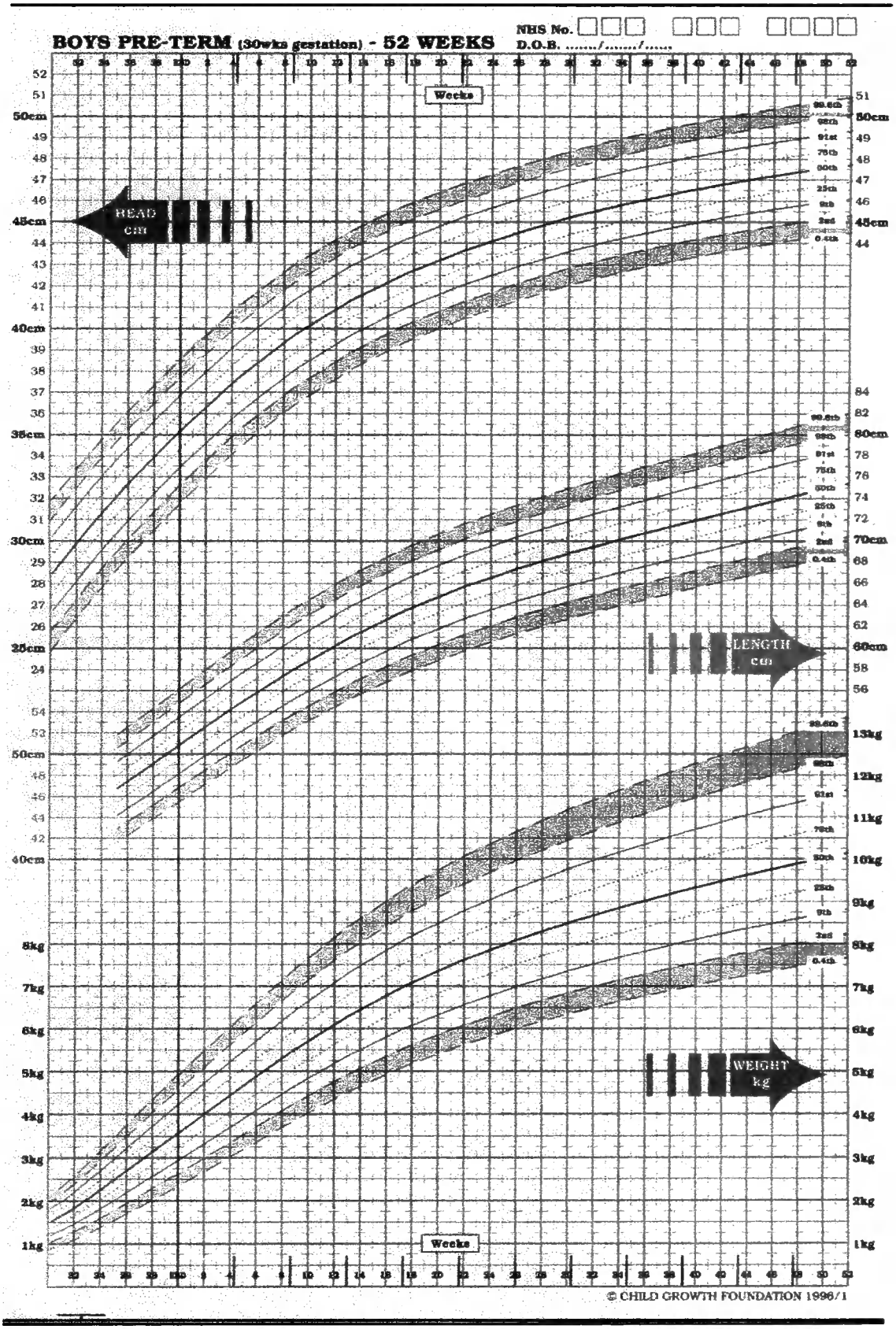
In chapter 4 we introduce the infancy data from the Newcastle growth and development study. This data frame includes routine weights on 3415 term infants resident in Newcastle in 1989, at the time aged 18-30 months. Within this chapter we carry out a preliminary analysis of the routine weight Z-scores and compare the Newcastle children to those children that contribute to the growth reference. The infancy data frame is the main focus of this thesis. This rich source of data provides the opportunity to develop the correlation structure of routine weight Z-scores in infancy. In chapter 5 we develop a model for this correlation structure. In chapter 7 we explore patterns in the conditional weight gain Z-scores obtained using the model derived for Newcastle correlations.

A sub-sample of the infants were followed-up at 7-9 years. These children belonged to three subgroups: (a) cases where children that were defined as failing to thrive in infancy using the 'thrive index' approach (see Chapter 3 for definition),

(b) controls were matched to cases in terms of age and level of deprivation and (c) a 20% systematic sample. In chapter 6, weight, height, BMI and head circumference data from the systematic sample are compared to the revised UK 1990 reference and the correlation of a child's height with reported parental heights will be established. The Z-scores for height, weight, BMI and head circumference of the case and control children are compared.

It is routine to monitor a child's height and weight at school entry. The school entry data of the sub-sample followed at 7-9 years were retrieved from school health records. In chapter 8, we perform a preliminary analysis of the routine childhood height and weight data. In chapter 9 the childhood data are analysed longitudinally. Systolic blood pressure at 7-9 years will be related to current and earlier size. Height at 7-9 years is contrasted with earlier height.

Figure 1.1: **Boys 4-in-1 growth chart** (pre-term to 52 weeks): revised UK 1990 growth reference for weight, length and head circumference (reproduced here with kind permission of Child Growth Foundation, copyright 1996)



# Chapter 2

# Statistical approaches used in the study of child growth

The intention of this chapter is not to provide an exhaustive review of all statistical approaches applied to the study of child growth. Instead it aims to give a brief overview falling mainly into two areas: deriving growth references (in section 2.1) and modelling the growth process at the individual or population level (in sections 2.2, 2.3, 2.4 and 2.5). In section 2.6 we introduce the concept of mid-parental height as an indicator of genetic potential that is often used in clinical practice for the assessment of short stature. A comprehensive review is provided for methods used within this thesis, such as the LMS method, and use of mid-parental heights. Discussion of mathematical approaches to the velocity of child growth is deferred to the next chapter, where the focus is on growth faltering.

## 2.1 Approaches to the construction of growth references

Age-related reference intervals are not only used in the monitoring of child growth, although the construction of 'growth charts' was probably the first example (Wright and Royston 1997). Age-related reference intervals are constructed for other variables such as CD4 counts, weight gain during pregnancy, serum cholesterol and blood pressure. They are commonly used in the routine monitoring of individuals, where interest is in detecting extreme values, such as those below the second centile or above the ninety eighth, possibly indicating abnormality (Wright and Royston 1997).

The quality of a growth reference depends on two factors, namely the data used to derive it and the statistical approach used to arrive at the centiles. The essence of the problem is captured by Cole (1993) in his summary of the criteria introduced by Waterlow et al. (1977):

> ... they require the reference population to be well nourished, the sampling procedure to be clearly defined and reproducible, the sample to be cross-sectional and of adequate size, the measurements to include all those that are relevant and to be of good quality, and finally the data and smoothing procedures to be available. (Cole 1993, pp36)

Good measurement techniques are of utmost importance in growth studies, especially if height or weight measures are then going to be used for the purpose of constructing velocity references (Tanner 1989) (see next chapter for discussion of velocity references). The purpose of a growth reference is to compare an individual's attained height or weight to other children of the same age and sex. Thus the quality of this comparison depends on how accurately the sample population is representative of the whole population and whether the sample is of adequate size (Healy 1986). Precise percentile estimates require large sample sizes and this is especially true for extreme percentiles such as the second and ninety eighth centile (Healy 1986; Guo et al. 2000). The requirement of cross-sectional data is not always met in the creation of growth references: from birth to 5.5 years and birth to three years within the Tanner-Whitehouse reference (Tanner et al. 1966) and NCHS reference (Hamill et al. 1977), respectively, there is a longitudinal component. Even in the current revised UK 1990 reference (Cole et al. 1998) there is longitudinal element in infancy for all anthropometric measures and over the full age-range for head-circumference. In infancy (4 weeks to 2 years) the majority of the anthropometric data is from the Cambridge infancy study (Whitehead et al. 1989b) and over the full age range (birth to 17/18 years) the head-circumference data is from the Edinburgh growth study (Ratcliffe et al. (1994) in Cole et al. (1998)).

> ... centile curves are constrained by links in both horizontal and vertical directions, the former representing consistency with changing age and the latter a well behaved distribution within age. The first of these constraints is explicit, in that the curves are smoothed across ages. The second constraint, that the frequency distribution of the measurement at each age should be

> consistent in some sense, and hence that the centile curves should be consistently spaced relative to their neighbours, is less generally accepted (Cole 1993, pp39).

Cole and Green (1992) used the term 'commonality' to describe centile curves being linked in position to their neighbours. The simplest example of this is when centiles arise from a normal distribution (Cole 1993). Without commonality centile curves can touch or even cross (Cole and Green 1992).

If the normal distribution is used in order to derive the centiles, then these can be defined by the mean and standard deviation (SD) of the distribution (Cole 1993):

$$\text{measurement centile} = \text{mean} + \text{SD} \times z \tag{2.1}$$

where $z$ is the normal equivalent deviate for the required centile. The main advantage in assuming a normal distribution in the construction of centiles is that (on the proviso that the normality assumption is reasonable) the standard errors of the estimated centiles are greatly reduced which is especially true in the tails of the distribution (Healy 1974). However if the assumption of normality is not reasonable, then the extreme centiles will be biased (Cole 1993).

If the assumption of normality is reasonable then the distribution at each age can be derived using the approach described above. However, although height is known to have a distribution that is close to normal during most of childhood (Cole 1993); Cole (1989b) found that during puberty there is a time-period when the height distribution of the NCHS and Dutch reference is skewed to the right. Although an additional consideration is that pooling over one year can create detectable skewness and kurtosis (Healy 1974), there is likely to be some element of grouping in the creation of both of these references.

In section 2.1.1 we focus on the LMS method and maximum penalized likelihood, the approach used in the creation of the revised UK 1990 reference (Cole et al. 1998). This reference will be used to convert anthropometric data from the Newcastle study to Z-scores. In section 2.1.2 we consider the three approaches that have been widely used for creating references: LMS method, the shifted log transformation and distribution-free HRY method. In section 2.1.3 we discuss some of the other approaches proposed for creating age-related references.

### 2.1.1 The LMS method

One approach to the creation of growth references for non-normal data is to use the LMS method. The LMS method assumes that anthropometric data can be transformed to normality using a power transformation, thus removing any skewness (Cole 1990). The anthropometic data is grouped according to age and the optimal power for each age group is calculated. The smoothed series of powers that vary with age is known as the **L** curve: the power of the transformation (Cole 1990). After transforming back to the original scale; the smoothed **M** and **S** curve summarise the trend with age of the median and coefficient of variation (Cole 1990), respectively. Given the L, M, and S curves, centile curves can be derived using equation (1.6).

$$C_{100\alpha}(t) = M(t)(1 + L(t)S(t)z_\alpha)^{1/L(t)} \tag{1.6}$$

where $C_{100\alpha}(t)$ is the centile curve plotted against $t$ and $z_\alpha$ is the normal equivalent deviate for that centile.

Cole (1988) expanded on the approach proposed by Van't Hof et al. (1985); who used the Box-Cox transformation to remove the skewness in skin-fold measures at each age. Van't Hof et al. (1985) suggested that the power of the transformation could vary smoothly with age and hence skewness would vary smoothly with age. Cole (1990) took this one step further and assumed that the transformation to remove the skewness resulted in a normally distributed variable. Thus the mean and standard deviation of the transformed distribution also change smoothly with age, allowing the construction of centile curves (Cole 1990).

Assuming that the variable of interest $y$, for example weight, is positive and has median $\mu$. Suppose that $y^\lambda$ (or $\log(y)$ if $\lambda = 0$) is normally distributed. Consider the family of transformations

$$x = \begin{cases} \frac{(\frac{y}{\mu})^\lambda - 1}{\lambda} & (\lambda \neq 0) \\ \ln(\frac{y}{\mu}) & (\lambda = 0) \end{cases} \quad y > 0 \tag{2.2}$$

proposed by Box and Cox (1964), where the value of $\lambda$ is to be estimated. If $\lambda = 1$ then the measurement is unchanged, if $\lambda = 0$ then we have the natural log transform of measurement and $\lambda = -1$ gives the reciprocal of measurement (Cole 1993). The maximum profile likelihood estimate of $\lambda$ both minimises the skewness and optimises the fit to normality (Cole 1988).

The transformation in equation (2.2) translates the median $\mu$ of $y$ to $x = 0$, and is continuous at $\lambda = 0$ (Cole and Green 1992). If $\lambda = 1$ the standard deviation of $x$ is exactly the coefficient of variation of $y$, which remains approximately true for all

moderate $\lambda$ (Cole and Green 1992). The optimal value of $\lambda$ minimises the standard deviation of $x$ (Cole and Green 1992).

The Z-score of $x$ and hence of $y$ is given by

$$z = \frac{x}{\sigma} = \begin{cases} \frac{(\frac{y}{\mu})^{\lambda}-1}{\lambda\sigma} & (\lambda \neq 0) \\ & y > 0 \\ \frac{\ln(\frac{y}{\mu})}{\sigma} & (\lambda = 0) \end{cases} \tag{2.3}$$

where $\sigma$ is the standard deviation of $x$ (therefore the coefficient of variation of $y$) and $z$ is assumed to have a standard normal distribution (Cole and Green 1992).

If we now assume that the distribution of $y$ varies with time $t$, and that $\lambda$, $\mu$ and $\sigma$ at $t$ are read off the smooth curves $L(t)$, $M(t)$ and $S(t)$, then it follows (Cole and Green 1992) that

$$z(t) = \begin{cases} ((\frac{y}{M(t)})^{L(t)} - 1)/L(t)S(t) & (L(t) \neq 0) \\ & y > 0 \\ \frac{\ln(\frac{y}{M(t)})}{S(t)} & (L(t) = 0) \end{cases} \tag{2.4}$$

If equation (2.4) is rearranged, then the $100\alpha$ centile of $y$ at $t$ is given by

$$C_{100\alpha}(t) = \begin{cases} M(t)(1 + L(t)S(t)Z_{\alpha})^{1/L(t)} & (L(t) \neq 0) \\ M(t)\exp[S(t)Z_{\alpha}] & (L(t) = 0) \end{cases} \tag{2.5}$$

where $Z_{\alpha}$ is the normal equivalent deviate. Thus if the L, M and S curves are smooth then the centiles derived from them will also be smooth (Cole and Green 1992).

To establish a growth reference using the LMS method, the data has to be divided into distinct age groups. Cole (1990) suggested that each age group should be as narrow as possible (ideally 1 year or less), with adequate numbers in each age group (at least 100 individuals). For each age group the best Box-Cox power is identified by selecting the one that gives the smallest coefficient of variation (Cole 1988). Once the powers for each age group are obtained the next issue is smoothing these, so that the L, M and S values change smoothly with age. Suggestions for how this smoothing could be achieved were to use: cubic splines, kernel methods, polynomials, or other functions for individual growth curves; see section 2.2 (Cole 1988) and exponential or fractional polynomials (Cole et al. 1998).

If the method of maximum penalized likelihood is used to provide smooth estimates of the L, M and S curves directly, then there is no need to group the data (Cole and Green 1992). The only decision to be made is the degree of smoothing of

the three curves (Cole and Green 1992). The penalized likelihood approach avoids identifying individual age groups that are good fits or outliers as it treats the entire data set as a single entity (Cole and Green 1992).

Assuming we have $n$ independent observations $y_1, \ldots, y_n$ at corresponding covariate values $t_1, \ldots, t_n$. The log-likelihood function derived from equation (2.4) (excluding an additive constant) is

$$l = l(L, M, S) = \sum_{i=1}^{n} \left( L(t_i) \log \frac{y_i}{M(t_i)} - \log S(t_i) - 0.5 z_i^2 \right) \tag{2.6}$$

where the $z_i$ are the Z-scores corresponding to the $y_i$. The curves L(t), M(t) and S(t) are estimated by maximising the penalized likelihood:

$$l - 0.5 \left( \alpha_\lambda \int L''(t)^2 dt + \alpha_\mu \int M''(t)^2 dt + \alpha_\sigma \int S''(t)^2 dt \right) \tag{2.7}$$

where $\alpha_\lambda$, $\alpha_\mu$ and $\alpha_\sigma$ are smoothing constants. The three integrals in equation (2.7) provide roughness penalties, so that maximising this equation strikes the balance between staying true to the data and smoothness of the L, M and S curves (Cole and Green 1992). Using this form of penalties leads to natural cubic splines with a knot at each value of $t$ (Cole and Green 1992). The complexity of each fitted cubic spline is measured by the 'equivalent degrees of freedom' (Cole and Green 1992) which is analogous to the number of terms in a polynomial (Cole et al. 1998).

The LMS method and maximum penalized likelihood were used to create the original UK 1990 growth reference for BMI (Cole et al. 1995), height and weight (Freeman et al. 1995). However, this reference was revised because of a sex bias for weights in infancy. The sex bias was a consequence of merging data sets separately for each sex[1] (Cole et al. 1998). In the revised reference, the data set adjustments made were the same for both sexes (Cole et al. 1998). The L curve, the power transformation, for weight starts from normality around birth (L=1) to a log transform (L=0) around 1 year (Cole et al. 1998), it then continues to decrease taking on negative values until maturity. The L-curve for height and head-circumference is fixed at a value of 1 over all ages.

Wade and Ades (1994) used parametric functions to describe the L, M and S curves; for example exponential models and maximum likelihood were used to create age-related references for CD4 lymphocyte counts for uninfected children born from mothers with HIV status. This approach guarantees that centiles are asymptotic

[1]The timing of the growth surveys (see section 1.2.5 for details) varied between 1978 and 1994 (Cole et al. 1998). All data sets were adjusted to be comparable to the NSHG survey which was one of the most recent (Freeman et al. 1995).

at adult values and avoids edge effects (Wade and Ades 1994). Wade and Ades (1994) initially randomly sampled a CD4 count from each child in order to achieve a cross-sectional sample. Later work by the same authors, explicitly modelled the correlation structure so that all CD4 counts on each child could be incorporated in the reference (Wade and Ades 1998).

In using the LMS method there is no guarantee that once the skewness is removed the resulting distribution will be normal. Cole (1993) states that after the power transformation the distribution will be nearer to normal, in particular, the mean and median will be closer together on the transformed scale than on the original scale. However, there is no certainty that the higher moments of the transformed distribution, such as the kurtosis, will coincide with those from the normal distribution (Cole 1993). Nevertheless, kurtosis tends to be less important than skewness as a contributor to non-normality (Cole and Green 1992).

### 2.1.2 Performance of the three main approaches to creating age-related references

The two main alternatives to the LMS method are the logarithmic-transformation (LOG method) of Royston (1991) and distribution-free method (HRY method) of Healy et al. (1988). The former transforms the anthropometric variable $Y$ to normality using a shifted logarithmic function $\log(Y - \tau)$ (Royston 1991), whereas the latter is a distribution free method that extends the scatterplot smoother approach of Cleveland (1979) to estimate the centiles empirically (Healy et al. 1988). However both these approaches use polynomials to achieve smooth centile estimates, which is a potential weakness because polynomials are not very good at modelling processes with an asymptote. Wright and Royston (1997) noted that time varying skewness is not easily accommodated by the shifted log-transform method.

Goldstein and Pan (1992) adapted the distribution-free method of Healy et al. (1988) to overcome the limitation generated by using polynomials in the HRY method: the empirically derived centiles were smoothed using piece-wise polynomials. Low order polynomials were derived for data that had already been split into contiguous age groups, these low-order polynomials were then smoothly joined ensuring that the derivatives of a desired order were continuous (Goldstein and Pan 1992).

Wright and Royston (1997) compared the shifted log-transform (Royston 1991), LMS method and HRY method on three data sets: serum cholesterol in men (rela-

tionship between serum cholesterol and age is weak), fetal abdominal circumference (relationship between abdominal circumference and age is strong but simple) and triceps skin-fold (relationship with age is complex). The centile curves derived for the cholesterol data were similar around the centre of the data using each of the approaches, but differences were evident in the extremes (Wright and Royston 1997). Wright and Royston (1997) concluded that all the approaches seemed to provide a good fit for the cholesterol data. The centile curves produced for the fetal abdominal circumference data using all three methods were again similar (Wright and Royston 1997). However, the distribution of the Z-scores for the abdominal circumference data had longer tails than the normal distribution (Wright and Royston 1997). The LMS method gave the most satisfactory centile curves for the tricep skinfold data (Wright and Royston 1997). The HRY and LOG method did not perform well on the skin-fold data with overfitting at higher ages (Wright and Royston 1997). Comparative papers of Pere (2000) and Healy (1992) also reached similar overall conclusions.

Overall, the LMS method seems best and it is the approach now adopted by numerous authors in the creation of growth references. The only downsides to this approach are (i) some non-normal kurtosis may remain and (ii) Z-scores have to be calculated from tabulated LMS values. The shifted log and Box-Cox transformation both include the normal and the log-normal as special cases (Cole 1993). The shifted log-normal, unlike the Box-Cox, can accommodate negative values such as velocity measurements (Cole 1993). However, the shifted Box-Cox method is also able to cope with negative values.

### 2.1.3 Other approaches to the creation of smooth centiles

In recent years numerous approaches have been put forward for deriving centile curves (Royston and Wright 1998; Rossiter 1991; Heagerty and Pepe 1999; Fatti et al. 1998; Shi et al. 1996; Sorribas et al. 2000; Tango 1998; Wellek and Merz 1995). These procedures usually suggest alternative ways of smoothing centiles, incorporating longitudinal element, transforming to normality or avoidance of distributional assumptions.

Royston and Wright (1998) proposed a parametric approach that uses an alternative transformation to normality and fractional polynomials for smoothing. Royston and Wright (1998) suggest the exponential normal transformation (or modulus exponential normal transformation) to remove non-normal skewness (and/or kurtosis).

> The shape parameter in this transformation is more directly related to skewness than is the power parameter in the Box-Cox transformation (Royston and Wright 1998, pp80).

Fractional polynomials, discussed in section 2.2.3, are used to model each parameter of the (modulus) exponential normal density (Royston and Wright 1998).

Heagerty and Pepe (1999) proposed a semi-parametric approach which does not use any underlying assumption for the distribution. The centiles are represented by a location function, scale function and a base-line distribution function (Heagerty and Pepe 1999). The scale and location functions are modelled using natural splines, whereas the base-line distribution function is estimated by local kernel smoothing of the empirical distribution function (Heagerty and Pepe 1999). Heagerty and Pepe (1999) illustrate how their semi-parametric approach can be used to condition weight on height-for-age and age.

Fatti et al. (1998) and Shi et al. (1996) both proposed approaches that took into the account the longitudinal nature of the individual profiles. Fatti et al. (1998) take a Bayesian perspective: an individual's past history (such as previous weight gain profile of a pregnant mother) forms a prior distribution, which when combined with the reference distribution at that age gives a conditional distribution for that individual's weight gain. This then forms the prior distribution for the next measurement occasion (Fatti et al. 1998). The philosophy of this approach has much in common with the conditional approaches discussed in Chapter 3. Shi et al. (1996) considered the longitudinal nature of the profiles of CD4 counts. In this approach the fixed and random effects are linear combinations of B-splines which are viewed to be more flexible than random effects models that use low order polynomials (Shi et al. 1996).

## 2.2 Parametric models for individual growth

### 2.2.1 Model requirements

Over the years there have been numerous models fitted to a child's individual growth in length or height. Some of the infancy models (see subsection 2.2.5) have also been fitted to weight data. In this section we discuss the parametric models that are most frequently cited in the growth literature. As Tanner (1989, pp8-9) puts it:

> Many attempts have been made to find mathematical curves

> which fit, and thus summarise, human and animal growth data. Most have ended in disillusion or fantasy; disillusion because fresh data failed to conform to them, fantasy because the system eventually contained so many parameters (or 'constants') that it became impossible to interpret them biologically. What is needed is a curve or curves with relatively few constants, each capable of being interpreted in a biologically meaningful way; yet the fit to actual data must be adequate, within the limits of measuring error. ... But fitting a curve to the individual values is the only way of extracting the maximum information about an individual's growth from the measurement data. ... The individual's consistency can only be measured by deviations from his own growth curve.

There are many advantages and disadvantages to modelling individual curves using parametric models. The main advantage in using a parametric function to describe the growth process is that the shape of the human growth curve can be summarised using relatively few parameters. Furthermore, predicted heights or weights at times other than measurement times can be obtained using the parametric equation: this was the approach use by Berkey et al. (1983a) to obtain correlations between heights and log-weights at the exact year time-points. Furthermore, velocity and acceleration curves can be obtained by differentiating the parametric equations with respect to time. However, if the growth data to be modelled is from a large data frame, then it can be time consuming to fit models individually. Especially when using nonlinear least squares and the Gauss-Newton method. When fitting a non-linear model parameter estimates are obtained iteratively and this can result in problems with convergence. Convergence may be very slow or the procedure may fail to converge at all, even when the starting values are close to the solution (Thisted 1988). Furthermore, the predicted values from nonlinear regression models will be biased and the extent of this bias depends upon the quantity known as the 'intrinsic nonlinearity' of the model or data set combination (Ratkowsky 1983). Providing a linear model is an 'adequate fit', it has a slight advantage over non-linear models because they are easy to fit and goodness-of-fit can be assessed using standard methods. If we were interested in deriving a mean growth curve from individual non-linear growth curves then the mean parameter curve[2] is not equivalent to the population mean curve. The mean parameter curve may systematically overestimate or underestimate mean size at each age, depending upon the curvature of the non-

[2]Mean parameter curve refers to the mean over parameter values

linear function (Merrell 1931).

A downside to a parametric model is that it imposes a fixed algebraic form on the growth process which may be too rigid to capture the true complexities of the process (Healy 1989a). This is especially true of curves based on few parameters (Healy 1989a), whereas models with a large number of parameters will be able to handle more complexities. A disadvantage in using a highly parameterised model is that a large number of height measures for each individual are required in order to achieve a reliable fit. However, even a highly parameterised model will not be able to take into account any unexpected short term variation if it is not designed to accommodate this (see section 2.5 for models developed to handle the fluctuations in the growth process, termed 'saltation and stasis' by Lampl et al. (1992)). Furthermore, interesting local variations, such as growth spurts, may be missed (Goldstein 1986a). Parametric models are also unable to account for growth variation attributable to other characteristics measured at each occasion (Goldstein 1986a), such as deprivation and seasonal variation.

Additional considerations are the assumptions made when fitting a model using ordinary least squares: namely, uncorrelated residuals with constant variance. Such requirements are unlikely to be met in a growth context (Healy 1989a). Negative correlations may occur if the time between measures is less than 1 year, because of seasonal variation (Healy 1989a). Large positive correlations could be a result of the continuity of the growth process (Healy 1989a): a child that is large on one occasion is likely to be large on the next. Other authors construe that positively correlated residuals indicate model mis-specification (Healy 1989a). However, non-constant variance and autocorrelation can be dealt with at the model fitting stage using weighted or generalised least squares. An alternative approach, discussed in section 2.3, is to use a multilevel model or random effects model that takes into account within and between individual parameter correlations (Goldstein 1986a; Laird and Ware 1982; Berkey and Laird 1986).

Some models are fitted to the whole growth curve from birth to maturity (see section 2.2.4). However, more often than not, different functions are fitted to different parts of an individual's growth curve, such as infancy and childhood models; see sections 2.2.5 and 2.2.6. In general, childhood models are highly parameterised and thus use requires a large number of height measurements during childhood. In a routine setting this is unlikely to be the case. However, the package AUXAL of Bock et al. (1994) employs Bayesian methods to fit the BTT (Bock et al. 1994) or JPA2 (Jolicoeur et al. 1992) model (see subsection 2.2.6) to incomplete height data.

### 2.2.2 Polynomials

A statistician's first approach might be to model attained height as a polynomial function of time, the complex shape of the growth process being accommodated by increasing the order of the polynomial. However, Joossens and Brems-Heyns (1975) (in Preece and Heinrich (1981)) found a polynomial of order 18 was necessary. Furthermore, polynomials are not very good in modelling processes with an asymptote, and it can be difficult to determine the appropriate order (Healy 1989a). Healy (1989a) views polynomials to be useful if growth is studied over a relatively short time. It is presumed that the time period studied does not include early infancy, the pubertal spurt or the asymptote reached at maturity.

### 2.2.3 Fractional polynomials

Some of the models outlined later in this section (Reed, Count and Guo models) are special cases of the family of fractional polynomials described by Royston and Altman (1994). A fractional polynomial is a regression function where the terms are powers of age but the choice of powers is widened to include negative as well as positive integers, with the addition of powers of 0.5 (square root), 0 (natural log) and -0.5 (inverse square root). This approach increases the flexibility of the form of the fitted curve in contrast to polynomials. Conventional low order polynomials do not always fit the data well and high order polynomials may follow the data closely but often fit badly at the extremes of the observed range (Royston and Altman 1994). Furthermore polynomials do not have asymptotes, thus they cannot fit data where limiting behaviour is expected (Royston and Altman 1994); for example, a child approaching final adult height.

Royston and Altman (1994) introduced the family of curves, fractional polynomials, whose power terms were restricted to a small predefined set of integer and non-integer values, and include conventional polynomials.

A fractional polynomial of degree $m$ is a function

$$\phi_m(X;\xi,p) = \xi_0 + \sum_{j=1}^{m} \xi_j X^{(p_j)} \tag{2.8}$$

where $X$ is a positive covariate, $m$ is a positive integer, $p_1,\ldots,p_m$ are powers with $p_1 < \cdots < p_m$, $\xi_0,\xi_1,\ldots,\xi_m$ are real coefficients, and

$$X^{(p_j)} = \begin{cases} X^{p_j} & \text{if } p_j \neq 0 \\ \ln X & \text{if } p_j = 0 \end{cases}$$

the Box-Tidwell transformation. Royston and Altman (1994) suggested that non-integer powers outside [-1,1] would not be found to be useful. Royston and Altman (1994) proposed a step-wise procedure for covariate selection and model fitting when several covariates were available. Cole (1995) chose to model the correlation structure of weight Z-scores in infancy using fractional polynomials, this application will be discussed in chapters 3 and 5.

### 2.2.4 Component-wise birth to maturity models

A parametric model covering the whole growth process has to accommodate the features of the whole growth process, described in section 1.1 in Chapter 1. In general, the model needs to capture: the steadily falling growth velocity in early life, the growth spurt in adolescence and plateau at maturity. Not an easy task for one model! The approach used in the past has been to model several adjoining segments of the growth curve. Some of the more recent childhood models, e.g. the JPPS, JPA1 and JPA2 models (Jolicoeur et al. 1992), can be applied to infancy data providing there is height data throughout childhood to maturity - these models are discussed in section 2.2.6. Two such component wise parametric models due to Karlberg (1987) and Reed and Berkey (1989) aim to model height from birth to maturity.

Karlberg (1987) devised the Karlberg ICP model on his beliefs about the hormonal regulation of the growth process. The approach of Karlberg (1987) is to split the growth process into three additive and super-imposed components; equations (2.9), (2.10) and (2.11). Each component of the model corresponds to known features of the endocrinological regulation of the growth process that can be considered in isolation from one another (Karlberg 1987).

1. The **Infancy component** is a constantly decelerating component which starts before birth and tails off by 3-4 years of age.

$$y = a_I + b_I(1 - \exp(-c_I t)) + \epsilon \tag{2.9}$$

2. The **Childhood component** commences during the first year of life at age $t_C$. The onset occurs usually between 6 and 12 months and is typically abrupt (Karlberg et al. 1987). The childhood component slowly decelerates to maturity at age $t_E$.

$$y = a_C + b_C t + c_C t^2 + \epsilon \tag{2.10}$$

3. The **Puberty component** represents the additional growth induced by puberty. It accelerates until age at peak velocity ($t_V$) and then decelerates until growth ceases ($t_E$).

$$y = \frac{a_P}{1 + \exp(-b_P(t - t_V))} + \epsilon \tag{2.11}$$

where $t$ is age in years in all 3 components (Karlberg 1987).

The fitting procedure for the ICP model involves multiple steps and is computer intensive: it requires the identification of age of onset of childhood component ($t_C$) and age when growth ceases ($t_E$). Karlberg (1987) determines the timing of the components by observing individual velocity curves. This approach has been used to create an ICP based growth chart (Fryer and Karlberg (1985) in Karlberg et al. (1987)). Karlberg et al. (1987) analysed supine length data from 191 individuals between birth and 3 years. Karlberg et al. (1987) found that the onset of the childhood component was earlier in girls than boys and is related to the magnitude of the infancy component.

Berkey and Reed (1987) originally developed their model to describe growth between birth and six years. This model, see equation (2.12), is linear in its parameters and is essentially the model of Count (1943) with the addition of increasing powers of reciprocal in age.

$$y = a + bt + c\log(t) + \frac{d}{t} + \frac{e}{t^2} + \frac{f}{t^3} + \cdots + \epsilon \tag{2.12}$$

Two versions of the general Reed model were proposed for early childhood, equations (2.12.1) and (2.12.2), which were not thought to suffer from the Count model age related bias (Berkey and Reed 1987).

First order Reed model

$$y = a + bt + c\log(t) + \frac{d}{t} + \epsilon \tag{2.12.1}$$

Second order Reed model

$$y = a + bt + c\log(t) + \frac{d}{t} + \frac{e}{t^2} + \epsilon \tag{2.12.2}$$

As log of zero is undefined it is usual to add a constant to the actual age when using the Reed (and Count) models. Reed and Berkey (1989) extended the approach used for early childhood to cover the whole period of growth. This involves joining two Reed models using splines, one covering childhood and the other adolescence (Reed and Berkey 1989). At the age where the two models join, the distance and

velocity curves should be continuous (Reed and Berkey 1989). The resulting model has between 6 and 8 parameters depending on the order of the Reed model required in childhood and adolescence. Reed and Berkey (1989) suggested the following Reed models for childhood and adolescence:
childhood

$$y = a_c + b_c x + c_c \log(x) + \frac{d_c}{x} + \frac{e_c}{x^2} + \epsilon \tag{2.13}$$

adolescence

$$y = a_A + c_A \log(x) + \frac{d_A}{x} + \frac{e_A}{x^2} + \frac{f_A}{x^3} + \epsilon \tag{2.14}$$

where $x$ is a linear transformation of chronological age and $y$ is length or weight. The term $b_A x$ is excluded in the adolescent part, this is because $b_A$ represents the asymptotic velocity which is zero at maturity (Reed and Berkey 1989). A reciprocal cubic was added to the adolescent component to improve flexibility (Reed and Berkey 1989). The transformed age variable has its origin $x = 0$ near conception (9 months subtracted from birth) and $x = 1$ is near the 'boundary' between childhood and adolescence. Equations (2.13) and (2.14) are combined by the use of an indicator variable $z$ which is zero during childhood and one when $x \geq 1$. The resulting equation is discontinuous where the two functions are joined, but Reed and Berkey (1989) resolved this by using the spline conditions for continuity in attained size and velocity, to give the 8 parameter equation:

$$\begin{aligned} y = {} & b_1 + b_2 \log(x) + \frac{b_3}{x} + \frac{b_4}{x^2} + b_5(z + x - zx + z\log(x)) \\ & + b_6(-2z - x + zx + \frac{z}{x}) + b_7(-3z - 2x + 2zx + \frac{z}{x^2}) \\ & + b_8(-4z - 3x + 3zx + \frac{z}{x^3}) \end{aligned} \tag{2.15}$$

Reed and Berkey (1989) illustrated this approach on 2 boys and 2 girls. For the examples given their model faired better than the triple logistic model of Bock and Thissen (1980); see section 2.2.6. Reed and Berkey (1989) suggested that for large data frames the boundary should be fixed at some reasonable age for children of the same sex. However children mature at different rates, so using this approach in practice may not be as straightforward as implied.

There are no published comparative papers on the Karlberg ICP model or Reed linear model from birth to maturity. One reason these models may not have subsequently appeared in the research literature is that they both contain a subjective element: determination of timing of components in ICP model and location of the join point of two Reed models.

### 2.2.5 Infancy models

Within this section we summarise the models derived primarily for infancy length data, but many have also been applied to weight data. In table 2.1 we present the infancy models in chronological order. Table 2.2 summarises the characteristics of the infancy models applied to length data. In general, it seems that between 3 and 5 parameters are adequate for modelling growth in infancy.

Two early models for length data were the non-linear Jenss and Bayley (1937) model and the linear Count (1943) model. Jenss and Bayley (1937) developed a model for length over the age range birth to six years and also applied the same model to weight data. Jenss and Bayley (1937) termed $\exp(d)$ the term 'growth/acceleration constant'. Byrad et al. (1991) applied the Jenss model to length and weight data from the Fels longitudinal study. They concluded that the Jenss model was designed to fit the usual pattern of growth and that minor variations from this pattern could result in extreme parameter values. Berkey (1982a) used the empirical Bayes approach to fit the Jenss model, using the prior distribution of growth-model parameters estimated from a large sample. Berkey (1982a) viewed this approach to be beneficial for infants that had missing observations and on average, to be less sensitive to measuring errors. Count (1943) originally proposed a three component model to cover the whole period of growth. However only the infancy component, see table 2.1, has been utilised in the literature. Count (1943) applied his model to several sources of data from differing nationalities. Often 1 is added to time to avoid taking the log of zero when fitting the Count model. Count

Table 2.1: **Infancy models**: $y$ denotes length or weight, $t$ denotes age (which is adjusted if model includes a $\log(t)$ term) and $\epsilon$ denotes the error term (adapted from Berkey and Reed (1987))

| | |
|---|---|
| Jenss-Bayley model (1937) | $y = a + bt - \exp(c + dt) + \epsilon$ |
| Count model (1943) | $y = a + bt + c\log(t) + \epsilon$ |
| Bialik model (1973) | $\log(y) = a + b\log(t) + c(\log(t))^2 + \epsilon$ |
| Kouchi model (1985) | $y = a + bt^c + \epsilon$ |
| 1st order Reed model (1987) | $y = a + bt + c\log(t) + \frac{d}{t} + \epsilon$ |
| 2nd order Reed model (1987) | $y = a + bt + c\log(t) + \frac{d}{t} + \frac{e}{t^2} + \epsilon$ |
| Karlberg infancy comp. (1987) | $y = a + b(1 - \exp(-ct)) + \epsilon$ |
| Guo model (1988) | $y = a + b\sqrt{t} + c\log(t) + \epsilon$ |

(1943) suggested that age at conception could be zero. Berkey and Reed (1987) explored the effect of varying the constant when fitting the Count and Reed models (discussed above) but came to no clear solution. However changing the constant added to age could affect the shape of the curve (Berkey and Reed 1987). Berkey and Reed (1987) suggested that $t = \frac{\text{months since birth}+9}{9}$ was an intuitively appealing transformation of time. Smith et al. (1983) found that the Count model was applicable to the weight patterns observed in 22 premature infants, and the only discrepancies were in the first few days when weight loss was most rapid and variable. Berkey (1982b) compared the fit of the Jenss and Count models on longitudinal weight and length data of 229 children from the Harvard Growth study. The Harvard children had up to 14 measures near targets ages of 3 to 12 months in 3 month intervals and every 6 months from 1 to 6 years (Berkey 1982b). Berkey (1982b) concluded that the Count model did not give a satisfactory fit for length or weight in this age-range and was found to have systematic age-related deficiencies. Berkey (1982b) found the Jenss model performed well for both length and weight. However use of the Jenss model may present 'minor problems' around the age of 6 months (Berkey 1982b).

The Reed first and second order models are essentially extensions of the Count model (Berkey and Reed 1987). Berkey and Reed (1987) found that the Reed model described the growth in length as well as the Jenss model. One advantage of the Reed model is that it can accommodate points of inflection and possibly provide a representation of some abnormal patterns of growth (Berkey and Reed 1987).

The model proposed by Bialik et al. (1973) is included in table 2.1 because it was derived from infancy weight data, the focus of the majority of the analysis in this thesis. However, since publication, it has not been utilised in the growth literature.

Table 2.2: Properties of infancy models

| Model | No. of parameters | Linear in parameters | Limit of Vel. as age increases |
|---|---|---|---|
| Jenss-Bayley model (1937) | 4 | N | $b$ |
| Count model (1943) | 3 | Y | $b$ |
| Kouchi model (1985) | 3 | N | $b$ |
| 1st order Reed model (1987) | 4 | Y | $b$ |
| 2nd order Reed model (1987) | 5 | Y | $b$ |
| Karlberg infancy comp. (1987) | 3 | N | 0 |
| Guo model (1988) | 3 | Y | 0 or 1 or $\infty$ (*) |

(*) depends on the magnitude and sign of $c - 1$

Bialik et al. (1973) applied their model to weight data from 449 infants in a rural area of Israel, with an average of 17.5 weights per child from birth to 1 year.

The model proposed by Kouchi et al. (1985b) was developed on length data from the Fels longitudinal study. The model was fitted to 441 individuals (229 males, 212 females), who had five to seven length measures recorded from 1 month to 2 years, born between 1931 and 1974. Kouchi et al. (1985a) also applied their model to weight data from the Fels longitudinal study: 344 infants (176 male, 168 female), born between 1930 and 1970, that had at least six data points including birth weight. The parameters of the Kouchi model were given interpretations, $a$ was birth weight or length at 1 month, $b$ was intrinsic growth rate and $c$ was pattern of growth (Kouchi et al. 1985b; Kouchi et al. 1985a). Kouchi et al. (1985a) and Kouchi et al. (1985b) explored relationship of the parameters within the Kouchi model and adult status. Two results of interest to this thesis were that head circumference at 18 and 30 years was found to be significantly correlated with rate of growth in infancy and correlation between rates of growth in infancy and adult stature were found to be as high as those between stature at 2 years and adult height (Roche 1992).

The model proposed by Guo et al. (1988) was also derived from data from the Fels longitudinal study. Head circumferences from 462 infants (247 boys, 215 boys) and weight (and length) data from 504 infants (265 males, 239 females) were used for the purpose of the construction of a 1 month increment reference for head-circumference (Guo et al. 1988), weight and length (Roche et al. 1989). The Guo model was used to interpolate monthly values for head-circumference, weight and length in order to create the increment reference; see section 3.3.3 in chapter 3.

Simondon et al. (1992) compared five of the infancy growth models: the Count model, the Kouchi model, Reed first- and second-order models and the Karlberg infancy component. Simondon et al. (1992) considered the fit to weight data of 95 Congolese infants, this was a selective sample from 2429 children measured on at least 7 occasions, between birth and 13 months of age. The Kouchi model provided a poor fit with an unacceptably high degree of collinearity and it failed to fit five curves (Simondon et al. 1992). In agreement with Berkey (1982b), the Count model was also found to provide a poor fit. The I-component of the Karlberg model was viewed to be the best three parameter model if it was used between the ages of 2 and 12 months. The Reed models provided the best fit, but the five parameter model was not viewed to be superior to the four parameter model (Simondon et al. 1992). Simondon et al. (1992) did not consider the Jenss and Guo model.

Peerson et al. (1993) compared nine models: the Count model, the Guo model,

the Jenss model, the Reed first- and second order models and the Karlberg model (six infancy models) and three polynomial models (quadratic, cubic and quartic). The models were fitted to length, weight, and head circumference of 39 breast-fed and 31 formula-fed infants from the DARLING study to compare breast and bottle fed infants over the age range 1 to 24 months. Peerson et al. (1993) considered the infancy and childhood component of the Karlberg model but excluded the quadratic term. Peerson et al. (1993) recommended the Karlberg model for length data of both breast-fed and bottle-fed infants. The Jenss model was suitable for describing weight growth in formula-fed infants and this model was thought to be appropriate for head circumference growth in both formula-fed and breast-fed infants (Peerson et al. 1993). Peerson et al. (1993) viewed none of the models to be adequate to describe weight curves of breast-fed babies. Peerson et al. (1993) did not consider the Kouchi model.

### 2.2.6 Childhood models

In this section we summarise models that have been applied to childhood height data, these models usually cover growth after infancy and are able to accommodate the pubertal growth spurt. In the literature, none of the more recent models appear to have been applied to weight data. In table 2.3 we summarise the parametric models applied to height data in childhood.

The earliest models applied to longitudinal height data were the Gompertz (1825) model (in Preece and Heinrich (1981)) and the logistic function. The logistic curve formed the third component of the component-wise model of Count (1943). Quo (1953) applied the logistic curve to average weight from 13 years to maturity. Laird (1967) applied a Gompertz style equation, in three phases, to weight data of males. Deming (1957) applied the Jenss model (infancy and childhood) and Gompertz model (adolescence) to length data of 24 boys and 24 girls. Marubini et al. (1971) compared the fit of the Gompertz and logistic models to height measures of 121 girls during adolescent growth and concluded that both models were adequate for describing the growth process. However, the logistic model gave the better quality of fit in terms of residual variance (Preece and Heinrich 1981). The Gompertz function is later used in the model developed by Shohoji and Sasaki (1987b), discussed below.

Preece and Baines (1978) adopted a different approach to the derivation of models for height data, they derived three models from the differential equation:

$$\frac{dy}{dt} = s(t) \times (y_1 - y) \tag{2.16}$$

Table 2.3: **Childhood models**: $y$ usually denotes height, $t$ usually denotes age from birth (unless stated otherwise) and $\epsilon$ denotes the error term

| | |
|---|---|
| Gompertz (1825) | $y = P + K\exp(-\exp(a - bt))$ |
| logistic | $y = P + \frac{K}{1+\exp(a-bt)}$ |
| | $P$ represents height at start of adolescence |
| | $K$ represents gain in height during adolescence |
| Pearl polynomial logistic (1925) | $y = \frac{A}{1+\exp[C+D_1t+D_2t^2+D_3t^3+D_4t^4+D_5t^5]} + \epsilon$ |
| Revised Pearl (1992) | $y = \frac{A}{1+\exp[\frac{B}{t}+C+D_1t+D_2t^2+D_3t^3+D_4t^4+D_5t^5]} + \epsilon$ |
| Preece-Baines (1978) | $y = y_1 - \frac{2(y_1-y_\theta)}{\exp[s_0(t-\theta)]+\exp[s_1(t-\theta)]}$ |
| Double logistic (1973) | $y = \frac{a_1}{1+\exp[-b_1(t-c_1)]} + \frac{f-a_1}{1+\exp[-b_2(t-c_2)]} + \epsilon$ |
| | $f$ is the height at maturity |
| Triple logistic (1976) | $y = a_1\left(\frac{1-p}{1+\exp[-b_1(t-c_1)]} + \frac{p}{1+\exp[-b_2(t-c_2)]}\right)$ |
| | $+\frac{f-a_1}{1+\exp[-b_3(t-c_3)]} + \epsilon$ |
| BTT (1994) | $y = \frac{a_1}{[1+\exp(-b_1(t-c_1))]^{d_1}} + \frac{a_2}{[1+\exp(-b_2(t-c_2))]^{d_2}}$ |
| | $+\frac{a_3}{[1+\exp(-b_3(t-c_3))]^{d_3}} + \epsilon$ |
| SS (1987) | $y = (1 - \exp(-\exp(A - Bt)))(C + Dt + E\log(t))$ |
| | $+U\exp(-\exp(A - Bt)) + \epsilon$ |
| KS6 (1990) | $y = C + Dt + E\log(1 + t)$ |
| | $+\exp(-\exp(A - Bt))[U - C - Dt - E\log(1 + t)] + \epsilon$ |
| | $U$ is adult height |
| SSC (1993) | $y = AW(t) + f(t)[1 - W(t)] + \epsilon$ |
| | $W(t) = \exp[-\exp B(G - t)]$ |
| | $f(t) = C + Dt + E\log(t)$ |
| JPPS (1988)* | $y = A\left(1 - \frac{1}{1+\left(\frac{t}{D_1}\right)^{C_1}+\left(\frac{t}{D_2}\right)^{C_2}+\left(\frac{t}{D_3}\right)^{C_3}}\right) + \epsilon$ |
| JPA1 (1992)* | $y = A\exp\left(\frac{-1}{C_1\log\left(\frac{t}{D_1}+1\right)+\left(\frac{t}{D_2}\right)^{C_2}+\left(\frac{t}{D_3}\right)^{C_3}}\right) + \epsilon$ |
| JPA2 (1992) | $y = A\left(1 - \frac{1}{1+\left(\frac{t+E}{D_1}\right)^{C_1}+\left(\frac{t+E}{D_2}\right)^{C_2}+\left(\frac{t+E}{D_2}\right)^{C_2}}\right) + \epsilon$ |

(*) $t$ is measured from conception rather than birth

where $y$ is height at time $t$, $y_1$ is final (or adult) height and $s(t)$ is a function of time which differs between models. All models were applied to height data from 35 boys and 23 girls belonging to the Harpenden Growth study. The three models derived by

Preece and Baines (1978) performed better than the double logistic model (discussed below), from the age 2 to maturity, and were found to be better than anything else available at that time. The best model overall was the Preece-Baines Model 1:

$$y = y_1 - \frac{2(y_1 - y_\theta)}{\exp[s_0(t-\theta)] + \exp[s_1(t-\theta)]} \tag{2.17}$$

where $s_0$ and $s_1$ are rate constants ($\frac{ds}{dt} = (s_1 - s)(s - s_0)$), $\theta$ is a time constant (related to age at peak velocity (Preece and Heinrich 1981)), $y_1$ is an estimate of adult size and $y_\theta$ is height at $t = \theta$. Hauspie et al. (1980) applied the Preece-Baines model 1, double logistic, logistic and Gompertz models to height data of 35 Belgian girls followed from birth to 18 years. Considering the height data between 2 and 18 years, the Preece Baines model 1 provided a better fit than the double logistic model (Hauspie et al. 1980). Neither model coped with height data at ages less than 1 year (Hauspie et al. 1980). During adolescence the double logistic curve fitted slightly better than the Gompertz curve (Hauspie et al. 1980).

Bock et al. (1973) proposed the double logistic model for recumbent length to cover the age range 1 year to maturity. In order to use this equation the height (or length) at maturity ($f$) is required. The first term represents the prepubertal component, this continues in a reduced degree until maturity, and the second represents the contribution of the adolescent spurt (Bock et al. 1973). The double logistic model was applied to length data of 56 boys and 51 girls from the Fels study (Bock et al. 1973). Bock and Thissen (1976) (in Bock and Thissen (1980)) then proposed the triple logistic model. The first two components (early- and mid- childhood) of the triple logistic model represent prepubertal growth. Bock and Thissen (1980) suggested using Bayes estimation if data were incomplete. Guo et al. (1992) applied the Preece-Baines model 1, the triple logistic model and kernel regression (discussed below) to heights of 143 boys and 84 girls over the age-range 2 to 18 years from the Fels study. Guo et al. (1992) found that the Preece-Baines model could not accommodate a mid-growth spurt. Guo et al. (1992) recommended the triple logistic model or kernel regression, with the latter been preferable as it required comparatively less data. Bock et al. (1994) (in Pan (1995)) later proposed the Bock-Toit-Thissen (BTT) model which is equivalent to the triple logistic model if $d_1 = d_2 = d_3 = 1$. The BTT model is employed in the growth package AUXAL, which also uses Bayesian methods of estimation to cope with incomplete data. AUXAL also fits the JPA2 model and also uses the non-structural approach of kernel smoothing (Gasser et al. 1984), both described below.

The model developed by Shohoji and Sasaki (1987b) (SS) was applied to height data of 2567 males and 971 females from the Hiroshima growth study over age-range

6 to 22 years. However, Shohoji and Sasaki (1987b) noted that the majority of the height data is in the age range 12 to 16 years. This function contains components that correspond to the Count model and Gompertz function. Cole (1993) suggested replacing the Count function with the Jenss function (SSC) and later put this into practice (Ledford and Cole 1998). The SS model has also been applied to average weight data of the Savannah Baboons (Shohoji and Sasaki 1987a; Shohoji and Sasaki 1985). The SS model is not invariant to change in time scale (Jolicoeur et al. 1992). The Kanefuji and Shohoji (1990) six parameter (KS6) model was a revision of the SS model to allow the inclusion of birth length, which was achieved by adding one to $t$ in the log terms. The KS6 model, JPPS model (Jolicoeur et al. 1988) and Preece-Baines model were fitted to height data of 365 Japanese girls. Kanefuji and Shohoji (1990) concluded that the KS6 model provided the best fit. They also demonstrated an application of empirical Bayesian methodology with their model.

Jolicoeur et al. (1988) proposed the JPPS model for height from infancy to maturity. Jolicoeur et al. (1988) illustrated the better fit of their model, over Preece-Baines model and Shohoji-Sasaki model, on 13 boys and 14 girls with data from 1 month to 19 years. The JPPS model was revised to give the JPA1 model (includes prenatal data) and JPA2 model (for postnatal data) (Jolicoeur et al. 1992). Jolicoeur et al. (1992) compared the fit of eight models: JPPS, JPA1, JPA2, KS6 and KS7 (adaptation by Jolicoeur et al. (1992)), Preece-Baines model 1, triple logistic and Pearl polynomial logistic (adapted by Jolicoeur et al. (1992)) applied to stature of 27 healthy children (41 height measures on each child) from the French Auxological survey. Jolicoeur et al. (1992) found that both the JPA1 and JPA2 provided better fits than the initial JPPS model. Overall Jolicoeur et al. (1992) found the JPA2 model provided the best fit to the French height data. Bock (1995) (in Ledford and Cole (1998)) compared an improved version of BTT, JPA2 and KS7, concluding that the JPA2 and BTT models were similar in performance. Abidi et al. (1996) demonstrated the application of empirical Bayes approach to the height records of 13 boys with four growth models (JPA1, JPA2, triple logistic and modified SS model), they viewed this approach to be of value when height data were incomplete, but emphasised that the quality of the prediction would depend on the quality of prior information.

Ledford and Cole (1998) compared three models (JPPS, Shohoji-Sasaki model adapted by Cole (SSC) and Preece-Baines model 1) applied to the stature of 91 individuals (48 males, 43 females) with 26 measurements, from 1 to 17 years, from the French Auxological study. Ledford and Cole (1998) have made comparisons of the childhood growth models on the same data-frame used by Jolicoeur et al. (1992)

but these individuals have fewer height measures. Ledford and Cole (1998) chose not to use the modifications of JPPS model as these would only be of value if height data at ages less than 1 year were to be used. Overall the JPPS model provided the best fit, closely followed by the SSC model (Ledford and Cole 1998). However the JPPS model may give an impression of a mid-growth spurt when none is present (Ledford and Cole 1998). They also highlighted the similarities between the SSC model and Karlbergs ICP model. However the former model is non-monotonic and can decline after reaching peak height. They also found that height data in early infancy was important in the fitting of both the SSC and JPPS model, because without this models for 3 children failed to converge.

## 2.3 Population approaches to growth data

In the last section we considered models that have been applied to, predominantly, an individual's height profile. The methods discussed in this section view individuals as belonging to a population. The overall aim is to describe the growth process of a population and use the information on the population to improve the estimates of an individual's own parameters. These approaches can be valuable for comparing growth in different populations and in considering the impact of other factors (Pan 1995). A series of height or weight observations are usually highly correlated from one occasion to the next. Therefore such correlation must be accounted for in the estimation and tests for parameters (Morrison 1976).

### 2.3.1 Generalised multivariate analysis of variance

Potthoff and Roy (1964) generalised multivariate analysis of variance (MANOVA) to encompass growth curve problems by adding a design matrix. Rao (1966) (in Geisser (1980)) demonstrated that the analysis of Potthoff and Roy (1964) was inadequate. MANOVA assumes that the expected values of individual serial growth data follow a multivariate normal distribution with mean values, polynomial functions of time, and a unstructured covariance matrix (Rao (1966) in Guo et al. (1992)). Grizzle and Allen (1969) incorporated covariates into the model proposed by Potthoff and Roy (1964) and Rao (1966). Geisser (1980) suggests that the Potthoff-Roy model was prolific because it provided a general format for a variety of growth situations. The 'growth curve model' is now a field of study in its own right and Von Rosen (1991) provides a review of developments since the model was established. A multi-dimensional extension of the Potthoff-Roy model has been established to deal with

the situation where several responses are simultaneously measured (Kshirsagar and Smith 1995). Often assumptions are required about the structure of the variance-covariance matrix if the number of individuals is not large enough (Kshirsagar and Smith 1995). Furthermore, most of the research literature on growth curve models deals with the situation where the covariance matrix of the observations on an individual at different time-points has a 'patterned' structure (Kshirsagar and Smith 1995).

Geisser (1970) considered the model proposed by Potthoff and Roy (1964) from a Bayesian viewpoint. Fearn (1975) adopted a different approach where individuals are treated as exchangeable with respect to their growth curve, making use of the theorem discussed by Lindley and Smith (1972). In this approach both the population trends and individual trends are modelled (Darby and Fearn 1979). Darby and Fearn (1979) applied the Bayesian growth curve model to longitudinal blood pressure data. Darby and Fearn (1979) also suggest that this approach allowed the inclusion of individuals with some missing observations.

A major limitation to this approach is the requirement that all children have the same number of height measures and be measured at the same time; even in a research study this is a tall order. Thus in order to apply this approach we have to case-wise delete children that don't have the required number of height measures or impute missing height measures. Furthermore, this model is unsuitable if individuals are measured at irregular times.

### 2.3.2 Random effects models

Two-stage random effects models can be applied to longitudinal data that is irregularly spaced and has missing observations (Laird and Ware 1982). Laird and Ware (1982) presented a general linear random-effects model that included growth models as a special case. In this formulation, the probability distribution for the multiple measurements has the same form for each individual, but the parameters of that distribution vary over individuals (Laird and Ware 1982). The distribution of these parameters, termed the 'random effects', in the population forms the second stage of the model (Laird and Ware 1982). Laird and Ware (1982) proposed a unified approach to fitting the two stage model based on a combination of empirical Bayes and maximum likelihood estimation of model parameters and using the EM algorithm. Berkey et al. (1989) applied the Reed model for adolescence to the heights of 62 boys from the Harvard growth study covering the age range 8 to 18 years. The random-effects model of Laird and Ware (1982) was then used to study the effect of

protein intake (Berkey et al. 1989). If a non-linear model is suitable for describing the growth curve of individuals then the two-stage approach described by Berkey and Laird (1986) is appropriate. Berkey and Laird (1986) explored the impact of gender and protein intake on parameters of the Jenss model fitted to recumbent length data of 229 children from the Harvard growth study covering the age range 3 months to 6 years.

Strenio et al. (1983) developed an approach related to the random effects model discussed above. The approach adopted was to incorporate additional information from the population to improve the estimation of individual growth curves (Strenio et al. 1983). It was suggested that any reasonable method for combining individual data with cross-sectional data ought to take into account reliability of measurements and the number of measurements per individual (Strenio et al. 1983). Strenio et al. (1983) felt that empirical Bayes methodology met these requirements and applied it to the problem of estimating growth curves. They also incorporated growth-related covariates into the model.

One of the advantages in using the random effects model to describe longitudinal growth over the Potthoff-Roy model (discussed above) is that it can still provide a growth curve for a child with only one height measure (Berkey et al. 1989). Furthermore, information from each individual contributes to the estimation of covariate effects and mean population curves.

### 2.3.3 Multilevel modelling

Multi-level modelling was initially developed in the educational context; pupils belong to classes that are within schools (Goldstein 1986b). The clustering of pupils within classes (within schools) imposes a correlation structure on the data that reflects the shared experiences of pupils within the same classroom (or school) (Rice and Leyland 1996). The philosophy behind multilevel modelling is to specify a model that takes the hierarchical structure of the data into account at the fitting stage. In a standard linear model there may be a high degree of collinearity between parameter estimates; whereas the multi-level model approach takes within- and between-individual parameter correlations into account.

Multi-level models can also be used in a repeated measures context; the measurement occasions are clustered within individuals (Goldstein 1995). The main advantage in using multi-level models for repeated measures data is that there is no requirement that all children have the same measurement ages (Goldstein 1986a).

The multi-level model incorporates the random effects model, discussed above, as a special case (Pan 1995).

Longitudinal data falls into a two-level hierarchy with the level 2 units representing the individual and the level 1 units representing the measurement occasions for that individual (Goldstein 1995). There is typically more variation at level 2 than level 1 (Goldstein 1995) as this represents say the variation in heights between individuals. The basic two-level model for individual $i$ on measurement occasion $t = 1, \ldots, n_i$ ($n_i$ is the number of measurement occasions for individual $i$) is (Goldstein 1986a):

$$y_{it} = \sum_j \beta_{ij} x_t^j + \sum_k \alpha_k z_{itk} + \epsilon_{it} \tag{2.18}$$

where

$$\beta_{i0} = \beta_0 + u_{i0}$$
$$\beta_{ij} = \beta_j + u_{ij}$$

In the above equation $x_t$ is the age at occasion $t$ and $j = 1, \ldots, p$ indexes the coefficients of the polynomial. The first summation in equation (2.18) represents the polynomial fitted to the set of say height measurements; whereas the second summation is over a set of further explanatory variables indexed by $k$ (Goldstein 1986a). The $\epsilon_{it}$ and $u_{ij}$ are the level 1 and level 2 residuals, respectively. The level 1 residuals have $\text{cov}(\epsilon_{it}, \epsilon'_{it}) = 0$ and $\text{var}(\epsilon_{it}) = \sigma_0^2$. The level 2 residuals have zero expectation but are not necessarily independent (Goldstein 1986a).

Instead of a polynomial in age it is possible to consider other linear models such as the Count (1943) model (Goldstein 1986a). A non-linear model in age may be more appropriate for growth data than a polynomial model; in particular in infancy where growth is rapid or late adolescence where growth reaches an asymptote at adult height (Goldstein 1995). For example, Steward (1994) applied the Jenss model (Jenss and Bayley 1937) using multi-level modelling to length data from two Welsh towns (collected 1973-75). Steward (1994) experienced convergence difficulties in fitting the non-linear Jenss model and suggested that further work was required to assess the reliability and accuracy of estimates obtained. Multi-level models were fitted to height and head circumference data from the Edinburgh longitudinal growth study, spanning birth to maturity, using extended spline models (Pan 1995; Pan and Goldstein 1998). These extended spline models were viewed to be more flexible than polynomials; they allow variable order functions and fractional polynomials (Pan 1995; Pan and Goldstein 1998). In the two-level model above we assume that the

level one residuals are uncorrelated. However if height measures are made close in time they are likely to be highly correlated. Goldstein et al. (1994) suggested using first or second order autoregressive models for the level 1 residuals. It is also possible to incorporate measurement error into the multi-level analysis (Goldstein 1995). A multivariate extension of multi-level models has also been developed to deal with the situation of modelling several responses as functions of the explanatory variables (Goldstein 1995). This approach is adopted in LGROW to establish longitudinal growth norms for height and weight (Pan and Goldstein 1997), see section 3.6 in chapter 3 for further discussion.

Multilevel models have also been applied to cross-sectional height and weight data. Pan et al. (1992) used a two-level model with grafted piece-wise polynomials for cross-sectional weight data. Goldstein (1989a) and Goldstein (1989b) discuss the use of two-level models for predicting adult heights from serial height measures.

Specialist software is required to estimate the coefficients within a multilevel model e.g. MLn uses iterative generalised least squares, and HLM and BUGS adopt a Bayesian approach to estimating coefficients. Unlike GMANOVA, multi-level models are able to deal with missing data and different timing in measurement occasions for each child. However, the multi-level approach assumes that measures are missing at random and that the timing of measurement is not related to actual measure. Thus multi-level models can not easily accommodate bias, such as light infants being weighed more often.

## 2.4 Non-parametric approaches

The disadvantage in using a parametric approach to model child growth is that the imposed functional form may be too rigid to capture the true complexities of the growth process (Healy 1989a). The overall aim of non-parametric approaches are to impose a less rigid structure upon the fitted curve but at an expense of introducing a larger number of parameters (Healy 1989a). The non-parametric approach is largely data driven and tend to be only suitable if there are a large number of measurements for each individual. The philosophy behind non-parametric approaches is that no assumptions are made about the form of the curve for each subject and the shape is determined locally by the data. It usually involves a decision to be made on the smoothing parameter; if this is too small we will be modelling random variation but if it is too large interesting local variations will be missed (Goldstein 1986a). Therefore, the process of smoothing is subjective, it involves some trade-off between

smoothness and goodness of fit (Cole 1993). Goldstein (1986a) suggests that one of its advantages is its ability to identify local events of interest, such as growth spurts. The disadvantages of non-parametric approaches are that they are computationally intensive and they do not yield compact expressions for prediction (Royston and Altman 1994); so usually have to be specified as tabulated values (Cole 1993). Furthermore, smoothing procedures tend to underestimate the peak velocity and the smoothed velocity pattern depends on the location of time-points and the time interval between measures (Milani 2000).

Here we discuss the following non-parametric approaches: spline functions, kernel estimation, longitudinal principal components and curve registration. Another alternative is to use the method of penalised likelihood, see discussion in section 2.1.1, which was used to produce LMS curves to obtain smooth centile curves for the revised UK 1990 reference (Cole and Green 1992; Cole et al. 1998).

### 2.4.1 Spline functions

A spline function is a curve made up of a series of piecewise polynomials (often cubics) joined at a sequence of points called knots and constrained so that both the slope and the curvature of adjacent pieces are equal at the knot (Healy 1989a). The smoothness of the splines is determined by the number and placing of the knots. Greater flexibility in the shape of the non-parametric curve can be achieved by increasing the number of knots (Healy 1989a). Healy (1989a) warns that the derivatives of spline curve are less smooth than the curve itself.

A cubic spline was used to smooth empirical percentile values for weight, height and head circumference in the construction of the NCHS reference (Hamill et al. 1979). Largo et al. (1978) applied cubic smoothing splines to the velocity curve of height in adolescence for children from the Zurich study and found that this approach gave results very similar to those of kernel estimation. Variable knot cubic splines have been used to fit height curves for children aged four to eleven years, but were found to be inadequate for early childhood and adolescence (Berkey et al. 1983b).

### 2.4.2 Kernel estimation

Kernel estimation is based on a weighted averaging of the observations within specified age intervals, in which the weights are obtained by integrating a pre-specified

kernel function (Guo et al. 1992). When applying kernel estimation, the kernel functions and the bandwidth have to be fixed (Gasser et al. 1989). The kernel function is usually derived from some mathematically optimal conditions (Guo et al. 1992). The lengths chosen for the specified age intervals determine the extent of smoothing and the goodness of fit; short intervals lead to less smoothing but result in a better fit. Gasser et al. (1989) suggests that the optimal choice of smoothing parameter is usually dependent on the particular growth pattern and the amount of variability. It is suggested that, depending on the problem, the minimum number of measurements per individual to employ kernel estimation lies in practice between 12 and 15 (Gasser et al. 1989).

Gasser et al. (1984) and Gasser et al. (1985) have used kernel estimation to determine the distance, velocity and acceleration curves for children from the Zurich study. Kernel estimation has been used to determine the timing of the mid-growth spurt and pubertal spurt from the acceleration curve (Gasser 1985). Gasser (1985) found that the timing of the mid-growth spurt was similar in boys and girls. Guo et al. (1992) compared kernel estimation to the triple logistic and Preece-Baines models applied to height data (2 to 18 years) from the Fels study. Guo et al. (1992) found that the kernel estimation and the triple logistic model provided similar descriptions of the pubertal spurt. Pan and Ratcliffe (1992) used kernel estimation to determine the distance (height) curve, numerical differentiation was used to derive the velocity and acceleration curves.

### 2.4.3 Shape-invariant modelling

The shape-invariant model (SIM) of Stützle et al. (1980) is a semi-parametric approach; it is intermediate between the parametric and nonparametric approaches. The SIM approach employs a mathematical algorithm, initially some guess about the functional form of the growth process is provided and this is improved iteratively using the data (Stützle et al. 1980). Stützle et al. (1980) chose to model the velocity curve over the distance curve because it was felt that this was closer to reflecting the dynamics of the growth process. It should be noted that the velocities were derived from the distances. The initial functions chosen were non-linear and this initial choice was then modified by the addition of a B-spline (Stützle et al. 1980). Stützle et al. (1980) applied their SIM approach to height velocity data of a random sample of 45 boys and 45 girls (longitudinal height data from 1 to 20 years). It was found that the same model structure was adequate for boys and girls (Stützle et al. 1980).

### 2.4.4 Longitudinal principal components

Kent (1975) (in Berkey and Kent (1983)) developed longitudinal principal component (LPC) analysis based upon a model suggested by Rao (1958). In using LPC analysis we are not restricted to any particular curve type or age range (Berkey and Kent 1983). The LPC approach defines a set of functions from the sample data and then expresses each individual's growth data as a linear combination of these functions (Berkey and Kent 1983). An extension of this approach summarises growth curves for several variables simultaneously (Berkey and Kent 1983).

In using the LPC approach, the need is to determine the number of principal components which account for as much of the variation in the data as possible. It was suggested that the variation unaccounted for could relate to pre-specified estimate of measurement error variance (Berkey and Kent 1983). In this approach the eigenvectors and factor scores are rescaled to simplify parameter interpretation (Kent (1975) in Berkey and Kent (1983)). A polynomial of degree seven was then fitted to each of the re-scaled eigenvectors (Berkey and Kent 1983).

Berkey and Kent (1983) applied LPC method to longitudinal height and weight data in early childhood (3 months to 6 years) from the Boston study. Berkey and Kent (1983) compared the fit of the four component LPC model with fit of the Jenss curve (Jenss and Bayley 1937). Berkey and Kent (1983) concluded that the four component LPC model was superior to the Jenss curve for length, but not for weight.

A disadvantage of the LPC approach is the requirement that every individual has the same set of target ages (Berkey and Kent 1983). The ability of the LPC approach, developed by Kent, to cope with incomplete data is viewed as an advantage (Berkey and Kent 1983). However, Berkey and Kent (1983) suggest that if there are numerous missing observations and irregular measurement ages then the Jenss curve should be used in preference for early childhood data.

### 2.4.5 Curve registration and structural averages

The aim of curve registration as applied to growth data is to display it in a way that highlights the characteristics of the growth process, i.e. spurts, troughs and levelling off. Once this is achieved, we can then look at important sources of pattern and variation. For example, do individuals that differ from the general form do so in some systematic way?

> The basic philosophy of functional data analysis is that we should think of observed data functions as single entities, rather than merely a sequence of individual observations (Ramsay and Silverman 1997, pp37).

A record of a functional observation $x$ consists of $n$ pairs $(t_j, y_j)$, where $y_j$ is a observation of $x(t_j)$, a snapshot of the argument value $t_j$ (often time). An essential preliminary to a functional data analysis is often the registration or alignment of salient curve features by suitable monotone transformations $h_i$ of the argument $t$, so that actual analyses are carried out on the values $x_i(h_i(t))$ (Ramsay and Li 1998). The motivation behind this curve registration is that the rigid metric of physical time might not be directly relevant to the internal dynamics of the real-life system (Ramsay and Silverman 1997).

> The Zurich team (Gasser et al. 1990) devised a process for averaging tempo between individuals over the whole curve, shifting individual curves on age continuously and non-linearly to an average development age scale. The result, called a 'structural average', or later, 'registration' curve is probably the best representation of the average or typical human growth curve that we have at present (Tanner 1999, pp9-10).

The aim of the 'structural average' approach devised by Gasser et al. (1990), was to arrive at an average curve that gives a parsimonious description of the data which ideally is able to show the typical pattern of growth, irrespective of individual peculiarities. This would then facilitate the comparison of groups, for example boys and girls.

A set of characteristic times or ages in development are defined which can be found in most individual curves, these are ages were maxima and minima occur (either in the curve itself or one of its derivatives), start and end point, and points of proportions of decrease or increase (Gasser et al. 1990; Ramsay and Silverman 1997).

> For growth, the distance curve reflects the dynamics less well than velocity and acceleration curves, and the maxima and minima of the latter are thus of prime importance (Gasser et al. 1990, pp463).

Kernel estimation was used by Gasser et al. (1990) to compute characteristic time points. Using a priori knowledge of height growth in childhood, i.e. the existence of a pubertal spurt (PS) and the mid-growth spurt (MS), the following characteristic ages were chosen:

- T0: age where infant velocity has dropped by 50% from its value at 1 year to a childhood baseline
- T1: age of minimal velocity prior to MS
- T2: age of maximal acceleration during the MS
- T3: age of maximal velocity during the MS
- T4: age of maximal deceleration at the end of MS
- T6: age of minimal velocity before the PS
- T7: age of maximal acceleration during the PS
- T8: age of maximal velocity during the PS
- T9: age of maximal deceleration during PS

The defining of characteristic points can be time consuming, because of presence of individual non-systematic fluctuations that mimic a spurt - so there is a need to differentiate between systematic and non-systematic fluctuations. Gasser et al. (1993) chose to use the characteristic points of height for skin-fold data because it was difficult to discern a clear pubertal and mid-growth spurt. Weight and skin-fold data are right skewed, so heavier children could dominate the pattern when forming a structural average. Thus skin-fold data was transformed before defining characteristic points and the 'structural median' curve was taken. Gasser et al. (1993) found that weight$^{(1/3)}$ pattern is similar to accumulation of adult height.

Once the curves are aligned on a transformed time-scale they look more similar since one type of inter-individual variability has been eliminated. The average of the transformed curves, i.e. the structural average corresponds to the pattern seen in individual curves. By construction, characteristic times should occur in the average curve at the average age with the average intensity.

> The definition of an average sample growth curve starts from the premises that the pattern of growth is fundamentally the same in all children, whereas the growth process proceeds with a different

> timing (or dynamic) and a different intensity from child to child (Gasser et al. 1989, pp25).

However it should be noted that this assumption may not be true for children with growth problems, girls with Turners syndrome grow at a slightly slower than average rate throughout childhood and fail to experience a pubertal growth spurt (Lyon et al. 1985).

Gasser et al. (1990) illustrated their methodology on height and weights of a subset of 112 girls and 120 boys from 413 Swiss children, enrolled in 1954 as part of the co-ordinated longitudinal studies of the Centre International se l'Enfance in Paris (Falkner 1960). These children were measured at 4, 13, 26, 39 weeks, then at 1, 1.5 and 2 years, then annually up to 9 years in girls and 10 years in boys, then every 6 months until the annual increment in height was less than 0.5cm in 2 years. Children excluded from this subset had failed to attend on more than two occasions or two successive occasions. 'Structural' average curves were also derived for other anthropometric measures; trunk, leg and arm lengths (Gasser et al. 1991); shoulder, hip and bicondylar widths (Gasser et al. 1991); circumferences and body mass indices (Gasser et al. 1994; Gasser et al. 1994). 'Structural averaging' was used to compare the registration curves for acceleration of the Fels, Zurich and Berkley growth study data (Ramsay et al. 1995), and to compare children with high BMI's, to children with low BMI's by taking a sample of 40 in each tail - by doing this the authors were primarily interested in dynamics of obesity (Gasser et al. 1994).

The non-parametric approach used in structural averaging is heavily reliant on having large quantities of data for each individual, however this is not really expanded on in recent literature. Ramsay and Silverman (1997) suggest that the sampling rate or resolution of the raw data is a key factor in determining the potential of functional data analysis. They also suggest that the most important aspect is essentially the local properties of the data relative to the amount of curvature in the data (note that the curvature depends on the amount of error). In the development of 'structural averages' it was suggested that both the number of subjects and number of time points should not fall below 10 to 12 (Gasser et al. 1989). Gasser et al. (1989) suggests that the 'structural averaging' approach copes reasonably well with missing observations. Recent research by James et al. (2000) has potential for curve registration where individuals are measured at irregular times and data are sparse.

## 2.5 Approaches used to address the short term variation in growth data

Initially hotly debated, it is now generally accepted that growth does exhibit short term variation. Historically growth was thought to be a continuous and regular process; only in recent years the consensus has changed. As the measurement interval in time decreases the pattern of height increments becomes more irregular (Hermanussen et al. 1998). Wales (1998) suggested short term growth was non-linear and unpredictable.

Butler et al. (1990) suggested that height velocity curves had a cyclical nature. Height data from the Edinburgh longitudinal growth study, where children were measured at six-monthly intervals, suggested that in addition to the mid-growth spurt there was evidence of a pre-school spurt and late childhood spurt (Butler et al. 1989).

Greco et al. (1990) observed weight gain patterns of a pulsatile nature in very low birth weight infants that were weighed on a daily basis. Gladstone et al. (1998) found that weight, length and head-circumference measures had an oscillatory nature for very low birth weight infants. Lampl et al. (1992) considered serial length measures of a small sample of infants that were measured at weekly, semi-weekly and daily intervals. They found that growth in length was discontinuous and proposed the saltation-stasis model for growth; namely:

> Individual serial growth data were modelled as a series of putative, distinct, stepwise (saltatory) increases or jumps separated by variable intervals of no change (Lampl et al. 1992, pp801).

They hypothesised that human length growth during the first two years occurs in intervals of less than 24 hours and were of an episodic nature. Similar patterns were observed for height measures of one adolescent boy measured daily (Lampl and Johnson 1993). However, Hermanussen (1995) suggested that there was no evidence of saltation, but there were periods of stasis within their data. Lampl and Johnson (1997) suggest that, in infancy, if the time between measurements exceeded 2 to 5 days then the chance of identifying saltatory growth patterns decreases. Tillmann et al. (1998) supported a similar model to Lampl et al. (1992), but suggested that height and weight velocities in a 1 year interval had a bi-phasic nature which were superimposed on strong seasonal trends (Thalange et al. 1996). Giani et al. (1996) suggested using a system dynamics approach to model the pulsatile nature of weight

gain in the first year of life.

## 2.6 The use of parental heights

### 2.6.1 The role of parental heights in a child's growth assessment

If a child is referred to a paediatrician for a suspected growth disorder, then at the time of assessment the heights of both parents will be measured. These are taken in order to assess whether the child is small/tall but appropriate for parental heights, if this is the case then the child will be said to have normal short/tall stature. As pointed out in community guidelines (Schilg and Hulse 1997), wherever possible, parents heights should be measured and recorded. It is also important to ensure that the heights recorded are those of the child's biological parents; i.e. make sure that the child is not adopted or fostered. These days there are also more single parent families. An additional consideration is that at the time of assessment one of the parents may be absent; thus the clinician may have to rely on the reported height of the absent parent.

In short, parental heights may provide a useful indication of a child's growth potential. This is assuming that the parents themselves do not have a growth disorder (Schilg and Hulse 1997) or failed to achieve their genetic potential in height because of economic problems (Tanner et al. 1970). In general there are two approaches to the use of parental height data. Firstly, the parents heights may be used to obtain the mid-parental height along with a target centile range for their child; see section 2.6.2. Alternatively, the parent's heights can be combined with current anthropometric data to predict final adult height; see section 2.6.3.

### 2.6.2 Mid-parental height

Galton (1886) coined the term mid-parental height; this is defined to be the mean of the heights of a child's mother and father. The child's target height is the mid-parental height adjusted for the child's sex, see equation (2.19). Tanner et al. (1970) used mid-parental height in their creation of a height growth standard for children, aged 2 to 9 years, allowing for their parents' heights. Tanner et al. (1970) suggests adjusting the parental heights if the parents are over 45 years to account for reduction in height at this age.

$$\text{target height} = \frac{\text{father's height} + \text{mother's height}}{2} \pm 6.5 \qquad (2.19)$$

where the convention is to add 6.5 cm if calculating for a boy and subtract 6.5 cm if calculating for a girl. Tanner (1986b) chose to add 6.5 centimetres to mid-parental height as this is half the difference observed in adult height between males and females.[3]

The target height can be converted to a Z-score ($Z_{mph}$) using the revised UK 1990 growth reference or plotted on a child's centile chart at 18 years. The target centile range for the child's adult height is then within ± 8.5 cm of target height (Tanner et al. 1970), although current charts suggest ± 10 cm e.g. Child Growth Foundation (1996a), Child Growth Foundation (1996c). Thus if a child's height falls within their target centile range then that child's height is appropriate for parents heights. However, Wright and Cheetham (1999) found that if the parents were unusually short or tall then target height[4] was a poor predictor of attained height. This is because short parents on average have less short children, the regression to the mean effect as first described by Galton (1886). Therefore use of mid-parental height could be misleading when used to assess short stature (Wright and Cheetham 1999).

Luo et al. (1998) used adult heights of 3560 Swedish children and their reported parental heights to derive two alternative equations to equation (2.19):

1. Boys target height $= 45.99 + 0.78\frac{\text{father's height+mother's height}}{2}$
2. Girls target height $= 37.85 + 0.75\frac{\text{father's height+mother's height}}{2}$

Luo et al. (1998) viewed their function to be better than equation (2.19) as the approach of Tanner et al. (1970) underestimates target height for children with parents of short stature.

Mid-parental height does not treat the two parents equally in centile terms; Cole (2000a) suggested increasing the mothers height by a factor of about 1.08 before averaging. A valid alternative to calculating target height Z-score using equation (2.19) is to average the two height Z-scores of the parents (Cole 2000a); see equation

[3] Cole (2000a) presented the difference in adult height from 17 national height references and found the mean to be 13.2cm.

[4] Wright and Cheetham (1999) used the gender correction of 7 cm with mid-parental height, this target height was then converted to a Z-score using the revised UK 1990 reference.

(2.20). The benefit in using this approach is that it attaches a greater weighting to the mothers height and the same formula can be used for both genders (Cole 2000a).

$$Z_{mph} = \frac{Z_m + Z_f}{2} \tag{2.20}$$

where $Z_{mph}$ is Z-score for mid-parental height, $Z_m$ is Z-score for mothers height, $Z_f$ is Z-score for fathers height.

The use of mid-parental height does not take into account that maternal height has a smaller variance than paternal height (Cole 1996). A better approach is to use equation (2.21) which takes into account the correlation between parental heights (Cole 1996). The correlation between parental heights is about 0.3 due to assortative mating (Cole 1996)[5], this approach is adopted in chapter 6 when considering reported parental height data.

mid-parental height Z-score

$$Z_{mph} = \frac{Z_m + Z_f}{\sqrt{2(1 + r(m, f))}} \tag{2.21}$$

where $Z_{mph}$, $Z_m$ and $Z_f$ are as defined above and $r(m, f)$ is the correlation between parental heights.

A conditional gain Z-score, equation (3.11) defined in section 3.4.2 of chapter 3, can be used to compare a child's height with their mid-parental height. Here $Z_2$ is the child's height Z-score ($Z_{ch}$) and $Z_1$ is the mid-parental height Z-score obtained using equation (2.21) and $r$ is the correlation between the child's height ($ch$) and the mid-parental height ($mph$). Alternatively the conditional gain Z-score could be used with just the Z-scores of the mothers, fathers or siblings height (aged 2-9 years) instead of mid-parental height Z-score.

The correlation ($r$) in equation (3.11) depends on the child's age. Tanner et al. (1970) used a correlation of 0.53 for boys and 0.49 for girls between the ages of 2 and 9. However Himes and Roche (1981) (in Cole (2000b)) found the values for correlation between height and mid-parental height in four studies varied between 0.4 and 0.55. Cole (2000b) has recently developed a chart to identify non-familial short stature. A value of 0.4 was used for the correlation because this provided a compromise between the correlations of child's height with mid-parental height

[5] Assortative mating: Designating mating which is not random, but correlated with the possession by the partners of certain similar (or dissimilar) characteristics (Oxford Talking Dictionary (CD-ROM), The Learning Company).

($r$=0.4-0.55); with one parents height ($r$=0.3) and with sibling (aged 2-9 years) of same age ($r$=0.45-0.52 Byrad et al. (1983))(Cole 2000b).

Cole (2000b)(appendix 1) shows that on theoretical grounds, the height Z-score correlation with either parent alone should be 0.8 times the mid-parental height correlation. This is outlined below.

$$r(mph, ch) = \frac{Cov(Z_{mph}, Z_{ch})}{\sqrt{Var(Z_{mph})Var(Z_{ch})}} \tag{2.22}$$

As $E(Z_m) = E(Z_f) = E(Z_{ch}) = 0$ and $Var(Z_{mph}) = Var(Z_{ch}) = 1$ then:-

$$r(mph, ch) = E\left[\left(\frac{Z_m + Z_f}{1.6}\right) Z_{ch}\right] = \frac{1}{1.6}[r(m, ch) + r(f, ch)] \tag{2.23}$$

Assuming that the correlation $r$ between child height and parental height is the same for both parents, then:

$$r(m, ch) = r(f, ch) = 0.8r(mph, ch) \tag{2.24}$$

### 2.6.3 Some examples of applications that use mid-parental height

There are three main approaches to the prediction of adult height: Bayley-Pineau (Bayley and Pinneau 1952), Roche-Wainer-Thissen (RWT) (Roche et al. 1975; Khamis and Guo 1993) and Tanner-Whitehouse (Tanner et al. 1975; Tanner et al. 1983) (all in Onat (1995)). Only the latter two incorporate parental height data, the original TW method allowed for mid-parental height but this was dropped in the revision (Onat 1995). Mid-parental height along with skeletal age, chronological age, weight and height are used in the revision of the RWT method. Wainer et al. (1978) illustrated that there was only a small increase in error of the prediction of adult height if the population mean values were substituted in for skeletal age or paternal height, if these were unknown.

Koziel (1997) fitted the Preece-Baines model 1 (Preece and Baines 1978) to the height data of 183 boys from the Wroclaw growth study. Koziel (1997) found that mid-parental height influenced the centile position: in particular, the height at take-off, height at peak velocity and adult height.

## 2.7 Discussion

In chapter 4 we introduce the data frame studied within this thesis, in which (i) we have routine infancy weight data from 3415 infants within a Newcastle birth cohort with a longitudinal element and (ii) in childhood there is a partial follow-up of this sample at school entry and 7-9 years. In section 2.1 we discussed approaches to the creation of cross-sectional references. As highlighted here, the 'quality' of the data is of utmost importance in the creation of a growth references. The Newcastle data is of a routine nature and as already pointed out in chapter 1 this makes the weight data unsuitable for creating a local reference. Therefore many of the methods discussed in section 2.1 are unsuitable. However, an approach that incorporates the correlation similar to Wade and Ades (1998) may be plausible and the Bayesian approach of Fatti et al. (1998) may be viable but these are not pursued here. There is also an issue of bias: children that are a cause for concern are likely to be monitored more frequently and to have lower weights.

The LMS method was discussed in more detail because the current UK 1990 reference was derived using the approach outlined in section 2.1.1. The revised UK 1990 reference is used within this thesis to convert anthropometric data to Z-scores. In chapters 4, 6 and 8 we assess the adequacy of the UK 1990 reference for monitoring attained growth in Newcastle children. There are numerous advantages to using a growth reference: providing the reference is adequate and adjusts for age, there is no real concern about the slightly different timings of measurement ages between individuals; a child's one off weight or height measure can be compared to the children that contribute to the growth reference and a conditional gain Z-score can be derived from weight Z-scores and this can be used to assess weight gain (see chapters 3, 5 and 7).

It would be possible to carry out a comparative study of the infancy models applied to routine weight data. This would be of value because (i) no study has compared all the models discussed in section 2.2.5 and (ii) comparative studies have always concentrated on a highly selective sample from a large study. However there is a danger, in doing this, that we might be trying to model abnormal growth which the Reed models can accommodate (Berkey and Reed 1987). This will not be pursued here but is a possibility for future research.

Childhood models have only been applied to height data. Whether they can be utilised on childhood weight data is an issue still to be addressed. However the data beyond infancy for the Newcastle data set is sparse. Thus, this issue can not be explored here.

The use of curve registration on infancy data is an area that has not been explored. This is beyond the scope of this thesis, but is the next natural approach. It would involve identifying characteristic points in infancy. Although we have a rich source of data; it still falls short in terms of using a non-parametric curve registration approach that is utilised in the package to do this. Sparse data can create numerous problems when a non-parametric approach is applied. To conclude the methods of Ramsay and Silverman (1997) and Gasser et al. (1990) are not directly applicable to the Newcastle data. However, the approach proposed by James et al. (2000) applied to the infancy weight data is a potential area of further research.

# Chapter 3

# Detection of growth faltering

In this chapter the focus is on growth faltering and the mathematical approaches to the problem. However much of the methodology discussed in this chapter is also applicable to excessive weight or height gain.

In infancy the focus is on monitoring weight gain, primarily for detecting infants at risk of failure-to-thrive. In childhood, this switches to height gain e.g. response of child that is growth hormone deficient to treatment with growth hormone. Weight monitoring is routinely carried out in infancy but this is in part due to the difficulty in obtaining accurate length measures in infancy, although Doull et al. (1995) do not support this. However it is generally accepted that weight is often the first thing affected if a child receives some nutritional or environmental insult during the growth process.

> ... the rate of growth during infancy, especially during early infancy, is rapid, and abnormalities in growth rate may often be detected in just a few months. There is little question that during infancy a decrease in growth rate is the earliest indication of nutritional failure. ... Acceptable data on increments in weight are more readily obtained than acceptable increments in length, and in nearly all circumstances in which nutritional deprivation influences growth, low gain in weight can be demonstrated before low gain in length (Fomon 1991, pp415).

In section 3.1 we concentrate on growth faltering in infancy, termed failure-to-thrive. We discuss the lack of consensus, in terms of a definition, with regard to identifying infants at risk of failure-to-thrive and briefly mention the various weight criterion used. Towards the end of this section we focus on the approach used in



Newcastle, termed the 'thrive index' methodology. This approach was developed on the infancy data considered in this thesis.

The cross-sectional charts, introduced in Chapter 1, provide no guidance in a longitudinal context. If we note the current weight or height of a child, but want to say something about a child's growth since the last weight and height measure: then we need to use a velocity/increment reference[1] or take a conditional approach to the problem, see sections 3.2 and 3.3, respectively. The conditional approach to growth monitoring will be one of the main focuses within this thesis. In section 3.4 we consider chart based approaches to assessing weight gain. We conclude this chapter by discussing the package LGROW that employs longitudinal growth norms.

## 3.1 Detection of growth faltering in infancy and failure-to-thrive

### 3.1.1 Failure to thrive

Failure to thrive and its definition is subject to ambiguity in the medical literature:

> The term 'failure to thrive' (FTT) has been used for more than 50 years to describe the infant or young child whose growth falls substantially behind that of his or her peers. Currently the predominant use of FTT is to describe a child's growth failure resulting from caloric or maternal deprivation or both. Despite widespread use, the term FTT lacks a clarified definition (Wilcox et al. 1989, pp391).

Although this statement was made over ten years ago, the same is true today; currently there is no accepted 'gold standard' method for identifying infants at risk of failing to thrive.

One thing authors do tend to agree on is some kind of criterion involving weight for defining the presence of FTT (Editorial 1990). All the various methods used involve children who fail to gain weight adequately and therefore do not achieve a

[1] A velocity and increment reference are essentially the same, the former term tends to be used in the creation of yearly velocity charts (units centimetres per year) whereas the latter term tends to be used for periods of less than a year (sometimes has same units as velocity but can be centimetres added in time period).

normal or expected rate of growth, according to whatever criteria is employed. The definition of FTT can sometimes also involve a slowing or disruption in acquiring emotional and social developmental milestones (Woolston 1991). Secondary characteristics are deceleration of linear growth and head circumference growth (Woolston 1991).

FTT is often (but not always) a consequence of insufficient calories, sometimes given the term undernutrition.

> Many studies have demonstrated the primal role of energy undernutrition in FTT (Pollitt and Leibel 1980; Berkowitz 1985; Skuse 1985) and that FTT infants will grow better if their intake is increased (Bithoney et al. 1989) (Wright and Talbot 1996, pp226).

There will be a variety of factors contributing to the undernutrition (e.g. illness, disease, weaning difficulties, late weaning, family disturbances, poor parenting, parent-child interactions (Boddy and Skuse 1994), socio-economic deprivation, child abuse and neglect (Taitz and King 1988)) and these are likely to interact (e.g. malnutrition leads to susceptibility to illness, which in turn may affect the child's ability to eat (Woolston 1991)). Wright and Birks (2000) suggest that undemanding behaviour, poor appetite and poor feeding skills may contribute to the onset and persistence of FTT.

It is generally accepted that FTT in infancy is associated with an increased risk of lasting deficits in growth (Kristiansson and Failstrom 1987; Dowdney et al. 1987; Dowdney et al. 1998; Drewett et al. 1999; Wright et al. 1998; Boddy et al. 2000). Kristiansson and Failstrom (1987) found that attained weight at 4 years was more effected than height for children with non-organic failure-to-thrive, but partial catch-up was possible for infants with low social scores that were subjected to intervention. Poor weight gain in infancy has also found to be associated with an increased risk of cardiovascular disease in adult life (Barker 1999; Berglund and Rabo 1973), developmental delay (Heptinstall et al. 1987) and problems with emotional and social adjustment (Hufton and Oates (1977) and Oates et al. (1984) in Dowdney et al. (1987)). Some population based studies have found that FTT in infancy was associated with poor cognitive development and reductions in Intelligence Quotient (IQ) (Dowdney et al. 1987; Skuse et al. 1994; Dowdney et al. 1998). Corbett (1994) in a study of cases selected by the criterion of Edwards et al. (1990)(defined in section 3.1.2) in two deprived areas of Newcastle found that 'thrive index' (defined in section 3.1.3) was associated with IQ. However, the Newcastle growth and development study (the data considered within this thesis) found that there was

no significant difference in IQ and reading abilities for 107 cases matched with 117 controls (Drewett et al. 1999). A similar conclusion was reached by Boddy et al. (2000). At the age of 6 years, they considered 42 cases matched with 42 controls identified from a year cohort resident in an inner-city area of South London in 1986. At the age of 15 months the case group showed significant deficits in cognitive ability but by the age of 6 years there was little evidence of cognitive disadvantage (Boddy et al. 2000). It is suggested that adverse effects of growth faltering in infancy on cognitive function may diminish over time (Boddy et al. 2000; Drewett et al. 1999). The research seems to suggest conflicting views on cognitive outcome, but the case populations are not directly comparable: the case children of Dowdney et al. (1987) had all experienced a long period of growth delay and there were strict exclusion criteria. Therefore, the major source of differences observed in childhood as a consequence of FTT in infancy is the wide range of selection criteria used to identify cases (see next section). A further issue, in early studies on FTT, is referral bias (Wright et al. 2000).

In addition to the problems of defining FTT and arriving at an appropriate weight gain criterion (see next section), there is much debate about the prevalence; differentiating between FTT associated with physical illness (organic failure to thrive) and FTT not associated with such illness (non organic failure to thrive - NOFTT); the contribution of emotional deprivation and that of malnutrition (i.e. food versus love debate).

In clinical practice, decisions are made on the basis of the whole clinical picture rather than from the weight chart alone. Currently, a child presenting themselves below the third centile or exhibiting slow weight gain might be diagnosed as FTT. In addition, the clinician would also take into consideration other factors such as: whether the child's birth weight was below the third centile, the height and build of the parents (although parental short stature should be interpreted with caution, parents may themselves have a growth disorder and/or suffered deprivation in their own childhood) and possibly a medical.

Other plausible reasons for growth faltering are (Maggioni and Lifshitz 1995):

- Familial short stature

  Most babies who gain weight slowly or whose weight gradually crosses centile lines downwards are simply adopting their own genetically determined growth trajectory. There is no reason to assume that every baby should continue on the same centile from birth onwards. About 50% of babies cross

> at least one channel on the weight chart between 6 weeks of age and 12 to 18 months. Up to 5 per cent fall across two channels. Babies who are large at birth are more likely to show falls of this magnitude (Hall 1996, pp112).

- constitutional growth delay This term typically refers to infants that are late maturers.

- intrauterine growth retardation(IUGR) This refers to infants whose growth was compromised in the uterus because of environmental, maternal, placental or fetal factors (Maggioni and Lifshitz 1995). They then fail to catch-up.

- breast-fed baby The Cambridge infant study, concluded breast-fed and artificially fed infants exhibited different growth patterns (Cole et al. 1989). The DARLING study found that breast-fed infants gained weight more slowly than formula-fed infants from similar socio-economic and ethnic backgrounds during the first 9 months of life (Dewey et al. (1992) in Peerson et al. (1993)).

There seems to be three important considerations when considering long term consequences, regardless of criterion used for detection (Skuse et al. 1994):

1. Age of onset Cole (1997, personal communication) felt it would be useful to show quantitatively the timing of detection of FTT, which he suspected was in the time region of 1 to 4 months. Retrospective study of the growth patterns in FTT children referred to the Parkin Service in Newcastle, revealed that the onset of growth faltering is usually within weeks of birth, and that 50% of the cases seen met their screening criteria (see section 3.1.3) by 6 months (Wright and Talbot 1996). Skuse et al. (1994) found that the first few months of life represented a 'sensitive period' for the relationship between growth and mental development.

2. Duration of episode of FTT The quantification of this will depend on criteria used to identify infants at risk. An additional issue is whether this is timed from the point where the infants growth falters or the time when the child's weight (gain) Z-score falls below the screening threshold. Wright (1996) used the latter to quantify duration.

3. Severity The severity is to some extent reflected in the age of onset of FTT and duration of FTT. However, degree of malnutrition is also an important factor (Woolston 1991). Wright (1996) used the lowest thrive index of a FTT infant as an indicator of severity.

Shrimpton et al. (2001) considered the timing of growth faltering in developing countries. Mean weights start to falter at about 3 months and length falters from birth in developing countries (Shrimpton et al. 2001).

### 3.1.2 Various weight criterion used for detecting FTT

In the medical research literature there is no consistent definition of the disturbance in weight gain indicative of FTT, so the diagnosis is at present arbitrary (Smith and Berenberg (1970) in Skuse (1985)). Most studies use different criteria for selecting infants with FTT in their study. Hufton and Oates (1977) study identified infants whose weight lay below the 10th centile, whereas Hannaway (1970) and Shaheen et al. (1968) used the 3rd centile (Dowdney et al. 1987). Boddy et al. (2000) used weight below the 3rd centile for at least 3 months. Less explicit criteria involve a rapid fall through centiles (Berwick et al. 1982) or an acceleration in weight gain after hospitalisation (Rosenn et al. (1980), Ayoub et al. (1979) and Ellerstein and Ostrov (1985) in Dowdney et al. (1987)). A rate of weight gain below -2 SD (Fomon 1974; Kristiansson et al. 1981) is also sometimes used. Some criteria also involve weight-for-height and height-for-age either alone or in addition to weight-for-age (Wilcox et al. 1989).

Dowdney et al. (1987) identified children that had failed-to-thrive in infancy retrospectively at the age of 4 years. The cases identified were white, full term singletons, whose weight and height lay below the tenth centile of the Tanner-Whitehouse standards (Tanner and Whitehouse 1976) at 4 years of age (their last clinic weight was also required to be below the tenth centile), with no medical disorder and height allowing for parental height (Tanner et al. 1970) was also below the tenth centile (Dowdney et al. 1987).

> Retrospective analysis of clinic weight data revealed that cases could have been discriminated on the basis of declining trajectory of weight gain in the first 12 months of life (Dowdney et al. 1987, pp538).

The strict criterion for inclusion meant that these children had experienced an ex-

tended period of growth retardation that had persisted beyond infancy.

Another alternative criterion was:

> 'A child whose weight deviates downwards across two or more major centiles from the maximum centile achieved at 4 to 8 weeks for a period of a month or more' (Edwards et al. 1990, pp1264).

where major centiles are 3rd, 10th, 25th, 50th, 75th, 90th and 97th.
Using this criterion, Edwards et al. (1990) identified 63 infants with poor weight gain out of 306 children attending child health clinics in two deprived areas of Newcastle. Corbett (1994) followed up these cases again at 6-7 years. However, Corbett et al. (1996) found that this screening criteria identified children with mainly borderline FTT, thus leading to the conclusion that definition supplied by Edwards et al. (1990) was over-inclusive. The 'thrive index' methodology discussed in the next section was motivated by the approach of Edwards et al. (1990).

### 3.1.3 The 'thrive index' methodology: the approach used in Newcastle

Wright and Talbot (1996) outline the process by which the Parkin Service identifies and manages children in the Newcastle area with different levels of severity of FTT. A child should have shown a substantial fall down the centile chart, since a baseline weight in the first 6 weeks of life, before the Parkin service are involved (Wright and Talbot 1996). The 'thrive index' approach is discussed below, this was developed from the approach used by Heimendinger and Laird (1983) to measure the effect of intervention on attained length.

- Choice of baseline weight Estimating an infant's expected growth pattern presents difficulties (Whitehead et al. 1989a). The birth weight centile may be used for this purpose, but it is determined largely by maternal factors (Edwards et al. 1990). Furthermore, the use of the birth weight centile has been discredited by Edwards et al. (1990); who found the maximum weight centile attained between age of 4 and 8 weeks to be superior in the prediction of weight centile at 12 months. Wright et al. (1998) used an average of weight Z-scores between birth and two months as a baseline weight Z-score, whereas weight Z-score at 6 weeks was used in development of the 'thrive index' (Wright et al. 1994).

- Diagnostic criteria Falling below a predetermined centile on a weight-for-age chart, such as the 2nd or 0.4th on the UK 1990 reference, is usually taken to be indicative of abnormality. Using the second centile may identify children that are constitutionally small (i.e. if the child has short parents) but will fail to identify children that have had a substantial fall from a high centile (Wright et al. 1994). Therefore, a measure of weight gain is needed rather than an indicator of atypical weight. The 'thrive index' methodology was developed by Wright et al. (1994) to address this issue.

There is a further issue, namely 'regression to the mean'. This phenomenon was first described by Galton (1886); he noted that there was a tendency for tall fathers to have less tall sons. It is essentially a statistical phenomenon, an individual measured once and then on a later occasion is more likely, on average, to be nearer then median on the second occasion than on the first (Cole 1995). Regression to the mean always occurs unless there is perfect correlation (Bland and Altman 1994a). Furthermore, if the two measures are weakly correlated then regression to the mean will have greater impact (Bland and Altman 1994b). Therefore the amount of regression to the mean depends critically on the magnitude of the correlation between these two measures (Cole 1995). In order to allow for regression to the mean we need to quantify the correlation between the two measures (Cole 1995).

In infancy, weight exhibits regression towards the mean (Cole 1994a). Thus a very light infant is more likely to exhibit upward centile crossing than an average or large infant (Cole 1994a). The reverse is true for a very heavy infant. In the literature the former is often termed 'catch-up' growth, whereas the latter is termed 'catch-down' (Cole 1994a). Thus a clinician may interpret this centile crossing as a cause for concern, when often it is not (Cole 1994a). Furthermore, not every light infant will 'catch-up'; just the majority will (Cole 1995).

If we consider two weight measurements with Z-scores $Z_1$ and $Z_2$ at $t_1$ and $t_2$, respectively. Then, if $Z_2$ is regressed on $Z_1$:

$$Z_2 = a + bZ_1 + \epsilon \tag{3.1}$$

where $a$,$b$ and $\epsilon$ are the intercept, gradient and error.

Routine infancy weight data from an annual cohort of 3418 full-term children (the data considered within this thesis - see Chapter 4) were converted to weight Z-scores using the Cambridge reference (Whitehead et al. 1989b). The Cambridge reference was used because at the time of publication the UK 1990 reference had

yet to be created. Furthermore, birth weights were not converted to weight Z-scores as the Cambridge reference did not cover this period. Wright et al. (1994) used equation (3.1) to regress last available weight Z-score at 9-24 months on weight Z-score at 6 weeks to give the following equation:

$$Z_{\text{predict}} = 0.225 + 0.62Z_{6\text{ weeks}}$$

which had a residual SD of 0.90. The 'thrive index' was then defined by:

$$TI = Z_{\text{actual}} - Z_{\text{predict}} = Z_{\text{actual}} - 0.62Z_{6\text{ weeks}} - 0.225$$

and a cut off of -1.48 ($0.9 \times -1.6449$) would identify the slowest growing 5% (Wright et al. 1994). Wright et al. (1994) chose to define the TI in this way to ensure that the TI would have zero mean and standard deviation equal to the residual standard deviation when fitting equation (3.1).

The thrive index methodology has subsequently been applied to other infants (Corbett et al. 1996; Wright et al. 1998), using the following equation for the TI:

$$TI = Z_{\text{late}} - 0.65Z_{\text{early}} \tag{3.2}$$

where $Z_{\text{early}}$ is the initial weight Z-score at an age <10 weeks and $Z_{\text{late}}$ is the later weight Z-score before the age of 2. Here the weight Z-scores are derived from the revised UK 1990 reference. Corbett et al. (1996) point out that the late weight Z-score could be recorded at any age between 6 and 18 months because the coefficient ($b$) in equation (3.1) is close to 0.65 in this age range. To identify the slowest growing 5%, using equation (3.2), a cut off of -1.26 was used (Corbett et al. 1996). Corbett et al. (1996) used the following cut-offs as indicators of severity of FTT: -0.9 for mild FTT, -1.26 for moderate FTT and -1.64 for severe FTT.

Two weight Z-scores below the screening threshold of the TI are required before the Parkin service becomes involved (Wright and Talbot 1996). However, although the TI approach identifies infants that are gaining weight slowly (an 'at risk' group), further assessment is then required (Wright and Talbot 1996). Wright and Talbot (1996) discuss this further assessment stage and suggest that NOFTT is best managed in the community.

Raynor and Rudolf (2000) compared five anthropometric methods of classifying under-nutrition in failure-to-thrive: % median weight-for-age, % median weight-for-height, % median weight/height ratio for age, BMI and thrive-index discussed above (early weight Z-score was birth weight Z-score so $b = 0.4$ in equation (3.1)).

Raynor and Rudolf (2000) found the TI approach identified the highest proportion of severe cases. In agreement with Wright et al. (1994), Raynor and Rudolf (2000) found that classification of under-nutrition varied widely between anthropometric indicators. Raynor and Rudolf (2000) also found no relationship between the five anthropometric indices and indicators of developmental delay or dietary intake. Raynor and Rudolf (2000) concluded that severity of FTT was best addressed by the clinician.

In conclusion, research is needed to establish:

> The definition, identification, and management of slow weight gain, FTT and NOFTT, and the role of conditional reference charts in managing these problems. The relevance of regular weighing to the detection of psychosocial deprivation needs further evaluation (Hall 1996, pp124).

Furthermore, there is a need to refine the definition of the syndrome of failure-to-thrive to identify those children for whom weight faltering in infancy is linked to poor cognitive outcomes (Boddy et al. 2000).

## 3.2 Tracking indices, distance charts and centile crossing

The term 'tracking' is often used in child growth or blood pressure context to refer to maintenance of rank order within a group of peers over time (Roche 1992). If a group of individual growth curves do not intersect then this is viewed to be perfect tracking (Foulkes and Davis 1981). The term tracking is sometimes used to mean the prediction of future values (Ware and Wu 1981). The correlation coefficient is sometimes used as a measure of tracking, but the disadvantage in using this is information on observed heights or weights is lost (Foulkes and Davis 1981). In 1981 several tracking indices were proposed: Foulkes-Davis index (Foulkes and Davis 1981), McMahan index (McMahan 1981) and Goldstein's growth constancy and separation indices (Goldstein 1981b). Ten years later Frongillo et al. (1990) developed a tracking score based on Goldsteins growth constancy index. None of these tracking indices have been applied extensively in the growth literature. Steward (1994) applied the Foulkes-Davis, McMahan, Goldsteins tracking indices plus other approaches to a cohort of pre-school children from two towns in Wales; he found that

all these tracking indices were of little value in identifying unusual growth patterns. Steward (1994) found that it was possible to achieve the necessary discrimination using linear regression of Z-scores.

Growth references are based on cross-sectional data, thus they are only really of use for one off measures. However, in clinical practice, it is common to interpret downward 'centile crossing', for example, attained height crossing down two (or more) major centiles during childhood. Cross-sectional growth references are not designed for this purpose! Recall that the Coventry consensus (see Chapter 1 for more details), concluded that routine growth monitoring to detect centile crossing had too low a sensitivity and specificity to be used for screening purposes (Hall 2000).

Park et al. (1997) and Li et al. (1998) used the term canal to refer to the area between the major centiles (5th, 10th, 25th, 50th, 75th, 90th and 95th) on the NCHS charts (Hamill et al. 1977). The term 'decanalization' was used to describe movement over a specified time interval to a non-neighbouring canal (Park et al. 1997; Li et al. 1998). Anthropometric data from the Fels longitudinal study was used to calculate probabilities of decanalization over: 6 or 12 month intervals in infancy (Park et al. 1997) and 1 or 2 year intervals in childhood (Li et al. 1998). Park et al. (1997) found that the probabilities of decanalization in infancy were larger for weight than for length or head-circumference and that decanalization was associated with birth weight. In childhood and adolescence decanalization was found to be less common than in infancy (Li et al. 1998). Li et al. (1998) found that the probability of decanalization for weights was generally larger than for heights, but movements towards the median were more common than movements away from median for both variables. Li et al. (1998) are describing 'regression to the mean' but no reference is made to this. In childhood transitions to levels outside the normal range were found to be more common than decanalization (Li et al. 1998).

Cole (1997a) worked on a Z-score scale to quantify the notion of centile crossing. If we consider the difference $Z_2 - Z_1$ (where $Z_1$ is the first measurement and $Z_2$ is the second and both are normally distributed with zero mean and variance 1), then:

$$E(Z_2 - Z_1) = E(Z_2) - E(Z_1) = 0 \tag{3.3}$$

$$\mathrm{Var}(Z_2 - Z_1) = \mathrm{Var}(Z_2) + \mathrm{Var}(Z_1) - 2\mathrm{Cov}(Z_1, Z_2) = 2 - 2r = 2(1 - r) \tag{3.4}$$

where $r$ is the correlation between $Z_2$ and $Z_1$. Therefore the change in Z-score (centile change) can be expressed as a Z-score by using:

unconditional gain Z-score

$$\text{Z-score for centile change} = \frac{Z_2 - Z_1}{\sqrt{2(1-r)}} \tag{3.5}$$

Equation (3.5) can be used to compare centiles derived from the UK 1990 growth reference, on the proviso that we know the relevant correlation coefficient. Cole (1997a) presented the age-on-age correlations for annual height measurements over the age range 2 to 9 years. These correlations were derived from 204 to 318 children (sexes combined) belonging to the French longitudinal study (Falkner (1961) and Sempé et al. (1979) in Cole (1994a)) (Cole 1997a). The correlation between height measures at different ages was higher between measurements made close in time and at older ages (Cole 1997a). This approach can also be used to compare a child's weight and height centile at the same age. Cole (1997a) proposed the use of the term 'centile bandwidth' as an alternative 'centile crossing', because the latter term is ambiguous.

> Centile crossing over time is not an effective screening instrument, because the chart provides no information to quantify it. How much centile crossing is acceptable? Does it depend on age? Does it depend on starting centile? (Cole 1998a, pp2698)

## 3.3 Velocity or increment charts

### 3.3.1 Velocity references and tempo-conditional references

The use of a longitudinal growth chart is preferable to the approach described in the last section. In particular, a child's height or weight velocity is thought to be a more sensitive indicator than height or weight distance of pathology (Tanner 1989).

> ... velocity represents what is happening *now*, whereas distance represents the sum of all that has happened in the past (Tanner 1989, pp188).

A velocity reference can be constructed using longitudinal data, but the spacing of the centile curves is influenced by the time between measurements (Cole 1993). Thus a velocity reference has to be derived and used on a set time interval, usually one year, because of seasonal variation and relative importance of measurement

error in comparison to actual growth (Tanner 1989). The effect of measuring error will be more important if an increment is observed to be less than the third centile (Tanner 1986a). We have already discussed how a child stays relatively close to their 'distance' centile after infancy until puberty. However, this is not the case with velocity references (Tanner 1986a). Tanner et al. (1966) suggested that velocity plots should be viewed to have a more episodic nature.

Tanner et al. (1966) derived height (weight) velocity standards from whole year increments of height (weight). These charts could only be used on the basis that the height (weight) measurements were made a year apart. They are used by calculating the change in height (weight) divided by the time elapsed and then plotted at the mean age of the two measurement ages.

> In monitoring individuals velocity or incremental standards are much more effective than 'distance' standards, and tempo-conditional standards are essential over the age of 9 (Tanner 1986a, pp131).

Tanner and Whitehouse (1976) updated these charts to create tempo-conditional references (or clinical longitudinal references) during puberty; a shaded region was added to the velocity charts to accommodate growth monitoring of early and late maturers. This shaded region corresponded to the area two standard deviations above and below the age of peak height (weight) velocity (Tanner and Whitehouse 1976). More recently, Buckler and Tanner (1997) updated the Tanner-Whitehouse clinical longitudinal reference using the amalgamated data set from Freeman et al. (1995) and longitudinal data from 198 Sheffield adolescents (Cameron 2002). Tanner and Davies (1985) also produced similar longitudinal height velocity charts for North American children based on the NCHS reference children using coloured lines for early and late maturers.

Berkey et al. (1993) used a different approach in their creation of longitudinal height velocity standards from 7 to 18 years. These were not created from a nationally representative sample but from children belonging to the Harvard six cities study from 1974-1989 (Berkey et al. 1993). This velocity reference was created by splitting the children into three groups: early, typical and late maturing. Peak height velocity for girls at 9, 11 and 13 years corresponded to the early, typical and late groups. Peak height velocity for boys at 11, 13 and 15 years corresponded to the early, typical and late groups. Centiles were created for each maturity group by first using an approach similar to the HRY method (Healy et al. 1988) (discussed in Chapter2) using four-month windows, then using the derivative of the triple logistic model (Bock and Thissen 1976) (also discussed in Chapter 2) to smooth the monthly

velocities (Berkey et al. 1993). Berkey et al. (1993) presented the 3rd, 50th and 97th centiles for early, typical and late maturing on the same chart. There is some debate in the research literature as to whether it is useful to monitor height velocity in childhood. Brook and Hindmarsh (1991) view the monitoring of height velocity to be a useful indicator of growth hormone secretion and advocate the Middlesex height velocity chart (Brook 1983) for this purpose. Brook (1998) later emphasised the importance of having the same trained personnel for both height measurements when using the Middlesex chart. Voss (1999) holds a conflicting view with regards to height velocity monitoring. Voss et al. (1991) suggested that monitoring height velocity was not useful, because a single height velocity even over 12 months lacks precision due to measuring error. Moreover, Voss et al. (1992) suggested that a single height measurement at school entry was the most sensitive indicator of silent disease (for example growth hormone deficiency, Turners Syndrome). Current guidelines (Hall 2000), support this latter view.

### 3.3.2 Increment charts: use of 'warning' and 'action' limits

Healy et al. (1988) considered longitudinal measurements of supine length of 427 infants (229 boys, 198 girls) from Sudan, taken at 4 week intervals. Four- and eight-weekly increments were obtained for each child and any negative increment more than 2 cm was excluded.[2] Healy et al. (1988) created the centiles empirically because of non-normality and then smoothed the resulting values by using the derivative of the Jenss-Bayley model (Jenss and Bayley 1937) (see chapter 2) to arrive at the increment chart. Healy et al. (1988) stated that this approach worked well for the 8-week interval but some 'manual smoothing' was needed for the 4 week intervals. Healy et al. (1988) suggested using these increment charts by setting an 'action limit' at the fifth centile and a 'warning limit' at a centile closer to the median. Where the latter could be the twenty-fifth centile and if a child falls below the 'warning limit' on two successive occasions[3] then this could be taken to indicate abnormality (Healy et al. 1988). Healy et al. (1988) viewed this approach to be appropriate in a velocity context as the correlation between one measurement and the next is quite low. Healy et al. (1988) found that by using this approach; measurements made over a short time interval were quicker to detect growth deficiency but at a cost;

[2]It is unlikely that a real decrease in length occurs over a short period (Healy et al. 1988). However negative increments are to be expected because of measurement error when the true growth in length is small (Healy et al. 1988)

[3]The chance of a false positive: a normal child with two successive increments below the 25th centile, is approximately 6.25%

with a substantial increase in the false alarm rate.

Zumrawi et al. (1992) used the same approach as Healy et al. (1988) on the weight measurements of the same children from Sudan. However Zumrawi et al. (1992) created two and four week increment charts for weight. Zumrawi et al. (1992) presented the correlation between successive weight increments and initial (birth) weight, and found the expected pattern: namely that early correlations are negative because large babies grow more slowly than small babies as the influence of maternal size dies away. However these correlations were small, indicating that little was lost in judging the growth velocity of a baby if birth weight is ignored (Zumrawi et al. 1992). Zumrawi et al. (1992) reached the same conclusion as Healy et al. (1988), namely that a given growth deficit will be picked up sooner by the shorter measurement interval (i.e. 2 weeks), but at a cost!

### 3.3.3 Increment tables

An alternative to the chart-based approach used in the USA is to publish increment tables (Baumgartner et al. 1986; Guo et al. 1988; Roche et al. 1989; Guo et al. 1991). These tables present age-range, for example 1-2 months, versus the mean and standard deviation of increment in weight (g/day) (recumbent length (cm/day)) along with tabulated 5th, 10th, 25th, 50th, 75th, 90th and 95th centiles for that age-range.

The initial increment charts were published for six month intervals (Roche and Himes 1980); the centiles were empirically derived from the Fels longitudinal study data and smoothed across age using low-term fourier transforms: recumbent length and head circumference between birth and 3 years, stature between 2 and 18 years and weight between birth and 18 years. These were later published as increment tables to allow increments to be compared directly with numerical values (Baumgartner et al. 1986). Later monthly increment tables were published for head circumference (cm/month), weight (kg/month) and length (cm/month) from 1 to 12 months (Guo et al. 1988; Roche et al. 1989). These were also in the form of charts with 5th, 50th and 95th centiles because at later ages the centiles were too close together to give a meaningful representation (Roche et al. 1989). In the creation of these increment tables, the Guo model fitted to each child's longitudinal data, was used to interpolate monthly values for head-circumference, weight and length (Guo et al. 1988; Roche et al. 1989). Roche et al. (1989) suggested that falling below the 5th centile for length gain or weight gain on two consecutive months was more likely to be due to poor nutrition or illness rather than biological variation.

One-monthly increments were published to allow earlier evaluation of growth velocity (Roche et al. 1989). However the timing of these increments was later revised to take into account apparent weight gain due to food intake and waste out and to minimise the effect of measuring error on length increment (Fomon 1991).

The increment tables for length (cm/day) and weight (g/day) covered the age range birth to 24 months (Guo et al. 1991). However the length and weight data from the Fels study was supplemented with data from the Iowa infant study (Guo et al. 1991). This lead to the final table being derived from the Iowa data from birth to 3 months, both between 3 and 6 months, and the Fels data from 6 to 24 months (Guo et al. 1991).

Up to the age of 6 months, increments in weight were presented at monthly intervals because after this age, the day-to-day variation in weight is large in relation to 1-month weight gains (Guo et al. 1991). One month intervals were not published for length increments because measurement error[4] of length is a relatively large proportion of the actual increment (Guo et al. 1991). Instead 2 monthly intervals were presented for length increments from birth to 6 months (Guo et al. 1991). Two monthly intervals were also presented for weight increments from birth to 1 year (Guo et al. 1991). Three monthly intervals for both weight and length increments were presented from birth to 24 months.

Guo et al. (1991) believed these increment tables to be useful in screening for deviations from normal growth, for example, identifying infants that may be at risk of failing-to-thrive or excessive weight gain.

Children from the Fels longitudinal study had their weights and lengths measured at 1, 3, 6, 9, 12, 18 and 24 months (Roche et al. 1989). Piwoz et al. (1992) used the monthly increment tables from 1 to 12 months (Roche et al. 1989) on monthly weight and length gains of 96 Peruvian infants from 2 to 12 months. Piwoz et al. (1992) found that the median weight and length of the Peruvian infants were close to the reference median at 2 months, but the median then dropped progressively lower than the increment reference. However after 2 months of age the variance of monthly weight and length gains of Peruvian infants were greater than the reference (Piwoz et al. 1992). Piwoz et al. (1992) used the same approach as Roche et al. (1989) to create the reference for Peruvian children, only using weight and length measures at 1, 3, 6, 9 and 12 months. The same pattern was observed, namely the variance of the raw gains were greater than the Peruvian reference (Piwoz et al. 1992). Piwoz et al. (1992) concluded that the curve-fitting and interpolation used in

[4]Measurement error of length in well-trained personnel is about 0.4 cm (Fomon 1991)

the creation of the increment reference produced an artificially narrow distribution of weight and length gains. The revised increment tables (Guo et al. 1991) may not fair quite as badly as the monthly reference, because this incorporated data from the Iowa study; these children had routine birth weight and were measured at 8, 14, 28, 42, 56, 84 and 112 days.

Kolsteren et al. (1997) carried out similar comparisons on the mean monthly length and weight velocities of infants from Madura, Indonesia (age-range 1 to 11 months). The focus in Kolsteren et al. (1997) was on the mean pattern of weight or length gain, but the tabulated standard deviations of the Indonesian sample are greater than the US reference. In addition, Kolsteren et al. (1997) also considered the published raw weight and length increments from the Wroclaw study (Borys-lawski 1988); they found that the Wroclaw mean curve had a different shape to the mean curve from the increment reference derived from the Fels study.

## 3.4 Conditional height gains and conditional gain Z-scores

### 3.4.1 Conditional height gains

The main restriction in using the velocity charts described in the last section, is the requirement of a set measurement interval. Another disadvantage of the velocity chart is that two charts are required ('distance' and 'velocity' standard) and the data need to be plotted twice. Healy (1974) first suggested a regression approach as an alternative to a height velocity reference. A reference is constructed for height at age $t + 1$ conditional on height at age $t$, where $t$ is the age in years.

> These should be more sensitive than the difference standards and would avoid ascribing abnormally low velocities to children who happen to be fortuitously above their expected stature at the start of the interval (Healy 1974, pp44).

The principle behind a 'conditional' reference is that a child's measurement is interpreted in the light of other covariates in addition to the usual age and sex. This type of approach has been used for height and height velocity during puberty, where tempo is adjusted for (discussed above) (Tanner and Whitehouse 1976), conditioning height on mid-parental height (discussed in section 2.6.2)(Tanner et al. 1970) and within-family standard for birth weight (Tanner et al. 1972) (Cole 1993).

Regression to the mean (discussed in section 3.1.3) occurs in child growth during infancy and puberty for both height and weight (Cole 1994a); in order to compensate for this a conditional reference is required. In this section we will consider the approach put forward by Healy (1974), namely current height conditional on height one year earlier. The methods discussed in this section are relevant to height monitoring in childhood and would be useful for detecting height faltering. In section 3.4.3 we consider conditional weight gain Z-scores, these will be the focus of this thesis, which are of interest in infancy.

Cameron (1980) was the first to put the suggestion made by Healy (1974) into action. Cameron (1980) published conditional standards for growth in the height of British children from 5 to 15.99 years of age, using data from the London County Council 1966-67 growth survey. Each age-sex group required a separate chart leading to a set of 22 charts (Cameron 1980), thus leading to a reference that was unwieldy to use in practice.

Berkey et al. (1983a) presented conditional standards for length and log weight in pre-school (3 months to 6 years) children of the Harvard longitudinal study in the USA. Berkey et al. (1983a) fitted the Jenss model (Jenss and Bayley 1937) to each individual's lengths and weights within the reference sample. The purpose of this was to interpolate the child's height and weight at yearly intervals after the age of 1 year. However the downside is that a structure is imposed on the child's growth curve. The interpolated yearly weights and lengths were then used to create the standard, with the underlying assumption that the population lengths and the natural logarithm of population weights were multivariate Normal at each age and each combination of ages (Berkey et al. 1983a). Berkey et al. (1983a) presented means, standard deviations and correlations at year intervals from one to six years.

If we consider height now (at time $t$) regressed on height one year earlier (at time $t-1$), then:

$$H_t = b_t H_{t-1} + c_t + \epsilon_t \tag{3.6}$$

where $b_t$ is the regression coefficient, $c_t$ is the intercept, and $\epsilon_t \backsim N(0, \sigma_t^2)$ (Cole 1994a). If equation (3.6) is rearranged then:

$$H_t - b_t H_{t-1} = c_t + \epsilon_t \tag{3.7}$$

The left hand side of equation (3.7) is defined as the 'conditional height velocity' and by definition is uncorrelated with $H_{t-1}$ (Cole 1994a).

The regression coefficient $b_t$ is much less than 1 in infancy and puberty but greater than 1 for most of childhood (Cole 1994a). Height velocity and conditional

height velocity are the same if it happen that $b_t = 1$ (Cole 1994a). The variability of conditional height velocity tends to be less than that of height velocity and is also unbiased (Cole 1994a). This latter point is of practical importance in the assessment of children with short or tall stature (Cole 1994a). Conditional height velocity is a function of attained height, leading to the requirement of different velocity centiles for children of different heights (Cole 1994a).

The conditional height velocity Z-score of a child growing from height $H_{t-1}$ to height $H_t$ is defined to be (Cole 1994a):

$$Z_{H_t|H_{t-1}} = \frac{H_t - E(H_t|H_{t-1})}{SD(H_t|H_{t-1})} = \frac{H_t - b_t H_{t-1} - c_t}{\sigma_t} \tag{3.8}$$

Thus for a child growing along the $100\alpha$th conditional velocity centile, height at age $t$ is given by (Cole 1994a):

$$H_t = b_t H_{t-1} + c_t + \sigma_t z_\alpha \tag{3.9}$$

where $z_\alpha$ is the normal equivalent deviate corresponding to centile $100\alpha$.

Cole (1994a) also extended equation (3.6) to incorporate height two years earlier ($H_{t-2}$); this equation was fitted to height data from the French longitudinal study. Cole (1994a) found that during childhood there was no real advantage in including $H_{t-2}$, but its use was of value either side of the peak height velocity in adolescence.

### 3.4.2 Conditional gain Z-score approach

Suppose that heights (weights) are converted to Z-scores, using for example, the revised UK 1990 reference. Now consider two height (weight) measurements with Z-scores $Z_1$ and $Z_2$, with a correlation between them of $r$. Then, if $Z_2$ is regressed on $Z_1$ we arrive at equation (3.1) given above. On the proviso that the Z-scores have not being derived from poorly matched references equation (3.1) can be simplified further (Wright et al. 1994):

$$Z_2 = rZ_1 + \epsilon \tag{3.10}$$

As $Z_1$ and $Z_2$ are Z-scores, the mean of both $Z_1$ and $Z_2$ is zero; so $a = 0$ by definition. Furthermore, the standard deviations of both $Z_1$ and $Z_2$ are one; so $b = r$ by definition. Therefore the expected value of $Z_2|Z_1$ is $rZ_1$, and the expression $(Z_2 - rZ_1)$ is a measure of change between $Z_2$ and its expected value. Furthermore the standard deviation of $(Z_2 - rZ_1)$ is given by $\sqrt{(1 - r^2)}$, which leads to the

conditional gain Z-score (Cole 1995):

$$Z_{2|1} = \frac{Z_2 - rZ_1}{\sqrt{(1-r^2)}} \tag{3.11}$$

The variable derived from equation (3.11) has a mean of zero and a standard deviation of one, and is a Z-score of $Z_2$ conditional on $Z_1$. Therefore a conditional gain Z-score greater (less) than 0 indicates a gain that is faster (slower) than expected.

Therefore the only information required to use equation (3.11) is a growth reference to convert the weight (height) measurements to Z-scores and the correlation $r$ between the two weight (height) Z-scores (Cole 1993). This correlation will depend on the age and sex of the child. Although Cole (1993) states this may well be available in the literature, this is not the case for weight Z-scores beyond infancy. However, Cole (1995) published correlations for weight Z-scores in infancy and yearly correlations are available for height in childhood (Cole 1997a; Cole 1994a; Cameron 1980; Bailey 1994; Berkey et al. 1983a).

### 3.4.3 Conditional weight gain Z-scores

Weight velocity is usually viewed to be a far more sensitive indicator of growth than weight attained in the detection of weight faltering (Cole 1989a). A conditional weight gain reference is a weight velocity reference that also compensates for regression to the mean (Cole 1995). A conditional reference addresses the following question: 'Knowing the infant's previous weight, what is his/her likely weight now?'(Cole 1995).

Cole (1995) generalised the approach used by Wright et al. (1994) in their development of the 'thrive index' outlined in section 3.1.3. Cole (1995) presented the correlation matrix for the weight Z-scores of individuals seen on all 15 occasions in the Cambridge infant study, see table 3.1. The purpose of the Cambridge infant growth study was to monitor the growth patterns of infants that were being fed in line with Department of Health guidelines (Whitehead et al. 1989b). The Cambridge infants were weighed and measured every 4 weeks from 4 to 52 weeks, and at 18 and 24 months (Cole 1995).

Cole (1995) originally avoided using the term velocity, opting instead to use the term 'gain'. If children are measured at 4 weekly intervals then the correlation can be obtained from table 3.1 and inputted into equation (3.11). If we now use $Z_{t1}$ and $Z_{t2}$ to denote the weight Z-scores at initial and later time and let $r$ be the correlation between these weight Z-scores. It is unlikely that a clinician will

Table 3.1: Cambridge correlation matrix excluding birth weight (N=105, n=223): original correlation matrix for weight Z-scores at 15 ages during early infancy, based on data for 223 infants (114 boys and 109 girls) seen on all 15 occasions

| week | 4 | 8 | 12 | 16 | 20 | 24 | 28 | 32 | 36 |
|---|---|---|---|---|---|---|---|---|---|
| 8 | 0.911 | | | | | | | | |
| 12 | 0.823 | 0.945 | | | | | | | |
| 16 | 0.752 | 0.881 | 0.958 | | | | | | |
| 20 | 0.659 | 0.788 | 0.892 | 0.959 | | | | | |
| 24 | 0.611 | 0.738 | 0.845 | 0.918 | 0.967 | | | | |
| 28 | 0.588 | 0.705 | 0.811 | 0.891 | 0.945 | 0.971 | | | |
| 32 | 0.565 | 0.680 | 0.779 | 0.859 | 0.915 | 0.947 | 0.971 | | |
| 36 | 0.535 | 0.651 | 0.745 | 0.830 | 0.880 | 0.919 | 0.950 | 0.981 | |
| 40 | 0.504 | 0.624 | 0.718 | 0.800 | 0.850 | 0.893 | 0.924 | 0.961 | 0.979 |
| 44 | 0.488 | 0.599 | 0.687 | 0.771 | 0.823 | 0.867 | 0.901 | 0.940 | 0.959 |
| 48 | 0.474 | 0.587 | 0.673 | 0.755 | 0.809 | 0.851 | 0.890 | 0.925 | 0.949 |
| 52 | 0.467 | 0.572 | 0.659 | 0.740 | 0.792 | 0.830 | 0.871 | 0.906 | 0.927 |
| 78 | 0.464 | 0.544 | 0.609 | 0.671 | 0.713 | 0.737 | 0.771 | 0.802 | 0.820 |
| 104 | 0.483 | 0.584 | 0.634 | 0.680 | 0.706 | 0.727 | 0.746 | 0.764 | 0.775 |
| week | 40 | 44 | 48 | 52 | 78 | | | | |
| 44 | 0.975 | | | | | | | | |
| 48 | 0.963 | 0.977 | | | | | | | |
| 52 | 0.946 | 0.965 | 0.973 | | | | | | |
| 78 | 0.849 | 0.877 | 0.890 | 0.910 | | | | | |
| 104 | 0.804 | 0.822 | 0.826 | 0.850 | 0.929 | | | | |

Table 3.2: Cambridge correlations excluding birth weight (N=105, n=223): Regression of Fisher's transformation of correlation coefficients on functions of the time difference (t2-t1) and mean time (t1+t2)/2

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| intercept | 2.98165 | 0.13201 | 22.586 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)$ | 0.58888 | 0.07957 | 7.401 | $4.50 \times 10^{-11}$ |
| $\log(t2-t1)$ | -1.66617 | 0.05805 | -28.701 | $< 2 \times 10^{-16}$ |
| $1/(t2-t1)$ | -2.05740 | 0.16158 | -12.733 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)\log(t2-t1)$ | 0.25100 | 0.01419 | 17.685 | $< 2 \times 10^{-16}$ |
| $(\log((t1+t2)/2))^2$ | -0.10351 | 0.01585 | -6.530 | $2.84 \times 10^{-9}$ |
| $R^2$=0.9938, $R^2(adj)$=0.9934, residual SE=0.03872 on 99 df | | | | |

see a child every 4 weeks. To compensate for this Cole (1995) chose to model the correlation between two weight Z-scores as a function of the mean ($\frac{t1+t2}{2}$) and the

time gap ($t2 - t1$) between the ages. Fisher's transformation was applied to the correlations in table 3.1. Cole (1995) fitted the model given in table 3.2 to the Cambridge infancy correlation matrix for weight Z-scores (table 3.1). Cole (1995) validated the equation in table 3.2 using pairs of weights from the Cambridge study and a subsample of 761 infants from the Newcastle data considered within this thesis. Cole (1998a) actually refitted this same model to the Cambridge correlations but this time including the correlation with birth weight Z-score, see chapter 5 for further details.

Cole (1995) also presented a conditional weight chart for boys and girls; this represented the median pattern of weight gain overlaid on the UK 1990 reference, see chapter 7 for further discussion of this chart.

More recently, Blair et al. (2000) used the conditional weight gain Z-score approach, using the model of Cole (1998a) for the correlation between weight Z-scores, to investigate growth patterns that may influence risk of sudden infant death syndrome (SIDS). Blair et al. (2000) found that the growth of SIDS infants from birth to final weight was significantly poorer than controls. Blair et al. (2000) also found that SIDS infants of normal birth weight had poorer weight gain than those of low birth weight. Blair et al. (2000) suggested that the conditional weight gain Z-score between birth and 6 weeks may provide a useful indication of infants at risk.

## 3.5 Conditional weight charts

### 3.5.1 Sheffield weight chart

The Sheffield chart was the first to combine distance and velocity information on the same chart. The Sheffield chart (Foundation for the study of infant deaths (1985) in Emery et al. (1985)) was devised for monitoring infants at possible risk of cot death[5], because previous studies had indicated that some babies have a 'substandard' weight gain before death (Emery et al. 1985). The Sheffield weight chart consists of a number of equidistant lines which form a series of channels. Over a period of two weeks an infant's attained weight was not expected to move up or down by more than one channel width (Emery et al. 1985). In addition, over a period of eight weeks a child was not expected to move up or down two channel widths (Emery et al. 1985). The frequent monitoring of weight (every 2 weeks) in order to assess

[5]For example, care of next infant (CONI), for the child born subsequent to an infant lost to cot death.

a child's weight gain was justified because the chart was designed for infants at risk of cot death (Cole 1998a). However, more recently, current guidelines recommend that infants be monitored no more frequently than fortnightly up to the age of 6 months (Wright 2000). This chart has its disadvantages: no allowance is made for regression to the mean and the channel widths are uninformative for weight measurements taken more than eight weeks apart (Cole 1995).

### 3.5.2 Cole's 3-in-1 weight-monitoring chart

Cole's 3-in-1 chart is designed for weight monitoring in infancy over 4 week intervals (Cole 1997b). It combines distance, velocity and conditional reference on one chart (Cole 1997b). The chart consists of the centiles for weight from the UK 1990 reference over the age range birth to one year with lines overlaid that cut across the centiles (Cole 1997b). Cole (1997b) gave these lines the term 'thrive lines' (the construction of these lines is described below); these provide a cut-off for identifying infants at risk of failing to thrive. Each thrive line rises with age (indicating weight gain), flattens off and then starts to fall again (indicating weight loss) (Child Growth Foundation 1996d). The downward trend with age of the thrive line after the age of 7 months means that an infant can lose weight without failing to thrive (Child Growth Foundation 1996d). The 3-in-1 weight monitoring chart for girls is reproduced in figure 3.1. In order to use this chart a child needs to have their weight monitored twice (with the two measurement occasions separated by a 4 week interval). After the attained weights are plotted, these two points are joined by a straight line. The gradient of this line is then compared to the nearest thrive line (Child Growth Foundation 1996d). If the slope (gradient) is greater than the thrive line then this child is no cause for concern, whereas if the slope (gradient) is less than the thrive line then this infant is at risk of failing-to-thrive (Cole 1997b). Cole (1998a) later suggested a similar procedure to that described in section 3.3.2 for using the 3-in-1 chart (Cole 1997b), namely assessing the weight gain over the past 4 weeks, and if it is below the fifth centile then to assess the past 8 weeks.

> Thus a child who grows parallel to or slower than the thrive line for 2 months or more is clearly failing to thrive (Cole 1998b, pp67).

Now consider two weight measurements with Z-scores $Z_1$ and $Z_2$, with a correlation between them of $r$ (this correlation is obtained from 223 Cambridge infants measured 4 weekly). Then if $Z_2$ is regressed on $Z_1$ we arrive at equation (3.2) and

the residual standard deviation of $Z_2$ from the regression is given by $\sqrt{1-r^2}$. Thus the conditional $100\alpha$th velocity centile is given by (Cole 1998a):

$$Z_{2|1,100\alpha} = rZ_1 + \sqrt{1-r^2}z_\alpha \tag{3.12}$$

where $z_\alpha$ is the normal equivalent deviate and takes on the value of 1.645 (for the conditional fifth centile) in the creation of the 'thrive lines' in Cole's 3-in-1 chart (Cole 1997b). Cole (1998a) has since created a weight chart with these 'thrive lines' plus the 95th conditional weight gain centile overlaid.

Cole's 3-in-1 weight chart has not been tested in clinical practice, the charts sensitivity and specificity still need to be assessed.

> The ability of a test to distinguish between the two groups is summarised by the sensitivity, specificity, predictive values and ROC curve for the test. However, the methodology is difficult to apply to growth charts, largely because the 'ill' group is hard to define (Cole 1998a, pp2705).

A disadvantage of Cole's 3-in-1 chart is the requirement of weight monitoring at 4 week intervals: at ages closer (further apart) in time more (fewer) than 5% of children will show weight faltering (Cole 1997b). It seems unlikely that children will always be routinely monitored at 4 weekly intervals, in the Newcastle infancy data (considered in this thesis) there is some evidence of small peaks around 4 week intervals. Cole (1997b) also raises the issue that data of poorer quality, such as routine weight data, is likely to lead to more than 5% of children exhibiting weight faltering. Cole (1998a) holds the view that using the slope of curves is mathematically correct and more flexible, than say the distance between curves used in both the Sheffield or Wright chart (see next section for details). However, it is suspected that there may be some difficulty in utilising the 3-in-1 chart for borderline cases, because the chart relies on comparing gradients.

### 3.5.3 Wright chart

The Wright chart (Wright et al. 1998) has non-equidistant channels, which are wider at the top of the normal range and progressively narrower at the lower end of the normal range. This reflects the phenomenon of regression to the mean (Healy and Goldstein 1978; Cole 1995; Wright et al. 1994) discussed above, namely large babies tend to fall and smaller babies tend to rise in terms of centile position.

The relationship between early and late centile positions (using the revised UK 1990 reference) was defined by using routine weight data from a cohort of 3418 full term infants from Newcastle (the data frame considered later in this thesis) (Wright et al. 1998). The 'early' Z-score ($Z_1$) of each child was obtained by taking the average of routine weight Z-scores at birth and the grouped ages of 1 and 2 months, termed the baseline Z-score (Wright et al. 1998). An average Z-score was used as a baseline to minimise the effect of possible distortion by individual measurement error or short term illness (Wright et al. 1998). The 'late' Z-score ($Z_2$) was defined to be the last available weight for each individual between the age 9 and 18 months (Wright et al. 1998). Using the same approach described in section 3.1.3 the following regression equation was arrived at (Wright et al. 1998)[6]:

$$Z_2 = 0.58Z_1 + 0.08 \tag{3.13}$$

After rounding the coefficient of $Z_1$ to 1 decimal place and dropping the constant term, the 'thrive index' ($TI = Z_2 - 0.6Z_1$) was calculated for every child. This allowed the normal range and lower limits of the TI to be calculated; suggesting 95% of children had values above -1.4 (Wright et al. 1998). A range of centile lines were selected; these had to meet the following requirement, namely that a fall through the equivalent of two adjoining channel widths (from any starting position within the normal range) would represent growth faltering that placed this child in the slowest gaining 5% (equivalent to fall in Z-score of 1.4) (Wright et al. 1998). However, it was not possible to achieve this whilst producing a chart that consistently represented more severe falls through the normal range (Wright et al. 1998). Although Wright et al. (1998) stated that if a child had a starting weight above the fiftieth centile, then a fall through three channel widths would place this child in the slowest gaining 1% of children, whereas a child whose starting weight was below the fiftieth centile would have to fall through four channel widths to be amongst the slowest 1%.

The Wright chart consists of 11 solid lines depicting 10 channels. The lines have starting Z-score values (with corresponding major centiles if they coincide) of 2.2, 1.1, 0 (50th), -0.67 (25th), -1.33 (9th), -1.8, -2.15, -2.4, -2.67 (0.4th), -2.85 and -3 (Wright et al. 1998). Additional guidance lines are added at Z-scores of -4 and -5 (Wright et al. 1998). The Wright chart for girls in the first year of life is reproduced in figure 3.2. The lower limit of the chart is at 32 weeks gestation, but Wright et al. (1998) state that the chart should be used with caution in this age range. The upper

[6]This is a different regression equation to that arrived at in section 3.1.3 because a different baseline Z-score was used and the revised UK 1990 reference has since been published.

age limit of the chart was 5 years because FTT often persists up to this age (Wright 1996). Although Wright et al. (1998) state that this chart should not be used to identify new cases after that age of 2.

Guidelines for using the Wright chart are: a fall through two channels is indicative of moderate failure to thrive, whereas a fall through 3 or 4 channels is indicative of severe failure to thrive (Wright 1998). Recovery is defined to be catch-up within 1 channel width of baseline (Wright 1998). Wright et al. (1998) evaluated the effectiveness of this chart and found that the new format significantly increased the proportion of correctly rated charts, with the greatest impact in severe cases.

The Wright chart is designed to monitor over longer periods of time than Cole's 3-in-1 chart. However it is only designed to monitor weight trends over periods of at least 6 months from baseline (Wright et al. 1998). The Wright chart does suffer from similar limitations discussed above for centile crossing, namely a child falling through two channels may have started near or far from one of the 11 solid lines.

## 3.6 Multilevel models for longitudinal growth norms and LGROW

As highlighted above, one of the problems with both the velocity and conditional reference charts is the restriction that a child must be measured at exactly yearly intervals for height during childhood or 4 weekly intervals for weight in infancy. It is unlikely, outside a research setting, that a child would be measured at such regular intervals. Furthermore a child is unlikely to have just have two measures. In the Newcastle data considered within this thesis, a child may have up to 11 weight measures in infancy plus weights and heights in childhood that are between 1.6 and 8.7 years apart.

Pan and Goldstein (1997) developed longitudinal growth norms, using a 2-level model fitted to repeated measures Z-scores, to address the issues outlined in the last paragraph. The development of longitudinal growth norm involves three stages. Initially a sample of repeated measurements, such as heights or weights, are converted to Z-scores using the LMS method (Cole 1988; Cole and Green 1992) as described in section 2.1.1. Assuming the LMS method provides normally distributed Z-scores, and treating $z_{ij}$ as a response, Pan and Goldstein (1997) construct the following

2-level model:

$$z_{ij} = \sum_{h=0}^{p} \beta_h t_{ij}^h + \sum_{h=0}^{q} u_{hj} t_{ij}^h + e_{ij} \tag{3.14}$$
$$E = e_{ij} \backsim N(0, \sigma_e^2 I)$$
$$U = u_{hj} \backsim N(0, D_u)$$

where $i = 1, \ldots, n_j; j = 1, \ldots, m$, $n_j$ is the number of measurement occasions (level 1) for the the $j$th individual and $m$ is the number of individuals. There are $p$ fixed coefficients and $q$ random coefficients (Pan and Goldstein 1997). Considering equation (3.14), the 'fixed' part of the model corresponds to the first summation, which represents the average polynomial trend with age (Pan and Goldstein 1997). It is assumed that the level 1 residuals ($e_{ij}$) in equation (3.14) are independently distributed with constant variance (Pan and Goldstein 1997). Pan and Goldstein (1997) do point out that this latter assumption was likely to be violated if measures were taken close in time and suggest that equation (3.14) could be extended to incorporate autocorrelated residuals. In the final stage the estimated parameters from equation (3.14) are used to provide longitudinal growth norms.

Equation (3.14) is used for serial height or weight Z-scores, if norms are to be constructed for weight Z-score conditional on current height Z-score or previous height and weight Z-scores then the bivariate extension given by equation (3.15) is required (Pan and Goldstein 1997).

$$z_{ij} = \delta\left[\sum_{h=0}^{p} \beta_h t_{ij}^h + \sum_{h=0}^{q} u_{hj} t_{ij}^h + e_{ij}\right] + (1-\delta)\left[\sum_{h=0}^{p'} \beta'_h t_{ij}^h + \sum_{h=0}^{q'} u'_{hj} t_{ij}^h + e'_{ij}\right] \tag{3.15}$$

where ($'$) represents random variables associated with weight. Pan and Goldstein (1997) illustrated use of equations (3.14) and (3.15) on longitudinal height and weight data of 91 boys, aged 2 to 18.5 years, from the Edinburgh growth study.

The package LGROW uses the multi-level approach described above to construct longitudinal growth norms from several European growth studies. The differences between these samples were adjusted for and some attempt was made at adjusting the growth rates (Goldstein and Pan 1998). LGROW can be used to assess patterns of height and weight measurements from 3 months of age to maturity (16 for girls and 18 for boys). The infancy period is avoided as the rapid change at this age was too difficult to model and there were some problems with stability of the function at both ends of the age-range (Goldstein and Pan 1998). However, the time period

between birth and three months is when a large number of infants are routinely monitored.

## 3.7 Discussion

In order to detect growth faltering, height (weight) velocity is a more sensitive indicator than attained height (weight). However height (weight) velocity depends on the initial height (weight), is subject to measurement error and makes no allowance for regression to the mean, although Carpenter (2000) views the effect of regression to the mean to be trivial over short time intervals. An alternative approach is to use a conditional reference. The conditional reference charts are subject to the same limitations as velocity charts; namely, they are restricted to a set age interval, such as 4 weeks or 1 year. However Cole's 3-in-1 chart and the Wright chart are superior to the attained weight chart because they provide some guidance in a longitudinal context. The multi-level approach employed in LGROW appears to be of value for weight or height measures taken several years apart during childhood but requires specialist software. A multi-level approach was also used by Royston (1995) for constructing conditional centiles of estimated foetal weight.

The approach preferred in this thesis is to use the conditional gain Z-score, given by equation (3.11) because this is flexible, with no restrictions been placed on time between weight (height) measures. An additional consideration is the time interval over which to calculate a conditional weight gain Z-score. If the interval is too short the conditional gain Z-score is likely to be more variable (due to a combination of measurement error and biological variation) than a conditional gain Z-score over a longer time interval (Cole 1995). For these reasons, Cole (1995) advocated the calculation of a conditional weight gain Z-score over the longest possible time interval as this also minimised the degree of negative correlation. However if a child's growth is faltering then we wouldn't want to wait over the longest time interval possible.

In order to obtain a conditional weight gain Z-score a correlation is required. Heimendinger and Laird (1983) stated that the correlations should be pertinent to the population under study. Cole (1995) had developed a model for the Cambridge correlations. However the Cambridge infancy weights may be unrepresentative of routine infancy weights in general; as research conditions may lead to higher correlations but this may be partly offset by the homogeneity resulting from the screening procedures employed by mid-wives to recruit Cambridge infants (Cole 1998a). In chapter 5 we explore the correlation structure of routine Newcastle infancy weight

data and develop a model for the correlation structure.

The conditional weight gain Z-score provides a way of comparing two weight measures. If there are more than two weight measures then this increases the number of conditional weight gain Z-scores that can be obtained, and it raises the question of how to interpret and utilise these as they evolve. A further issue is how many previous weight Z-scores is it useful to condition on? This has been addressed for height in childhood by Cole (1994a) and Pan and Goldstein (1997). Pan and Goldstein (1997) concluded that conditioning on 2 or 3 height measures was different to conditioning on 1, but Cole (1994a) suggested that it was only useful to condition on two previous heights around puberty. These issues are considered further in reference to routine infancy weight data in chapter 7.

Figure 3.1: Cole's 3-in-1 weight-monitoring chart for girls (reproduced here with kind permission of the Child Growth Foundation, copyright 1996)

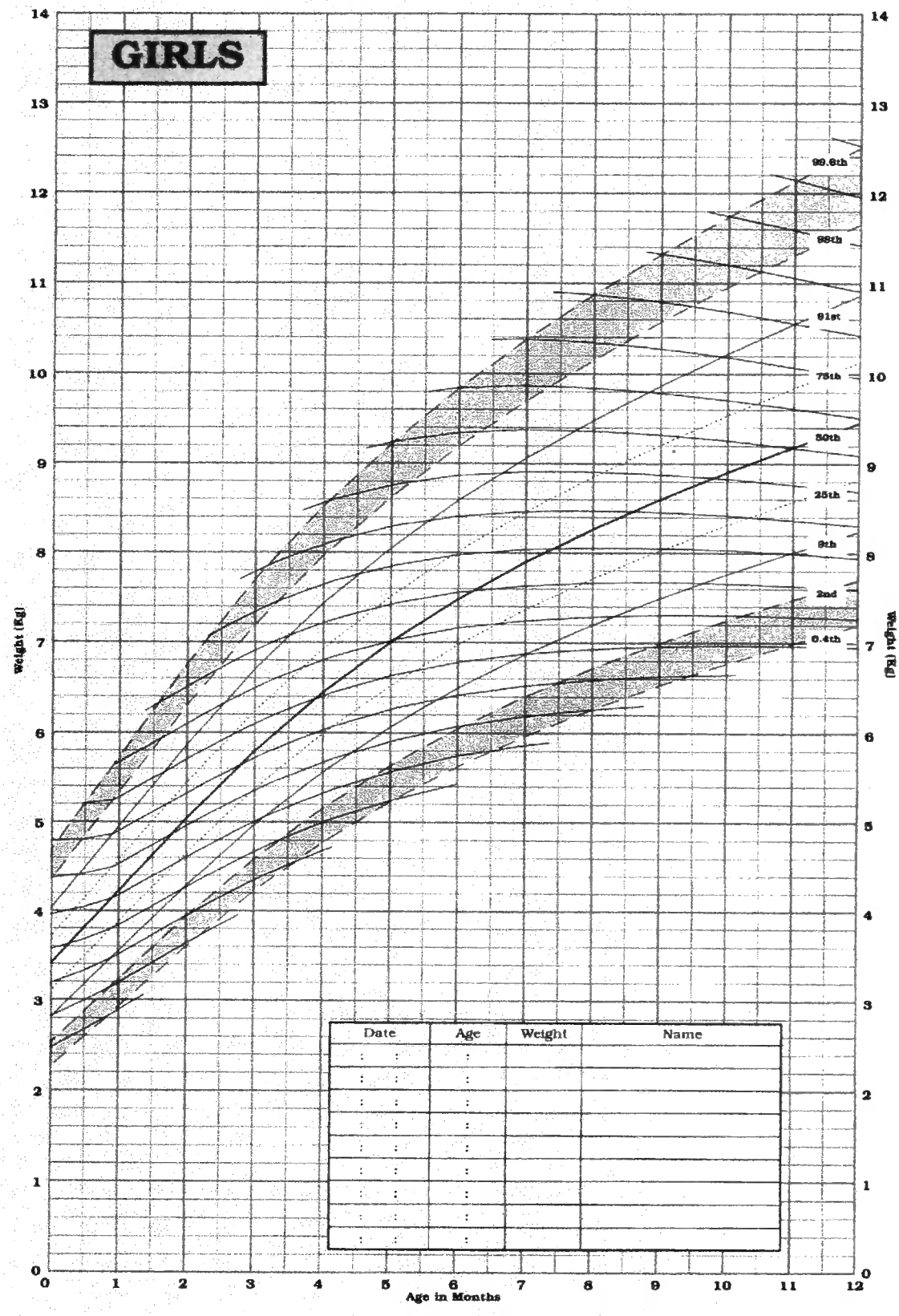
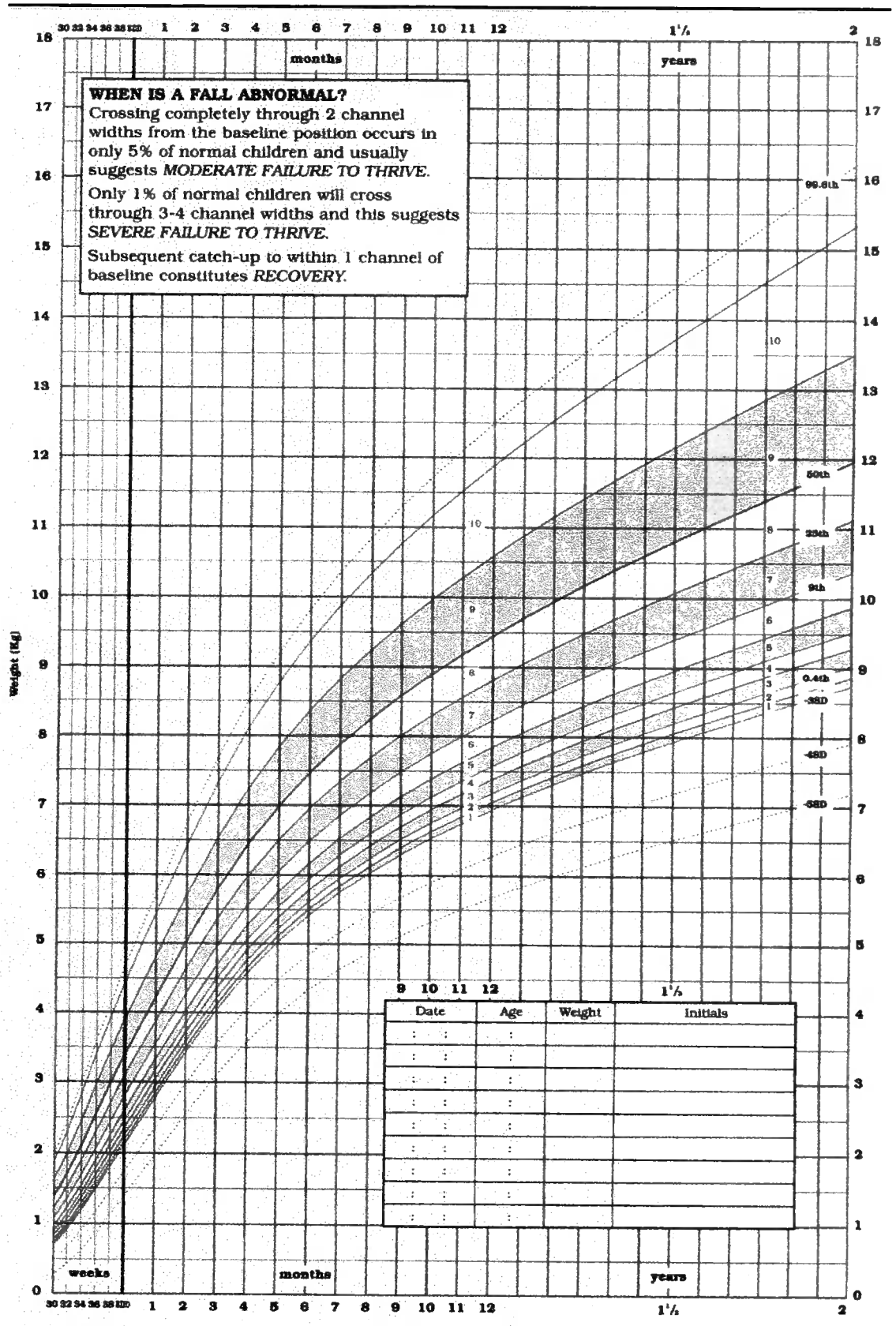


Figure 3.2: Wright chart for girls over age range pre-term to two years (reproduced here with kind permission of Dr C.M. Wright)



# Chapter 4

# The Newcastle infancy data

This chapter concentrates on a year birth cohort of routine infancy weight data collected in Newcastle. In section 4.1 the role of weight monitoring in infancy is briefly discussed. Section 4.2 provides a summary of the possible sources of error within routine infancy weights. A description of all variables within the Newcastle infancy data frame can be found in section 4.3. Section 4.4 summarises the main publications on the infancy data. In section 4.5 we obtain summary statistics and histograms for the weight Z-scores at grouped ages. The adequacy of the revised UK 1990 reference for attained weight of infants from Newcastle is also explored. In section 4.6, the approach used in the research study to identify children at risk of failure-to-thrive is described. In section 4.7 we consider all the routine weight Z-scores in the infancy data frame. Within this section there are many exploratory plots of weight Z-scores versus age with a lowess curve (Cleveland 1979) overlaid. This allows us to explore the trends within the infancy weight data. In this thesis graphical displays are placed at the end of chapters.

## 4.1 Weight monitoring in infancy

A child's first weight is recorded within minutes of being born. After birth, weights are routinely monitored by health visitors, usually when a mother visits the baby-clinic for a review appointment or an immunisation. However, a baby is likely to be monitored more frequently if there is concern about the child's attained weight or weight gain on the clinician's or mother's part. In monitoring a child's weight the main concern is to determine whether it is within the normal range.

Current guidelines (Wright 2000), recommend that a child need only be weighed

at birth, at immunisation times and at surveillance checks until the age of 1 year. Furthermore, a baby should not be weighed more than once every 2 weeks under the age of 6 months and no more than once a month thereafter. However, a child should be weighed whenever there is clinical concern and weight checks should be continued beyond 1 year in this instance.

> Paediatricians and primary care workers throughout the world use weight gain as one of their most basic measures of health and well-being in early childhood (Wright et al. 1994, pp356).

Growth monitoring in infancy may facilitate detection of a variety of disorders, such as organic and non-organic failure-to-thrive. Parents value weight monitoring in infancy and the babies weight gain may provide an opportunity to discuss nutrition and other aspects of child rearing (Hall 1996). Poor weight gain and growth faltering in infancy have been shown to be associated with cot death (Sinclair-Smith et al. (1976) in Wright et al. (1994)), developmental delay (Dowdney et al. 1987; Dowdney et al. 1998) and ischaemic heart disease (Barker et al. 1989). However, in the past, doubts have been expressed about the usefulness of weight monitoring as a preventative or diagnostic method (Hall 1996; Editorial 1992).

> As a routine procedure there seems little justification for regular weighing after the first few months of life once the parent and primary care team are satisfied that the baby is feeding normally and has begun to gain weight. We are not convinced that the advantages conferred by regular weighing after the first few months of life justify the resources required or anxiety generated by uncertain or inexpert interpretation of growth charts (Hall 1996, pp114).

More recently, in a systematic review, Garner et al. (2000) concluded that growth monitoring has been inadequately evaluated.

A child's birth weight is influenced by maternal stature and conditions during pregnancy. However birth weight along with gestational age is essential for growth monitoring and is also an important epidemiological marker (Wright 2000). In particular, Barker et al. (1989) and Barker (1999) found an association between size and shape at birth and subsequent risk of ischaemic heart disease, obesity, hypertension and glucose intolerance.

Weight is a measure of the varying combination of height, body fat, water and muscle bulk. In addition, it is subject to error due to biological variations and clothing weight (see next section for further details). Therefore, weight is a less straightforward measure of growth than height (Wright et al. 1994). However, the practical value of weight monitoring is its widespread availability, the ease of measurement and its accuracy in infancy and early childhood (Wright et al. 1994). In infancy, until the age of 2 years, the measurement of supine length requires two observers, specialist equipment and training. Thus, length measurements taken by the skilled professional are still subject to high levels of measurement error (Wright et al. 1994). However, Doull et al. (1995) measured 38 infants and found that the standard deviation of a single length measurement in infants younger than 1 year was comparable to the standard deviation of a single height measurement in older children. As a result of the Coventry Consensus, current guidelines suggest that there is no justification for routine monitoring of length before the age of 2 years and length should only be measured when there is concern about a child's growth or weight gain (Wright 2000).

## 4.2 Accuracy of routine weights in infancy

Accurate weighing requires that the baby be weighed nude unless there are special circumstances (Hall 1996). Davies and Williams (1983) drew attention to poor weighing methods, cold rooms, babies weighed in different states of undress, inaccurate weighing scales and insufficient use of growth charts. In addition, observers were found to have inadequate understanding of normal variations in weight gain and a poor understanding of the nutritional influences on weight gain.

Although it is generally accepted that routine weight measures in infancy will contain some element of error. There is actually very little written about this topic in the literature, Alsop-Shields and Alexander (1997) carried out a study of the errors that can occur when weighing infants, but this only involved 7 babies.

The following list outlines issues of accuracy and reproducibility:

1. instrument error Each clinic will have a different set of scales, although these should be routinely checked for accuracy. Modern scales, if properly maintained, weigh to within 10-20 grams (Hall 1996).

2. intra-observer error Repeat weight measurements made on the same child are rarely the same, even when the same instrument is used by

the same observer. Alsop-Shields and Alexander (1997) found that for each observer, the difference between repeat weights was between 0 and 10g.

3. test-retest stability This is the weighing of the same infant at different times. Willett (1990) found this to be less than the source of variation in weight measurements from biological factors such as intake and hydration.

4. inter-observer error Weight measurements made by different members of staff on the same infant will vary. Alsop-Shields and Alexander (1997) found that between observers this was a maximum of 20g.

5. bias Readings of weight are subject to bias if the observer knows previous readings or has preconceived ideas as to whether or not the child is growing 'normally' (Hall 1996).

6. biological variation Within a day the babies weight is effected by the timing of feeding, urination and defecation, i.e. intake and output. Weight may fluctuate by several hundred grams, depending on the contents of bowel, bladder, and stomach, as well as minor fluctuations due to intercurrent illness (Hall 1996). Fomon (1991) suggests that the weight of a single feeding is about 180g.

7. weight of clothing Ideally a baby would be stripped down before weighing, but in practice the infant could be weighed in their nappies and possibly in light clothing. Unused disposable nappies can weigh between 30 and 60 grammes, towelling nappies weighed between 128 and 141 grammes (Alsop-Shields and Alexander 1997). Obviously soiled nappies would weigh more than this.

8. short term variability Growth rates vary with the season of the year, intercurrent illnesses and other factors. Therefore even the most precise measurements made over a short period of time (6 months or less) are likely to be misleading (Hall 1996).

9. errors in recording weights The babies weight is recorded on their 'well-baby' card. Errors may occur at this stage due to recording error or if the weight is only noted on the chart. Ideally, weights should be noted in the records and plotted on a growth chart.

## 4.3 Newcastle Infancy data

An annual birth cohort of 3653 children, resident in Newcastle in November 1989 and aged 18-30 months, were identified using the Newcastle-upon-Tyne Child Health computer system. These infants were born between 1st April 1987 and 31st March 1988. All the baby clinics in the city were visited and the child health records of these infants were reviewed by Dr C.M. Wright to obtain data on birth weight, gestation and all routine weights together with limited medical information. Only 47 children were identified in their records as having a chronic medical disorder that might have affected their growth (Wright et al. 1993).

There were 3418 term infants within the birth cohort and children born before 37 weeks of gestation were excluded. The full infancy data frame contains birth weight and up to 10 additional routine weights. For each infant the routine weights nearest six target ages (6 weeks, 3, 6, 9, 12 and 18 months) were identified. Wright (1997) defined the child's health record to be inadequate if there was less than one clinic weight before or after the age of 6 months. There were 703 (20.6%) children with inadequate records, 257 (7.5%) of these had no clinic weights (i.e. weights recorded after birth weight) and 446 (13.0%) only had weights in the first or second half of the first year (Wright 1997). Wright (1997) found that the children with the most weights were consistently lighter while children with inadequate records, if weighed, were heavier.

In the process of carrying out the follow-up study at 7-9 years (see Chapter 6), it was discovered that one of these records was actually a duplicate (ID's 1873 and 2987, data on 2987 was retained) and two individuals (ID's 1090 and 2715) were actually born pre-term (32 and 36 weeks, respectively). Therefore, the infancy data frame now contains weight data on 3415 term infants, 1711 males and 1704 females. There are 1055 (30.9%) infants with birth weight and weights in all six age groupings and 90 (2.6%) infants with no recorded weights.

The infancy data frame consists of the following variables for each individual:

1. **ID** An identification number was allocated to each infant from the birth cohort.

2. **Ges** Gestational age is the term used for age in prenatal period, this tends to be measured from the first day of the last menstrual period (Tanner 1989). Lengths of gestation of 37 to 42 weeks are regarded as normal. In the Newcastle infancy data gestational age varied be-

tween 37 and 43 weeks, and 784 (23.0%) infants had no gestational age recorded.

3. **SES** Every child in the birth cohort was matched using their postcode to a predefined neighbourhood. The level of deprivation was then calculated using the Townsend score, which summarises census data based on car and home ownership, overcrowding and unemployment rates (Wright et al. 1994). **SES** is a categorical variable that has three levels, deprived, intermediate and affluent, coded 3, 2 and 1, respectively. This variable provides a crude indicator of the child's socio-economic status at the end of infancy. There are 379 (11.1%), 2353 (68.9%), and 683 (20.0%) infants defined as affluent, intermediate and deprived, respectively.

4. **Sex** Coding: 1 for males and 2 for females.

5. **Age1** This is a column of zeroes corresponding to age for birth weight.

6. **Age2**, **Age3**, **Age4**, **Age5**, **Age6** and **Age7** is the actual age in months which fell into the six age-groupings of 6 weeks, 3, 6, 9, 12 and 18 months.

7. **Weight1** is the birth weight in Kilograms.

8. **Weight2**, **Weight3**, **Weight4**, **Weight5**, **Weight6** and **Weight7** are the routine weights in Kilogrammes at ages defined in point 6.

9. **Zscore1**, **Zscore2**, **Zscore3**, **Zscore4**, **Zscore5**, **Zscore6** and **Zscore7** are the weights defined in points 7 and 8 converted to Z-scores using the L, M and S values from the revised UK 1990 growth reference (Freeman et al. 1995; Preece et al. 1996; Cole et al. 1998). The birth weight Z-score is calculated using L, M and S values at age zero in the reference, because within this data frame all infants are born at term.

## 4.4 Previous research on the Newcastle infancy data

The purpose of retrieving weight data from a routine whole year birth cohort resident in Newcastle was to explore limits of normal variation in weight gain in infancy.

At the time of collecting the infancy data, there was no modern UK growth reference. The Tanner-Whitehouse charts (Tanner et al. 1966) and the Gairdner-Pearson charts (Gairdner and Pearson 1971; Gairdner and Pearson 1985) for premature and other infants were used in child-health clinics. Wright et al. (1993) explored the validity of using the Tanner and Whitehouse standards (Tanner et al. 1966), the National Center for Health Statistics (NCHS) growth standards (Hamill et al. 1977) and the Cambridge standards (Whitehead et al. 1989b) for the Newcastle weight data. The main conclusions where that both the NCHS and Tanner-Whitehouse Standards were unsuitable for use before the age of 1 year, as the Newcastle children show a rise above both of these standards up to 3 months, then a downward fall to 18 months. Thus, both the NCHS and Tanner-Whitehouse standards would give the impression of growth faltering after the age of 3 months. The Cambridge standards showed less of a discrepancy overall. The early weights of Newcastle children were similar to the Cambridge children but then showed a gradual gain on the standards. Wright et al. (1993) combined the Z-scores of boys and girls when they compared Newcastle children to the NCHS, Tanner-Whitehouse and Cambridge standards.

Wright et al. (1994) developed the 'thrive index' (TI), which has already been discussed in Chapter 3. The last available weight Z-score for each individual in the time period 9-24 months was regressed on their Z-score at 6 weeks. The pair-wise correlations derived from the Newcastle infancy data at grouped ages were used to derive expected and lower tabulated limits for an hypothetical child on the major centiles (3rd, 10th, 25th, 50th, 75th, 90th and 97th) of the Cambridge standards at 6 weeks.

Wright et al. (1994) explored the relationship between deprivation and weight gain. The TI was calculated for each child using the last available weight Z-score between 9 and 18 months and weight Z-score at 6 weeks. Wright et al. (1994) found that children from deprived areas were lighter at all ages and by one year they were three times as likely as affluent children to be below the third centile. Twice as many deprived children were found to be below the 5% screening threshold compared to intermediate. However, there was also an excess of affluent children below the 5% threshold compared to intermediate. Thus, the overall conclusion was that cases of failure-to-thrive could come from a wide social background.

## 4.5 Preliminary analysis of the Newcastle infancy weight data at grouped ages

### 4.5.1 Data summaries

In this section, the infancy data frame with birth weight and up to 6 routine weights allocated to grouped ages of 6 weeks, 3, 6, 9, 12 and 18 months will be considered. All the publications discussed in the last section are concerned with this data frame. The full infancy data frame will be summarised later in this chapter. The number of routine weight measures per individual by sex can be found in table 4.1. The distribution of the number of weight measures is similar for males and females ($\chi^2$=6.6225, p=0.4692). The number of routine weight measures per individual by level of deprivation can be found in table 4.2. The distribution of the number of weight measures is related to the level of deprivation ($\chi^2$=36.8653, p=0.0008). In particular, there are fewer than expected affluent infants having 5 weights, there are fewer than expected deprived children with 7 weights and more than expected deprived children with 2 weights.

Table 4.3 summarises the age distribution in each of the age groupings for all infants. In all age groups the mean is pretty similar to grouped age, but the variances are larger for age groups 6 months and older, especially in the last age grouping of '18 months'. There are also more 'missing' weights in the later age groups, post 9 months a third of the birth cohort have no weight measurements taken in that particular age grouping.

Table 4.1: Infancy: Summary of number of weight measures in six age-groupings and at birth by sex

| No. | male | female | Total |
|---|---|---|---|
| seven | 503 | 552 | 1055 |
| six | 393 | 355 | 748 |
| five | 300 | 283 | 583 |
| four | 171 | 189 | 360 |
| three | 112 | 112 | 224 |
| two | 77 | 68 | 145 |
| one | 110 | 100 | 210 |
| none | 45 | 45 | 90 |
| Total | 1711 | 1704 | 3415 |

Table 4.2: Infancy: Summary of number of weight measures in six age-groupings and at birth by SES

| | Number of weights | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| SES | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | total |
| 1 | 12 | 29 | 19 | 32 | 31 | 42 | 87 | 127 | 379 |
| 2 | 62 | 146 | 83 | 141 | 246 | 415 | 511 | 749 | 2353 |
| 3 | 16 | 35 | 43 | 51 | 83 | 126 | 150 | 179 | 683 |

Table 4.3: Infancy: Summary statistics of distributions of actual age within each age grouping

| Age | Min. | LQ | Median | Mean | UQ | Max. | SD | Missing |
|---|---|---|---|---|---|---|---|---|
| Ges | 37 | 39 | 40 | 39.58 | 40 | 43 | 1.25 | 786 |
| 6 weeks | 0.750 | 1.380 | 1.740 | 1.641 | 1.930 | 2.100 | 0.3497 | 595 |
| 3 months | 2.070 | 2.890 | 3.020 | 3.048 | 3.180 | 4.490 | 0.3328 | 689 |
| 6 months | 4.520 | 5.670 | 5.970 | 5.949 | 6.260 | 7.480 | 0.5466 | 750 |
| 9 months | 7.510 | 8.590 | 9.050 | 9.059 | 9.570 | 10.490 | 0.6883 | 1310 |
| 12 months | 10.52 | 11.74 | 12.03 | 12.04 | 12.36 | 13.48 | 0.5749 | 1221 |
| 18 months | 13.51 | 15.69 | 17.64 | 17.56 | 18.82 | 23.97 | 2.3375 | 1694 |

Standard deviation scores (Z-scores) by definition should be symmetrically distributed, with mean zero and a standard deviation of one. At birth and 6 weeks, the centre of the Z-score distribution for both boys and girls is shifted to the left of zero; see figure 4.1. If we consider the upper panel of figure 4.3: the histograms and quantile-quantile plots for weight Z-scores at 9 months for boys and girls. This figure illustrates the typical features observed for weight Z-scores of boys and girls at grouped ages in infancy (see figures 4.2, 4.3 and 4.4). The histogram for boys weight Z-scores at 9 months is reasonably symmetric with some outlying high negative Z-scores. The histogram for girls weight Z-scores at 9 months indicates that the mean is shifted to the left of zero and we have a long tail to the left. The quantile-quantile plots indicate that distribution of weight Z-scores in infancy are reasonably normal however they may have slightly fatter tails, which in part could be due to cases defined later (see section 4.6). The histograms and summary statistics of weight Z-scores for age-grouping by sex given in table 4.4 indicate that in early infancy, children in Newcastle tend to be lighter than those children that contribute

to the reference, this will be discussed further in the next section. A general feature that will be discussed in more detail below, is the mean Z-score for girls weight is consistently below zero at around -0.2.

### 4.5.2 The adequacy of the revised UK 1990 reference for Newcastle infants

Wright et al. (1996) used the Newcastle infancy data to illustrate that the original UK 1990 reference (Freeman et al. 1995) had a sex bias in infancy. It was found that two and half times as many girls as boys had weights below the third centile during the first year. The UK 1990 reference was revised (Preece et al. 1996) and according to Cole et al. (1998) there is no sex bias in the current reference. In the first year of life the UK reference is derived predominantly from the Cambridge infant growth study and the only other weights are provided by the HUMAG infants (this study was designed to be nationally representative).

The revised UK 1990 reference (Freeman et al. 1995; Preece et al. 1996; Cole et al. 1998) was used to convert the Newcastle infancy weight data to Z-scores.

Table 4.4: Summary statistics of Z-scores for weight by sex, for birth and age groupings 6 weeks, 3, 6, 9, 12 and 18 months

| | Boys | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Age | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
| birth | -3.6300 | -0.9690 | -0.2560 | -0.2695 | 0.4004 | 3.0040 | 1.0244 | 1588 |
| 6 weeks | -4.4800 | -0.6575 | -0.0700 | -0.0602 | 0.5600 | 3.0900 | 0.9551 | 1402 |
| 3 mths | -4.6100 | -0.6300 | 0.0200 | 0.0003 | 0.6600 | 3.7700 | 0.9870 | 1358 |
| 6 mths | -4.9000 | -0.6600 | 0.0300 | 0.0454 | 0.73000 | 4.0000 | 1.0222 | 1328 |
| 9 mths | -4.6400 | -0.6000 | 0.0800 | 0.0826 | 0.7725 | 5.0400 | 1.0653 | 1060 |
| 12 mths | -4.3100 | -0.6000 | 0.1100 | 0.1035 | 0.8000 | 4.2400 | 1.0529 | 1081 |
| 18 mths | -4.6700 | -0.6275 | 0.0350 | 0.0527 | 0.7500 | 3.9600 | 1.0513 | 846 |
| | Girls | | | | | | | |
| Age | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
| birth | -4.0560 | -0.9452 | -0.2486 | -0.2749 | 0.4262 | 3.1050 | 1.0835 | 1580 |
| 6 weeks | -4.0400 | -0.8300 | -0.1850 | -0.2081 | 0.4100 | 2.9400 | 0.9748 | 1418 |
| 3 mths | -3.8100 | -0.8825 | -0.1800 | -0.2027 | 0.4600 | 3.5000 | 1.0021 | 1368 |
| 6 mths | -3.7600 | -0.8800 | -0.1500 | -0.1522 | 0.5700 | 2.9700 | 1.0618 | 1337 |
| 9 mths | -4.0500 | -0.8400 | -0.1700 | -0.1184 | 0.6300 | 2.9700 | 1.0894 | 1045 |
| 12 mths | -3.5400 | -0.8100 | -0.1200 | -0.0857 | 0.6800 | 3.2300 | 1.0795 | 1113 |
| 18 mths | -4.4000 | -0.8825 | -0.2200 | -0.1919 | 0.5300 | 3.6000 | 1.1115 | 876 |

However, it is important to be aware of the effect of using an inappropriate reference. For example, if the mean weight Z-score of Newcastle children was consistently below zero then we may find more children than expected below the third centile. Cole (1993) defined the mismatch between data and a reference in terms of the offset and trend. The offset is defined as the mean Z-score of the weight data, while the trend is the regression coefficient of the Z-score of the weight data on age. Cole (1993) states that if the offset and trend are near zero then the data and reference are well matched. No guidance is provided in terms of the size of the age range this approach could be applied to. It is suspected that in infancy and puberty, when growth changes are rapid, intervals of much less than one year are required to assess the adequacy of the growth reference for any anthropometric data. During childhood an interval of a year is probably appropriate as there are seasonal variations in height growth, with children growing more in spring and summer than in autumn and winter.

In the approach described above, if the offset is non-zero but we have no linear trend with age, then the reference adjusts the data appropriately for age. However we should be cautious in using cut-offs as there will be either a deficit or excess of children below say the third centile. A non-zero linear trend with age indicates that the reference fails to adjust for age and the offset will be meaningless as this will also vary with age. If the age trend is assumed to have been removed when it has not then this may lead to spurious correlations between anthropometry and other age related variables (Cole 1993). In infancy or puberty there is the possibility of a quadratic age trend, but the approach defined above only detects a linear trend.

We used a t-test to assess whether or not the mean Z-score is zero within each sex-age grouping, the resulting t-values can be found in table 4.5. There may be some reason to doubt the null hypothesis that the mean weight Z-score is zero for boys at birth and for the age groupings 6 weeks, 9 and 12 months. There is strong evidence against the null hypothesis that the mean weight Z-score is zero in all age groups for girls, with Newcastle girls being consistently lighter than the reference children.

Results of testing that the variance of weight Z-scores within each sex-age grouping is one can be found in table 4.6. The F-test was used to test the null hypothesis that the variance is one. However, the sampling theory for $s^2$ is sensitive to non-normality so results from this test should be interpreted cautiously if there are serious departures from normality (Armitage and Berry 1987). There may be some reason to doubt that the variance of the weight Z-scores is one at birth, 6, 9, 12 and 18 months for girls. It is suggested that the variance of weight Z-scores in Newcas-

tle girls, from 6 months onwards, is greater than those girls that contribute to the revised UK 1990 reference. There may be some reason to doubt that the variance of the weight Z-scores is one at 6 weeks, 9, 12 and 18 months for boys. It is suggested that the variance of weight Z-scores in Newcastle boys, from 9 months onwards, is greater than those boys that contribute to the revised UK 1990 reference. It should be noted that infant weight Z-scores within the '18 months' group have an age range of 10 months.

Table 4.7 summarises the results of linear regression of Z-score for weight on age in months at 6 weeks, 3, 6, 9, 12 and 18 months. The regression is not done for the birth weight Z-scores as this is not an age grouping as the Z-scores are calculated for all term individuals assuming that age at birth is zero. The results from table 4.7 indicate that there is no reason to doubt the null hypothesis of zero slope, indicating that the infancy data may be appropriately adjusted for age by the revised UK 1990 growth reference.

Table 4.5: Infancy: Results of testing that the mean weight Z-score is zero in each age-group (by sex)

| | Boys weight Z-scores | | |
|---|---|---|---|
| Age group | t | p | 95% CI |
| Birth | -10.4826 | $< 2.2 \times 10^{-16}$ | [-0.3199, -0.2190] |
| 6 weeks | -2.3618 | 0.0183 | [-0.1103, -0.0102] |
| 3 months | -0.0096 | 0.9923 | [-0.0528, 0.05228] |
| 6 months | 1.6193 | 0.1056 | [-0.0096, 0.1004] |
| 9 months | 2.5248 | 0.0117 | [0.0184, 0.1468] |
| 12 months | 3.2322 | 0.0013 | [0.0407, 0.1663] |
| 18 months | 1.4592 | 0.1449 | [-0.0182, 0.1237] |
| | Girls weight Z-scores | | |
| Age group | t | p | 95% CI |
| Birth | -10.0852 | $< 2.2 \times 10^{-16}$ | [-0.3284, -0.2214] |
| 6 weeks | -8.038 | $1.908 \times 10^{-15}$ | [-0.2588, -0.1573] |
| 3 months | -7.4807 | $1.317 \times 10^{-13}$ | [-0.2558, -0.1495] |
| 6 months | -5.2417 | $1.847 \times 10^{-7}$ | [-0.2092, -0.0952] |
| 9 months | -3.513 | 0.0005 | [-0.1845, -0.0523] |
| 12 months | -2.647 | 0.0082 | [-0.1491, -0.0222] |
| 18 months | -5.1103 | $3.95 \times 10^{-7}$ | [-0.2656, -0.1182] |

Table 4.6: Infancy: Results of testing that the variance of weight Z-score is one in each age-group (by sex)

| | Boys weight Z-scores | | |
|---|---|---|---|
| Age group | $\frac{(n-1)s^2}{\sigma^2}$ | approx. p-value | 95% CI |
| Birth | 1665.32 | 0.1645 | [0.9800, 1.1264] |
| 6 weeks | 1277.98 | 0.0201 | [0.8482, 0.9837] |
| 3 months | 1321.86 | 0.5000 | [0.9048, 1.0518] |
| 6 months | 1386.59 | 0.2474 | [0.9697, 1.1292] |
| 9 months | 1201.87 | 0.0019 | [1.0441, 1.2381] |
| 12 months | 1197.22 | 0.0117 | [1.0207, 1.2083] |
| 18 months | 933.96 | 0.0305 | [1.0070, 1.2187] |
| | Girls weight Z-scores | | |
| Age group | $\frac{(n-1)s^2}{\sigma^2}$ | approx. p-value | 95% CI |
| Birth | 1853.72 | $1.0160\times10^{-6}$ | [1.0962, 1.2604] |
| 6 weeks | 1346.36 | 0.1845 | [0.8839, 1.0242] |
| 3 months | 1372.69 | 0.9133 | [0.9329, 1.0839] |
| 6 months | 1506.12 | 0.0010 | [1.0465, 1.2180] |
| 9 months | 1238.93 | $1.9909\times10^{-5}$ | [1.0911, 1.2955] |
| 12 months | 1295.85 | $9.6819\times10^{-5}$ | [1.0742, 1.2686] |
| 18 months | 1081.04 | $8.4203\times10^{-7}$ | [1.1274, 1.3599] |

Table 4.7: Slope coefficients from regression of weight Z-scores in infancy on age

| **Boys** | Estimate | Std. Error | t value | Pr(> \|t\|) |
|---|---|---|---|---|
| Z(6 weeks) | 0.01313 | 0.07259 | 0.181 | 0.857 |
| Z(3 months) | -0.04735 | 0.08291 | -0.571 | 0.568 |
| Z(6 months) | -0.06079 | 0.05029 | -1.209 | 0.227 |
| Z(9 months) | -0.03321 | 0.04653 | -0.714 | 0.476 |
| Z(12 months) | -0.05008 | 0.05486 | -0.913 | 0.361 |
| Z(18 months) | -0.02588 | 0.01534 | -1.687 | 0.0919 |
| **Girls** | Estimate | Std. Error | t value | Pr(> \|t\|) |
| Z(6 weeks) | 0.02786 | 0.07446 | 0.374 | 0.7083 |
| Z(3 months) | -0.05445 | 0.07920 | -0.687 | 0.492 |
| Z(6 months) | -0.03959 | 0.05429 | -0.729 | 0.466 |
| Z(9 months) | 0.01233 | 0.05021 | 0.245 | 0.806 |
| Z(12 months) | -0.01253 | 0.05717 | -0.219 | 0.827 |
| Z(18 months) | 0.004663 | 0.016179 | 0.288 | 0.773 |

## 4.6 Regression on a baseline Z-score

The routine Newcastle infancy data was also grouped differently in early infancy. Instead of a 6 week grouping there is a 1 and 2 month grouping. One of the 1 or 2 month weights will have previously been the 6 week weight. The summary statistics of the weight Z-scores for 1 and 2 month groupings can be found in table 4.8. The mean weight Z-score for girls at 1 and 2 months is shifted to the left of zero at about -0.2. A negative offset is also observed for the boys weight Z-scores at 1 month but there is some indication of a catch up with the reference at 2 months.

### 4.6.1 Original analysis

A baseline weight Z-score ($Z$(base)) was derived for each individual using the average of weight Z-scores at birth and at age groupings of 1 and 2 months. So for example, if an infant only had weights at birth and one weight in the age grouping of 2 months, then:

$$Z(\text{base}) = \frac{Z(\text{birth}) + Z(2 \text{ mths})}{2}$$

An average baseline weight Z-score was chosen over one early weight Z-score, say the 6 week weight Z-score, as this was viewed to be less prone to distortion by individual measurement error or short term illness (Wright et al. 1998).

In earlier work Wright et al. (1994) used weight Z-score at 6 weeks in the development of the 'thrive index'. There were two reasons for this. At the time of

Table 4.8: Infancy: Summary statistics of Z-scores for weight by sex; for age groupings 1 and 2 months

| | **Boys** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Age | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
| 1 mth | -3.3690 | -0.7282 | -0.1630 | -0.1514 | 0.4775 | 2.6940 | 0.9464 | 1354 |
| 2 mths | -4.4840 | -0.6320 | -0.0200 | -0.0310 | 0.6025 | 3.0890 | 0.9743 | 1263 |
| | **Girls** | | | | | | | |
| Age | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
| 1 mth | -4.0380 | -0.8770 | -0.2070 | -0.2479 | 0.4060 | 3.3140 | 0.9897 | 1363 |
| 2 mths | -4.3390 | -0.8095 | -0.2090 | -0.2034 | 0.4265 | 2.9440 | 0.9874 | 1259 |

publication only the Cambridge standards were available and these did not include L, M and S values for birth and weight Z-score at 6 weeks was more highly correlated than birth weight Z-score with later weight Z-scores (Wright 1996).

In the original analysis, carried out by Dr C.M. Wright, Z-scores at 3, 6, 9, 12 and 18 months were regressed on baseline weight Z-score, giving regression coefficients for $Z(\text{base})$ of 0.82, 0.70, 0.62, 0.63 and 0.60 at 3, 6, 9, 12 and 18 months, and corresponding thrive index (TI) equations:

$$\begin{aligned} TI(3 \text{ mths}) &= Z(3 \text{ mths}) - 0.82Z(\text{base}) \\ TI(6 \text{ mths}) &= Z(6 \text{ mths}) - 0.70Z(\text{base}) \\ TI(9 \text{ mths}) &= Z(9 \text{ mths}) - 0.62Z(\text{base}) \\ TI(12 \text{ mths}) &= Z(12 \text{ mths}) - 0.63Z(\text{base}) \\ TI(18 \text{ mths}) &= Z(18 \text{ mths}) - 0.60Z(\text{base}) \end{aligned}$$

Thrive indices for each infant were then calculated, and the 5% threshold was derived empirically from the thrive indices at grouped ages of 3, 6, 9, 12 and 18 months. This gave cut-offs of -0.95 at 3 months, -1.19 at 6 months, -1.36 at 9 months, -1.33 at 12 months and -1.46 at 18 months. The constant in the regression equation was not included in the calculation of the thrive index, the effect of which is that the mean thrive index will not be zero but close to the constant in the regression equation.

A 5% threshold for each thrive index was chosen for two reasons (Corbett 1998). Firstly, earlier population based work in Newcastle (Corbett 1994) suggested that falls of this magnitude were associated with significant cognitive deficits, and secondly that the threshold observes recommended convention (Drotar (1990) in Corbett (1998)).

An infant was classified as a case if its 'thrive index' value was below the 5% threshold in 2 or more of the age bands, between 3 and 18 months.

> The requirement that the thrive index be below the 5th centile on two or more occasions ensured that the weight faltering persisted over at least 3 months for a child to be screened as a case, and ensured that a child could not be identified as failing to thrive on the basis of a single erroneously recorded weight (Drewett et al. 1999, pp 553).

This approach identified 136 cases, a high risk group that in practice would merit further clinical investigations, rather than a definite diagnostic group (Wright 2000).

### 4.6.2 Regression analysis after correction to birth Z-score

The above regression analysis was repeated as there are now 3415 term infants and all the birth weight Z-scores in the original analysis were slightly out because of a numerical error in the FORTRAN code used by Dr C.M. Wright. As the baseline Z-score is based on up to 3 early weights the impact of this was expected to be small. The baseline Z-score for each individual was obtained by taking the average of the weight Z-score at birth, 1 month and 2 months, if present. Summary statistics for the baseline Z-scores by sex can be found in table 4.9. One hundred and seventy (5%) individuals (90 boys, 80 girls) had no baseline weight Z-score. Table 4.10 cross-tabulates the number of early weights against number of late weights. Only the individuals highlighted in **bold** within this table could be identified as cases using the above approach as we need at least two late weights that are less than the 5% threshold.

The Z-scores at 3 months, 6 months, 9 months, 12 months and 18 months were then regressed on this baseline Z-score. The results of these regressions with an intercept can be found in table 4.11. The intercept term in all of these regressions is significantly different from zero, which means that when we use equations (4.1), given below, to obtain TI's, the mean $TI(j)$ ($j$=3, 6, 9, 12 and 18 months) will be non-zero.

Table 4.9: Summary statistics of baseline Z-scores for weight by sex

| Z(base) | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| Boys | -3.9390 | -0.7581 | -0.1801 | -0.1739 | 0.4587 | 2.6080 | 0.9308 | 1621 |
| Girls | -3.9900 | -0.8435 | -0.2118 | -0.2407 | 0.3832 | 2.9770 | 0.9726 | 1624 |

Table 4.10: Table of number of early versus late weight measurements, where early is birth and groupings of 1 and 2 months, and late is groupings of 3, 6, 9, 12 and 18 months. Figures highlighted in **bold** have enough weights to be detected as cases if two or more late weights fall below the 5% threshold, those highlighted in *italics* don't.

| no. early weights | no. late weights | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| 0 | *90* | *43* | *27* | *3* | *7* | *0* |
| 1 | *159* | *60* | **48** | **34** | **37** | **11** |
| 2 | *46* | *78* | **120** | **163** | **129** | **94** |
| 3 | *15* | *78* | **177** | **388** | **612** | **996** |

$$\begin{aligned} TI(3 \text{ mths}) &= Z(3 \text{ mths}) - 0.85Z(\text{base}) \\ TI(6 \text{ mths}) &= Z(6 \text{ mths}) - 0.73Z(\text{base}) \\ TI(9 \text{ mths}) &= Z(9 \text{ mths}) - 0.64Z(\text{base}) \\ TI(12 \text{ mths}) &= Z(12 \text{ mths}) - 0.65Z(\text{base}) \\ TI(18 \text{ mths}) &= Z(18 \text{ mths}) - 0.63Z(\text{base}) \end{aligned} \tag{4.1}$$

Table 4.12 contains the summary statistics and empirical 'cut-offs' for the calculated thrive indices. As discussed above, a child is defined as a case if that child's thrive indices at 3, 6, 9, 12 and 18 months were below the 5% thresholds (defined in the second from the right column in table 4.12) two or more times. Table 4.13 contains counts of the number of individuals that fell below the cut-offs $x$ amount of times. This approach identified 136 cases, the ID's of these children can be found in table 4.14.

The error in the birth weight Z-score has had little impact, again we have 136 cases, but now individuals with ID's 507, 1120, 1168 and 2780 are no longer cases by the protocol definition and these are replaced with individuals with ID's 863, 1889, 3192 and 3558.

Table 4.11: Regression of weight Z-score at grouped ages of 3, 6, 9, 12 and 18 months on baseline weight Z-score

Weight Z-score at 3 months

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.07027 | 0.01188 | 5.916 | $3.71 \times 10^{-9}$ |
| Z(base) | 0.85209 | 0.01243 | 68.555 | $< 2 \times 10^{-16}$ |

$R^2$=0.6337, $R^2$(adj)=0.6335, residual SE=0.6055 on 2717 df

Weight Z-score at 6 months

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.10223 | 0.01582 | 6.462 | $1.23 \times 10^{-10}$ |
| Z(base) | 0.73200 | 0.01649 | 44.382 | $< 2 \times 10^{-16}$ |

$R^2$=0.4269, $R^2$(adj)=0.4267, residual SE=0.794 on 2644 df

Weight Z-score at 9 months

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.13435 | 0.02038 | 6.593 | $5.46 \times 10^{-11}$ |
| Z(base) | 0.64312 | 0.02112 | 30.455 | $< 2 \times 10^{-16}$ |

$R^2$=0.3085, $R^2$(adj)=0.3082, residual SE=0.901 on 2079 df

Weight Z-score at 12 months

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.13961 | 0.01928 | 7.242 | $6.1 \times 10^{-13}$ |
| Z(base) | 0.65257 | 0.02012 | 32.436 | $< 2 \times 10^{-16}$ |

$R^2$=0.3275, $R^2$(adj)=0.3272, residual SE=0.8763 on 2160 df

Weight Z-score at 18 months

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.08610 | 0.02311 | 3.726 | 0.000201 |
| Z(base) | 0.62918 | 0.02381 | 26.428 | $< 2 \times 10^{-16}$ |

$R^2$=0.2951, $R^2$(adj)=0.2947, residual SE=0.9121 on 1668 df

## 4.7 A preliminary analysis of all the routine weight Z-scores

### 4.7.1 Comparing attained weights of Newcastle infants with the revised UK 1990 growth reference

The offset and linear trend of the Newcastle weight Z-scores for boys and girls was addressed in section 4.5.2. However, as mentioned earlier, infancy is a period of rapid growth and there is the possibility of a curvilinear trend. An appealing approach is to use a scatterplot smoother such as lowess (Cleveland 1979) to arrive at a pictorial

Table 4.12: Summary statistics and cut-offs for thrive indices at 5th and 10th percentile

| New TI | Min. | LQ | Median | Mean | UQ | Max. | NA's | 5% | 10% |
|---|---|---|---|---|---|---|---|---|---|
| 3 mths | -4.022 | -0.308 | 0.070 | 0.070 | 0.455 | 3.528 | 696 | -0.88 | -0.66 |
| 6 mths | -4.116 | -0.446 | 0.069 | 0.102 | 0.634 | 3.254 | 769 | -1.13 | -0.87 |
| 9 mths | -3.965 | -0.479 | 0.106 | 0.134 | 0.735 | 3.681 | 1334 | -1.31 | -1.00 |
| 12 mths | -2.706 | -0.468 | 0.137 | 0.139 | 0.724 | 3.173 | 1253 | -1.27 | -0.96 |
| 18 mths | -3.641 | -0.536 | 0.063 | 0.086 | 0.695 | 3.360 | 1745 | -1.41 | -1.08 |

Table 4.13: Frequencies for the number of times an individual fell below a cut-off

| no.($x$) | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Frequency | 3087 | 192 | 65 | 47 | 17 | 7 |

Table 4.14: Identification numbers of cases

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 53 | 348 | 576 | 973 | 1382 | 1764 | 1973 | 2260 | 2596 | 2824 | 3144 | 3366 | 3624 |
| 127 | 407 | 592 | 975 | 1412 | 1766 | 1983 | 2266 | 2638 | 2828 | 3184 | 3371 | 3633 |
| 149 | 419 | 631 | 1006 | 1432 | 1798 | 2030 | 2289 | 2657 | 2833 | 3192 | 3394 | 3662 |
| 165 | 423 | 633 | 1008 | 1575 | 1813 | 2046 | 2333 | 2680 | 2834 | 3204 | 3426 | 3679 |
| 175 | 450 | 688 | 1040 | 1599 | 1833 | 2128 | 2371 | 2687 | 2844 | 3272 | 3442 | |
| 193 | 451 | 720 | 1089 | 1609 | 1870 | 2132 | 2397 | 2695 | 2960 | 3288 | 3499 | |
| 203 | 471 | 863 | 1104 | 1614 | 1881 | 2139 | 2398 | 2705 | 2987 | 3289 | 3512 | |
| 269 | 494 | 884 | 1163 | 1629 | 1889 | 2153 | 2421 | 2717 | 3054 | 3290 | 3533 | |
| 301 | 535 | 918 | 1251 | 1674 | 1910 | 2157 | 2430 | 2726 | 3091 | 3325 | 3550 | |
| 339 | 559 | 956 | 1258 | 1731 | 1927 | 2188 | 2448 | 2786 | 3119 | 3339 | 3558 | |
| 345 | 561 | 967 | 1366 | 1761 | 1947 | 2255 | 2472 | 2792 | 3121 | 3352 | 3569 | |

representation of any curvilinear trend. This would allow us to compare the growth of Newcastle children to those infants that contribute to the growth reference. It would also permit us to assess the adequacy of the revised UK 1990 growth reference for Newcastle children.

Lowess is a procedure that uses robust locally linear fits. A window is placed at each point $x$. The data points that lie within this window are weighted so that nearby points get the most weight and a robust weighted regression is used to predict the value $y(x)$ at $x$ (Venables and Ripley 1997).

The full infancy data frame (3415 individuals with up to 11 weights) was utilised. A scatterplot of weight Z-score versus age for boys and girls can be found in figure 4.5. Before the age of 1 year a large proportion of children are weighed leading to a dense scatter in this region. Above 1 year markedly fewer children are weighed which results in a sparse scatter in this age region. A lowess curve[1] was added to both of these plots to see how children in Newcastle compare to the growth reference. The upper plots in figure 4.5 for boys indicate that at birth Newcastle boys tend to be lighter than the children that contribute to the reference, but from about 13 weeks the boys are slightly heavier than those children that contribute to the reference. First impressions of the same plots for girls indicate that girls in Newcastle are lighter than the reference at all ages.

Figures 4.6 and 4.7 contain scatterplots of weight Z-score versus age for boys and girls, respectively. These plots are all drawn on the same scale to aid comparisons. The upper panel in figure 4.6 show a scatterplot of weight Z-scores for case boys versus age with a lowess curve overlaid. This shows that the trajectory of weight Z-scores for case boys dips from birth to about -2 and attained weight status slowly improves but appears to be below zero at the end of infancy. The lower panel in figure 4.6 shows a scatterplot of weight Z-scores for all boys excluding cases versus age with a default lowess curve overlaid. This plot has a similar trend to the plot for all boys in figure 4.5. There is one boy with an outlying high negative weight Z-score (Z=-5.61) within the plot for boys excluding cases. This boy has ID 1773 and only has one recorded weight at 2.89 years.

The upper panel for case girls in figure 4.7 shows a similar trend to case boys. The lower panel in figure 4.7 of weight Z-score versus age for girls excluding cases, indicates that Newcastle girls are lighter at birth than those children that contribute to the reference. The apparent negative offset discussed above and in section 4.5.2 for girls is largely due to the case children dragging the mean Z-score of the population downwards in a negative direction. The downward trend towards the end of infancy in all of the plots in figures 4.6 and 4.7 reflects that weight data after 1 year is sparser and that infants weighed after this age tend to be lighter. Figure 4.8 looks at the effect on the lowess curve by varying the values of the span. The default

[1]The default span (f) is 2/3. The span controls the window size and is the proportion of the data which is included.

span is 2/3, with a span of 1/3 the lowess curves are less regular. However with a span of 1, some of the features are lost. The default span seems to be a reasonable compromise.

Figure 4.9 contains a summary of all the default lowess curves discussed above. The lines in red and blue correspond to the girls and boys, respectively. The fitted values from lowess at birth and 6 grouped ages can be found in table 4.15. The case boys and girls trajectory is almost identical with case boys falling from a slightly higher weight Z-score at birth and being weighed until slightly older ages. The pattern for girls (excluding cases) and boys (excluding cases) are fairly similar, with girls mean weight Z-scores tending to be just below the zero line and boys mean weight Z-scores just above the zero line. The effect of excluding the case children is to shift the trend curves in an upward direction, which leads to the impression that Newcastle boys on average seem to be heavier than the reference boys in infancy and Newcastle girls on average are slightly lighter than the reference girls.

The noticeable feature of the Newcastle infancy weight data is that the mean weight Z-score is below zero for both boys and girls at birth. So it seems reasonable to explore if allowing for gestational age would reduce this negative offset because we know the gestational age of 2516 infants. The weight Z-score at birth was obtained allowing for gestational age if known. The distribution of gestational ages can be found in table 4.16, the majority of infants have a gestational age between 38 and 41 weeks. If no gestational age is recorded then it was assumed that this was 40 weeks. The summary statistics for the birth weight Z-score allowing for gestational age can be found in table 4.17. Incorporating the gestational age increases the mean weight Z-score for both sexes, however the mean birth weight Z-score is still significantly different from zero; see table 4.18. Thus the negative offset at birth (and possibly in the early weeks) could be partly due to length of gestation. In table 4.19 we regress the birth weight Z-score allowing for gestation on gestational age. There appears to

Table 4.15: **Infancy data (excluding cases)** Fitted values from lowess of Z-scores for weight by sex at birth and ages: 6 weeks, 3, 6, 9, 12 and 18 months

| Age (days) | 0 | 42 | 91 | 182 | 351 | 364 | 546 |
|---|---|---|---|---|---|---|---|
| Boys | -0.237 | -0.110 | 0.027 | 0.103 | 0.135 | 0.134 | 0.097 |
| Girls | -0.266 | -0.208 | -0.142 | -0.073 | -0.022 | -0.024 | -0.048 |

be a negative trend in the birth weight Z-scores with gestational age, see figure 4.10. Thus suggesting that the revised UK 1990 reference may not adjust appropriately for age at birth. Gestational age only provides an indication of length of gestation and is recorded to the nearest week. In constructing the revised UK 1990 reference, the LMS values in the age range around birth are derived from splining prenatal, birth weight and postnatal data (personal communication, Dr C.M. Wright, April 2002). Therefore, weight Z-scores should be interpreted cautiously at birth and within the first month of life.

Table 4.16: **Infancy data**: Summary of gestational ages by sex

| Ges (weeks) | 37 | 38 | 39 | 40 | 41 | 42 | 43 |
|---|---|---|---|---|---|---|---|
| Boys | 68 | 216 | 249 | 480 | 211 | 77 | 3 |
| Girls | 76 | 210 | 288 | 491 | 186 | 71 | 5 |

Table 4.17: Summary statistics of birth weight Z-scores (after allowing for gestation) by sex

| | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| Boys | -3.6300 | -0.7786 | -0.1112 | -0.1191 | 0.5101 | 4.2230 | 0.9969 | 1588 |
| Girls | -3.8650 | -0.7655 | -0.1040 | -0.0952 | 0.5544 | 3.6340 | 1.0255 | 1580 |

Table 4.18: Birth weight Z-score (allowing for gestation): Results of testing that the mean is zero and variance is one (by sex)

| | t | p | 95% CI |
|---|---|---|---|
| Boys | -4.7614 | $2.098 \times 10^{-6}$ | [-0.1682, -0.0700] |
| Girls | -3.6883 | 0.0002 | [-0.1458, -0.0446] |
| | $\frac{(n-1)s^2}{\sigma^2}$ | approx. p-value | 95% CI |
| Boys | 1577.08 | 0.8603 | [0.9281, 1.0667] |
| Girls | 1660.68 | 0.1461 | [0.9821, 1.1291] |

Table 4.19: Slope coefficients from regression of birth weight Z-score on gestational age

| Z(Ges) | Estimate | Std. Error | t value | $Pr(>\|t\|)$ |
|---|---|---|---|---|
| Boys | -0.15919 | 0.02136 | -7.454 | $1.48 \times 10^{-13}$ |
| Girls | -0.09975 | 0.02227 | -4.480 | $8.00 \times 10^{-6}$ |

### 4.7.2 Trend curves of weight Z-scores by level of deprivation and number of weights

In an ideal world, displaying a connected plot of the weight Z-scores for each individual would be desirable. However there are 3415 individuals within the infancy data set so an effective display of all these growth curves is not feasible. An alternative approach is to produce scatterplots for subsets of the infancy data and add summary lowess curves. In the last section lowess curves were produced for boys, girls and case children. The boys and girls are at different levels on the Z-score scale, with mean Z-score of girls tending to be below zero and the mean Z-score for boys above zero. Hence if we were interested in the subgroups with different levels of deprivation, say, we would split these subgroups by sex.

Table 4.20 contains a summary of the number of boys and girls within each level of deprivation. The distribution of level of deprivation is similar for boys and girls ($\chi^2$=0.6044, p=0.7392) and the majority of individuals are within the intermediate category. Scatterplots of the weight Z-score versus age for the different levels of deprivation can be found in figure 4.11. Again these are all drawn on the same scale to aid comparison. Figure 4.12 contains a summary of the default lowess curves for affluent, intermediate and deprived children. The red and blue lines represent the girls and boys, respectively. The intermediate group for both sexes have a similar trend curves to those displayed in figure 4.5. At birth the affluent children tend to have higher weight Z-scores than the intermediate, and the intermediate have higher weight Z-scores than the deprived. The lowess curves for the girls exhibit the expected features: the deprived children have consistently lower weight Z-scores than the intermediate children, although there is some indication of catch up in attained weight status by the end of infancy. The trend curve for the affluent girls doesn't extend as far into infancy and seems to agree reasonably well with the growth reference, indicating that possibly affluent girls may be similar to the Cambridge

girls. The boys as a whole all overtake the reference in early infancy and attained weight status for both affluent and deprived children is higher than intermediate up until the age of 1 year. Its at the end of infancy that level of deprivation appears to influence attained weight status for boys. There is a downward trend for affluent and deprived boys at the end of infancy, although how much weight this observation carries is debatable as the data is sparse in this region. The unexpected steep downward trend for affluent boys appears to be largely due to two individuals with weights at just under 3 years (these individuals are not cases).

Table 4.21 contains a summary of the number of boys and girls with $y$ routine weights in infancy. The distribution of number of routine weights is similar for boys and girls ($\chi^2$=11.2983, p=0.4186). Figure 4.13 contains the default lowess curves of weight Z-scores split according to number of routine weights and by sex. We may have expected that infants weighed more frequently may be lighter or show a downward trend towards the end of infancy. The upper panel represents individuals with 2 to 5 weights, there are a lot of fluctuations within these trend curves and no consistent pattern. The lower panel represents individuals with 7 or more weights, the patterns within these plots are more regular and the majority of curves exhibit a downward trend towards the end of infancy, with the exception that girls with 10 or 11 routine weights and boys with 9 routine weights exhibit an upward trend towards the end of infancy. In particular, the girls with 10 or more weights exhibit the opposite behaviour to what we'd expect.

## 4.8 Discussion and Conclusions

The routine Newcastle infancy weight data were converted to weight Z-scores using the revised UK 1990 growth reference (Freeman et al. 1995; Preece et al. 1996; Cole et al. 1998). A growth reference is required to say anything about the attained growth status of a child. An advantage of using a growth reference is that the growth

Table 4.20: Summary table of level of deprivation by sex

| SES | 1 | 2 | 3 |
|---|---|---|---|
| Boys | 188 | 1189 | 334 |
| Girls | 191 | 1164 | 349 |

Table 4.21: Counts of the number of routine weight measures for each individual by sex

| No. of weights ($y$) | male | female | Total |
|---|---|---|---|
| None | 41 | 42 | 83 |
| 1 | 102 | 91 | 193 |
| 2 | 67 | 66 | 133 |
| 3 | 70 | 72 | 142 |
| 4 | 91 | 85 | 176 |
| 5 | 111 | 124 | 235 |
| 6 | 116 | 127 | 243 |
| 7 | 161 | 128 | 289 |
| 8 | 161 | 147 | 308 |
| 9 | 199 | 177 | 376 |
| 10 | 201 | 200 | 401 |
| 11 | 391 | 445 | 836 |

status of children can be compared even if the children are not weighed at the same age because a growth reference should adequately account for the age of the child. It is important to be aware of the impact of using an inappropriate growth reference. As demonstrated by Wright et al. (1993) use of the Tanner-Whitehouse reference (Tanner et al. 1966) would lead to the impression that after 3 months the majority of Newcastle childrens growth would appear to falter. The original UK 1990 growth reference (Freeman et al. 1995) was found to have a sex bias when applied to the routine Newcastle weight data (Wright et al. 1996), with an excess of girls with weight Z-scores below the third centile. A similar observation was made for the weight Z-scores of female Sheffield infants (Wales 1996). It is now claimed that there is now no longer a sex bias in the revised UK 1990 growth reference (Preece et al. 1996; Cole et al. 1998).

Initial preliminary analysis indicated that the sex bias towards girls appeared to be still present in the revised UK 1990 reference. However, the mean weight Z-score of girls is about -0.2 throughout infancy but this is not of clinical significance[2]. It would appear that the 84 case girls exert a downward pull on the mean weight Z-score of the Newcastle girls as a whole, leading to an exaggeration in the negative offset. Newcastle children are significantly lighter at birth than children that contribute to the reference even after allowing for gestational age if known. In general, for

[2]The general consensus is that if the difference was more than half a centile space (i.e. mean Z-score above 0.335 or below -0.335) then this would be of clinical significance (personal communication, Dr C.M. Wright, April 2002).

Newcastle infants, the girls tend to be slightly lighter and the boys slightly heavier than the reference children. There is a negative linear trend with age if the birth weight Z-score is obtained allowing for gestational age. There is no strong linear trend of the weight Z-scores with age after the age of 6 weeks, indicating that the reference adequately adjusts for age. However, there is some evidence of a curvilinear trend but this is in part expected because the Newcastle weights are routine data, so infants that are lighter are weighed over a longer time period leading to a downward trend at the end of infancy.

In infancy the revised UK 1990 reference is derived mainly from infants that belong to the Cambridge growth study (Whitehead et al. 1989b). The Cambridge data is likely to include a high proportion of educated mothers as Cambridge is a University City and these children are likely to belong to affluent families. The rationale for using the Cambridge data is that a high proportion of children are breast fed and weaned according to current guidelines. In Newcastle a high proportion of children are bottle-fed which could explain why the boys appear to be slightly heavier than the reference children. It is well known that breast and bottle fed infants show different growth patterns (Dewey et al. 1992). An interesting feature of the data for Newcastle girls was that the more deprived children had lower weight Z-scores and the affluent girls followed a trend that was similar to the growth reference. The effect of deprivation seemed to be less problematic for boys in terms of attained weight status, only at birth were more deprived boys lighter. It appears that affluent and deprived boys both tend to be heavier than intermediate boys from about 12 weeks until 1 year.

Overall, it is not entirely unreasonable to use the revised UK 1990 reference to convert Newcastle weight data to Z-scores. The revised UK 1990 growth reference seems to adequately account for age and it appears that girls tend to be slightly lighter and boys slightly heavier than the reference children. The biases described in this chapter may affect children in the tails of the distribution. However, the case children show a downward trend from birth whereas the rest of the children show an upward trend. As long as we are aware of these patterns there are no problems with using the revised UK 1990 growth reference until the age of 1 year. This is an agreement with Savage et al. (1999) findings. A sample of 127 *healthy* children from the Glasgow infant growth study were assessed monthly to 6 months, then at 9, 12, 18 and 24 months. Savage et al. (1999) found small differences for weight between the revised UK 1990 reference and their data, but viewed these as clinically not important. They also found a small excess of girls (but a deficit of boys) below the tenth centile. The mean weight Z-scores in the Glasgow sample follow a similar

pattern to the Newcastle data, boys and girls mean weight Z-score tending to be above and below the zero line, respectively. The mean Z-scores of Glasgow infants look marginally higher than the Newcastle infants, in particular at birth and at the end of infancy.

The approach used in the Newcastle research study to identify case children at risk of failure-to-thrive selects a subset of children with a markedly different growth pattern to the rest of the cohort. The trajectory of weight Z-scores for case children dips from birth to about -2 and attained weight status slowly improves but is still below zero at the end of infancy. The growth pattern of case boys and girls is remarkably similar. A later cohort of 229 children in Newcastle (Wright et al. 1998) were identified as failing to thrive using the 'thrive index' approach by contrasting an early and late weight. 120 of these children were within intervention practices and 109 in control practices.

> Both groups, however, reached the same lowest points at similar ages, suggesting that this represented the same underlying growth trajectory (Wright et al. 1998, pp572)

Future work could focus on the characterisation of this growth trajectory, which in itself could possibly lead to earlier detection of growth faltering. Identification of the age when children are more at risk of failing-to-thrive would be clinically valuable. The approach used for identifying potential cases, the number of times the child falls below the threshold, could also provide a crude index of severity and this may merit further research. For example, is the number of times the thrive index falls below the threshold related in any way to severity of future outcome; that is, does a TI less than the threshold 5 times lead to more severe stunting or wasting in childhood. Other research has shown that the severity and duration of failure-to-thrive are important in terms of future outcome (Woolston 1991). Another possibility would be to incorporate errors detailed in section 4.2 into the analysis, as a child's weight gain will encompass both real weight gain, biological variation, change in clothing and measurement error.

As shown in the preliminary work within this chapter, the patterns of growth in infancy show some interesting features. Past research has avoided working extensively with weights in infancy as they are harder to characterise, Cole (1995) and Wright et al. (1994) being the exceptions.

> Is it not high time that methods that have been applied to the height-growth curve were extended and modified to cope with

> more awkward measurements such as weight and skinfold thickness? (Healy 1989a, pp20)

is equally applicable today, especially beyond infancy. Does canalisation, the power to stabilise and return to predetermined growth curve after being pushed off trajectory, occur in infancy. Further work could look at issues of variability and movement from higher to lower centiles or vice-versa. Is it possible that more variability is observed for children growing 'abnormally'? In the Cambridge infancy study, they observed shifts rarely greater than 0.3 on Z-score scale in either direction. In early infancy the Newcastle weight data are likely to show shifts of this magnitude because all the children growing well are catching up with the reference children at this stage. So future work could concentrate on weight Z-scores after the age of 3 months.

To conclude, it seems appropriate to convert weights of infants from Newcastle to weight Z-scores using the revised UK 1990 reference. In general, female infants from Newcastle tend to be slightly lighter than those children that contribute to the reference. Whereas male infants from Newcastle tend to be slightly heavier than those children that contribute to the reference. The variance of weight Z-scores in late infancy appears to be slightly greater than one. Therefore, in late infancy we should be cautious in the interpretation of weight Z-scores in the tails of the distribution. The Lowess procedure of Cleveland (1979) was found to be a valuable tool in assessing trends within scatterplots of the weight Z-scores. It appears that infants that fail-to-thrive in infancy have a characteristic growth curve.

Figure 4.1: **Upper panel** Histogram and quantile-quantile plot of weight Z-scores at birth for boys and girls, on left and right, respectively. **Lower panel** Histogram and quantile-quantile plot of weight Z-scores at 6 weeks for boys and girls, on left and right, respectively

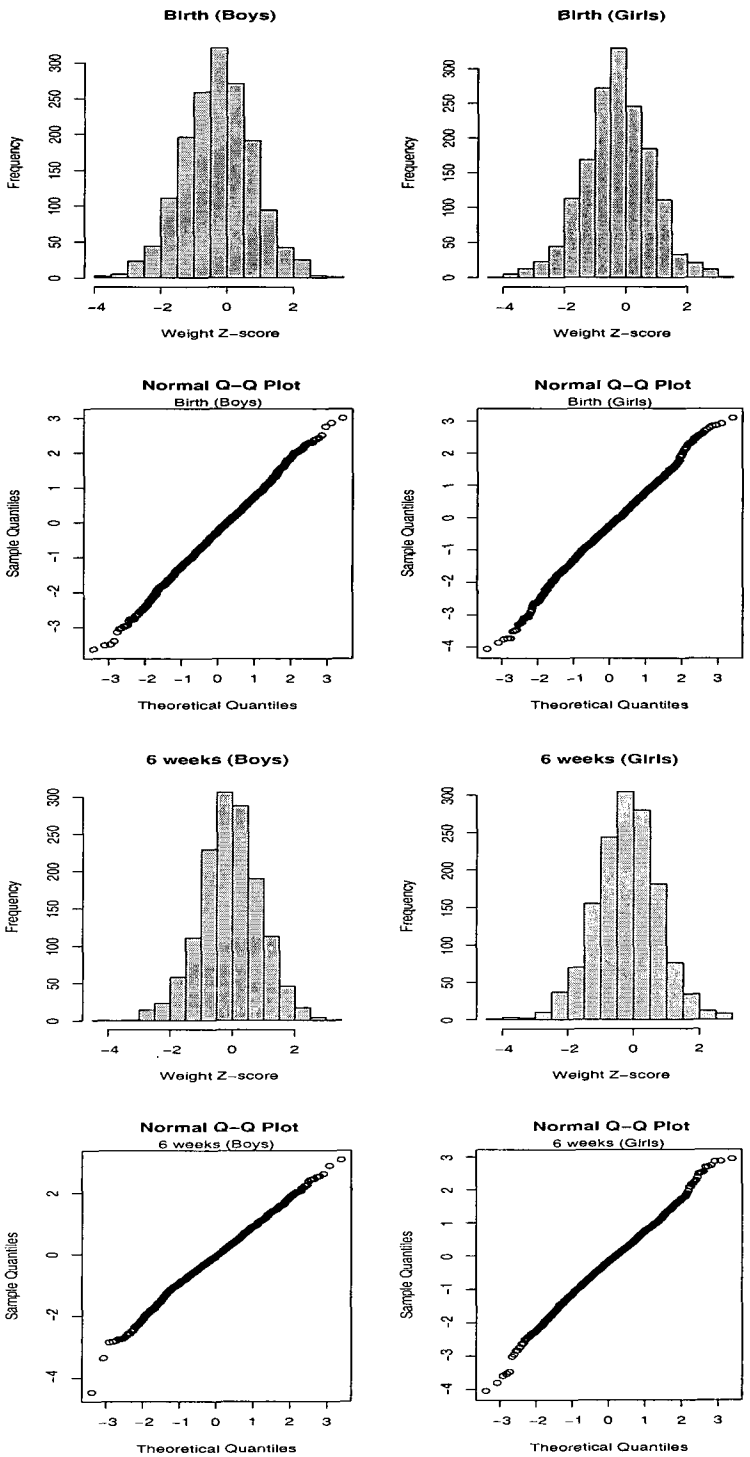


Figure 4.2: **Upper panel** Histogram and quantile-quantile plot of weight Z-scores at 3 months for boys and girls, on left and right, respectively. **Lower panel** Histogram and quantile-quantile plot of weight Z-scores at 6 months for boys and girls, on left and right, respectively

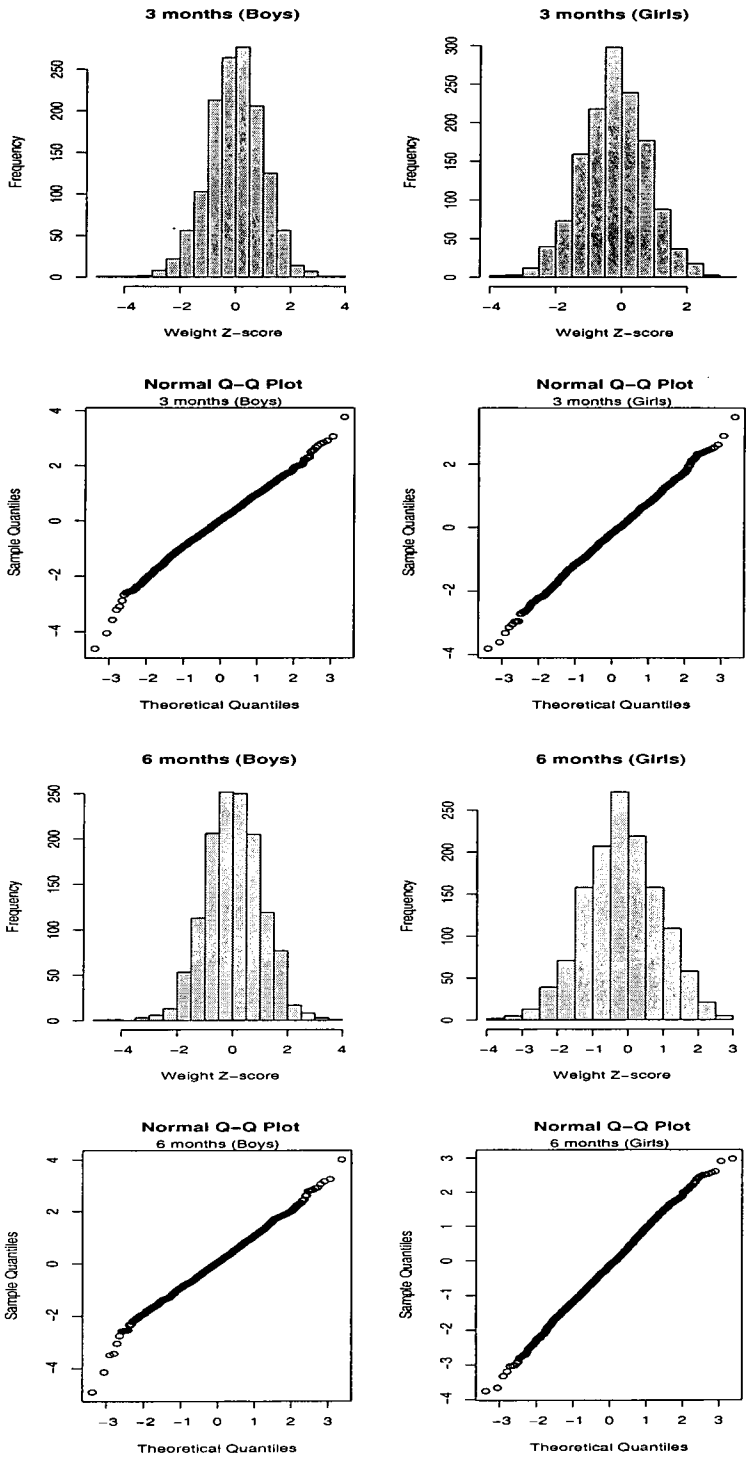


Figure 4.3: **Upper panel** Histogram and quantile-quantile plot of weight Z-scores at 9 months for boys and girls, on left and right, respectively. **Lower panel** Histogram and quantile-quantile plot of weight Z-scores at 12 months for boys and girls, on left and right, respectively.

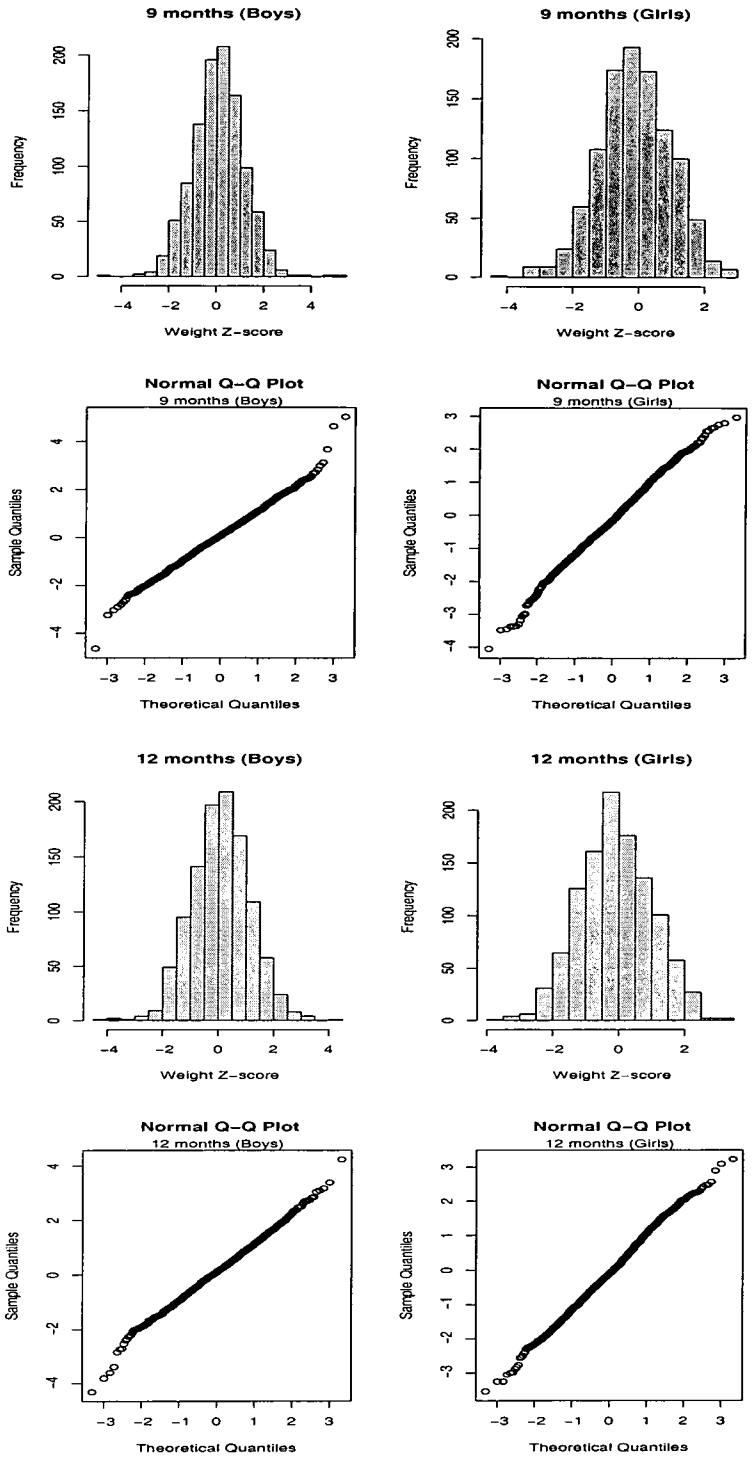


Figure 4.4: Histogram and quantile-quantile plot of weight Z-scores at 18 months for boys and girls, on left and right, respectively

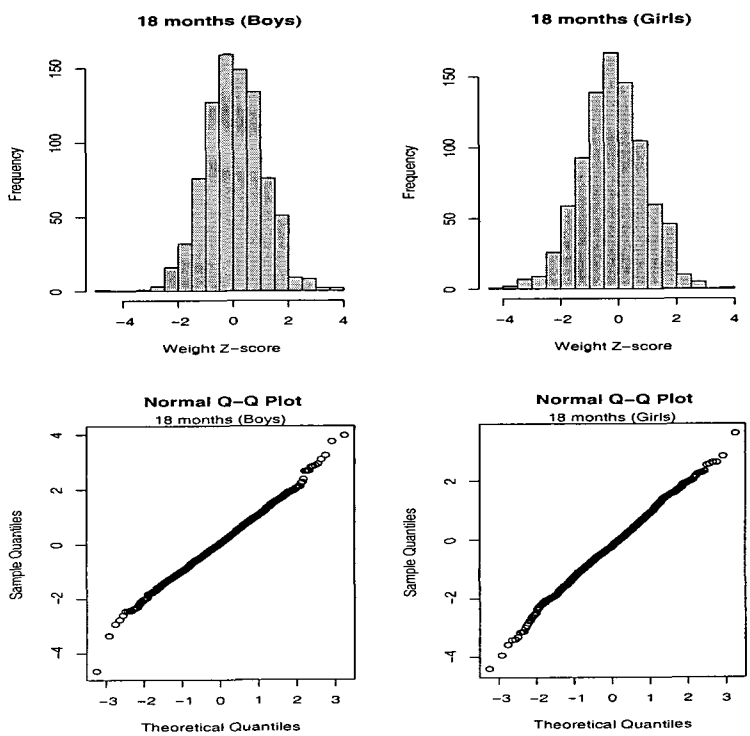


Figure 4.5: **Plots of weight Z-score versus age** Upper panel Plot of weight Z-score versus age for all boys with default lowess curve. Lower panel Plot of weight Z-score versus age for all girls with default lowess curve.

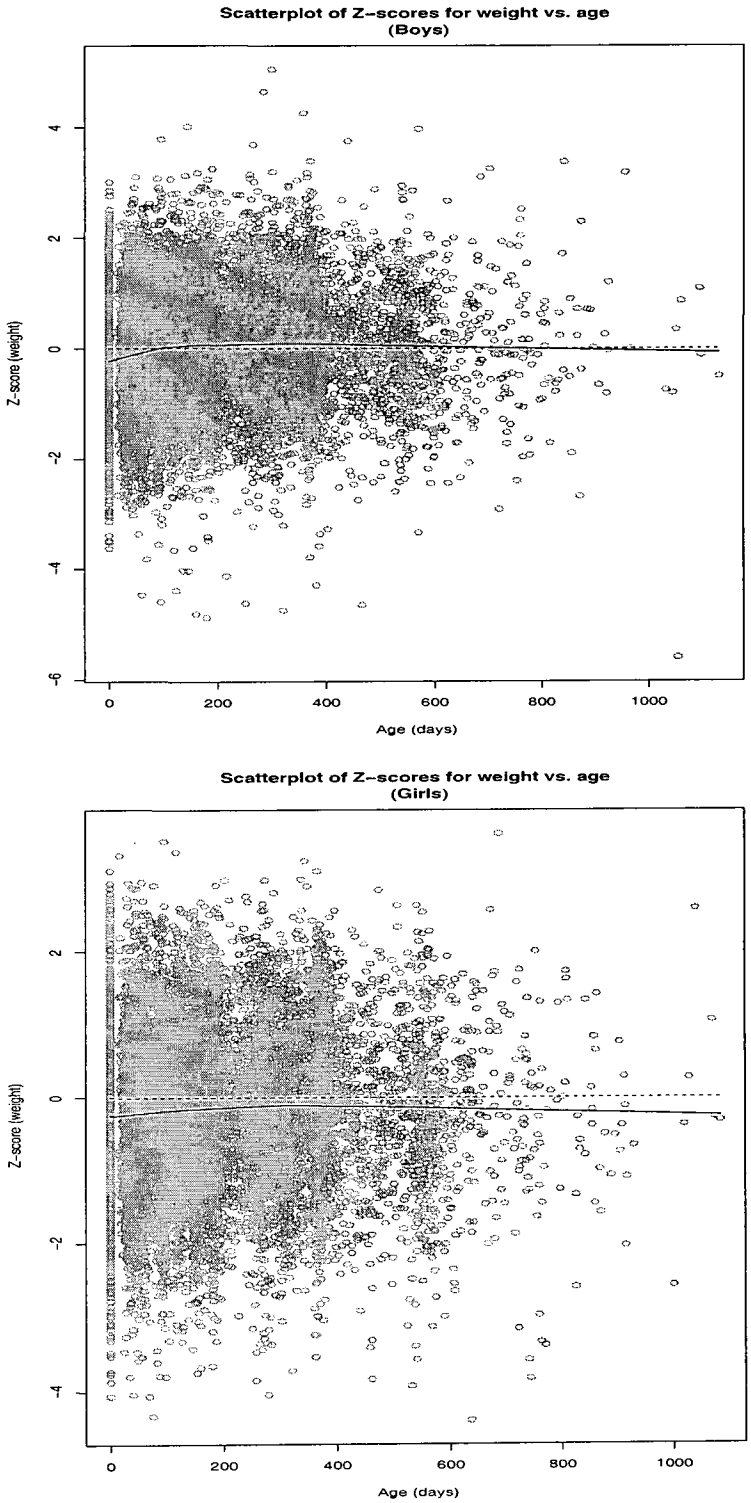


Figure 4.6: **Boys** Upper panel Plot of weight Z-score versus age for case boys with default lowess curve. Lower panel Plot of weight Z-score versus age for all boys excluding cases with default lowess curve.

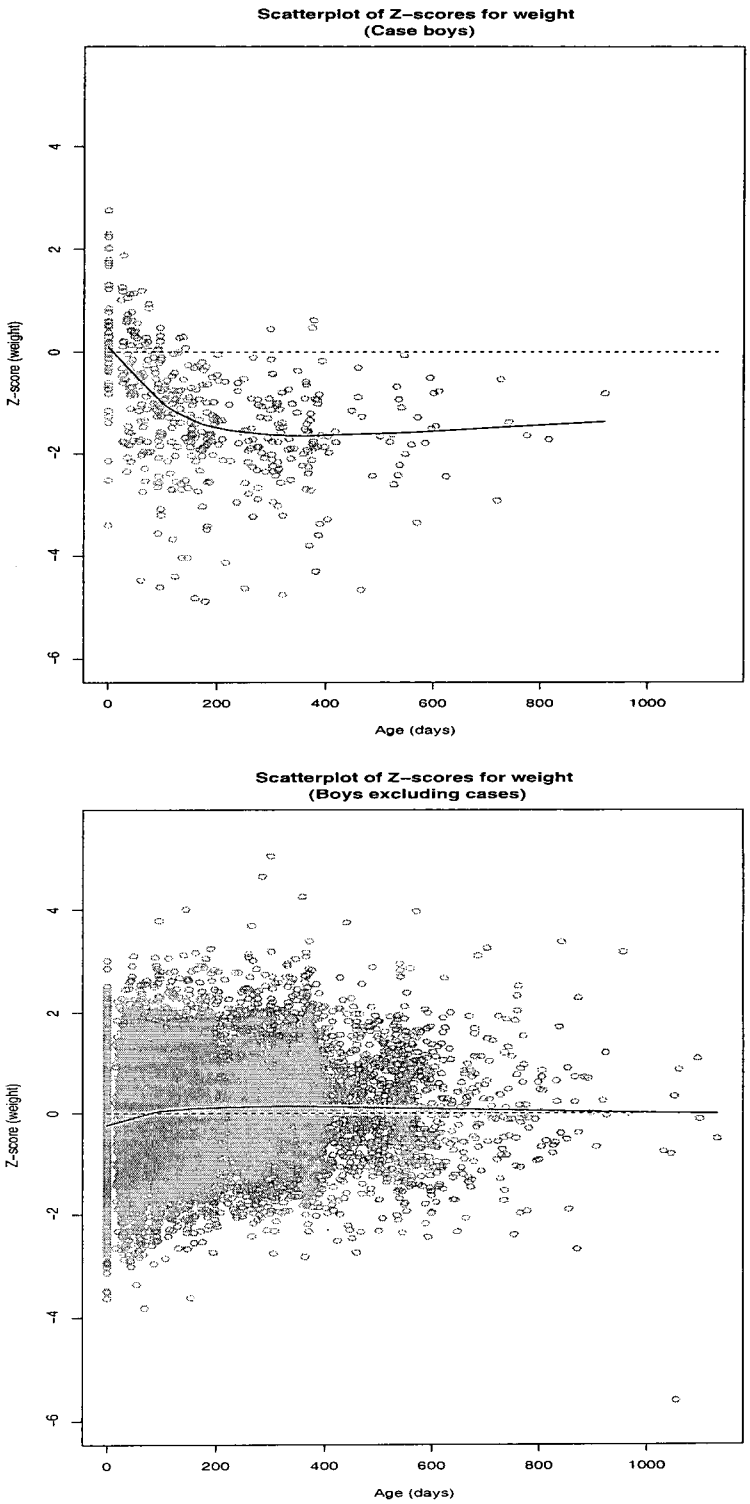


Figure 4.7: **Girls** Upper panel Plot of weight Z-score versus age for case girls with default lowess curve. Lower panel Plot of weight Z-score versus age for all girls excluding cases with default lowess curve.

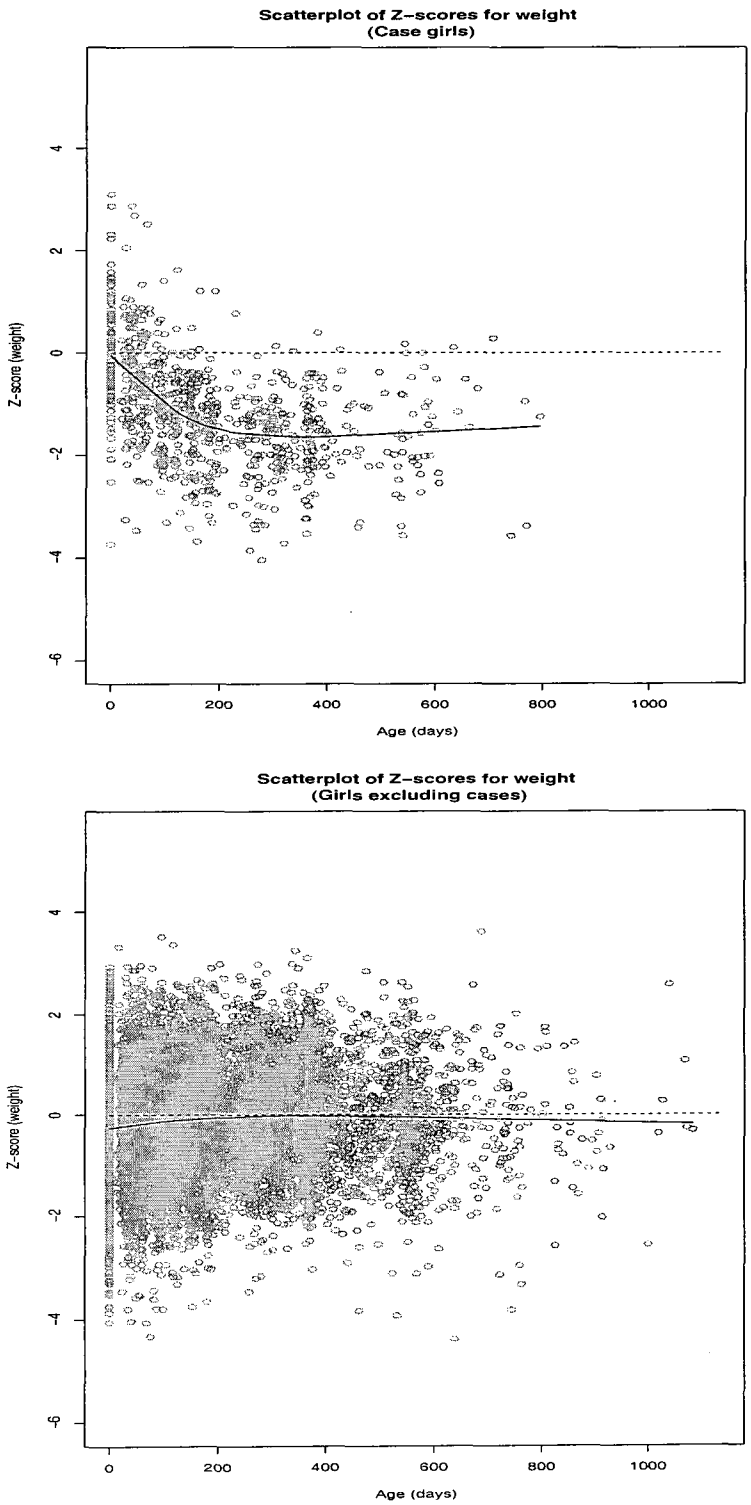


Figure 4.8: **Lowess curves** Upper panel Lowess curve for case and other boys with varying span. Lower panel Lowess curve for case and other girls with varying span.

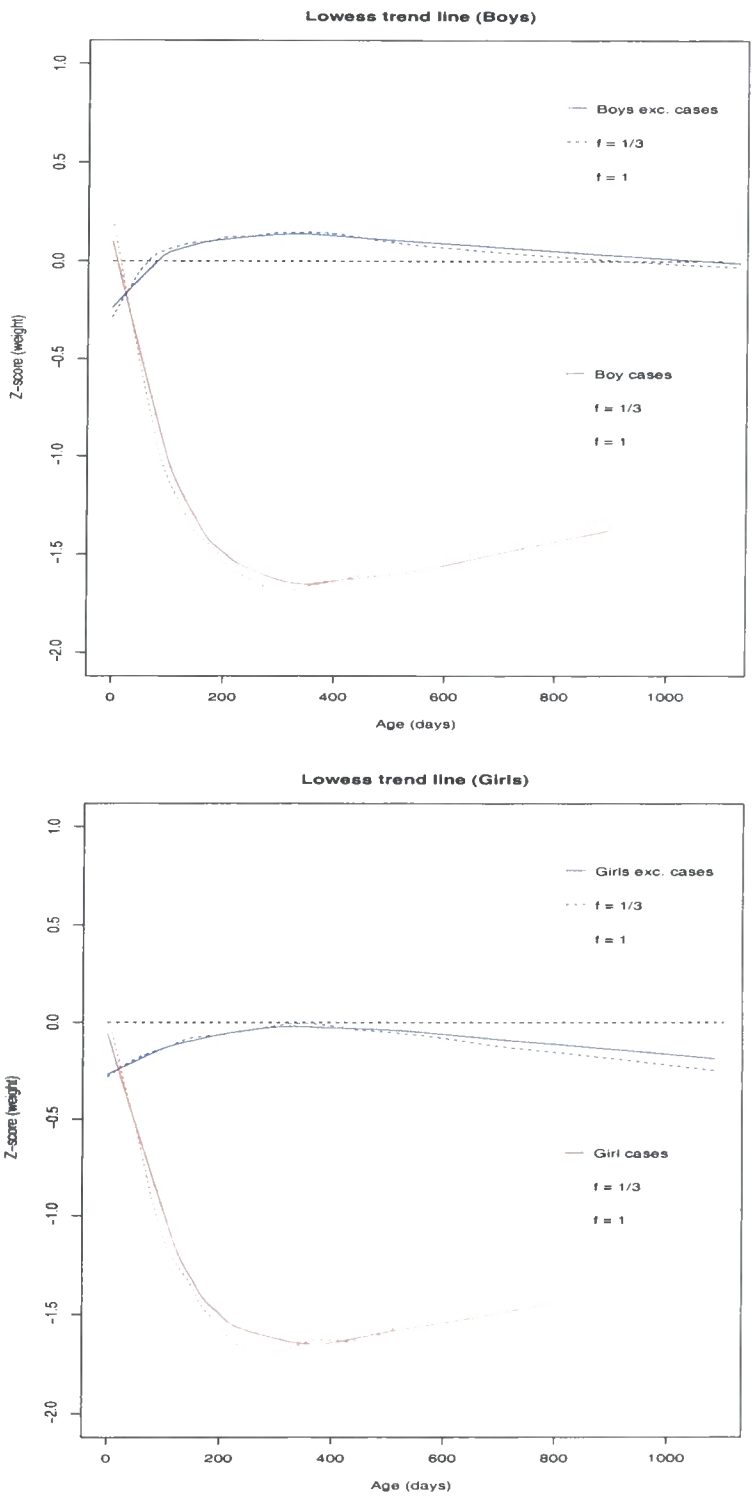


Figure 4.9: Default Lowess curves for case boys, case girls, boys excluding case boys and girls excluding case girls

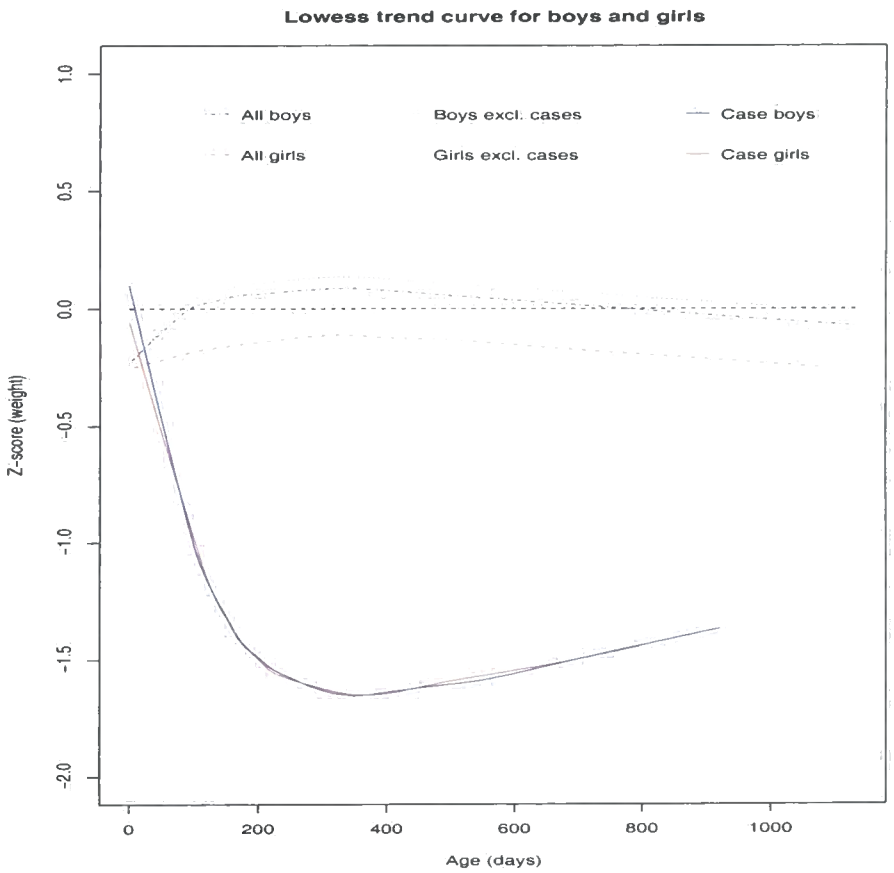


Figure 4.10: **Infancy data** Plot of birth weight Z-score (allowing for gestation) versus gestational age

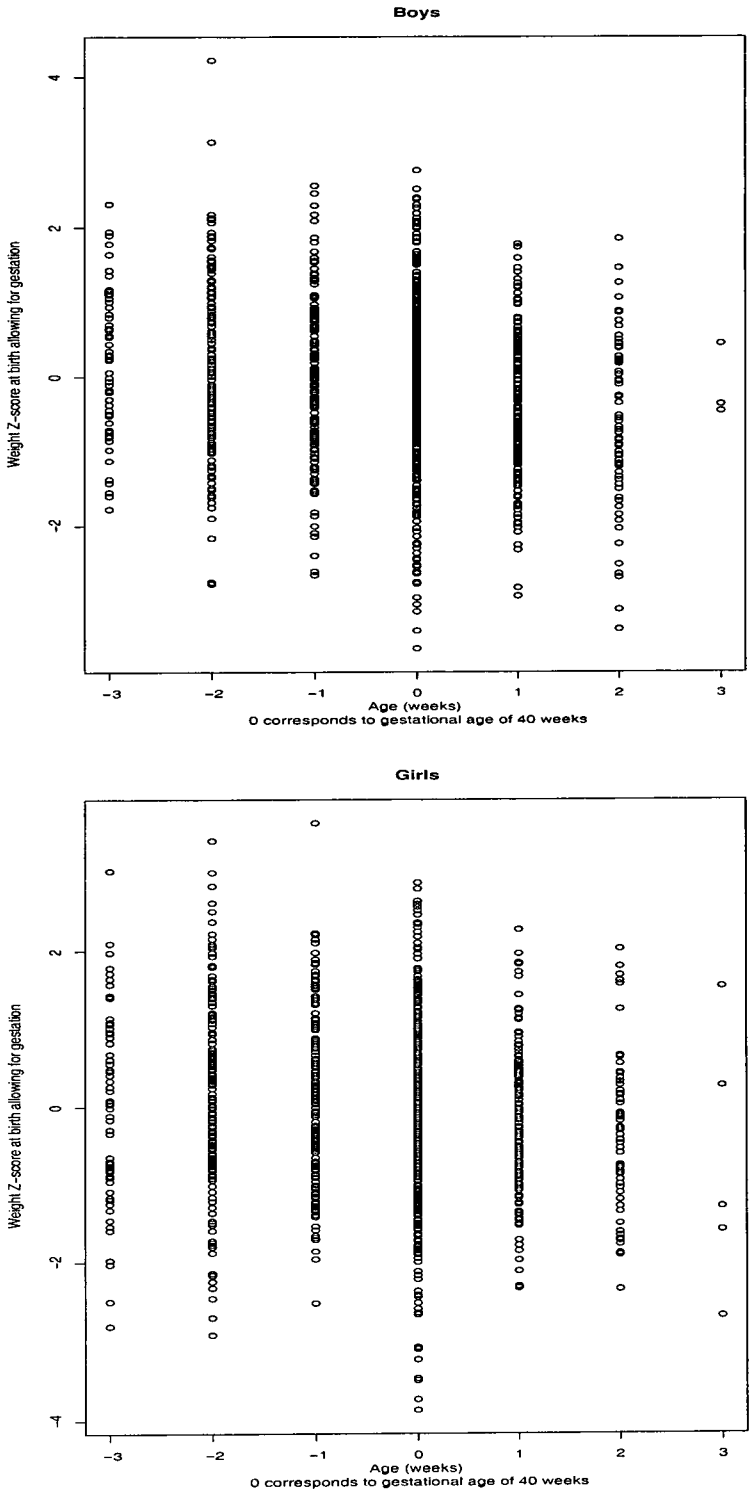


Figure 4.11: Left Scatterplots of weight Z-score versus age for boys by level of deprivation Right Scatterplots of weight Z-score versus age for girls by level of deprivation

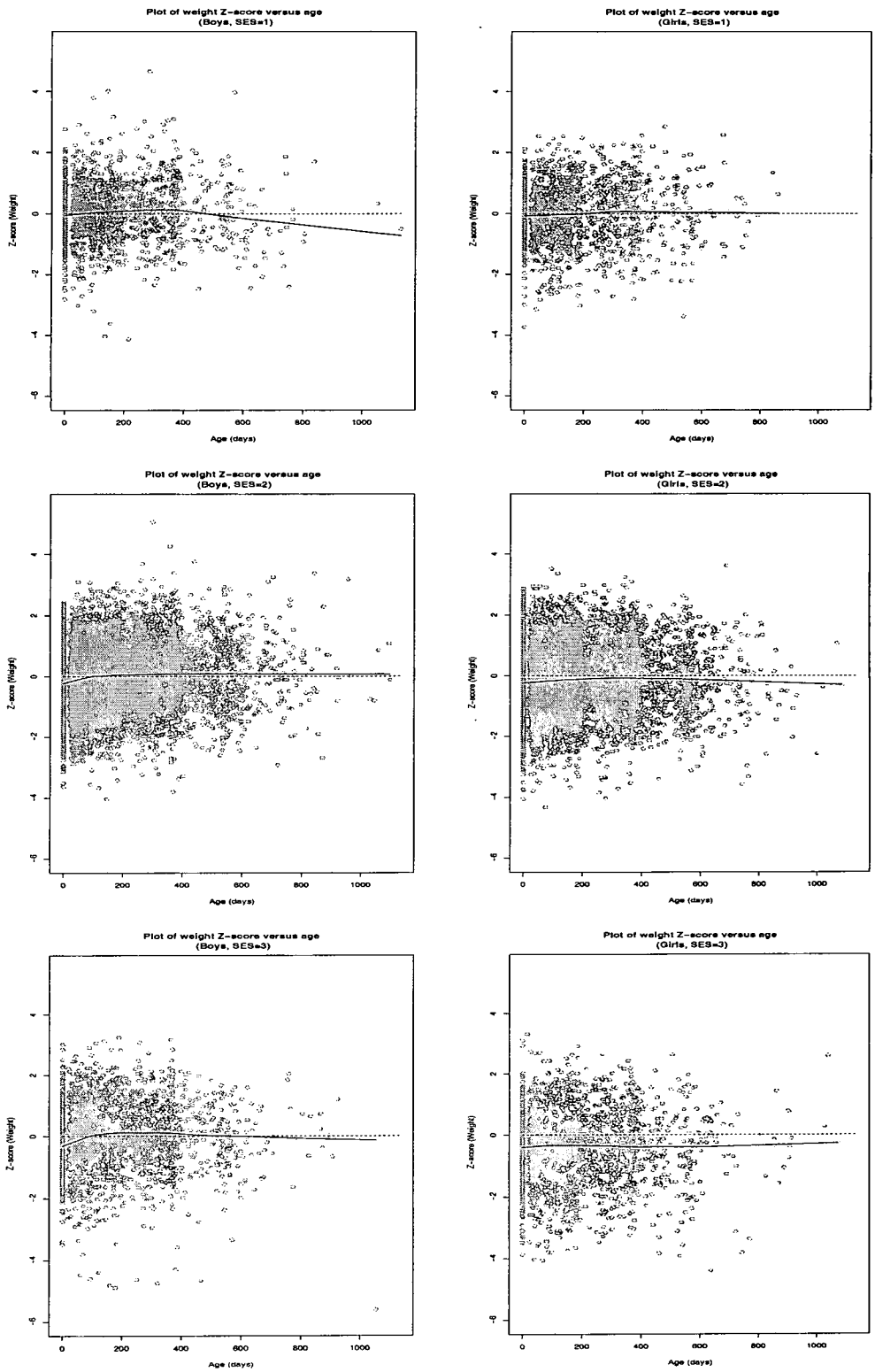


Figure 4.12: Default Lowess curves for affluent, intermediate and deprived boys and girls

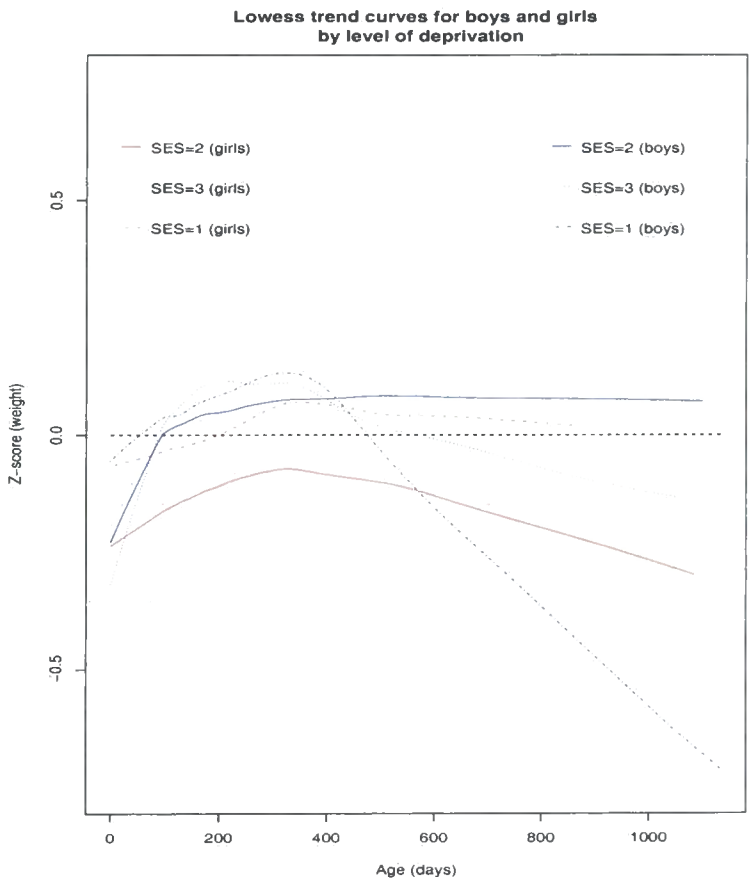
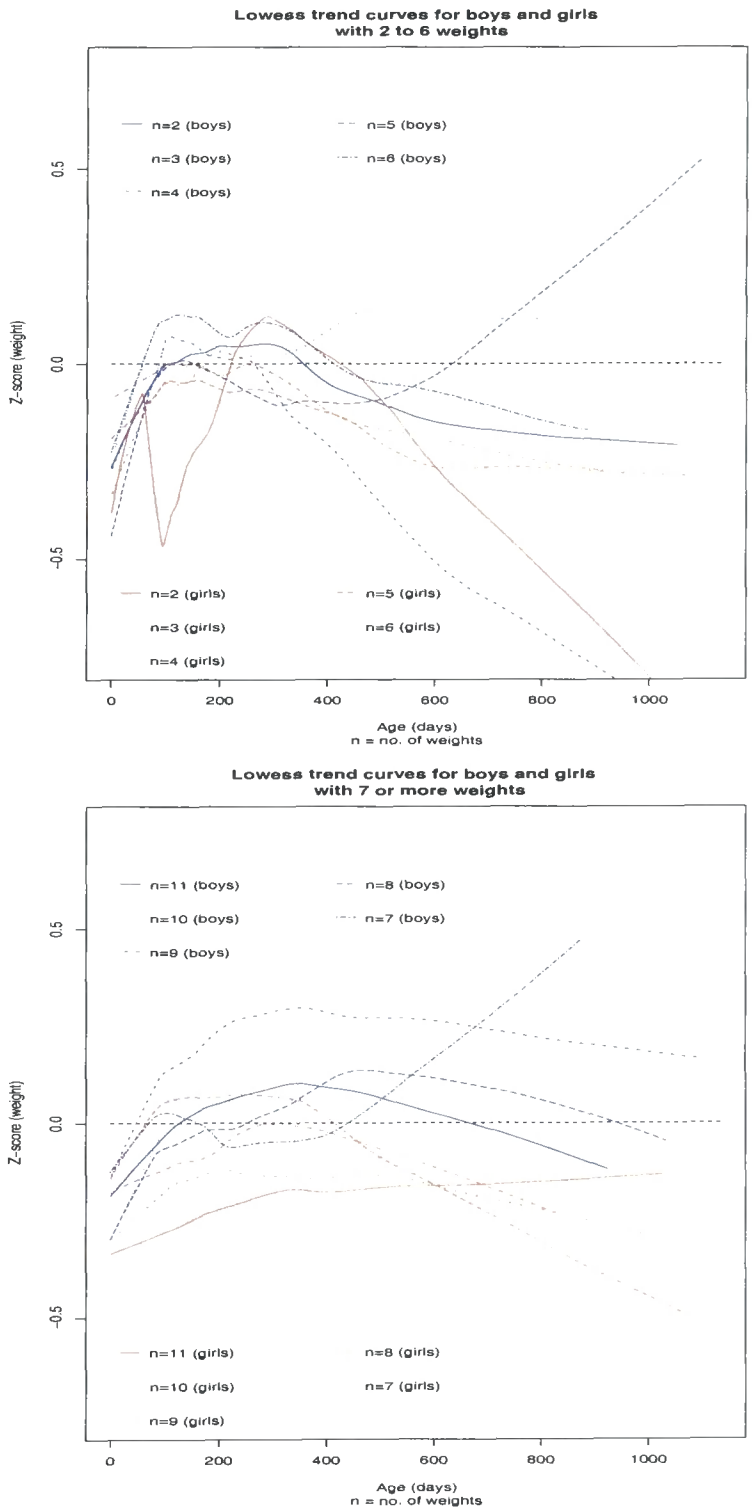


Figure 4.13: Default Lowess curves for number of routine weights by sex



# Chapter 5

# Correlation structure of Newcastle infancy weight Z-scores

The first two-thirds of this chapter concentrate on early exploratory work of the correlation matrix (based on 1055 children) for weight Z-scores at birth and the six grouped ages. It tells the story of the process of model development for the correlation structure. In section 5.1 the role of the correlation between weight Z-scores, in weight monitoring in infancy is described. In section 5.2, the correlation matrix for weight Z-scores at birth and the six grouped ages is presented and the scaled inverse correlation matrix is obtained. Within section 5.2, principal components analysis is carried out on the correlation matrix. In section 5.3 we explore the possibility that the correlation matrix has a pattern assuming that the weight measures are equally spaced. The developments in section 5.3 provide the inspiration for the model developed using regression techniques in section 5.4. The model derived in section 5.4 was initially from the correlations within the $(7 \times 7)$ correlation matrix. In section 5.5 the original routine weights are regrouped to provide finer detail of the correlation surface we wish to model. The model derived in section 5.4 is then fitted to the correlations generated for the weight Z-scores grouped to the nearest week or fortnight. In section 5.6 we explore whether the derived functional form is applicable to the correlation matrix derived from the weight Z-scores in the Cambridge infant study (Cole 1995). We conclude by applying the derived functional form to the combined Newcastle and Cambridge correlations; see section 5.7.

## 5.1 The role of the correlation matrix in monitoring weight gain in infancy

An infant's weight is usually monitored regularly in the first year of life and the attained weight is compared to a cross-sectional reference such as the revised UK 1990 reference (Freeman et al. 1995; Preece et al. 1996). Centile crossing is often interpreted by the clinician as a sign of abnormal growth under the premise that a child growing normally stays close to his or her weight centile. It is not appropriate to use growth references in this way, because they cannot quantify changes in weight as they are derived from cross-sectional data (Cole 1995). In addition, the definition of 'centile crossing' is somewhat arbitrary. For example, consider two children whose attained weight crosses two major centiles. One may have been just above the 50th centile and fallen just below the 25th centile, whereas the other child may have started initially nearer the 75th centile and fallen to a weight just above the 9th centile. Both children have crossed 2 major centiles, however we'd be more concerned about the second child's growth.

It is generally accepted that to detect factors causing weight faltering, weight velocity is a far more sensitive indicator than weight attained (Cole 1989a). Longitudinal data is required to create a weight velocity chart, consisting of mean and standard deviation (SD) of weight increments between specified ages. Both the mean and SD change with age, but in addition the SD depends on the time between measurements (Cole 1995). The measurement error is greater over shorter time intervals (Cole 1993). In addition, a saltatory pattern (Lampl et al. 1992) or weight gain of a pulsatile nature (Greco et al. 1990) would increase the variability over short time periods. However one of the main disadvantages in using weight velocity is that in order to use velocity charts, the child is required to be weighed at set times. For example, the USA increment weight tables for infancy are for 1 month intervals from birth to 6 months, 2 month intervals from birth to 12 months and 3 month intervals from birth to 24 months (Guo et al. 1991). The child's weight velocity also depends on the starting weight. Light children have a higher expected velocity than heavy children (Cole 1995). An alternative is to use a conditional reference, this is a weight velocity reference that compensates for regression to the mean (Cole 1995).

One of the questions to address when assessing a child's growth, is 'Given we know the infant's previous weight, what is his/her likely weight now?' Recall from Chapter 3, that one of the ways a conditional gain Z-score (equation (3.11)) can be

used is to evaluate a child's weight gain between two time-points.

$$Z_{t2|t1} = \frac{Z_{t2} - rZ_{t1}}{\sqrt{(1-r^2)}} \tag{3.11}$$

where $t1 < t2$, $Z_{t1}$ is the initial weight Z-score at $t1$, $Z_{t2}$ is the later weight Z-score at $t2$ and $r$ is the correlation between weight Z-scores at $t1$ and $t2$.

The only information needed to use equation (3.11) is a growth reference to convert weight measures to Z-scores and the correlation between the two weight Z-scores (Cole 1993). The required correlation will depend on the age and sex of the particular child (Cole 1993).

> ... it would be straightforward to construct a mathematical model for $r$, using existing longitudinal data, which would then predict the appropriate correlation for any pair of ages. (Cole 1993, pp35)

It would be advantageous if a simple model could be found for the correlation. The clinician would be able to convert both weights to Z-scores then obtain the correlation between these two weight Z-scores by just inputting the two measurement ages into the functional form. The two weight Z-scores can then be contrasted using equation (3.11) and a judgement made as to whether this child is growing 'normally' or not. Currently this facility does exist within the Child Growth Foundation (1996b) disk: a SDSGAIN function is available for Microsoft Excel users. This uses the functional form published by Cole (1995). This function is for use on weights at any 2 ages between 4 weeks and 2 years, and the later age should be at least 4 weeks greater than the initial age. This same function was revised by Cole (1998a) (see section 5.6.2) to incorporate birth weight, which was initially excluded because it was not obtained by the research study team.

## 5.2 Correlation matrix for infancy weight data

### 5.2.1 The correlation matrix

There are 3415 term infants within the birth cohort (3418 minus duplicate and two pre-terms discovered during follow-up study). The routine weights of each child were assigned by Dr C.M. Wright to target ages of 6 weeks; 3, 6, 9, 12 and 18 months. Table 5.1 contains the correlations observed between weights and weight Z-scores for 1055 children (503 boys and 552 girls) that had weights at birth and in all 6

Table 5.1: **Correlation matrix for 1055 individuals with weights in 7 age-groupings** (above diagonal - correlation between raw weights, below diagonal - correlation between weight Z-scores in **bold**)

| | birth | 6 wks | 3 mths | 6 mths | 9 mths | 12 mths | 18 mths |
|---|---|---|---|---|---|---|---|
| birth | 1.000 | 0.643 | 0.596 | 0.499 | 0.466 | 0.464 | 0.446 |
| 6 wks | **0.724** | 1.000 | 0.757 | 0.672 | 0.594 | 0.587 | 0.521 |
| 3 mths | **0.605** | **0.885** | 1.000 | 0.854 | 0.778 | 0.753 | 0.638 |
| 6 mths | **0.487** | **0.734** | **0.863** | 1.000 | 0.898 | 0.861 | 0.734 |
| 9 mths | **0.452** | **0.644** | **0.779** | **0.909** | 1.000 | 0.927 | 0.798 |
| 12 mths | **0.439** | **0.622** | **0.742** | **0.865** | **0.935** | 1.000 | 0.858 |
| 18 mths | **0.442** | **0.578** | **0.672** | **0.775** | **0.849** | **0.898** | 1.000 |

age groupings. As can be seen from table 5.1, in general the correlation between the weight Z-scores is higher than the observed correlation between the raw weight data. This correlation matrix also indicates that the correlation is higher between consecutive weight Z-scores i.e. weight Z-scores at birth and 6 weeks, 6 weeks and 3 months, 3 months and 6 months etc. The highest correlation is observed between weight Z-scores at 9 and 12 months. A slightly lower correlation is observed between weight Z-scores at 12 and 18 months because this correlation is over a longer time period of 6 months. Surprisingly the correlation between weight Z-scores at birth and 18 months is marginally higher than the correlation between weight Z-scores at birth and 12 months. The lowest correlations within this matrix tend to be between the weight Z-score for birth and later weight Z-scores (i.e. 6 weeks to 18 months), this is not entirely unexpected as it is usually recognised that birth weight is representative of weight gained in the uterus and therefore is largely dependent on the mother (Tanner 1989).

Wright et al. (1994) presented the pair-wise correlations within table 5.2 (excluding correlation with birth weight). In 1994 there was no UK 1990 growth reference: Wright et al. (1994) used the Cambridge growth reference (Whitehead et al. 1989b). However the correlations in table 5.2 agree with those published. The correlations highlighted in **bold** within table 5.2 were used to define lower limits for expected weight at 3, 6, 9, 12 and 18 months for boys and girls with weights on the major centiles (i.e. 3rd, 10th, 25th, 50th, 75th, 90th and 97th percentiles).

The correlation matrix in table 5.2 is included here to accompany the matrix of scatterplots in figure 5.1. In figure 5.1, within plots of later weight Z-score versus birth weight Z-score, the points are positively correlated but widely dispersed.

Within this matrix of scatterplots there are a handful of outlying individuals that have large positive or large negative Z-scores. The matrix of scatterplots illustrates the general feature that weight Z-scores taken close in time are highly correlated.

### 5.2.2 Inverse correlation matrix

The inverse of the correlation matrix within the lower diagonal of table 5.1 can be found in table 5.3. All the elements within table 5.3 are non-zero. If an element of the inverse correlation matrix was found to be zero then this would indicate that the corresponding variables were conditionally independent given the remaining variables (Whittaker 1990). Each diagonal element of the inverse correlation matrix is related to the proportion of the variation in the corresponding variable explained by regressing on the remaining variables. In fact, each diagonal element of the inverse correlation matrix equals $1/(1 - R^2)$ where $R$ is the multiple correlation coefficient between that variable and the rest. So for example, the proportion of explained variation for the weight Z-score at 18 months given the rest of the weight

Table 5.2: Infancy: Correlation matrix of weight Z-scores for pair-wise complete observations (sample sizes are in parenthesis)

| | birth | 6 wks | 3 mths | 6 mths | 9 mths | 12 mths | 18 mths |
|---|---|---|---|---|---|---|---|
| birth | 1.000 | *0.708* (2747) | *0.567* (2654) | *0.460* (2579) | *0.408* (2032) | *0.429* (2107) | *0.432* (1635) |
| 6 wks | | 1.000 | **0.868** (2547) | **0.716** (2476) | **0.604** (1961) | **0.611** (2037) | **0.568** (1551) |
| 3 mths | | | 1.000 | 0.869 (2473) | 0.761 (1953) | 0.733 (2029) | 0.672 (1542) |
| 6 mths | | | | 1.000 | 0.903 (1935) | 0.861 (2033) | 0.773 (1531) |
| 9 mths | | | | | 1.000 | 0.930 (1654) | 0.847 (1326) |
| 12 mths | | | | | | 1.000 | 0.900 (1380) |
| 18 mths | | | | | | | 1.000 |

Z-scores is:

$$R^2(\text{Z-score at 18 months; rest of Z-scores}) = (5.268 - 1)/5.268 = 0.810$$

i.e. 81.0%. The right hand column of table 5.3 contains the multiple correlation coefficient for the Z-score at each of the grouped ages given the rest. Therefore the Z-score at 9 months is the most predictable and the Z-score at birth is the least. Here we are assuming that we have all the other variables, but in practice would be difficult to predict weight Z-score at 3 months from the weight Z-score at 18 months.

Table 5.4 contains the inverse correlation matrix in table 5.3, scaled so that there are units along the diagonal. The off diagonal elements of this scaled inverse correlation matrix are the negatives of the partial correlation coefficients between the corresponding pair of variables given the remaining variables (Whittaker 1990). The notable feature is that all the correlations are near zero, except the leading off-diagonal where the correlation between consecutive weight Z-scores persists. This suggests that the original correlation matrix has a near Markov correlation structure.

### 5.2.3 Principal components analysis of the correlation matrix

Principal components analysis is a dimension reduction technique for a large number of interdependent variables which tries to retain most of the variation in the original data in fewer dimensions. Principal components are uncorrelated linear combinations of the original variables. The first few principal components contain most of the variation in the original variables and are often useful for revealing the structure

Table 5.3: **Inverse correlation matrix** of correlation matrix for infancy weight Z-scores with weights in all 7 age groupings

| Z-score | birth | 6 wks | 3 mths | 6 mths | 9 mths | 12 mths | 18 mths | $R^2$ |
|---|---|---|---|---|---|---|---|---|
| birth | 2.170 | | | | | | | 0.539 |
| 6 wks | -1.876 | 6.369 | | | | | | 0.843 |
| 3 mths | 0.277 | -4.893 | 8.537 | | | | | 0.883 |
| 6 mths | 0.407 | -0.258 | -3.575 | 9.088 | | | | 0.890 |
| 9 mths | -0.280 | 1.009 | -0.646 | -4.821 | 11.866 | | | 0.916 |
| 12 mths | 0.214 | -0.150 | 0.146 | -1.042 | -6.370 | 11.655 | | 0.914 |
| 18 mths | -0.329 | -0.086 | 0.158 | 0.354 | -0.644 | -4.352 | 5.268 | 0.810 |

Table 5.4: Scaled inverse correlation matrix for weight Z-scores in infancy

| Z-score | birth | 6 wks | 3 mths | 6 mths | 9 mths | 12 mths | 18 mths |
|---|---|---|---|---|---|---|---|
| birth | 1.000 | | | | | | |
| 6 wks | -0.505 | 1.000 | | | | | |
| 3 mths | 0.064 | -0.664 | 1.000 | | | | |
| 6 mths | 0.092 | -0.034 | -0.406 | 1.000 | | | |
| 9 mths | -0.055 | 0.116 | -0.064 | -0.464 | 1.000 | | |
| 12 mths | 0.042 | -0.017 | 0.015 | -0.101 | -0.542 | 1.000 | |
| 18 mths | -0.097 | -0.015 | 0.024 | 0.051 | -0.081 | -0.555 | 1.000 |

in the data. In this instance the original variables are the weight Z-scores at birth and grouped ages: 6 weeks, 3, 6, 9, 12 and 18 months.

Letting $V$ represent the sample correlation matrix of a data matrix X $(p \times 1)$. Then $V$ is a $p \times p$ matrix that is symmetric and non-negative definite. The spectral decomposition of $V$ is given by equation (5.1).

$$V = U^T \Lambda U \tag{5.1}$$

where $U^T = (\underline{u}_1, \ldots, \underline{u}_p)$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$. So $U^T$ is a matrix whose columns are eigenvectors and $\Lambda$ is a diagonal matrix of eigenvalues.

Assuming $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ then the $j$th principal component is given by equation (5.2).

$$Y_j = \underline{u}_j^T X \tag{5.2}$$

and $Var(Y_j) = \lambda_j, (j = 1, \ldots, p)$ and $Cov(Y_j, Y_k) = 0, j \neq k$ (uncorrelated). Therefore $Y_1$ has the largest variance and $Y_p$ has the smallest variance. If the first two principal components explain the 'majority' of the variation then a scatterplot of the objects on these two dimensions will give a reasonable representation of the overall distribution of the data (Mardia et al. 1979).

The correlation matrix in table 5.1 (based on 1055 infants) was subjected to principal components analysis. A summary of the results can be found in table 5.5 and a scree plot of the variances of the principal components is produced in figure 5.2. The first principal component was found to be an 'average' of all weight Z-scores, possibly representing some measure of 'overall size'. The second principal component contrasted late weight Z-scores (6 to 18 months) with early weight Z-scores (birth to 3 months), this is in effect some measure of 'weight gain'. The remaining principal

components are less easy to interpret but only make a contribution of 11% to the variation explained. In figure 5.3 we plot the second principal component (PC2) versus the first principal component (PC1) labelled by case status. Only 63 cases had weights in all 7 age groupings, but all of these individuals are clustered in the bottom left hand corner of the upper panel in figure 5.3. This is not entirely surprising as the thrive index approach used to identify cases contrasts late weight Z-scores with an average of up to three weight Z-scores in the first two months. Plots of PC2 versus PC1 labelled by sex, SES and Ges can be found in the lower panel of figure 5.3, upper panel of figure 5.4 and lower panel of figure 5.4, respectively. The plot labelled by sex looks reasonably random and there is no evidence of clustering. In the plot labelled by level of deprivation, there is a preponderance of deprived children with high negative values on PC1 but there are some deprived children with high positive values on PC1. In the plot labelled by gestational ages, the ages were grouped according to whether gestational age was less than 40 weeks, 40 weeks or greater than 40 weeks. This was done as labelling by each gestational age overloaded the scatterplot with too much information. The gestational age was categorised in this way because 40 weeks is considered to be the expected gestational age and also the mode value for gestation. If a baby was born before 40 weeks we might expect that the infant might not have quite reached their full growth potential and babies born post 40 weeks may have had their growth restricted because of intrauterine environment (Tanner 1989). However around expected delivery date all babies could be subject to intrauterine growth restriction regardless of their gestational age. There are a preponderance of infants born at a gestational age of less than 40 weeks on the left and upper parts of the scatter, indicating possibly that growth may have been slowed by early than expected delivery but there was some degree of catch-up after delivery.

## 5.3 Covariance matrices with pattern

### 5.3.1 Proposed pattern for correlation matrix

Recall that, in section 5.2.2 we considered the scaled inverse correlation matrix and concluded from tables 5.3 and 5.4, that the original correlation matrix is a near Markov correlation matrix. So if we assume for the moment that the weight Z-scores are equally spaced. Then a correlation matrix with pattern (equation (5.3)) could be appropriate, as correlations for two weight Z-scores close in time are higher than those further apart in time.

Table 5.5: Results from principal component analysis of correlation matrix for weight Z-scores in infancy at birth and six grouped ages (prop. var. and cum. prop. denote 'Proportion of Variance' and 'Cumulative Proportion' respectively)

| | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 |
|---|---|---|---|---|---|---|---|
| Z(birth) | 0.285 | -0.667 | 0.607 | 0.299 | 0.108 | -0.061 | 0.036 |
| Z(6 wks) | 0.367 | -0.446 | -0.321 | -0.435 | -0.471 | 0.383 | -0.068 |
| Z(3 mths) | 0.397 | -0.189 | -0.474 | -0.106 | 0.379 | -0.654 | -0.008 |
| Z(6 mths) | 0.407 | 0.139 | -0.283 | 0.441 | 0.390 | 0.564 | 0.266 |
| Z(9 mths) | 0.403 | 0.291 | 0.030 | 0.384 | -0.315 | -0.098 | -0.704 |
| Z(12 mths) | 0.398 | 0.333 | 0.184 | 0.026 | -0.469 | -0.272 | 0.635 |
| Z(18 mths) | 0.376 | 0.325 | 0.434 | -0.606 | 0.390 | 0.145 | -0.157 |
| SD's | 2.304 | 0.960 | 0.623 | 0.418 | 0.294 | 0.263 | 0.227 |
| prop. var. | 0.758 | 0.132 | 0.056 | 0.025 | 0.012 | 0.010 | 0.007 |
| cum. prop. | 0.758 | 0.890 | 0.945 | 0.970 | 0.983 | 0.993 | 1.000 |

$$R = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 \\ & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ & & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ & & & 1 & \rho & \rho^2 & \rho^3 \\ & & & & 1 & \rho & \rho^2 \\ & & & & & 1 & \rho \\ & & & & & & 1 \end{pmatrix} \tag{5.3}$$

This form of covariance matrix plays an important role in stochastic processes and time-series analysis (Morrison 1976). The partial correlation of the $(i-1)$st and $(i+1)$st variates with the $i$th held constant is

$$\rho_{i-1,i+1,i} = \frac{\rho^2 - \rho\rho}{1-\rho^2} = 0$$
$$i = 2, \ldots, 6$$

In the time series context this implies that a variate $X_i$ is dependent on its predecessors with smaller subscripts only through its immediate neighbour $X_{i-1}$. Fixing the value of that random variable leaves $X_i$ and $X_{i-2}$ uncorrelated.

### 5.3.2 Maximum likelihood approach to estimation of $\rho$ from the correlation matrix

If the Newcastle correlation matrix took on the form given by equation (5.3), then we would only need to have an estimate of $\rho$ in order to ascertain the correlation structure. The approach used below is to assume that the weight Z-scores are a random sample from a multivariate Normal distribution with mean $\mu$ and variance $\sigma^2$. We will also assume that the covariance matrix $\Sigma = \sigma^2 R$. Then starting with the log-likelihood for the multivariate normal distribution, we will maximise this log-likelihood and arrive at maximum likelihood estimates of $\hat{\rho}$ and $\widehat{\sigma^2}$.

If we now consider the maximum likelihood estimation of a mean vector $\mu$ and the variance matrix $\Sigma$ in a multivariate normal population. Given a random sample $x_1, \ldots, x_n$ from $N_p(\mu, \Sigma)$, then the log-likelihood for $\mu$ and $\Sigma$ is:

$$L(\mu, \Sigma) = -\frac{np}{2}\log 2\pi - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T\Sigma^{-1}(x_i - \mu) \tag{5.4}$$

It can be shown that the maximum likelihood estimate of $\mu$ is $\bar{x}$, $\hat{\mu}$ can then be substituted in to the log-likelihood and after some algebra equation (5.4) can be written as:

$$L(\hat{\mu}, \Sigma) = C + \frac{n}{2}\left[\log\left|\Sigma^{-1}S^*\right| - \text{trace}(\Sigma^{-1}S^*)\right] \tag{5.5}$$

where

$$C = -\frac{n}{2}\left(p\log 2\pi + \log|S^*|\right)$$

and $S^*$ is the modified sample variance matrix $S^* = \frac{n-1}{n}S$. As $n$ is large, $\frac{n-1}{n} \approx 1$ and hence $S^* \approx S$.

If we now consider the general case of a $k \times k$ pattern covariance matrix:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{k-1} \\ & 1 & \rho & \ldots & \rho^{k-2} \\ & & \ddots & \ddots & \vdots \\ & & & 1 & \rho \\ & & & & 1 \end{pmatrix} \tag{5.6}$$

Then the inverse of this covariance matrix is:

$$\Sigma^{-1} = \frac{1}{(1-\rho^2)\sigma^2} \begin{pmatrix} 1 & -\rho & 0 & \ldots & 0 \\ -\rho & 1+\rho^2 & -\rho & \ldots & 0 \\ & \ddots & \ddots & \ddots & \vdots \\ & & -\rho & 1+\rho^2 & -\rho \\ & & & -\rho & 1 \end{pmatrix} \tag{5.7}$$

Therefore:

$$|\Sigma^{-1}| = \frac{1}{(1-\rho^2)^{k-1}(\sigma^2)^k} \tag{5.8}$$

and:

$$\log(|\Sigma^{-1}S|) = \text{constant} - (k-1)\log(1-\rho^2) - k\log\sigma^2 \tag{5.9}$$

Suppose we write the general $k \times k$ matrix $S^*$ as:

$$S^* = \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1k} \\ S_{21} & S_{22} & \dots & S_{2k} \\ \vdots & \vdots & & \vdots \\ S_{k1} & S_{k2} & \dots & S_{kk} \end{pmatrix} \tag{5.10}$$

then:

$$\text{trace}(\Sigma^{-1}S^*) = \frac{\text{trace}(S) + \rho^2\sum_{i=2}^{k-1} S_{ii} - 2\rho\sum_{i=1}^{k-1} S_{i,i+1}}{(1-\rho^2)\sigma^2} \tag{5.11}$$

So substituting results from equations (5.9) and (5.11) into equation (5.4), we then have to maximise the following function:

$$L(\rho,\sigma^2) = \frac{n}{2}\Bigg[\text{constant} - k\log\sigma^2 - (k-1)\log(1-\rho^2) - \frac{\text{trace}(S) + \rho^2\sum_{i=2}^{k-1} S_{ii} - 2\rho\sum_{i=1}^{k-1} S_{i,i+1}}{(1-\rho^2)\sigma^2}\Bigg] \tag{5.12}$$

So differentiating (5.12) with respect to $\sigma^2$ gives:

$$\frac{\partial L}{\partial\sigma^2} = -\frac{k}{\sigma^2} + \frac{\text{trace}(S) + \rho^2\sum_{i=2}^{k-1} S_{ii} - 2\rho\sum_{i=1}^{k-1} S_{i,i+1}}{(1-\rho^2)(\sigma^2)^2} \tag{5.13}$$

Therefore setting (5.13) to be zero gives the following maximum likelihood estimate of $\widehat{\sigma^2}$:

$$\widehat{\sigma^2} = \frac{\text{trace}(S) + \rho^2\sum_{i=2}^{k-1} S_{ii} - 2\rho\sum_{i=1}^{k-1} S_{i,i+1}}{(1-\rho^2)k} \tag{5.14}$$

Equation (5.12) is now differentiated with respect to $\rho$ to give:

$$\frac{\partial L}{\partial\rho} = \frac{2\rho(k-1)}{(1-\rho^2)} - \frac{(2\rho\sum_{i=2}^{k-1} S_{ii} - 2\sum_{i=1}^{k-1} S_{i,i+1})(1-\rho^2)\sigma^2}{(1-\rho^2)^2(\sigma^2)^2} - \frac{2\rho\sigma^2(\text{trace}(S) + \rho^2\sum_{i=2}^{k-1} S_{ii} - 2\rho\sum_{i=1}^{k-1} S_{i,i+1})}{(1-\rho^2)^2(\sigma^2)^2} \tag{5.15}$$

If (5.15) is now set to zero and substitute $\widehat{\sigma^2}$ defined in equation (5.14) then after some algebra get:

$$\rho^3\left(1-\frac{1}{k}\right)\sum_{i=2}^{k-1}S_{ii}+\rho^2\left(\frac{2}{k}-1\right)\sum_{i=1}^{k-1}S_{i,i+1}+\rho\left(-\frac{1}{k}\text{trace}(S)-\sum_{i=2}^{k-1}S_{ii}\right)+\sum_{i=1}^{k-1}S_{i,i+1}=0 \quad (5.16)$$

Equation (5.16) can then be solved for $\rho$ to obtain the maximum likelihood estimate of $\rho$.

### 5.3.3 Determining $\rho$ from the Newcastle covariance matrix

Table 5.6 contains the covariance matrix for the Newcastle weight Z-scores; this is based on 1055 individuals with weights at birth and in 6 age groupings defined by Dr C.M. Wright. If the revised UK 1990 reference was 'perfectly matched' to the Newcastle infancy data then the correlation and covariance matrix would be identical. Now assuming for the moment that the weight data is equally spaced, then using equation (5.16) we arrive at the cubic on the left in figure 5.5. The cubic only crosses the zero line once in the range -1 to 1, leading to value of $\rho \approx 0.87$ and $\sigma^2 \approx 1.07$. The matrix derived from this value of $\rho$ and $\sigma^2$ is given in table 5.7.

The value of $\rho$ derived above assumes that the data is equally spaced when in fact the difference between the grouped ages varies between 6 weeks and 6 months. Another alternative which would give approximately equally spaced weight Z-scores, is to take age groupings, created by Dr C.M. Wright, from birth to 1 year in steps

Table 5.6: Covariance matrix for weight Z-scores at birth and six grouped ages (1055 infants)

| Z-score | birth | 6 wks | 3 mths | 6 mths | 9 mths | 12 mths | 18 mths |
|---|---|---|---|---|---|---|---|
| birth | 1.058 | | | | | | |
| 6 wks | 0.718 | 0.930 | | | | | |
| 3 mths | 0.618 | 0.847 | 0.987 | | | | |
| 6 mths | 0.524 | 0.741 | 0.897 | 1.094 | | | |
| 9 mths | 0.506 | 0.676 | 0.842 | 1.035 | 1.185 | | |
| 12 mths | 0.484 | 0.642 | 0.789 | 0.970 | 1.090 | 1.148 | |
| 18 mths | 0.493 | 0.605 | 0.724 | 0.880 | 1.003 | 1.043 | 1.177 |

Table 5.7: Estimate of covariance matrix, assuming weight Z-score data is equally spaced with $\rho = 0.87$ and $\sigma^2 = 1.07$, for weight Z-scores at birth and six grouped ages (1055 infants)

| Z-score | birth | 6 wks | 3 mths | 6 mths | 9 mths | 12 mths | 18 mths |
|---|---|---|---|---|---|---|---|
| birth | 1.070 | | | | | | |
| 6 wks | 0.931 | 1.070 | | | | | |
| 3 mths | 0.810 | 0.931 | 1.070 | | | | |
| 6 mths | 0.705 | 0.810 | 0.931 | 1.070 | | | |
| 9 mths | 0.613 | 0.705 | 0.810 | 0.931 | 1.070 | | |
| 12 mths | 0.533 | 0.613 | 0.705 | 0.810 | 0.931 | 1.070 | |
| 18 mths | 0.464 | 0.533 | 0.613 | 0.705 | 0.810 | 0.931 | 1.070 |

of 3 months. Table 5.8 contains the covariance matrix for 1526 individuals with weight Z-scores in the 5 age groupings. The plot on the right in figure 5.5 contains the cubic derived from this covariance matrix. Again the cubic only crosses the zero line once in the range -1 to 1, leading to value of $\rho \approx 0.83$ and $\sigma^2 \approx 1.12$.

In order to compare the estimated covariance matrix with the sample covariance matrix, the null hypothesis, $\Sigma = \Sigma_0$ was considered. In this instance $\Sigma_0$ represents the sample covariance matrix and $\Sigma = \hat{\Sigma}$, where $\hat{\Sigma}$ is the estimated covariance matrix for maximum likelihood estimates of $\hat{\rho}$ and $\widehat{\sigma^2}$.

Consider a single random sample of $n$ observation vectors from the $p$-dimensional multinormal population with mean vector $\mu$ and positive definite covariance matrix $\Sigma$. Then the test statistic is given by equation (5.17) (Morrison 1976, pp248).

Table 5.8: Covariance matrix for weight Z-scores at birth and grouped ages 3, 6, 9 and 12 months (1526 infants)

| Z-score | birth | 3 mths | 6 mths | 9 mths | 12 mths |
|---|---|---|---|---|---|
| birth | 1.051 | | | | |
| 3 mths | 0.601 | 1.021 | | | |
| 6 mths | 0.514 | 0.927 | 1.123 | | |
| 9 mths | 0.487 | 0.856 | 1.055 | 1.200 | |
| 12 mths | 0.470 | 0.799 | 0.982 | 1.096 | 1.152 |

Table 5.9: Estimate of covariance matrix, assuming weight Z-score data is equally spaced with $\rho = 0.83$ and $\sigma^2 = 1.12$, for weight Z-scores at birth and grouped ages 3, 6, 9 and 12 months (1526 infants)

| Z-score | birth | 3 mths | 6 mths | 9 mths | 12 mths |
|---|---|---|---|---|---|
| birth | 1.120 | | | | |
| 3 mths | 0.930 | 1.120 | | | |
| 6 mths | 0.772 | 0.930 | 1.120 | | |
| 9 mths | 0.640 | 0.772 | 0.930 | 1.120 | |
| 12 mths | 0.532 | 0.640 | 0.772 | 0.930 | 1.120 |

$$L = \nu(log|\Sigma_0| - log|\hat{\Sigma}| + \text{trace}(\hat{\Sigma}\Sigma_0^{-1}) - p) \tag{5.17}$$

where $\nu$ is the degrees of freedom parameter for $\hat{\Sigma}$. When $n$ is large then $L$ is distributed as a chi-squared variate with $p(p+1)/2$ degrees of freedom if the null hypothesis is true.

When comparing the estimated covariance matrix in table 5.7 with sample covariance matrix in table 5.6:

$$|\Sigma_0| = 1.74 \times 10^{-4} \quad |\hat{\Sigma}| = 3.31 \times 10^{-4} \quad \text{tr}(\hat{\Sigma}\Sigma_0^{-1}) = 8.305$$
$$\nu = 1054 \quad p = 7 \quad L = 1054 \times 0.663 = 698.8$$

$L$ is much greater than 41.34, the 5% cut-off point for chi-square distribution with 28 degrees of freedom, leading us to reject the null hypothesis that the two covariance matrices are equal.

When comparing the estimated covariance matrix in table 5.9 with sample covariance matrix in table 5.8:

$$|\Sigma_0| = 6.2 \times 10^{-3} \quad |\hat{\Sigma}| = 1.65 \times 10^{-2} \quad \text{tr}(\hat{\Sigma}\Sigma_0^{-1}) = 7.112$$
$$\nu = 1525 \quad p = 5 \quad L = 1525 \times 1.133 = 1728.1$$

$L$ is much greater than 25.00, the 5% cut-off point for chi-square distribution with 15 degrees of freedom, leading us to reject the null hypothesis that the two covariance matrices are equal.

In both instances the agreement between the sample covariance matrix and the estimated covariance matrix is poor. It would seem that the fit of the suggested pattern matrix, given by equation (5.6), is not that good.

## 5.4 Exploratory regression modelling of the Newcastle correlations

### 5.4.1 Proposed functional form

As a result of work discussed in the last section it is plausible that the correlation may be a function of the form:

$$r_{t_i t_j} = \rho^{f(t_i,t_j)} \tag{5.18}$$

where $r_{t_i t_j}$ is the actual correlation in row $i$ and column $j$ of matrix given in table 5.1, $\rho$ is an unknown constant and $f(t_i, t_j)$ is a function of $t_i$ and $t_j$ where $t_i < t_j$.

If we log both sides of equation (5.18) then:

$$\log(r_{t_i t_j}) = f(t_i, t_j)\log(\rho) \tag{5.19}$$

One possibility would have been to fit an appropriate analytic model $f(t_i, t_j)$ and use a similar approach to the previous section. However this is not feasible in practice, because using the likelihood approach would be labour intensive and there is no satisfactory way of comparing the estimated covariance matrix with the sample covariance matrix. An additional consideration is that the weight data are not in fact all measured at the same time. A reasonable alternative is to model the correlation by regressing $\log(r_{t_i t_j})$ on some function of $t_i$ and $t_j$. The fit of any derived model can then be assessed using standard regression techniques. At this stage it should be noted, that Cole (1995) and Cole (1998a) do not propose this form for the correlation structure.

### 5.4.2 Properties of the correlation coefficient

The greater the number of observations, the better the sample correlation approximates the true correlation. In the normal case the variance of the sample correlation coefficient is given by equation (5.20) (Kendall and Stuart 1979). However, equation (5.20) is of little value in practice as the distribution of $r$ tends to normality so slowly that it is unwise to use it for $n < 500$ (Kendall and Stuart 1979).

$$Var(r) = \frac{(1-\rho^2)^2}{n} \tag{5.20}$$

Fisher's transformation (1921)

$$z_r = \frac{1}{2}\log\frac{1+r}{1-r} = \tanh^{-1}(r) \tag{5.21}$$

Fisher's transformation (equation (5.21)) of $r$ tends to normality very much faster than $r$ and when $n > 50$ has approximate variance:

$$Var(z_r) \approx \frac{1}{n-3} \tag{5.22}$$

which is independent of $\rho$. However, higher order terms of the $Var(z_r)$ involve $\rho$.

If $n > 50$, then $\sqrt{\frac{1}{n-3}}$ is adequate to estimate the standard deviation of $z_r$ (Kendall and Stuart 1979). Therefore, an approximate confidence interval (CI) for a correlation (Sheskin 1997) is:

$$CI_{z_{r(1-\alpha)}} = z_r \pm z_{\frac{\alpha}{2}}\sqrt{\frac{1}{n-3}} \tag{5.23}$$

where $z_r$ is the Fisher transformation of $r$ and $z_{\frac{\alpha}{2}}$ is the Normal equivalent deviate for tail area $\frac{\alpha}{2}$. For fixed $n$ the width of the CI increases with decreasing $r$. For fixed $r$ the width of the CI decreases with increasing $n$. Therefore, provided $n$ is large or the correlation is high, then we have reasonably narrow confidence intervals. However, if the sample correlation coefficient is obtained from a sample with less than about 50 pairs of observations, then use of equation (5.23) to obtain the confidence interval for the correlation is unsafe. An alternative would be to use (5.24) to construct a confidence interval (Sheskin 1997).

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, df = n-2, \tag{5.24}$$

so that a confidence interval for a correlation is given by:

$$CI_r = r \pm t_{n-2,1-\frac{\alpha}{2}}\sqrt{\frac{1-r^2}{n-2}} \tag{5.25}$$

Using small samples ($n \leq 50$) results in very large confidence intervals for the calculated correlation. Furthermore the upper and lower bounds of confidence intervals given by equations (5.23) and (5.25) may be outside the (-1,1) region.

### 5.4.3 The general linear regression model and weighted least squares

As already discussed in section 5.4.1, we suspect that log correlation is in some way related to the measurement ages, $t1$ and $t2$. Linear regression is a means of

arriving at some functional form for the correlations. In using linear regression, it is possible to explore various functional forms of $t1$ and $t2$ and then assess the goodness of fit of the proposed model. However, as already discussed in section 5.4.2 we have wider confidence intervals around low correlation coefficients or correlation coefficients derived from small samples. In effect, the variance of the correlations are not all equal. In this situation, it is appropriate to use weighted least squares.

### The general linear regression model

Suppose the model under consideration can be written in the form given by:

$$y = X\beta + \epsilon \tag{5.26}$$

where $y$ is an $(n \times 1)$ vector of observations, $X$ is an $(n \times p)$ matrix of known form, $\beta$ is a $(p \times 1)$ vector of unknown parameters and $\epsilon$ is an $(n \times 1)$ vector of errors. The $E(\epsilon) = 0$ and $Var(\epsilon) = I\sigma^2$, so the elements of $\epsilon$ are uncorrelated.

The error sum of squares is given by equation (5.27) and the least squares estimate of $\beta$ is the value $b$, which when substituted in (5.27) minimises $\epsilon'\epsilon$.

$$\epsilon'\epsilon = (y - X\beta)'(y - X\beta) = Y'Y - 2\beta'X'Y + \beta'X'X\beta \tag{5.27}$$

The value of $b$ can be determined by differentiating equation (5.27) with respect to $\beta$, substituting $b$ for $\beta$ and setting the resulting matrix equation to zero. This leads to the normal equations given by (5.28). If the $p$ normal equations are independent, $X'X$ is non-singular and its inverse exists then the solution to the normal equations is given by equation (5.29).

$$(X'X)b = X'Y \tag{5.28}$$

$$b = (X'X)^{-1}X'Y \tag{5.29}$$

The solution $b$ has the following properties (Draper and Smith 1998):

1. $b$ is an estimate of $\beta$ that minimises the error sum of squares irrespective of any distribution properties of the errors.

2. The elements of $b$ are linear functions of the observations $Y_1, \ldots, Y_n$ and provide unbiased estimates of the elements of $\beta$ which have minimum variance irrespective of any distribution properties of the errors.

3. If the errors are independent and $\epsilon_i \sim N(0, \sigma^2)$, then $b$ is the maximum likelihood estimate of $\beta$.

The assumption that the errors are normally distributed is not needed to estimate $b$. However, it is required in order to make tests that depend on assumption of normality, such as t and F-test and in the obtaining of confidence intervals based on the t and F distributions.

### Weighted least squares

If it happens that the variances of the observations are not all equal or the observations are correlated, then we might assume $E(\epsilon) = 0$ and the $Var(\epsilon) = V\sigma^2$ for some appropriate $V$, and so assume $\epsilon \sim N(0, V\sigma^2)$. In this instance the ordinary least squares estimate (equation (5.29)) does not apply and instead weighted or generalised least squares can be used to obtain estimates. Draper and Smith (1998) uses the terms:

- 'weighted least squares' to describe the situation when $V$ is a diagonal matrix with unequal diagonal elements.
- 'generalized least squares' to describe the situation when the off diagonal elements of $V$ are non-zero, this approach is used when the observations themselves are correlated.

The basic idea behind weighted least squares is to transform the observations $Y$ to other variables $Z$ which do appear to satisfy the usual tentative assumptions (Draper and Smith 1998). It is often possible to find a unique non-singular symmetric matrix $P$ such that:

$$P'P = PP = P^2 = V$$

If we now write $f = P^{-1}\epsilon$, then $E(f) = 0$ and $f \sim N(0, I\sigma^2)$. Equation (5.26) is premultiplied by $P^{-1}$, to arrive at the new model:

$$P^{-1}Y = P^{-1}X\beta + P^{-1}\epsilon \tag{5.30}$$

or

$$Z = Q\beta + f \tag{5.30}$$

The residual sum of squares is:

$$f'f = \epsilon' V^{-1} \epsilon = (Y - X\beta)' V^{-1} (Y - X\beta) \tag{5.31}$$

and

$$b = (X' V^{-1} X)^{-1} X V^{-1} Y \tag{5.32}$$

If weighted least squares were called for but ordinary least squares was used instead, then the estimates would be still unbiased but would not have minimum variance (Draper and Smith 1998).

We are proposing to model the log correlation, where the variances of each log correlation are not equal. Therefore we need to use weighted least squares in order to fit the model. If we can approximate the $Var(\log(r))$ we will be able to get a handle on the form of the weight matrix $V$. An alternative approach would be to use ordinary least squares and use the resulting residual plots to arrive at some approximation for $V$ (Draper and Smith 1998).

Consider a variable $X$ with mean $\mu$ and variance $\sigma^2$. If we wish to obtain the expectation and variance of some function of $X$, say $Y = f(X)$, then one approach is to use the method called propagation of error or $\delta$ method (Rice 1995). If we use a Taylor series expansion of $f$ about $\mu$, then to first order:

$$Y = f(X) \approx f(\mu) + (X - \mu) f'(\mu) \tag{5.33}$$

So $E(Y) \approx f(\mu)$ and $Var(Y) = \sigma^2 [f'(\mu)]^2$. The adequacy of the above approximation depends on how non-linear $f$ is in a neighbourhood of $\mu$ and on the size of $\sigma$ (Rice 1995).

In this instance $f(X)$ is $\log(r)$, so equation (5.34) provides an approximation of the variance of log correlation. In order to estimate $Var(\log(r))$ we also need to estimate $\sigma^2$, this could be approximated by using equation (5.20).

$$Var(\log(r)) \approx \frac{\sigma^2}{r^2} \tag{5.34}$$

Logically it seems reasonable in the first instance to choose $V$ to be:

$$V = \begin{pmatrix} \frac{(1-r_1^2)^2}{n_1 r_1^2} & 0 & \dots & \dots & 0 \\ 0 & \frac{(1-r_2^2)^2}{n_2 r_2^2} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \frac{(1-r_m^2)^2}{n_m r_m^2} \end{pmatrix} \tag{5.35}$$

as we have shown that approximately $Var(\log(r)) \approx \frac{(1-r^2)^2}{nr^2}$.

### 5.4.4 Summary plots of Newcastle correlation matrix

Table 5.10 contains a data frame of the 21 correlations derived from 1055 individuals with weight Z-scores at birth and at the six grouped ages (weights allocated to target ages by Dr C. M. Wright). The grouped ages are now in weeks and assume that $Time2$ $(t2)$ is always greater than $Time1$ $(t1)$. In figure 5.6 we produce a three dimensional representation of the correlation matrix viewed from 4 perspectives.

A two dimensional representation of the correlation matrix can be found in figure 5.7; here the red line corresponds to correlations with birth weight Z-score, yellow line to correlations with weight Z-score at 6 weeks, green line to correlations with weight Z-score at 13 weeks, cyan line to correlations with weight Z-score at 26 weeks, indigo line to correlations with weight Z-score at 39 weeks and magenta point to correlation between weight Z-scores at 52 and 78 weeks. Weight Z-scores close in time are highly correlated but as time elapsed between weight measurement increases the correlation between weight Z-scores decreases. The correlation between birth weight Z-score and weight Z-score at 1 year and at 18 months are of similar magnitude indicating that the correlation with birth weight Z-score may have reached a plateau. However, the weight Z-scores in the 18 month grouping cover a wide age range.

Table 5.10: Data frame used to model correlations derived from 1055 individuals

| $r(t1,t2)$ | $Time1(t1)$ | $Time2(t2)$ | $r(t1,t2)$ | $Time1(t1)$ | $Time2(t2)$ |
|---|---|---|---|---|---|
| 0.7236 | 0 | 6 | 0.8635 | 13 | 26 |
| 0.6048 | 0 | 13 | 0.7789 | 13 | 39 |
| 0.4868 | 0 | 26 | 0.7416 | 13 | 52 |
| 0.4525 | 0 | 39 | 0.6722 | 13 | 78 |
| 0.4391 | 0 | 52 | 0.9090 | 26 | 39 |
| 0.4415 | 0 | 78 | 0.8654 | 26 | 52 |
| 0.8846 | 6 | 13 | 0.7754 | 26 | 78 |
| 0.7342 | 6 | 26 | 0.9350 | 39 | 52 |
| 0.6436 | 6 | 39 | 0.8491 | 39 | 78 |
| 0.6217 | 6 | 52 | 0.8977 | 52 | 78 |
| 0.5780 | 6 | 78 | | | |

### 5.4.5 Identification of potential functions of the initial ($t1$) and later ($t2$) time points

Log correlation is likely to be modelled by some function of $t1$, $t2$ and $t2 - t1$, i.e. it depends on the age at measurement and time elapsed between measurement occasions. A standard approach used in regression to select suitable functions of $t1$, $t2$ and $t2 - t1$, is to produce scatterplots of log correlation versus functions of $t1$, $t2$ and $t2 - t1$. However, in this instance these plots are not very informative. The purpose of these scatterplots is to look for linear associations between the response and functions of the variables of interest. An alternative was to look at the linear regression of $\log(r(t1, t2))$ on a function of $t1$ and then to compare the multiple correlation coefficients for the various transformations. A similar process would then be repeated for log correlation on functions of $t2$ and log correlation on functions of $t2 - t1$. Instead of examining the multiple correlation coefficient, other alternatives would have been to compare the residual mean squares ($s^2$) or Mallow's $C_p$ (Draper and Smith 1998).

In transforming $t1$ and $t2$ we are assuming that the predictor variables are not subject to some error. This is not strictly true as the correlations are derived from grouped ages. The usual transformations considered are square roots, squares, logs and inverses. In addition the negative exponential transformation was considered as the decrease in correlation with increasing time elapsed is similar to an exponential decay. For convenience unity was arbitrarily added to $t1$ and $t2$ to avoid the complications of logging zero or inverting zero (the effect of varying this constant is explored further in section 5.4.7). It is important that the correlation between birth weight Z-score and later weight Z-scores is modelled because preliminary plots in Chapter 4 indicated that there was a downward trend from birth for infants that experienced growth faltering.

In table 5.11 we present the results of regressing log correlation on single predictors. The single predictors were various functions of $t1$, $t2$ and $t2 - t1$. For comparative purposes the results of regressing log correlation on functions of the average of the time points is included because this term was within the model for correlation proposed by Cole (1995). It appears that any function of $t1$ is likely to be a good predictor, with $\log(t1 + 1)$ being the best.

Taking $\log(t1 + 1)$ as the best single predictor, the effect of adding other terms to this model was explored. $\log(r(t1, t2))$ was regressed on $\log(t1 + 1)$ and other functions of $t2$, $t2 - t1$ and $(t1 + t2)/2$; see table 5.12. In table 5.12, AIC refers to the Akaike Information Criterion and is defined as minus twice log likelihood plus $2p$

Table 5.11: Multiple correlation coefficient ($R^2$) for the linear regression of $\log(r(t1,t2))$ on single predictors. The single predictors were functions of $t1$, $t2$, $t2-t1$ and $(t1+t2)/2$ (where $f(t1)$ refers to functions of $t1$, $f(t2)$ denotes functions of $t2$, $f(t2-t1)$ denotes functions of $t2-t1$ and $f((t1+t2)/2)$ denotes functions of $(t1+t2)/2$)

| $f(t1)$ | $R^2$ | $f(t2)$ | $R^2$ |
|---|---|---|---|
| $t1$ | 0.517 | $t2$ | 0.002 |
| $\log(t1+1)$ | 0.701 | $\log(t2+1)$ | 0.004 |
| $\sqrt{t1}$ | 0.674 | $\sqrt{t2}$ | 0.003 |
| $1/(t1+1)$ | 0.619 | $1/(t2+1)$ | 0.005 |
| $\exp(-t1)$ | 0.572 | $\exp(-t2)$ | 0.002 |
| $f(t2-t1)$ | $R^2$ | $f((t1+t2)/2)$ | $R^2$ |
| $t2-t1$ | 0.324 | $(t1+t2)/2$ | 0.083 |
| $\log(t2-t1)$ | 0.279 | $\log[(t1+t2)/2]$ | 0.028 |
| $\sqrt{t2-t1}$ | 0.313 | $\sqrt{(t1+t2)/2}$ | 0.056 |
| $1/(t2-t1)$ | 0.170 | $2/(t1+t2)$ | 0.001 |
| $\exp(t1-t2)$ | 0.015 | $\exp[-(t1+t2)/2]$ | 0.002 |

where $p$ is the rank of the model (the number of effective parameters). The model which has the lowest AIC value and largest F-value in table 5.12 provides the best fit.

The two-variable model, equation (5.36), provides the 'best' fit. At this stage, for simplicity and convenience, we will now call models of the form given by (5.36), the 'Argyle model'.

Argyle model

$$\log(r(t1,t2)) = A\log(t1+1) + B\log(t2+1) + C + \epsilon \tag{5.36}$$

A summary of the fit of the Argyle model using ordinary least squares can be found in the upper table of table 5.13, this model explains 98% of the variation within the data. The coefficients within this model are of similar magnitude and the intercept term in equation (5.36) is not significantly different from zero. So in effect, this model involves a difference of time on a log scale.

The correlations in table 5.10 are all derived from the same number of individuals but the correlations between weight Z-scores are of differing magnitude. Therefore it is appropriate to use weighted least squares to fit equation (5.36) to the Newcastle

Table 5.12: Impact of adding various functions of $t2$, $t2 - t1$ and $(t1 + t2)/2$ to linear model $\log(r(t1, t2)) = a \log(t1 + 1) + \epsilon$ (AIC refers to the Akaike Information Criterion)

| variable added | Sum of Sq | RSS | AIC | F |
|---|---|---|---|---|
| none | | 0.375 | -80.550 | |
| $\log(t2 + 1)$ | 0.350 | 0.024 | -136.063 | 260.41 |
| $\log(t2 - t1)$ | 0.341 | 0.034 | -128.941 | 180.33 |
| $\log[(t1 + t2)/2]$ | 0.336 | 0.039 | -126.056 | 154.88 |
| $1/(t2 + 1)$ | 0.326 | 0.048 | -121.550 | 121.49 |
| $1/(t2 - t1)$ | 0.340 | 0.035 | -128.576 | 176.92 |
| $2/(t1 + t2)$ | 0.278 | 0.097 | -106.999 | 51.763 |
| $\sqrt{(t2)}$ | 0.317 | 0.057 | -117.994 | 99.764 |
| $\sqrt{t2 - t1}$ | 0.314 | 0.061 | -116.820 | 93.356 |
| $\sqrt{(t1 + t2)/2}$ | 0.286 | 0.089 | -108.835 | 58.135 |
| $\exp(-t2)$ | 0.135 | 0.239 | -87.963 | 10.181 |
| $\exp(t1 - t2)$ | 0.198 | 0.176 | -94.364 | 20.222 |
| $\exp[-(t1 + t2)/2)]$ | 0.141 | 0.234 | -88.421 | 10.802 |

correlations. So a $V$ matrix of the following form may be appropriate:

$$V = \begin{pmatrix} \frac{(1-r_1^2)^2}{r_1^2} & 0 & \dots & \dots & 0 \\ 0 & \frac{(1-r_2^2)^2}{r_2^2} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \frac{(1-r_{21}^2)^2}{r_{21}^2} \end{pmatrix} \tag{5.37}$$

as discussed in section 5.4.3 the $Var(\log(r)) \propto \frac{(1-r^2)^2}{r^2}$. In equation (5.37), $r_1$ corresponds to the correlation between weight Z-scores at birth and 6 weeks, $r_2$ to the correlation between weight Z-scores at birth and 13 weeks, ... and $r_{21}$ is the correlation between weight Z-scores at 52 and 78 weeks. Therefore:

$$P^{-1} = \begin{pmatrix} \frac{r_1}{1-r_1^2} & 0 & \dots & \dots & 0 \\ 0 & \frac{r_2}{1-r_2^2} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \frac{r_{21}}{1-r_{21}^2} \end{pmatrix} \tag{5.38}$$

The weights in the call to lm in R correspond to the diagonal elements of $P^{-1}$ and these vary between 0.54 (for correlation between weight Z-scores at birth and 52

weeks) and 7.4380 (for correlation between weight Z-scores at 39 and 52 weeks). The resulting fit from using weighted least squares can be found in the lower table within table 5.13: there is a slight improvement in the multiple correlation coefficient, the coefficients have marginally increased and the intercept term is now just significantly different from zero.

In figure 5.8 we plot the standardised residuals versus fitted values for the models given in table 5.13. When the model is fitted by weighted least squares, the standardised residuals were obtained by multiplying the residuals by $P^{-1}$ and then dividing by the standard deviation (SD=0.0404) of the weighted least squares residuals. Using weighted least squares has improved the residual plot, although there is still a slight excess of negative residuals.

### 5.4.6 Other model that provides a reasonable fit

In table 5.12 the regression of log correlation on the $\log(t1+1)$ and $\log(t2-t1)$ also provided a reasonable fit. Table 5.14 details the fit of this model, the upper table is the result from using ordinary least squares and the lower table from using weighted least squares. Surprisingly using weighted least squares leads to a reduction in the multiple correlation coefficient but this is because in using weighted least squares

Table 5.13: **Original Newcastle correlations [N=21]** (based on 1055 individuals with weight Z-scores in 7 age groupings): Regression of log transformation of correlation coefficients on functions of initial time $(t1+1)$ and later time $(t2+1)$

| **unweighted** | $\log(r(t1,t2)) = a\log(t1+1) + b\log(t2+1) + c + \epsilon$ | | | |
|---|---|---|---|---|
| | Value | Std. Error | t-value | Pr(>\|t\|) |
| (Intercept) | 0.076429 | 0.047922 | 1.595 | 0.128 |
| $\log(t1+1)$ | 0.198595 | 0.006588 | 30.144 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.226499 | 0.014036 | -16.137 | $3.77 \times 10^{-12}$ |
| $R^2$=0.9807, $R^2$(adj)=0.9785, residual SE=0.03669 on 18 df | | | | |

| **weighted** | $\log(r(t1,t2)) = a\log(t1+1) + b\log(t2+1) + c + \epsilon$ | | | |
|---|---|---|---|---|
| | Value | Std. Error | t-value | Pr(>\|t\|) |
| (Intercept) | 0.095707 | 0.034143 | 2.803 | 0.0118 |
| $\log(t1+1)$ | 0.205850 | 0.006089 | 33.806 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.235010 | 0.011308 | -20.784 | $4.95 \times 10^{-14}$ |
| $R^2$=0.9845, $R^2$(adj)=0.9828, residual SE=0.036 on 18 df | | | | |

we now have a strong linear trend and some indication of heteroscedasticity in the residual plot. This leads us to conclude that this model doesn't provide as good a fit as the Argyle model, equation (5.36).

### 5.4.7 Effect of varying the constant added to time points

In the above regressions of log correlation on initial and later time, one was added arbitrarily to the time points to avoid the complication of logging zero. It might seem reasonable to add 39 weeks (length of gestation is approximately 9 months) to both time points, so that age at conception is zero. In table 5.15 we present a summary of the fit of the Argyle model when 39 weeks is added to the time points. The addition of 39 weeks to the time point results in a reduction in the multiple correlation coefficient and a change in magnitude of the coefficients and intercept term.

Argyle model with constant $c$ added to measurement age

$$\log(r(t1,t2)) = A\log(t1+c) + B\log(t2+c) + C + \epsilon \tag{5.39}$$

Table 5.14: **Original Newcastle correlations [N=21]** (based on 1055 individuals with weight Z-scores in 7 age groupings): Regression of log transformation of correlation coefficients on functions of initial time $(t1+1)$ and time elapsed $(t2-t1)$

| **unweighted** | $\log(r(t1,t2)) = a\log(t1+1) + b\log(t2-t1) + c + \epsilon$ | | | |
|---|---|---|---|---|
| coeff. | Value | Std. Error | t-value | Pr(> \|t\|) |
| (Intercept) | -0.07699 | 0.04669 | -1.649 | 0.116 |
| $\log(t1+1)$ | 0.14692 | 0.00685 | 21.448 | $2.86 \times 10^{-14}$ |
| $\log(t2-t1)$ | -0.17791 | 0.01325 | -13.429 | $8.09 \times 10^{-11}$ |
| $R^2$=0.9728, $R^2$(adj)=0.9698, residual SE=0.04346 on 18 df | | | | |

| **weighted** | $\log(r(t1,t2)) = a\log(t1+1) + b\log(t2-t1) + c + \epsilon$ | | | |
|---|---|---|---|---|
| | Value | Std. Error | t-value | Pr(> \|t\|) |
| (Intercept) | -0.096201 | 0.044726 | -2.151 | 0.0453 |
| $\log(t1+1)$ | 0.131161 | 0.007746 | 16.932 | $1.67 \times 10^{-12}$ |
| $\log(t2-t1)$ | -0.160250 | 0.013428 | -11.934 | $5.51 \times 10^{-10}$ |
| $R^2$=0.9566, $R^2$(adj)=0.9518, residual SE=0.0603 on 18 df | | | | |

Table 5.15: **Original Newcastle correlations [N=21]** (based on 1055 individuals with weight Z-scores in 7 age groupings): Regression of log transformation of correlation coefficients on functions of initial time $(t1+39)$ and later time $(t2+39)$

| **unweighted** | $\log(r(t1,t2)) = a\log(t1+39) + b\log(t2+39) + c + \epsilon$ | | | |
|---|---|---|---|---|
| | Value | Std. Error | t-value | Pr(> \|t\|) |
| (Intercept) | -2.14529 | 0.42308 | -5.071 | $7.97 \times 10^{-5}$ |
| $\log(t1+39)$ | 0.99505 | 0.10719 | 9.283 | $2.77 \times 10^{-8}$ |
| $\log(t2+39)$ | -0.48400 | 0.09517 | -5.086 | $7.71 \times 10^{-5}$ |
| $R^2$=0.8277, $R^2$(adj)=0.8086, residual SE=0.1095 on 18 df | | | | |

| **weighted** | $\log(r(t1,t2)) = a\log(t1+39) + b\log(t2+39) + c + \epsilon$ | | | |
|---|---|---|---|---|
| | Value | Std. Error | t-value | Pr(> \|t\|) |
| (Intercept) | -1.29086 | 0.30573 | -4.222 | 0.000512 |
| $\log(t1+39)$ | 0.82671 | 0.09096 | 9.089 | $3.81 \times 10^{-8}$ |
| $\log(t2+39)$ | -0.52093 | 0.08789 | -5.927 | $1.31 \times 10^{-5}$ |
| $R^2$=0.8213, $R^2$(adj)=0.8014, residual SE=0.1224 on 18 df | | | | |

In order to explore how changing the constant $(c)$ affected the fit of model given by equation (5.39), the value of $c$ was allowed to vary between 0.5 and 10 (in steps of 0.5). The resulting deviance, which is proportional to the residual sum of squares, and coefficients were plotted against the constant term. The best fit is obtained if the constant added is around 1.5 weeks; see figure 5.10. As the constant added increases beyond 2 weeks, the deviance and the coefficient of $\log(t1+c)$ increase, whereas the intercept and coefficient of $\log(t2+c)$ decrease. This process was repeated by adding a constant between 1 and 2.5 (in steps of 0.1) in order to locate the constant that provided the optimum fit. It would appear that the addition of about 1.6 weeks is optimum; see figure 5.10. We summarise the fit of the Argyle model with constant 1.6 weeks in table 5.16. The effect of changing the constant from 1 to 1.6, leads to a decrease in the residual standard error and an increase in the magnitude of the coefficients.

Table 5.16: **Original Newcastle correlations [N=21]** (based on 1055 individuals with weight Z-scores in 7 age groupings): Regression of log transformation of correlation coefficients on functions of initial time $(t1+1.6)$ and later time $(t2+1.6)$ using weighted least squares

| **weighted** | $\log(r(t1,t2)) = a\log(t1+1.6) + b\log(t2+1.6) + c + \epsilon$ | | | |
|---|---|---|---|---|
| | Value | Std. Error | t-value | Pr(> \|t\|) |
| (Intercept) | 0.056211 | 0.028084 | 2.002 | 0.0606 |
| $\log(t1+1.6)$ | 0.233038 | 0.005569 | 41.846 | $< 2 \times 10^{-16}$ |
| $\log(t2+1.6)$ | -0.248243 | 0.009485 | -26.171 | $8.88 \times 10^{-16}$ |
| $R^2$=0.9899, $R^2$(adj)=0.9887, residual SE=0.02917 on 18 df | | | | |

## 5.5 Development of model for correlation structure on full Newcastle infancy data frame

### 5.5.1 The individuals with measurements in all 7 age groupings

In section 5.4, the correlations modelled were based on 1055 children, just less than a third of the birth cohort. This raises the possibility that these children may not be representative. Table 5.17 summarises the level of deprivation of the 1055 children (503 boys, 552 girls) that contribute to the correlation matrix and of the remaining 2360 children (1208 boys, 1152 girls) of the birth cohort. The socio-economic class distribution for these two populations are significantly different (chi-square test: $X^2$=9.2085, df=2, p=0.010), because there is a large deficit in deprived children within the group of 1055 children with weights in all 7 age groupings and an excess of deprived children in the remaining 2360 children. The sex distribution within the two populations is not significantly different (chi-square test: $X^2$=3.4513, df=1, p=0.063).

It may well be that children with weights in all seven age groupings are lighter. Figure 5.11 contains variable width notch box plots for the 1055 individuals with weights in all 7 age groupings and the remaining 2360 infants from the whole birth cohort. There appears to be a slight tendency for the median weight Z-score for individuals with weights in all 7 age groupings to be slightly lower, but this only reaches statistical significance at the grouped ages of 3 and 6 months. Results of

testing that that the mean weight Z-score in each age group is the same for both groups can be found in table 5.18. With the exception of birth and 18 months, the children that contribute to the correlation matrix are significantly lighter than the rest of the birth cohort. However, the scatterplot in figure 5.12 indicates that the case children within the correlation matrix may be unduly influencing the mean Z-score in age groups after birth.

The correlations in table 5.10 were obtained using weight Z-score data grouped into six age groupings, these being 6 weeks, 3, 6, 9, 12 and 18 months. When data are grouped, information on individual values is lost. Table 5.19 summarises the age range these age groupings represent for the 1055 individuals that contribute to the correlation matrix. The maximum range is 10.5 months for the 18 month grouping and within the first year there is an age range of 2.5 to 3 months for each age-grouping.

Healy (1962) discusses the effect of age grouping in the construction of height

Table 5.17: Level of deprivation for individuals that contribute to the correlation matrix and rest of birth cohort (where 1, 2 and 3 denote affluent, intermediate and deprived, respectively)

| | 1 | 2 | 3 |
|---|---|---|---|
| 7 weights | 127 (12%) | 749 (71%) | 179 (17%) |
| Rest | 252 (11%) | 1604 (68%) | 504 (21%) |

Table 5.18: Results of test that mean weight Z-score of two groups (1055 children with 7 weights, remaining 2360 children) is same in each age-grouping

| Grouped age | mean 7 weights | mean Rest | t | p | CI |
|---|---|---|---|---|---|
| birth | -0.2997 | -0.2584 | -1.0524 | 0.2927 | [-0.118, 0.036] |
| 6 wks | -0.1973 | -0.0971 | -2.6671 | 0.0077 | [-0.174, -0.027] |
| 3 mths | -0.2002 | -0.0397 | -4.0996 | $4.286 \times 10^{-5}$ | [-0.237, -0.084] |
| 6 mths | -0.1272 | -0.0056 | -2.9358 | 0.0034 | [-0.203, -0.040] |
| 9 mths | -0.0657 | 0.0316 | -2.0644 | 0.0391 | [-0.190, -0.005] |
| 12 mths | -0.0587 | 0.0690 | -2.7959 | 0.0052 | [-0.217, -0.038] |
| 18 mths | -0.0699 | -0.0746 | 0.0885 | 0.9295 | [-0.101, 0.111] |

Table 5.19: Summary statistics of distributions of actual age (months) within each age interval for 1055 individuals that contribute to correlation matrix

| Age | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | range |
|---|---|---|---|---|---|---|---|
| 6 weeks | 0.750 | 1.440 | 1.840 | 1.709 | 1.970 | 2.100 | 1.35 |
| 3 months | 2.070 | 2.890 | 3.020 | 3.014 | 3.110 | 4.490 | 2.42 |
| 6 months | 4.690 | 5.800 | 6.000 | 6.017 | 6.230 | 7.410 | 2.72 |
| 9 months | 7.510 | 8.620 | 9.020 | 9.079 | 9.570 | 10.490 | 2.98 |
| 12 months | 10.56 | 11.84 | 12.07 | 12.07 | 12.36 | 13.48 | 2.92 |
| 18 months | 13.51 | 16.07 | 17.70 | 17.62 | 18.72 | 23.97 | 10.46 |

growth standards. If over the time-period the height of a child increases then:

> the sample standard deviation will be larger than the 'instantaneous' figure (Healy 1962, pp49).

Therefore, the effect of grouping ages in calculating the correlation coefficient will be to inflate the variance of a growth measurement when calculated from a sample spanning an age range during which the average value of a measurement changed (Healy 1962). Healy (1962) presented a correction factor that could be calculated to obtains an estimate of the 'instantaneous' variance, under the assumptions that measurements were normally distributed and that the mean and variance increased linearly with age. Goldstein (1981a) expanded on this work to give the correction for covariance and correlation due to age-grouping, assuming that ratio of the variances is constant and the correlation is constant over the age range.

In the current situation we are working on a Z-score scale, so if the data and growth reference were perfectly matched then the mean and variances of the Z-score over that age range would be expected to be zero and one respectively. However, exploratory data analysis in Chapter 4 indicated that in general Newcastle infants are substantially lighter than the reference in the early weeks of life and after the age of 1 year the weight Z-scores have a downward trend. So for example, in the age grouping '6 weeks' both the boys and girls will be catching up with the reference, resulting in an inflation in the variance of weight Z-scores in this age-range. The correlation of earlier weight Z-scores with the '18 months' weight Z-score is likely to be affected the most as this is over an age range of 10 months and there is a downward trend for both boys and girls in this time-period.

In an attempt to get a better handle on the correlation structure of weight Z-

scores the full infancy data frame with age in days was utilised, as described in the next subsection. If we can reduce the impact of grouping on the correlation coefficient by calculating correlations based on individuals whose ages differ only by days rather than months, there is then no need for corrections based on assumptions.

### 5.5.2 The full infancy data frame

There are 3415 term infants in the birth cohort, within this data frame an infant has birth weight and up to ten routine weights in infancy. This data frame was refined to select only infants that had two or more routine weights, as we need at least two weights to be able to say anything about an infants growth, for example to detect growth faltering. There were 3139 individuals (91.9%) that had two or more weights. The individuals with no weights or one weight may be in some way different to the individuals that have two or more weights. Table 5.20 summarises the sex, level of deprivation and gestational age of the 3139 individuals with 2 or more weights, and the 276 individuals with less than two weights. The two groups are similar in terms of sex distribution, level of deprivation and gestational age (chi-square test for sex: $X^2$=0.3192, df=1, p=0.5721; chi-square test for level of deprivation: $X^2$=2.395, df=2, p=0.3019 and chi-square test for indicator of gestational age: $X^2$=1.2727, df=2, p=0.5292). One hundred and forty eight children with less than two weights have a birth weight, this group has a mean birth weight Z-score of -0.278. This mean was compared to the mean Z-score of -0.149, from the 3020 individuals that had a birth weight within the 3139 individuals that had two or more weights (t-test t=-1.3071, p=0.1931). Therefore, there is no reason to doubt that the mean birth weight Z-score is the same in the two groups. To conclude, it appears that the individuals dropped in considering children with two or more weights do not differ significantly; in terms of level of deprivation, sex, gestational age and birth weight; from the individuals that we will study from this point.

A further data frame was then created with all pair wise combinations of ages and corresponding weights for all individuals that had two or more weights. This led to a data frame with 99856 rows and this was constructed in such a way that $t1$ was always less than $t2$. Using $t1$ and $t2$ to denote initial age and later age, respectively. There are 31538 distinct ($t1$,$t2$) pairs when the age is in days format. The number of individuals within a particular ($t1$,$t2$) combination varied between 1 and 227. The peaks in the distribution of values of $t1$ and $t2$ are at birth and 1 year respectively. The initial time-point ($t1$) can vary between birth and 1 year 4 months. The later time point ($t2$) can vary between 12 days and 3 years 1 month.

Table 5.20: Sex distribtion, distribution of level of deprivation, distribution of gestational age of individuals that have 2 or more weights and individuals that have less than 2 weights (sex: 1 males, 2 females. Level of deprivation: 1, 2 and 3 denote affluent, intermediate and deprived respectively)

| | sex | | level of deprivation | | | Gestational age indicator | | |
|---|---|---|---|---|---|---|---|---|
| no. of weights | 1 | 2 | 1 | 2 | 3 | $< 40$ | 40 wks | $> 40$ |
| $\geq 2$ | 1568 | 1571 | 343 | 2160 | 636 | 1089 | 955 | 540 |
| $< 2$ | 143 | 133 | 36 | 193 | 47 | 18 | 16 | 13 |

The time elapsed $(t2 - t1)$ between the two of the measurement instances can vary between 4 days and 3 years.

### 5.5.3 Effect of age grouping on obtaining correlations between weight Z-scores

At this stage we introduce the following notation. Let $n$ denote the number of individuals used to calculate the correlation between weight Z-scores for a given pair ($t1$,$t2$) and $N$ be the number of correlations calculated.

The age in days, $t1$ and $t2$, were grouped to the nearest 4, 7, 14 and 28 days. The data were grouped to 28 days to explore whether this affected the smoothness of the correlation rather than for the purpose of modelling. Age grouping was achieved by rounding the age in days divided by the smoother (4, 7, 14 or 28). For example when grouping to the nearest week, a value of $t1$ or $t2$ between 4 and 10 would fall into the grouped age of 1 week. The initial program rejected any groups with less than 10 weight Z-score pairs before calculating a correlation. However, as discussed in section 5.4.2, if the sample size is less than 50 the sample correlation coefficients obtained are likely to have large confidence intervals. Furthermore, there are still a large number of correlations generated for samples with greater than 50 individuals; see table 5.21. Table 5.21 contains a summary of the number of correlations obtained and the percentage of data they represent with age groupings of 4, 7, 14 and 28 days. Therefore, we will concentrate only on correlations obtained for samples with more than 50 individuals.

Figures 5.13 and 5.14 contain a graphical display of the number of individuals ($n > 50$) that contribute to the correlations versus the initial time-point, later time-

point and time elapsed. With the exception of the data obtained by grouping age to 28 days, the first three plots are on the same scale as the fortnightly correlations. Table 5.21, figure 5.13 and figure 5.14 indicate that grouping to the nearest fortnight seems to be a compromise between representing a high percentage of data and obtaining a large number of correlations. Grouping to the nearest fortnight also provides reasonable coverage of $t1$ and $t2$ values.

The upper plots in figure 5.15 provide a two dimensional graphical representation of the correlations calculated with age smoothed to nearest fortnight versus later time point ($t2$) and difference ($t2 - t1$). Points within these plots that have the same $t1$ value are connected, so the red line represents the correlation between later weight Z-scores with birth weight Z-score. The lower plots in figure 5.15 are plots of the Fisher transformed correlations versus $t2$ and $t2 - t1$. The Fisher's transformed correlations are less variable than the untransformed correlations. As the time elapsed between weight Z-scores increases, the connected plot of correlation exhibits a downward trend with irregular spikes. This could be due to different groups of individuals contributing to the correlation, the variation in the number of individuals used to calculate the correlation or measurement error. The Newcastle weights are routine weights, so the weights were obtained on different scales, in different states of undress and recorded by different observers.

Similar plots for correlations obtained from age to nearest week were quite crowded. A first attempt to improve these graphical representations of the weekly correlation structure was to obtain lowess curves (Cleveland 1979) (see Chapter 4 for description of lowess procedure) for each set of points with the same value of $t1$. The resulting lowess curves can be found in figure 5.16, some of these exhibit an upward trend towards the end of the infancy period (if $t1$ is birth or 5 weeks). The upward turn could in part be due to the scatterplot smoothers default span of two-thirds used in the lowess procedure. In general, the lowess curves do illustrate that there tends to be a general downward trend with increasing $t2$.

Table 5.21: Table of the number of correlations (N) obtained and percentage of data they represent with age groupings of 4, 7, 14 and 28 days

| | 4 days | 7 days | 14 days | 28 days |
|---|---|---|---|---|
| N | 300 | 496 | 385 | 212 |
| % of data | 33.7 | 62.4 | 86.8 | 96.8 |

Similar plots for the correlation between weight Z-scores calculated with age grouped to 4 and 28 days were produced. In contrast to the weekly correlations, clearer plots are obtained for grouping age to nearest 4 days, this is largely due to there being fewer points for values of t1 greater than 0. The plots for correlations obtained for age grouped to nearest 4 days indicated that the coverage of t1 and t2 values is poor. Therefore, it would not be worth modelling these correlations. In addition, it is also rare that infants are monitored this frequently, with the possible exception being in a hospital setting. The plots of correlation for age grouped to 28 days were a little more regular but still had a spiky appearance. In addition, the effect of grouping the data to the nearest 28 days extends the $t2$ time scale. However, these correlations will not be considered any further and were only included for comparison purposes.

### 5.5.4 Modelling of correlations derived from Newcastle data with age grouped to nearest fortnight

Equation (5.36) was fitted using ordinary least squares to the correlations derived from the Newcastle weight Z-scores with age grouped to nearest fortnight and the upper table in table 5.22 details the fit. A plot of the standardised residuals versus fitted values for the Argyle model fitted by ordinary least squares can be found on the left of figure 5.17. There is some evidence of heteroscedasticity within this residual plot with more variability observed for low correlations, this is largely due to the larger confidence intervals around the calculated correlations for the lowest correlations. Therefore, using weighted least squares to fit equation (5.36) should lead to an improvement in the fit of this model. Weighted least squares with a $V$ matrix of the form given by equation (5.35) was then used to fit the Argyle model to the fortnightly correlations. The weights vary from a minimum of 3.555 (correlation between weight Z-scores at 6 and 80 weeks, $r$=0.397, $n$=57) to a maximum of 362.700 (correlation between weight Z-scores at 22 and 26 weeks, $r$=0.967, $n$=578). The use of weighted least squares improves the fit of the model in terms of increasing the adjusted multiple correlation coefficient, see lower table in table 5.22. The magnitude of the coefficients of the log terms have slightly increased and we now have a significant intercept. The coefficients are also similar in magnitude to those obtained in table 5.13 for the original Newcastle correlations. A plot of the standardised residuals versus fitted values for the model fitted by weighted least squares can be found on the right of figure 5.17. This plot looks reasonably random, although there is some indication of slightly more variability for high correlations (the

heteroscedasticity is now in the reverse direction), see upper panel in figure 5.18. A plot of standardised residuals versus fitted values, labelled by an indicator of sample size can be found in the lower panel of figure 5.18. A cut-off point of 500 was chosen as the approximation for variance, given by equation (5.20) is only really suitable for $n \geq 500$. This residual plot indicates a slight excess of negative residuals for sample sizes greater than 500.

There are two extreme standardised residuals with values greater than 4, these correspond to the correlation between weight Z-scores at birth and 4 weeks ($r = 0.811$, $n = 1413$), and at birth and 6 weeks ($r = 0.755$, $n = 1308$), the fitted values are lower than the actual correlations. One method of detecting influential observations, is to calculate the Cook's distance. The influence of the $i$th data point may be measured by the squared scaled distance:

$$D_i = \frac{(\hat{Y} - \widehat{Y(i)})'(\hat{Y} - \widehat{Y(i)})}{ps^2} \tag{5.40}$$

where, $\hat{Y} = Xb$ is the usual vector of predicted values, $\widehat{Y(i)} = Xb(i)$ is the vector of predicted values from a least squares fit when the $i$th data point is deleted, $p$ is the number of parameters and $s^2$ is the residual mean square. A plot of Cook's distance versus index can be found on the left of figure 5.19. This plot indicates that points within the first few rows are influential, namely the correlation between weight Z-scores at birth and 4 weeks, and at birth and 6 weeks. Table 5.23 summarises the fit of the Argyle model if we exclude each of these two correlations identified by Cook's distance. Excluding the correlation corresponding to birth and 4 weeks leads to the largest improvement in the adjusted multiple correlation coefficient. There is a small improvement in the adjusted multiple correlation coefficient if the correlations corresponding to (0,4) and (0,6) weeks are excluded. Notice that excluding both influential correlations leads to the log terms having coefficients of similar magnitude.

Using the same approach as described in the section 5.4.7, the constant added to the time-point was varied (see figure 5.20), and it appears that the addition of 2.3 weeks to the time-point appears to be optimal. Table 5.24 details the model fit when 2.3 weeks is added to the time-points, the adjusted multiple correlation coefficient improves slightly and the coefficients of log terms have increased in magnitude to about 0.27. The plot of standardised residuals versus fitted values can be found on the left of figure 5.22, the two outlying standardised residuals are no longer as extreme but there is still slightly more variability for higher correlations. However using Cook's distance, see plot on right of figure 5.19, the same two correlations are still influential, along with the correlation between birth weight Z-score and 76 weeks ($r$=0.519). Excluding the two most influential observations leads to a slight

Table 5.22: **Newcastle correlations based on data grouped to nearest fortnight [N=385]**: Regression of log correlation coefficients on $\log(t1+1)$ and $\log(t2+1)$. Upper table presents results from unweighted least squares and lower table presents results from weighted least squares

| **unweighted** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.003302 | 0.033973 | 0.097 | 0.923 |
| $\log(t1+1)$ | 0.219986 | 0.004564 | 48.195 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.231024 | 0.009279 | -24.898 | $< 2 \times 10^{-16}$ |
| $R^2$=0.8697, $R^2$(adj)=0.8691, residual SE=0.09043 on 382 df | | | | |

| **weighted** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.037439 | 0.016539 | 2.264 | 0.0242 |
| $\log(t1+1)$ | 0.227741 | 0.003757 | 60.623 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.240101 | 0.005687 | -42.217 | $< 2 \times 10^{-16}$ |
| $R^2$=0.9083, $R^2$(adj)=0.9078, residual SE=0.3284 on 382 df | | | | |

improvement in the adjusted multiple correlation coefficient; see lower table of table 5.24.

If we now consider the impact of excluding the two most influential observations on the constant added to time in the Argyle model. It is found that the addition of 2 weeks appears to be optimum; see figure 5.21. A summary of the fit of the Argyle model with $c = 2$ can be found in table 5.25. There is a very slight improvement in the adjusted multiple correlation coefficient and the coefficients of the log terms are slightly reduced in magnitude. A plot of the standardised residuals versus fitted values can be found on the right of figure 5.22.

### 5.5.5 Modelling of correlations derived from Newcastle data with age grouped to nearest week

A similar approach to the previous section was used but now the correlations are derived from age smoothed to nearest week. In general the weekly correlations are based on fewer individuals. The fit of equation (5.36) using weighted least squares can be found in table 5.26, again using weighted least squares leads to an improvement in the adjusted multiple correlation coefficient. The weights used varied from 1.599 (correlation of 0.191 between weight Z-scores at birth and 59 weeks derived from a sample of 65) to 272.900 (correlation of 0.972 between weight

Table 5.23: **Newcastle correlations based on data grouped to nearest fortnight [N=385]**: Regression of log correlation coefficients on $\log(t1+1)$ and $\log(t2+1)$ using weighted least squares. Upper table presents results from excluding correlation corresponding to birth and 4 weeks, next table presents results from excluding correlation corresponding to birth and 6 weeks and final table presents results from excluding correlation corresponding to (0,4) and (0,6)

| **excl. (0,4)** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.018744 | 0.016746 | 1.119 | 0.264 |
| $\log(t1+1)$ | 0.230031 | 0.003712 | 61.973 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.236790 | 0.005614 | -42.176 | $< 2 \times 10^{-16}$ |
| $R^2$=0.9124, $R^2$(adj)=0.912, residual SE=0.3211 on 381 df | | | | |

| **excl. (0,6)** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.026179 | 0.016528 | 1.584 | 0.114 |
| $\log(t1+1)$ | 0.229789 | 0.003732 | 61.566 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.238685 | 0.005603 | -42.597 | $< 2 \times 10^{-16}$ |
| $R^2$=0.9114, $R^2$(adj)=0.9109, residual SE=0.3228 on 381 df | | | | |

| **excl. (0,4) & (0,6)** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.005186 | 0.016702 | 0.311 | 0.756 |
| $\log(t1+1)$ | 0.232420 | 0.003677 | 63.215 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.235018 | 0.005511 | -42.647 | $< 2 \times 10^{-16}$ |
| $R^2$=0.9161, $R^2$(adj)=0.9157, residual SE=0.3143 on 380 df | | | | |

Z-scores at 17 and 21 weeks derived from a sample of 238). The coefficients of the model for the weekly correlations are of a similar magnitude to those derived from the fortnightly correlations. A plot of the standardised residuals versus fitted values for the Argyle model using ordinary least squares can be found on the left of figure 5.23, this has an unusual pattern but there is some evidence of heteroscedasticity. A plot of the standardised residuals versus fitted values for the Argyle model using weighted least squares can be found on the right of figure 5.23, there is slightly more variability at low and high correlations but less variability at medium correlations. A plot of Cook's distance versus index can be found on the left of figure 5.24. There are four influential points and these correspond to the correlation between weight Z-scores at birth and 3 weeks ($r = 0.857$, $n = 173$), birth and 4 weeks ($r = 0.809$, $n = 771$), birth and 5 weeks ($r = 0.799$, $n = 773$) and birth and 6 weeks ($r = 0.764$, $n = 840$). These influential points are in the same time region as the extreme values identified for fortnightly Newcastle correlations.

Table 5.24: **Newcastle correlations based on data grouped to nearest fortnight [N=385]**: Upper table Regression of log correlation coefficients on $\log(t1 + 2.3)$ and $\log(t2 + 2.3)$ using weighted least squares Lower table Impact of excluding two most influential observations

$\log(r(t1, t2)) = a\log(t1 + 2.3) + b\log(t2 + 2.3) + c + \text{error}$

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.028695 | 0.016300 | -1.76 | 0.0791 |
| $\log(t1 + 2.3)$ | 0.279877 | 0.004212 | 66.45 | $< 2 \times 10^{-16}$ |
| $\log(t2 + 2.3)$ | -0.268651 | 0.005630 | -47.72 | $< 2 \times 10^{-16}$ |

$R^2$=0.9225, $R^2$(adj)=0.9221, residual SE=0.3018 on 382 df

| **excl. (0,4) & (0,6)** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.04792 | 0.01693 | -2.831 | 0.00489 |
| $\log(t1 + 2.3)$ | 0.28240 | 0.00421 | 67.071 | $< 2 \times 10^{-16}$ |
| $\log(t2 + 2.3)$ | -0.26550 | 0.00562 | -47.240 | $< 2 \times 10^{-16}$ |

$R^2$=0.9249, $R^2$(adj)=0.9245, residual SE=0.2975 on 380 df

Table 5.25: **Newcastle correlations based on data grouped to nearest fortnight [N=383]**: Regression of log correlation coefficients on $\log(t1 + 2)$ and $\log(t2 + 2)$ using weighted least squares after excluding the two most influential points

$\log(r(t1, t2)) = a\log(t1 + 2) + b\log(t2 + 2) + c + \text{error}$

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.036832 | 0.016628 | -2.215 | 0.0273 |
| $\log(t1 + 2)$ | 0.272855 | 0.004056 | 67.267 | $< 2 \times 10^{-16}$ |
| $\log(t2 + 2)$ | -0.259777 | 0.005521 | -47.050 | $< 2 \times 10^{-16}$ |

$R^2$=0.9252, $R^2$(adj)=0.9249, residual SE=0.2967 on 380 df

Using the same approach as described in section 5.4.7, the constant added to the time-point was varied; see figure 5.25. It appears that the addition of 2.9 weeks to the time-point appears to be optimal. This constant is greater than the constant identified for the fortnightly correlations. The upper table within table 5.27 details the fit of the Argyle model when 2.9 weeks is added to the time-points, there is an improvement in the adjusted multiple correlation coefficient and the magnitude of the coefficients of the log terms increase. The plot of standardised residuals versus

Table 5.26: **Newcastle correlations based on data grouped to nearest week [N=496]**: Regression of log correlation coefficients on log(t1+1) and log(t2+1). Upper table presents results from unweighted least squares and lower table presents results from weighted least squares.

| **unweighted** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.043054 | 0.037751 | 1.14 | 0.255 |
| $\log(t1+1)$ | 0.229727 | 0.005652 | 40.65 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.252773 | 0.010811 | -23.38 | $< 2 \times 10^{-16}$ |
| $R^2$=0.7994, $R^2$(adj)=0.7986, residual SE=0.1223 on 493 df | | | | |

| **weighted** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.049215 | 0.018411 | 2.673 | 0.00776 |
| $\log(t1+1)$ | 0.228679 | 0.004149 | 55.122 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.243298 | 0.006459 | -37.670 | $< 2 \times 10^{-16}$ |
| $R^2$=0.866, $R^2$(adj)=0.8654, residual SE=0.358 on 493 df | | | | |

fitted values can be found on the left in figure 5.27. This residual plot has a similar pattern to that observed when one was added to the time-points, but the outlying residuals are not as extreme. A plot of Cook's distance versus index can be found on the right of figure 5.24. The same four points are identified as influential along with the correlation between weight Z-scores at birth and 79 weeks ($r = 0.555$, $n = 80$). In figure 5.26 we consider the impact on the constant added to time-point if the most influential point is excluded (the correlation between weight Z-scores at birth and 5 weeks). This plot suggests that the addition of 2.8 weeks to time-point appears to be optimal. Therefore, excluding the most influential observation has had little impact on the constant added to age. A summary of the fit of the Argyle model with $c = 2.8$ can be found in the lower table of table 5.27. A plot of the standardised residuals versus fitted values for this model can be found on the right of figure 5.27. This model provides a pretty similar fit to the Argyle model with $c = 2.9$. When age is grouped to the nearest week the optimal constant added to the time-points is greater than when age is grouped to the nearest fortnight.

Table 5.27: **Newcastle correlations based on data grouped to nearest week [N=496]**: Upper table Regression of log correlation on $\log(t1+2.9)$ and $\log(t2+2.9)$ using weighted least squares Lower table Regression of log correlation on $\log(t1+2.8)$ and $\log(t2+2.8)$ using weighted least squares (after excluding one influential observation)

$\log(r(t1,t2)) = a\log(t1+2.9) + b\log(t2+2.9) + c + \text{error}$

| | Value | Std. Error | t-value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.066204 | 0.018965 | -3.491 | 0.000525 |
| $\log(t1+2.9)$ | 0.317833 | 0.005198 | 61.142 | $< 2\times10^{-16}$ |
| $\log(t2+2.9)$ | -0.288520 | 0.006591 | -43.775 | $< 2\times10^{-16}$ |

$R^2$=0.8884, $R^2$(adj)=0.888, residual SE=0.3265 on 493 df

$\log(r(t1,t2)) = a\log(t1+2.8) + b\log(t2+2.8) + c + \text{error}$

| **excl. (0,5)** | Value | Std. Error | t-value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.069870 | 0.019092 | -3.66 | 0.00028 |
| $\log(t1+2.8)$ | 0.315450 | 0.005139 | 61.39 | $< 2\times10^{-16}$ |
| $\log(t2+2.8)$ | -0.285154 | 0.006552 | -43.52 | $< 2\times10^{-16}$ |

$R^2$=0.8897, $R^2$(adj)=0.8893, residual SE=0.3249 on 492 df

## 5.6 Cambridge infant study and model proposed by Cole (1995)

### 5.6.1 The Cambridge infant study and correlation matrices

The Cambridge infant growth study was set up in 1983 to monitor the growth patterns of infants being fed according to Department of Health guidelines (Whitehead et al. 1989b). These infants had their weights and other anthropometric measures taken every 4 weeks from 4 weeks to 52 weeks and at 18 months and two years. These measurements were all taken within $\pm 3$ days of target age. The Cambridge infants were all at least 35 weeks gestation and no adjustment was made by Cole (1995) for gestational age.

In the original analysis of Cole (1995), birth weight was not included in the correlation matrix as this was not taken by the study team. Two hundred and twenty three infants (114 boys and 109 girls) were seen on all 15 occasions, see table 3.1 in chapter 3 for original correlation matrix. The two sexes were not significantly different, so their data was combined in the formation of the correlation matrix

(Cole 1995). Cole (1998a) later considered the correlation matrix for the Cambridge infants including routine birth weight; see table 5.28. There are now 221 infants with weights on all 16 occasions. The analysis within this section will concentrate on the Cambridge correlation matrix in table 5.28, because preliminary analysis in Chapter 4 indicated that children that exhibited growth faltering, did so soon after birth. Therefore it is important that the correlation with birth weight Z-score is incorporated in any model of the correlation structure.

Plots of the Cambridge correlations, given in table 5.28, versus $t2$ and $(t2 - t1)$ can be found in the upper plots of figure 5.28. Points within these plots that have the same t1 value are connected, so the red line represents the correlation between later weight Z-scores with birth weight Z-score, similarly the orange line corresponds to correlation between later weight Z-scores with weight Z-score at 4 weeks, ... and the uppermost violet line corresponds to the correlation between later weight Z-

Table 5.28: **Cambridge correlation matrix including birth weight** [N=120,n=221]: Correlation matrix for weight Z-scores at 16 ages during early infancy, based on data for 221 boys and girls seen on all 16 occasions

| Age | 0 | 4 | 8 | 12 | 16 | 20 | 24 | 28 | 32 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1.000 | | | | | | | | |
| 4 | 0.771 | 1.000 | | | | | | | |
| 8 | 0.646 | 0.911 | 1.000 | | | | | | |
| 12 | 0.589 | 0.829 | 0.945 | 1.000 | | | | | |
| 16 | 0.545 | 0.759 | 0.880 | 0.957 | 1.000 | | | | |
| 20 | 0.496 | 0.667 | 0.786 | 0.889 | 0.959 | 1.000 | | | |
| 24 | 0.482 | 0.635 | 0.747 | 0.849 | 0.922 | 0.968 | 1.000 | | |
| 28 | 0.463 | 0.608 | 0.710 | 0.811 | 0.892 | 0.945 | 0.970 | 1.000 | |
| 32 | 0.446 | 0.583 | 0.682 | 0.777 | 0.860 | 0.915 | 0.945 | 0.970 | 1.000 |
| 36 | 0.426 | 0.550 | 0.652 | 0.742 | 0.830 | 0.880 | 0.917 | 0.950 | 0.981 |
| 40 | 0.411 | 0.525 | 0.630 | 0.719 | 0.803 | 0.851 | 0.892 | 0.924 | 0.961 |
| 44 | 0.400 | 0.497 | 0.596 | 0.682 | 0.770 | 0.822 | 0.863 | 0.900 | 0.939 |
| 48 | 0.392 | 0.485 | 0.585 | 0.667 | 0.753 | 0.807 | 0.847 | 0.889 | 0.924 |
| 52 | 0.379 | 0.468 | 0.561 | 0.646 | 0.733 | 0.786 | 0.823 | 0.867 | 0.903 |
| 78 | 0.383 | 0.460 | 0.526 | 0.590 | 0.658 | 0.701 | 0.725 | 0.762 | 0.795 |
| 104 | 0.363 | 0.476 | 0.563 | 0.613 | 0.665 | 0.693 | 0.713 | 0.736 | 0.755 |

| | 36 | 40 | 44 | 48 | 52 | 78 | 104 |
|---|---|---|---|---|---|---|---|
| 36 | 1.000 | | | | | | |
| 40 | 0.980 | 1.000 | | | | | |
| 44 | 0.959 | 0.975 | 1.000 | | | | |
| 48 | 0.948 | 0.962 | 0.977 | 1.000 | | | |
| 52 | 0.925 | 0.945 | 0.965 | 0.972 | 1.000 | | |
| 78 | 0.814 | 0.844 | 0.873 | 0.887 | 0.908 | 1.000 | |
| 104 | 0.768 | 0.797 | 0.818 | 0.822 | 0.848 | 0.928 | 1.000 |

scores and weight Z-score at 1 year. There is a downward trend in the connected plots of correlation versus $t2$ until $t2$ takes on the value of one year. In general we might expect that as the time between two weight measures increases that the correlation between the two weight Z-scores decreases or plateaus. However the plot of correlation versus $t2$ indicates that this is not always the case for the Cambridge correlations. When $t1$ takes on values of 4, 8, 12 and 16 the correlation between weight Z-score at $t1$ and 2 years is higher than the correlation between weight Z-score at $t1$ and 18 months. No reference is made to this feature within Cole (1995) or Cole (1998a). The lower plots in figure 5.28 contain plots of the Fisher's transformed correlations versus $t2$ and $(t2-t1)$. The upward trend post 18 months is less marked in these plots.

### 5.6.2 Model proposed by Cole (1995) applied to Cambridge correlations

Cole (1995) chose to model the Cambridge Fisher's transformed correlations using fractional polynomials (Royston and Altman 1994). The transformed correlations were then modelled as a function of the two ages, mean $(t1 + t2)/2$ and difference $t2 - t1$. A summary of the fit of the original 5 variable model plus intercept can be found in chapter 3, table 3.2. Cole (1995) does not give details of how the final model choice was arrived at.

The conditional weight gain Z-score should have a mean of zero, standard deviation of 1 and ought to be uncorrelated with the initial weight Z-score. Cole (1995) validated the model in table 3.2 using data from the Cambridge infancy study (223 infants that derived Cambridge correlation matrix plus 183 infants from later cohorts) and a sub-population of 761 infants from the Newcastle growth and development study (described in chapter 4). Considering all possible age pairs in the Cambridge and Newcastle infancy data frame led to 26264 and 15405 pairs, respectively. Weight gain Z-scores were obtained using equation (3.11) and the model in table 3.2. The calculated conditional weight gain Z-scores were then split into 16 groups which corresponded roughly with the 15 grouped ages. Excluding pair combinations with fewer than 50 points for the Newcastle data, led to 90 and 108 age gap combinations for the Newcastle and Cambridge data, respectively. For each group the mean and SD of the conditional weight gain Z-score, and its correlation with initial weight Z-score were calculated. Cole (1995) concluded that on the whole there was reasonable agreement between Newcastle and Cambridge, suggesting that the conditional reference could be applied to other data.

Table 5.29 summarises the model Cole (1998a) fitted to the Cambridge correlations (including birth weight). The coefficients in this model are of the same sign although the magnitude of all the coefficients have changed slightly. The quadratic term in the model proposed by Cole (1995) appears to serve the purpose of fitting the upward trend at the end of infancy. A plot of standardised residuals versus fitted values for this model can be found in top left hand corner of figure 5.29, this looks reasonably random with a few extreme residuals. A plot of Cook's distance versus index can be found on the right of figure 5.29. There are two points identified as influential, the correlation between weight Z-scores at: 48 and 52 weeks ($r$=0.972) and 32 and 36 weeks ($r$=0.981). These two influential points are derived from weight Z-scores 4 weeks apart and both correlations are of similar magnitude.

### 5.6.3 Model proposed by Cole (1995) applied to Newcastle correlations

In this section we consider the fit of the model proposed by Cole (1995) to the Newcastle fortnightly and weekly correlations. Weighted least squares was used to fit the model proposed by Cole (1995) to the Newcastle correlations derived from age grouped to the nearest fortnight. In this instance, it seems reasonable to choose

Table 5.29: Model proposed by Cole (1995,1998) fitted to Cambridge correlations (including birth weight)

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| intercept | 3.177449 | 0.099103 | 32.062 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)$ | 0.326505 | 0.037146 | 8.790 | $1.8 \times 10^{-14}$ |
| $\log(t2-t1)$ | -1.482101 | 0.047966 | -30.899 | $< 2 \times 10^{-16}$ |
| $1/(t2-t1)$ | -2.023555 | 0.159455 | -12.690 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)\log(t2-t1)$ | 0.199544 | 0.011698 | 17.058 | $< 2 \times 10^{-16}$ |
| $(\log((t1+t2)/2))^2$ | -0.046462 | 0.009176 | -5.063 | $1.6 \times 10^{-6}$ |

$R^2$=0.9936, $R^2$(adj)=0.9933, residual SE=0.04159 on 114 df

$V$ to be:

$$V = \begin{pmatrix} \frac{1}{n_1-3} & 0 & \dots & \dots & 0 \\ 0 & \frac{1}{n_2-3} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \frac{1}{n_m-3} \end{pmatrix} \tag{5.41}$$

as the $Var(z_r) \approx \frac{1}{n-3}$ for $n > 50$. Hence:

$$P^{-1} = \begin{pmatrix} \sqrt{n_1-3} & 0 & \dots & \dots & 0 \\ 0 & \sqrt{n_2-3} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \sqrt{n_m-3} \end{pmatrix} \tag{5.42}$$

Table 5.30 summarises the fit of the model proposed by Cole (1995) to the Newcastle fortnightly correlations using unweighted and weighted least squares. The weights vary between 6.928 (correlations derived from samples of size 51, namely correlations between weight Z-scores at 14 and 66 weeks, and at 22 and 74 weeks) and 39.050 (correlation between weight Z-scores at birth and 8 weeks derived from sample size of 1528). Using weighted least squares improves the fit of the model proposed by Cole (1995). The coefficient of the quadratic log mean term is not significantly different from zero. With the exception of the coefficient for the log mean term, the coefficients are of the same sign and have a slightly smaller magnitude than those derived from the Cambridge correlations. The quadratic term of the model proposed by Cole (1995) was excluded and the model refitted. The resulting model can be found in the lowest table of table 5.30, this leads to a very similar adjusted multiple correlation coefficient.

A plot of the standardised residuals versus fitted values for the model proposed by Cole (1995) can be found on the left of figure 5.31. In this plot there is some evidence of curvature possibly due to the model overfitting the upturn. There is also some indication of slightly more variability for high correlations. A plot of Cook's distance versus index can be found on the left of figure 5.32, this indicates that there are two points that are influential, the correlation between weight Z-scores at: 0 and 4 weeks ($r$=0.811) and 4 and 8 weeks ($r$=0.852). Both of these influential points are in early infancy.

Similarly the model proposed by Cole (1995) was fitted to the Newcastle weekly correlations using ordinary and weighted least squares; see table 5.31. Using weighted least squares, weights vary between 6.928 (for samples of size 51) and 32.880 (for

Table 5.30: **Newcastle correlations based on data grouped to nearest fortnight [N=385]**:Regression of Fisher's transformation of correlation coefficients on functions of the time difference (t2-t1) and mean time (t1+t2)/2. Upper table presents results from unweighted least squares and middle table presents results from weighted least squares. Lowest table presents model proposed by Cole (1995) without quadratic term of log((t1+t2)/2)

| **unweighted** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (intercept) | 3.15689 | 0.16125 | 19.577 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)$ | 0.26654 | 0.06344 | 4.201 | $3.31 \times 10^{-5}$ |
| $\log(t2-t1)$ | -1.48806 | 0.08713 | -17.079 | $< 2 \times 10^{-16}$ |
| $1/(t2-t1)$ | -1.91155 | 0.18268 | -10.464 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)\log(t2-t1)$ | 0.22579 | 0.02257 | 10.006 | $< 2 \times 10^{-16}$ |
| $(\log((t1+t2)/2))^2$ | -0.05019 | 0.01621 | -3.097 | 0.0021 |
| $R^2$=0.9302, $R^2$(adj)=0.9293, residual SE=0.1201 on 379 df | | | | |

| **weighted** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (intercept) | 3.09387 | 0.14234 | 21.736 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)$ | 0.15993 | 0.05315 | 3.009 | 0.00279 |
| $\log(t2-t1)$ | -1.35966 | 0.07533 | -18.048 | $< 2 \times 10^{-16}$ |
| $1/(t2-t1)$ | -1.78715 | 0.17360 | -10.295 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)\log(t2-t1)$ | 0.18945 | 0.01983 | 9.556 | $< 2 \times 10^{-16}$ |
| $(\log((t1+t2)/2))^2$ | -0.01481 | 0.01401 | -1.057 | 0.29123 |
| $R^2$=0.9389, $R^2$(adj)=0.9381, residual SE=0.4106 on 379 df | | | | |

| **excl. quadratic** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (intercept) | 3.05914 | 0.13851 | 22.086 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)$ | 0.11534 | 0.03232 | 3.568 | 0.000405 |
| $\log(t2-t1)$ | -1.30373 | 0.05363 | -24.309 | $< 2 \times 10^{-16}$ |
| $1/(t2-t1)$ | -1.71525 | 0.15974 | -10.738 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)\log(t2-t1)$ | 0.17402 | 0.01341 | 12.980 | $< 2 \times 10^{-16}$ |
| $R^2$=0.9387, $R^2$(adj)=0.9381, residual SE=0.4106 on 380 df | | | | |

the correlation of 0.561 between weight Z-scores at birth and 13 weeks derived from a sample of size 1084). Notice that the quadratic term in this model is significantly different from zero, but the signs of the linear and quadratic terms of log mean age are opposite to those derived from the Cambridge correlations. A plot of the standardised residuals versus fitted values for the model proposed by Cole (1995) can be found on the right of figure 5.31. This plot looks reasonably random, there is still slightly more variability at higher correlations. A plot of Cook's distance versus index can be found on the right of figure 5.32, there are 2 points identified as influential, the correlation between weight Z-scores at: 5 and 7 weeks ($r$=0.831,$n$=65)

and 7 and 9 weeks ($r$=0.969, $n$=60). Both these correlations are in the early weeks of life and are between weight Z-scores that are two weeks apart.

### 5.6.4 The relationship between Fisher's transformed correlation and correlation

Fisher's transformation is a variance stabilising transformation, see section 5.4.2. In order to look at the relationship between Fisher's and log transformation of correlation, the Cambridge correlations were compared on the two scales, see figure 5.30. The box-plot in the top left of figure 5.30, indicates that Fisher's transformation does disperse the transformed correlations over a larger range than the log of the correlation. Both the distribution of log correlation and correlation have a long tail to the left, whereas Fisher's transformed correlation has a long tail to the right. The scatterplots in figure 5.30 do show a strong curvilinear relationship between Fisher's transformation of the correlation and the original correlation or the log

Table 5.31: **Newcastle correlations based on data grouped to nearest week [N=496]**:Regression of Fisher's transformation of correlation coefficients on functions of the time difference (t2-t1) and mean time (t1+t2)/2. Upper table presents results from unweighted least squares and lower table presents results from weighted least squares.

| **unweighted** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (intercept) | 2.93747 | 0.21545 | 13.634 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)$ | -0.09173 | 0.07943 | -1.155 | 0.2487 |
| $\log(t2-t1)$ | -1.06170 | 0.11281 | -9.411 | $< 2 \times 10^{-16}$ |
| $1/(t2-t1)$ | -1.78338 | 0.27433 | -6.501 | $1.97 \times 10^{-10}$ |
| $\log((t1+t2)/2)\log(t2-t1)$ | 0.09736 | 0.02181 | 4.464 | $9.99 \times 10^{-6}$ |
| $(\log((t1+t2)/2))^2$ | 0.07944 | 0.03088 | 2.572 | 0.0104 |
| $R^2$=0.8991, $R^2$(adj)=0.8981, residual SE=0.1398 on 490 df | | | | |

| **weighted** | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (intercept) | 2.96336 | 0.18473 | 16.042 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)$ | -0.14928 | 0.05884 | -2.537 | 0.01150 |
| $\log(t2-t1)$ | -1.02862 | 0.09774 | -10.524 | $< 2 \times 10^{-16}$ |
| $1/(t2-t1)$ | -1.69516 | 0.25240 | -6.716 | $5.18 \times 10^{-11}$ |
| $\log((t1+t2)/2)\log(t2-t1)$ | 0.07061 | 0.02695 | 2.621 | 0.00905 |
| $(\log((t1+t2)/2))^2$ | 0.11118 | 0.01824 | 6.096 | $2.21 \times 10^{-9}$ |
| $R^2$=0.9144, $R^2$(adj)=0.9135, residual SE=0.4175 on 490 df | | | | |

transformation.

### 5.6.5 Fit of Argyle model to Cambridge correlations

Table 5.32 summarises the fit of the Argyle model to the Cambridge correlations. Using the weights given by equation (5.38), these vary between 0.418 (correlation of 0.363 between weight Z-scores at birth and 104 weeks and 26.060 (correlation of 0.981 between weight Z-scores at 32 and 36 weeks). In terms of the multiple correlation coefficient, the Argyle model explains 97% of variation in the original data. It is not meaningful to compare models in tables 5.32 and 5.29 because the correlations have been subjected to different transformations, namely log and Fisher's transformation, respectively. A plot of the standardised residuals versus fitted values for the Argyle model can be found on the left of figure 5.33, there is some evidence of heteroscedasticity within this plot. However, the standardised residuals of the model proposed by Cole (1995) are spread over a wider range of values than the standardised residuals from the Argyle model. A plot of Cook's distance versus index can be found on the right of figure 5.33, this indicates that there is one influential observation, this being the correlation between weight Z-scores at birth and 4 weeks ($r$=0.771).

Using the same approach as described in section 5.4.7, the constant added to the time-point was varied (see figure 5.34), and it appears that the addition of 1.9 weeks to the time-point appears to be optimal. The lowest table in table 5.32 details the model fit when 1.9 weeks is added to the time-points, multiple R-squared improves slightly, the magnitude of the intercept term has decreased and the coefficients have increased in magnitude to about 0.29. The plot of standardised residuals versus fitted values and plot of Cook's distance versus index for the Argyle model with $c$=1.9 can be found in figure 5.35. There are several extreme negative residuals and using Cook's distance identifies 3 influential observations, these being the correlation between weight Z-scores at: 4 and 8 weeks ($r$=0.991), 48 and 52 weeks ($r$=0.972) and 8 and 12 weeks ($r$=0.945). These are different influential observations to those identified by Cook's distance when we fitted the Argyle model with $c$=1. These are points close in time with high correlations. Therefore the final model for the Cambridge correlations is:

$$\log(r(t1, t2)) = -0.0372 + 0.2863 \log(t1 + 1.9) - 0.2923 \log(t2 + 1.9) \qquad (5.43)$$

where $t1$ and $t2$ are the initial age and later age (given in weeks). This model is similar to that derived for the Newcastle fortnightly correlations, a similar constant

Table 5.32: **Cambridge correlation matrix including birth weight** [N=120,n=221]: Regression of log transformation of correlation coefficients on log initial time $(t1 + c)$ and log later time $(t2 + c)$

| **unweighted** | $\log(r(t1,t2)) = a\log(t1+1) + b\log(t2+1) + c + \epsilon$ | | | |
|---|---|---|---|---|
| | Value | Std. Error | t-value | Pr(> \|t\|) |
| (Intercept) | 0.072342 | 0.032106 | 2.253 | 0.0261 |
| $\log(t1+1)$ | 0.241906 | 0.004697 | 51.507 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.264686 | 0.009321 | -28.397 | $< 2 \times 10^{-16}$ |
| $R^2$=0.9584, $R^2$(adj)=0.9577, residual SE=0.05386 on 117 df | | | | |

| **weighted** | $\log(r(t1,t2)) = a\log(t1+1) + b\log(t2+1) + c + \epsilon$ | | | |
|---|---|---|---|---|
| | Value | Std. Error | t-value | Pr(> \|t\|) |
| (Intercept) | 0.081295 | 0.017884 | 4.546 | $1.34 \times 10^{-5}$ |
| $\log(t1+1)$ | 0.252039 | 0.004095 | 61.544 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.272435 | 0.006111 | -44.582 | $< 2 \times 10^{-16}$ |
| $R^2$=0.9706, $R^2$(adj)=0.9701, residual SE=0.05277 on 117 df | | | | |

| **weighted** | $\log(r(t1,t2)) = a\log(t1+1.9) + b\log(t2+1.9) + c + \epsilon$ | | | |
|---|---|---|---|---|
| | Value | Std. Error | t-value | Pr(> \|t\|) |
| (Intercept) | 0.037194 | 0.015761 | 2.36 | 0.0199 |
| $\log(t1+1.9)$ | 0.286294 | 0.003939 | 72.69 | $< 2 \times 10^{-16}$ |
| $\log(t2+1.9)$ | -0.292317 | 0.005437 | -53.77 | $< 2 \times 10^{-16}$ |
| $R^2$=0.9787, $R^2$(adj)=0.9784, residual SE=0.04485 on 117 df | | | | |

is added to age and the coefficients of the log terms in the Cambridge model are of a slightly greater magnitude.

### 5.6.6 A comparison of a subset of Newcastle weekly correlations with Cambridge correlations

Research data correlations for weight Z-scores are likely to be higher than for routine weight measurements (Cole 1997a). Cole (1993) stated that modelling correlation in a research study and then applying to routinely collected data may lead to poorly specified SD scores and that

> in such cases it may be possible to 'shrink' the correlation to take into account increased measurement error (Cole 1993, pp36).

Cole (1997a) discussed the effect of measuring error in relation to height measurements and suggested that each correlation $r$ could be scaled down to $r'$ using:

$$r' = r\left(1 - \left[\frac{\delta^2 - \epsilon^2}{\sigma^2}\right]\right) \tag{5.44}$$

where $\sigma$ is the standard deviation of height, $\epsilon$ is the likely measurement error of height in research studies and $\delta$ is the likely measurement error in routinely collected data. Cole (1997a) suggested that $\sigma$ took on a value of about 5cm, $\epsilon$ was between 0.1 and 0.4 cm, and $\delta$ is currently uncertain, but made suggestion of 0.8. However the error in weights can be difficult to quantify in infancy and there is no generally agreed value for measurement error or standard deviation. Furthermore the distribution of weight is skewed, so the correlation between weight measurements is not the same as the correlation between weight Z-scores.

Below an informal comparison is made between the correlations from routine Newcastle weights and research Cambridge weights. The Newcastle correlations with age grouped to nearest week were used for these comparison purposes as the Cambridge weights are taken within $\pm 3$ of target age. A set of correlations were extracted from the weekly Newcastle correlations, these were at the same time points as the Cambridge data (table 5.28), see table 5.33. In the early weeks of life the Newcastle correlations are based on a greater number of individuals; for ($t1$,$t2$) values of (0,4),(0,8) and (4,8); than the Cambridge data. The Newcastle correlations are only available up to the later age of 18 months (with the exception of birth and 104 weeks). It will be apparent from table 5.33 that the sample sizes for some of the correlations are small, so equation (5.25) was used to calculate confidence intervals for the Newcastle correlations. With the exception of the correlation between weight Z-scores at birth and 8 weeks, the correlations from the Cambridge study are within the confidence intervals of the Newcastle weekly correlations (see tables 5.34 and 5.35).

Table 5.33: Derived correlation matrix for Newcastle data (Age is in weeks), number of individuals that correlation is derived from is in brackets and NA means correlation not available for Newcastle data

| Age | 0 | 4 | 8 | 12 | 16 | 20 | 24 | 28 | 32 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1.000 | | | | | | | | |
| 4 | 0.8085 | 1.000 | | | | | | | |
| | (771) | | | | | | | | |
| 8 | 0.6974 | 0.8748 | 1.000 | | | | | | |
| | (762) | (271) | | | | | | | |
| 12 | 0.6326 | 0.7895 | 0.9376 | 1.000 | | | | | |
| | (467) | (158) | (191) | | | | | | |
| 16 | 0.5145 | 0.7688 | 0.8466 | 0.9308 | 1.000 | | | | |
| | (371) | (102) | (133) | (100) | | | | | |
| 20 | 0.5089 | 0.7284 | 0.8272 | 0.9018 | 0.9583 | 1.000 | | | |
| | (252) | (71) | (73) | (50) | (50) | | | | |
| 24 | 0.4478 | 0.6185 | 0.8268 | 0.9122 | 0.949 | 0.9441 | 1.000 | | |
| | (273) | (68) | (78) | (59) | (54) | (31) | | | |
| 28 | 0.4587 | 0.6492 | 0.6837 | 0.8476 | 0.9052 | 0.9332 | 0.9683 | 1.000 | |
| | (262) | (71) | (68) | (42) | (39) | (28) | (29) | | |
| 32 | 0.4337 | 0.5593 | 0.6941 | 0.7315 | 0.857 | 0.8513 | 0.9576 | NA | 1.000 |
| | (114) | (36) | (30) | (25) | (16) | (19) | (11) | | |
| 36 | 0.3817 | 0.6044 | 0.747 | 0.7347 | 0.8416 | 0.9267 | 0.8811 | 0.9344 | NA |
| | (227) | (67) | (74) | (43) | (35) | (25) | (20) | (29) | |
| 40 | 0.3816 | 0.3436 | 0.5274 | 0.7958 | 0.7573 | 0.7874 | 0.8755 | 0.8039 | 0.8892 |
| | (236) | (58) | (67) | (41) | (29) | (23) | (26) | (24) | (10) |
| 44 | 0.3663 | 0.4048 | 0.4545 | 0.5357 | 0.6936 | 0.7449 | 0.8399 | 0.903 | 0.954 |
| | (226) | (69) | (72) | (43) | (30) | (18) | (21) | (25) | (10) |
| 48 | 0.4865 | 0.4748 | 0.6906 | 0.8154 | 0.8283 | 0.8510 | 0.8516 | 0.9011 | NA |
| | (188) | (51) | (56) | (40) | (31) | (15) | (16) | (22) | |
| 52 | 0.4574 | 0.559 | 0.6191 | 0.6099 | 0.7865 | 0.7747 | 0.9266 | 0.8947 | 0.9068 |
| | (462) | (108) | (130) | (68) | (58) | (29) | (49) | (42) | (11) |
| 78 | 0.2957 | 0.5298 | 0.479 | 0.6245 | 0.3208 | 0.7271 | 0.43 | 0.7857 | NA |
| | (97) | (34) | (36) | (15) | (10) | (11) | (11) | (14) | |
| 104 | 0.0801 | NA | NA | NA | NA | NA | NA | NA | NA |
| | (10) | | | | | | | | |

| | 36 | 40 | 44 | 48 | 52 | 78 | 104 |
|---|---|---|---|---|---|---|---|
| 36 | 1.0000 | | | | | | |
| 40 | 0.9854 | 1.0000 | | | | | |
| | (15) | | | | | | |
| 44 | 0.9415 | 0.8807 | 1.0000 | | | | |
| | (29) | (24) | | | | | |
| 48 | 0.9616 | 0.9618 | NA | 1.0000 | | | |
| | (12) | (20) | | | | | |
| 52 | 0.9343 | 0.9435 | 0.9604 | 0.9849 | 1.0000 | | |
| | (37) | (46) | (66) | (31) | | | |
| 78 | 0.946 | 0.9584 | 0.9501 | NA | 0.9395 | 1.0000 | |
| | (10) | (17) | (17) | | (19) | | |
| 104 | NA | NA | NA | NA | NA | NA | 1.0000 |

Table 5.34: Comparison of derived correlation matrix for Newcastle data (Age is in weeks) with Cambridge correlation matrix

| t1 | t2 | Newcastle correlation | n | Cambridge correlation | lower | upper |
|---|---|---|---|---|---|---|
| 0 | 4 | 0.8085 | 771 | 0.771 | 0.7668 | 0.8502 |
| 0 | 8 | 0.6974 | 762 | 0.646 | 0.6464 | 0.7484 |
| 0 | 12 | 0.6326 | 467 | 0.589 | 0.562 | 0.7032 |
| 0 | 16 | 0.5145 | 371 | 0.545 | 0.4267 | 0.6023 |
| 0 | 20 | 0.5089 | 252 | 0.496 | 0.4017 | 0.6161 |
| 0 | 24 | 0.4478 | 273 | 0.482 | 0.3409 | 0.5547 |
| 0 | 28 | 0.4587 | 262 | 0.463 | 0.3502 | 0.5672 |
| 0 | 32 | 0.4337 | 114 | 0.446 | 0.265 | 0.6024 |
| 0 | 36 | 0.3817 | 227 | 0.426 | 0.2603 | 0.5031 |
| 0 | 40 | 0.3816 | 236 | 0.411 | 0.2626 | 0.5006 |
| 0 | 44 | 0.3663 | 226 | 0.4 | 0.2438 | 0.4888 |
| 0 | 48 | 0.4865 | 188 | 0.392 | 0.3601 | 0.6129 |
| 0 | 52 | 0.4574 | 462 | 0.379 | 0.3759 | 0.5389 |
| 0 | 78 | 0.2957 | 97 | 0.383 | 0.1011 | 0.4903 |
| 0 | 104 | 0.0801 | 10 | 0.363 | -0.7326 | 0.8928 |
| 4 | 8 | 0.8748 | 271 | 0.911 | 0.8166 | 0.933 |
| 4 | 12 | 0.7895 | 158 | 0.829 | 0.6924 | 0.8866 |
| 4 | 16 | 0.7688 | 102 | 0.759 | 0.6419 | 0.8957 |
| 4 | 20 | 0.7284 | 71 | 0.667 | 0.5639 | 0.8929 |
| 4 | 24 | 0.6185 | 68 | 0.635 | 0.4254 | 0.8116 |
| 4 | 28 | 0.6492 | 71 | 0.608 | 0.4665 | 0.8319 |
| 4 | 32 | 0.5593 | 36 | 0.583 | 0.2704 | 0.8482 |
| 4 | 36 | 0.6044 | 67 | 0.55 | 0.4071 | 0.8017 |
| 4 | 40 | 0.3436 | 58 | 0.525 | 0.0922 | 0.595 |
| 4 | 44 | 0.4048 | 69 | 0.497 | 0.1818 | 0.6278 |
| 4 | 48 | 0.4748 | 51 | 0.485 | 0.2221 | 0.7275 |
| 4 | 52 | 0.559 | 108 | 0.468 | 0.3993 | 0.7187 |
| 4 | 78 | 0.5298 | 34 | 0.46 | 0.2244 | 0.8352 |
| 8 | 12 | 0.9376 | 191 | 0.945 | 0.8877 | 0.9875 |
| 8 | 16 | 0.8466 | 133 | 0.88 | 0.7546 | 0.9386 |
| 8 | 20 | 0.8272 | 73 | 0.786 | 0.6942 | 0.9602 |
| 8 | 24 | 0.8268 | 78 | 0.747 | 0.6983 | 0.9553 |
| 8 | 28 | 0.6837 | 68 | 0.71 | 0.5044 | 0.863 |
| 8 | 32 | 0.6941 | 30 | 0.682 | 0.4154 | 0.9728 |
| 8 | 36 | 0.747 | 74 | 0.652 | 0.5908 | 0.9032 |
| 8 | 40 | 0.5274 | 67 | 0.63 | 0.3169 | 0.7379 |
| 8 | 44 | 0.4545 | 72 | 0.596 | 0.2422 | 0.6668 |
| 8 | 48 | 0.6906 | 56 | 0.585 | 0.4933 | 0.8879 |
| 8 | 52 | 0.6191 | 130 | 0.561 | 0.4818 | 0.7564 |
| 8 | 78 | 0.479 | 36 | 0.526 | 0.1731 | 0.7849 |
| 12 | 16 | 0.9308 | 100 | 0.957 | 0.8575 | 1.0041 |
| 12 | 20 | 0.9018 | 50 | 0.889 | 0.7764 | 1.0272 |
| 12 | 24 | 0.9122 | 59 | 0.849 | 0.8035 | 1.0209 |
| 12 | 28 | 0.8476 | 42 | 0.811 | 0.678 | 1.0172 |
| 12 | 32 | 0.7315 | 25 | 0.777 | 0.4374 | 1.0256 |
| 12 | 36 | 0.7347 | 43 | 0.742 | 0.5207 | 0.9487 |
| 12 | 40 | 0.7958 | 41 | 0.719 | 0.5997 | 0.9919 |
| 12 | 44 | 0.5357 | 43 | 0.682 | 0.2694 | 0.802 |
| 12 | 48 | 0.8154 | 40 | 0.667 | 0.6253 | 1.0055 |
| 12 | 52 | 0.6099 | 68 | 0.646 | 0.4151 | 0.8047 |
| 12 | 78 | 0.6245 | 15 | 0.59 | 0.1565 | 1.0925 |

Table 5.35: Comparison of derived correlation matrix for Newcastle data (Age is in weeks) with Cambridge correlation matrix (continued)

| t1 | t2 | Newcastle correlation | n | Cambridge correlation | lower | upper |
|---|---|---|---|---|---|---|
| 16 | 20 | 0.9583 | 50 | 0.959 | 0.8754 | 1.0412 |
| 16 | 24 | 0.949 | 54 | 0.922 | 0.8613 | 1.0367 |
| 16 | 28 | 0.9052 | 39 | 0.892 | 0.7636 | 1.0468 |
| 16 | 32 | 0.857 | 16 | 0.86 | 0.5616 | 1.1524 |
| 16 | 36 | 0.8416 | 35 | 0.83 | 0.6503 | 1.0329 |
| 16 | 40 | 0.7573 | 29 | 0.803 | 0.4994 | 1.0152 |
| 16 | 44 | 0.6936 | 30 | 0.77 | 0.4147 | 0.9725 |
| 16 | 48 | 0.8283 | 31 | 0.753 | 0.6155 | 1.0411 |
| 16 | 52 | 0.7865 | 58 | 0.733 | 0.6212 | 0.9518 |
| 16 | 78 | 0.3208 | 10 | 0.658 | -0.4514 | 1.093 |
| 20 | 24 | 0.9441 | 31 | 0.968 | 0.8189 | 1.0693 |
| 20 | 28 | 0.9332 | 28 | 0.945 | 0.7883 | 1.0781 |
| 20 | 32 | 0.8513 | 19 | 0.915 | 0.5828 | 1.1198 |
| 20 | 36 | 0.9267 | 25 | 0.88 | 0.7646 | 1.0888 |
| 20 | 40 | 0.7874 | 23 | 0.851 | 0.5077 | 1.0671 |
| 20 | 44 | 0.7449 | 18 | 0.822 | 0.3913 | 1.0985 |
| 20 | 48 | 0.851 | 15 | 0.807 | 0.5363 | 1.1657 |
| 20 | 52 | 0.7747 | 29 | 0.786 | 0.525 | 1.0244 |
| 20 | 78 | 0.7271 | 11 | 0.701 | 0.2094 | 1.2448 |
| 24 | 28 | 0.9683 | 29 | 0.97 | 0.8697 | 1.0669 |
| 24 | 32 | 0.9576 | 11 | 0.945 | 0.7404 | 1.1748 |
| 24 | 36 | 0.8811 | 20 | 0.917 | 0.6469 | 1.1153 |
| 24 | 40 | 0.8755 | 26 | 0.892 | 0.6719 | 1.0791 |
| 24 | 44 | 0.8399 | 21 | 0.863 | 0.5793 | 1.1005 |
| 24 | 48 | 0.8516 | 16 | 0.847 | 0.5511 | 1.1521 |
| 24 | 52 | 0.9266 | 49 | 0.823 | 0.8163 | 1.0369 |
| 24 | 78 | 0.43 | 11 | 0.725 | -0.2508 | 1.1108 |
| 28 | 36 | 0.9344 | 29 | 0.95 | 0.7937 | 1.0751 |
| 28 | 40 | 0.8039 | 24 | 0.924 | 0.5409 | 1.0669 |
| 28 | 44 | 0.903 | 25 | 0.9 | 0.7177 | 1.0883 |
| 28 | 48 | 0.9011 | 22 | 0.889 | 0.6988 | 1.1034 |
| 28 | 52 | 0.8947 | 42 | 0.867 | 0.752 | 1.0374 |
| 28 | 78 | 0.7857 | 14 | 0.762 | 0.3966 | 1.1748 |
| 32 | 40 | 0.8892 | 10 | 0.961 | 0.5162 | 1.2622 |
| 32 | 44 | 0.954 | 10 | 0.939 | 0.7096 | 1.1984 |
| 32 | 52 | 0.9068 | 11 | 0.903 | 0.5889 | 1.2247 |
| 36 | 40 | 0.9854 | 15 | 0.98 | 0.8834 | 1.0874 |
| 36 | 44 | 0.9415 | 29 | 0.959 | 0.8084 | 1.0746 |
| 36 | 48 | 0.9616 | 12 | 0.948 | 0.7682 | 1.155 |
| 36 | 52 | 0.9343 | 37 | 0.925 | 0.812 | 1.0566 |
| 36 | 78 | 0.946 | 10 | 0.814 | 0.6817 | 1.2103 |
| 40 | 44 | 0.8807 | 24 | 0.975 | 0.6713 | 1.0901 |
| 40 | 48 | 0.9618 | 20 | 0.962 | 0.8262 | 1.0974 |
| 40 | 52 | 0.9435 | 46 | 0.945 | 0.8428 | 1.0442 |
| 40 | 78 | 0.9584 | 17 | 0.844 | 0.8013 | 1.1155 |
| 44 | 52 | 0.9604 | 66 | 0.965 | 0.8908 | 1.03 |
| 44 | 78 | 0.9501 | 17 | 0.873 | 0.7784 | 1.1218 |
| 48 | 52 | 0.9849 | 31 | 0.972 | 0.9191 | 1.0507 |
| 52 | 78 | 0.9395 | 19 | 0.908 | 0.7642 | 1.1148 |

## 5.7 Combining the Newcastle and Cambridge correlations

We wish to explore whether the Cambridge and Newcastle correlations can essentially be described by the same function. The fortnightly correlations from the Newcastle infancy data were merged with the correlations from the Cambridge study. These correlations need to be assigned to two levels in order to take into account their different sources (Cambridge versus Newcastle, research versus routine). A dummy variable (ind) was created that was 0 if the correlation was from Cambridge study and 1 if the correlation was from the Newcastle study.

### 5.7.1 Argyle model applied to combined Newcastle and Cambridge correlations

In this section we consider the interaction of the Newcastle-Cambridge indicator variable with the Argyle model. Hence we are interested in the following model:

$$\begin{aligned}\log(r) = A + B\text{ind} + C\log(t1 + c) + D\log(t2 + c) + E\text{ind} : \log(t1 + c) \\ + F\text{ind} : \log(t2 + c) + \epsilon \end{aligned} \tag{5.45}$$

Weighted least squares was used to fit the Argyle model, given by equation (5.45), to the combined Cambridge and Newcastle correlations. The weights vary from a minimum of 3.555 (correlation between Newcastle weight Z-scores at 6 and 80 weeks, $r$=0.397, $n$=57) to a maximum of 387.5 (correlation between Cambridge weight Z-scores at 32 and 36 weeks, $r$=0.981, $n$=221). Table 5.36 summarises the fit of the Argyle model to combined Cambridge and Newcastle correlations. A plot of Cook's distance versus index for the model given in table 5.36 can be found on the left of figure 5.36. This plot indicates that there are two influential Newcastle correlations: correlation between weight Z-scores at birth and 4 weeks, and birth and 6 weeks. These two influential observations were excluded and the impact of varying the constant added to the initial and final time point was explored; see figure 5.37. It would appear that the addition of 2 weeks appears to optimise the fit of the Argyle model, which is the same constant that was added when considering Newcastle correlations alone. A summary of the fit of this model can be found in table 5.37, this model explains about 94% of the variation in the original data. This model would suggest that the effects of $\log(t1 + 2)$ and $\log(t2 + 2)$ are different for Cambridge and Newcastle and that the intercept term for both models is different. The coefficients of the log terms for Newcastle correlations are of smaller magnitude

than the Cambridge correlations. Suggesting that the Cambridge weight Z-scores are more correlated than the Newcastle weight Z-scores. A plot of the standardised residuals versus fitted values labelled by Newcastle or Cambridge can be found on the right of figure 5.36. This residual plot has a similar pattern to that for the Newcastle fortnightly correlations and there appears to be slightly more variability in the residuals for the Newcastle correlations.

Table 5.36: **Cambridge & Newcastle correlations [N=505]**: Regression of log correlation coefficients on $\log(t1+1)$, $\log(t2+1)$ interacted with Newcastle-Cambridge indicator variable (ind) using weighted least squares

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.081295 | 0.026696 | 3.045 | 0.002448 |
| $\log(t1+1)$ | 0.252039 | 0.006113 | 41.230 | $< 2 \times 10^{-16}$ |
| $\log(t2+1)$ | -0.272435 | 0.009122 | -29.867 | $< 2 \times 10^{-16}$ |
| ind | -0.043856 | 0.030768 | -1.425 | 0.154677 |
| $\log(t1+1)$:ind | -0.024298 | 0.007032 | -3.455 | 0.000596 |
| $\log(t2+1)$:ind | 0.032334 | 0.010530 | 3.071 | 0.002252 |

$R^2$=0.9257, $R^2$(adj)=0.925, residual SE=0.3037 on 499 df

Table 5.37: **Newcastle & Cambridge correlations [N=503]**: Regression of log correlation coefficients on $\log(t1+2)$ and $\log(t2+2)$ interacted with Newcastle-Cambridge indicator variable (ind) using weighted least squares (after excluding influential Newcastle correlations corresponding to (0,4) and (0,6))

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.032920 | 0.024954 | 1.319 | 0.187704 |
| $\log(t1+2)$ | 0.289351 | 0.006278 | 46.092 | $< 2 \times 10^{-16}$ |
| $\log(t2+2)$ | -0.294040 | 0.008610 | -34.150 | $< 2 \times 10^{-16}$ |
| ind | -0.069752 | 0.029262 | -2.384 | 0.017513 |
| $\log(t1+2)$:ind | -0.016496 | 0.007301 | -2.259 | 0.024296 |
| $\log(t2+2)$:ind | 0.034263 | 0.009995 | 3.428 | 0.000658 |

$R^2$=0.9402, $R^2$(adj)=0.9396, residual SE=0.2727 on 497 df

### 5.7.2 Model proposed by Cole (1995) applied to combined Newcastle and Cambridge correlations

In this section we consider the interaction of the Newcastle-Cambridge indicator variable with the model proposed by Cole (1995). Hence we are interested in the following model:

$$\begin{aligned} z_r = {} & A + B\text{ind} + C\log(m) + D\log(d) + E\frac{1}{d} + F\log(m)\log(d) + G(\log(m))^2 \\ & + H\text{ind} : \log(m) + I\text{ind} : \log(d) + J\text{ind} : \frac{1}{d} + K\text{ind} : \log(m)\log(d) \\ & + L\text{ind} : (\log(m))^2 + \epsilon \end{aligned} \tag{5.46}$$

where $z_r$ is the Fisher's transformed correlation, $m = (t1 + t2)/2$ and $d = t2 - t1$ $(t1 < t2)$. Weighted least squares was used to fit the model proposed by Cole (1995), given by equation (5.46), to the combined Cambridge and Newcastle correlations. The weights vary from a minimum of 6.928 (sample size of 51) to a maximum of 39.050 (sample size of 1528). A summary of the fit of the model proposed by Cole (1995) applied to the Newcastle and Cambridge correlations can be found in table 5.38. A plot of Cook's distance versus index can be found on the left of figure 5.38, this indicates that there is one influential point: the correlation between Newcastle weight Z-scores at birth and 4 weeks ($r = 0.811$, $n = 1413$). A summary of the fit of the model proposed by Cole (1995) after excluding this one influential observation can be found in table 5.39, this model explains about 95.5% of the variation in the original data. As the coefficient of the Cambridge-Newcastle indicator and the coefficients of the Cambridge-Newcastle indicator interacted with terms of model proposed by Cole (1995) are not significant, it would appear that the Cambridge and Newcastle correlations could be described by the same model. A plot of the standardised residuals versus fitted values, labelled by Newcastle or Cambridge, can be found on the right of figure 5.38. There appears to be slightly more variability for high correlations.

## 5.8 Discussion and Conclusions

A selection of routine weights for infants in the whole birth cohort were allocated to target ages of 6 weeks; 3, 6, 9, 12 and 18 months by Dr C.M. Wright. This chapter has dealt with initial exploratory work carried out on the correlation matrix obtained for 1055 individuals with weights at all 7 occasions. This correlation matrix exhibits the expected behaviour, that weight measurements made close in time were highly

Table 5.38: Cambridge & Newcastle correlations [N=505]: Summary of fit of model proposed by Cole (1995,1998)

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| intercept | 3.17745 | 0.22826 | 13.920 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)$ | 0.32650 | 0.08556 | 3.816 | 0.000153 |
| $\log(t2-t1)$ | -1.48210 | 0.11048 | -13.416 | $< 2 \times 10^{-16}$ |
| $1/(t2-t1)$ | -2.02355 | 0.36726 | -5.510 | $5.79 \times 10^{-8}$ |
| $\log((t1+t2)/2)\log(t2-t1)$ | 0.19954 | 0.02694 | 7.406 | $5.67 \times 10^{-13}$ |
| $(\log((t1+t2)/2))^2$ | -0.04646 | 0.02113 | -2.198 | 0.028385 |
| ind | -0.08357 | 0.26151 | -0.320 | 0.749417 |
| $\log((t1+t2)/2)$:ind | -0.16658 | 0.09793 | -1.701 | 0.089575 |
| $\log(t2-t1)$:ind | 0.12244 | 0.12949 | 0.946 | 0.344804 |
| $1/(t2-t1)$:ind | 0.23640 | 0.39888 | 0.593 | 0.553674 |
| $\log((t1+t2)/2)\log(t2-t1)$:ind | -0.01009 | 0.03228 | -0.313 | 0.754706 |
| $(\log((t1+t2)/2))^2$:ind | 0.03166 | 0.02458 | 1.288 | 0.198468 |

$R^2$=0.9556, $R^2$(adj)=0.9546, residual SE=0.3681 on 493 df

Table 5.39: Cambridge & Newcastle correlations [N=504]: Summary of fit of model proposed by Cole (1995,1998) after excluding one influential observation

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| intercept | 3.17745 | 0.22781 | 13.948 | $< 2 \times 10^{-16}$ |
| $\log((t1+t2)/2)$ | 0.32650 | 0.08539 | 3.824 | 0.000148 |
| $\log(t2-t1)$ | -1.48210 | 0.11026 | -13.442 | $< 2 \times 10^{-16}$ |
| $1/(t2-t1)$ | -2.02355 | 0.36654 | -5.521 | $5.47 \times 10^{-8}$ |
| $\log((t1+t2)/2)\log(t2-t1)$ | 0.19954 | 0.02689 | 7.421 | $5.15 \times 10^{-13}$ |
| $(\log((t1+t2)/2))^2$ | -0.04646 | 0.02109 | -2.203 | 0.028078 |
| ind | -0.16694 | 0.26548 | -0.629 | 0.529759 |
| $\log((t1+t2)/2)$:ind | -0.11376 | 0.10247 | -1.110 | 0.267464 |
| $\log(t2-t1)$:ind | 0.12493 | 0.12924 | 0.967 | 0.334197 |
| $1/(t2-t1)$:ind | 0.24526 | 0.39813 | 0.616 | 0.538154 |
| $\log((t1+t2)/2)\log(t2-t1)$:ind | -0.01070 | 0.03222 | -0.332 | 0.739997 |
| $(\log((t1+t2)/2))^2$:ind | 0.02328 | 0.02502 | 0.930 | 0.352625 |

$R^2$=0.9558, $R^2$(adj)=0.9549, residual SE=0.3674 on 492 df

correlated but less so the greater the time elapsed between weight measurements. Examination of the scaled inverse correlation matrix indicated that the correlation

between weight Z-scores may have an approximate Markov correlation structure.

Principal component analysis was used as an exploratory approach on the correlation matrix for the weight Z-scores. The first principal component was found to be a measure of overall size. The second principal component contrasted late weights with early weights, so a tentative conclusion was that this represented some measure of weight gain. Principal component analysis provided the opportunity to arrive at a 2-dimensional representation of the data which captured a large proportion of the total variation. When the second principal component was plotted against the first principal component, the case children were clustered in the lower left hand corner of this plot. However the case cluster was not distinct from the rest of the data. This observation was not surprising because the 'thrive index' approach used to select cases contrasts late weights with early weights and the case children are markedly lighter at the end of infancy.

Preliminary analysis using the log-likelihood of the Normal distribution indicated that a correlation matrix with pattern might not be entirely unreasonable. The correlation function in a Markov correlation structure is $\rho(t_i, t_j) = \rho^{|t_i - t_j|}$. Thus modelling log correlation seemed to be a suitable starting point. We suspected that log correlation may be modelled by some function of $t1$, $t2$ or the difference between the two time-points $t2 - t1$. Preliminary exploratory analysis concentrated on the $(7 \times 7)$ correlation matrix at grouped ages, allocated by Dr C.M. Wright. Log correlation was regressed on individual functions of $t1$ and $t2$. The Argyle model (equation (5.39)) was put forward as an alternative to the model proposed by Cole (1995) and Cole (1998a) for modelling the correlation between weight Z-scores in infancy.

The infants contributing to the $(7 \times 7)$ correlation matrix had fewer deprived children than expected and had a lower median weight Z-score at grouped ages of 3 and 6 months. Therefore, these infants may not be typical of the Newcastle population as a whole. The full set of Newcastle routine weights were then utilised, this had birth weight and up to 10 routine weights. Individuals that had two or more weights were considered further. The infants that had less than two weights were not found to be different to those remaining in the data frame in terms of sex, level of deprivation, gestational age if known and birth weight Z-score. A new data frame was created with all pair wise combinations for each infant that had two or more weights. Correlations between weight Z-scores were then obtained by regrouping the age in days to the nearest 4 days; 1, 2 and 4 weeks. Correlations were retained for sample sizes that were greater than 50. Grouping to the nearest fortnight seemed to be optimal in terms of coverage of $t1$ and $t2$, sample size and retaining a large

proportion of the original data frame. In addition, current guidelines recommend that infants should not be weighed more frequently than every fortnight in the first 6 months (Wright 2000).

The Argyle model was fitted to the correlations generated from the weekly and fortnightly Newcastle weight Z-scores using weighted least squares. The weights used were derived from the first order approximation of the variance of the correlation. The Argyle model provided a good fit in terms of the multiple correlation coefficient, however even after using weighted least squares there was some evidence of heteroscedasticity. Further exploration revealed that this dispersion was related to the size of the correlation, with more variability for high correlations. This has been noted elsewhere (Ghosh 1966), the variance approximation works well for large sample sizes and $|\rho| < 0.9$. Further improvements in terms of heteroscedasticity might be possible if higher order approximations for the variance were used for the weights. There is also the possibility that the correlations themselves might be correlated, this could also be accounted for using weighted least squares. However this facility is not available in R and there is no strong evidence of autocorrelation in the residual plots. The correlations between birth weight Z-score and a later weight Z-score in the age region of 4-6 weeks are influential observations for both weekly and fortnightly correlations. Further work could concentrate on the scatter-plots of these subset of correlations to see whether these are unusual in some way. However it may be that over this time period, the Newcastle children are catching up with the reference children leading to elevated variances for these age groupings.

In the Argyle model, one was added to the time-point to avoid logging zero. The effect of varying this constant was explored. Constants of 2 and 2.8 weeks; for fortnightly and weekly correlations, respectively; were found to be optimal. Berkey and Reed (1987) had previously explored the effect of changing the constant in the fit of the Count (1943) and Reed models (Berkey and Reed 1987) on length data from Harvard growth study. Berkey and Reed (1987) didn't arrive at any clear-cut answer and suggested assigning $t = 0$ to conception and $t = 1$ to birth using the following transformation of time scale: $t = (12x + 9)/9$ where $x$ is the age in years. For data where the age is in weeks, this equates to $t = (y + 39)/39$ where $y$ is age in weeks. However, we effectively fitted this model in table 5.15 and found that this transformation of time scale did poorly when modelling log-correlations.

The correlation structure of weight Z-scores in infancy has only been modelled explicitly by Cole (1995). Cole (1995) modelled the correlations from 223 infants in the Cambridge infant study. Fisher's transformation of correlation was modelled using 5 variables: log mean age and its quadratic, log age difference, inverse of

age difference and interaction of log mean age and log difference in age, plus an intercept. The model proposed by Cole (1995) was also fitted to the weekly and fortnightly Newcastle correlations. This model also provided a good fit, in terms of the multiple correlation coefficient, however there was some evidence of curvature in the residual plots for fortnightly correlations. Moreover, the coefficient for the quadratic term of log mean age was not significantly different to zero. Thus the model proposed by Cole (1995) without the quadratic term provides a better fit for the fortnightly correlations. This is likely to be due to the model proposed by Cole (1995) over-fitting the upturn at the end of infancy. The model proposed by Cole (1995) provided a better fit to the weekly Newcastle correlations, which is largely due to there being some evidence of a upward turn for the correlations with birth weight Z-score at the end of infancy. In addition, the Cambridge correlations are in effect derived from age grouped to nearest week. However, when the model proposed by Cole (1995) is fitted to Newcastle weekly correlations, the coefficients for the log mean age are of opposite sign to those in the same functional form fitted to the Cambridge correlations.

The Argyle model was fitted to the Cambridge correlations. This model provided a reasonable fit, in terms of the multiple correlation coefficient, however again there was evidence of heteroscedasticity within the residual plots. The effect of changing the constant added to the time-points to 1.9 weeks in the Argyle model was to improve the adjusted multiple correlation coefficient but led to several very large negative residuals for high correlations. However, no monotone decreasing function of $t1$ and $t2$ can hope to model the Cambridge tails.

The assumptions behind the linear model are homogeneity of variance, simplicity of structure for the expected value of the response and at least approximate normality of the additive errors (Atkinson 1985). If these three requirements are not met on the original scale of measurement of the response, then it may be that a non-linear transformation of the response may lead to homogeneity of variance and at least approximate normality (Atkinson 1985). Fisher's transformation of correlation used by Cole (1995) aims to do the latter. An empirical indicator that a transformation might be helpful is if the response is non-negative, which is certainly the case for correlations of weight Z-scores. If the response is non-negative it cannot follow a normal distribution. Atkinson (1985) also points out that if all values of the response are far from zero and the scatter in the observations is relatively small, the transformation will have little effect. The correlations between weight Z-scores are all positive so using Fisher's transformation will not map them onto negative values, it just disperses them over a wider range of positive numbers. However, the

approach used here was to model log correlation as we believed that the correlation matrix was near Markov. Log correlation maps the high positive correlations onto negative values but does not disperse them over a wider age range. The downside to modelling log correlation is that we need to derive weights for use in weighted least squares to have homoscedasticity in the residual plots. The advantages in modelling log correlation are that the model derived has fewer parameters than that derived by Cole (1995). Considering the Argyle model fitted to fortnightly correlations with constant 2 weeks added to age, leads to the following model:

$$\log(r) = -0.037 + 0.273\log(t1+2) - 0.260\log(t2+2) + \epsilon \qquad (5.47)$$

Exponentiating the above equation, the model for correlation on the original scale, gives:

$$r = 0.964\epsilon\frac{(t1+2)^{0.273}}{(t2+2)^{0.260}}$$

Note that we are assuming a multiplicative error on the original scale. An alternative one variable version of the Argyle model would be to regress $\log(r)$ on $\log(t1+c) - \log(t2+c)$, if $c$ is chosen in such a way that the coefficients of the log terms are similar.

Contrast this with the model proposed by Cole (1998a):

$$\begin{aligned}\text{Fisher}(r) = 3.18 + 0.33\log\left(\frac{t1+t2}{2}\right) - 1.48\log(t2-t1) - \frac{2.03}{t2-t1} \\ + 0.20\log\left(\frac{t1+t2}{2}\right)\log(t2-t1) - 0.05\left(\log\left(\frac{t1+t2}{2}\right)\right)^2\end{aligned}$$

Transforming this back to a correlation gives:

$$r = \frac{a-1}{a+1}$$

where

$$a = 578.25 \times \frac{\exp(-\frac{4.06}{t2-t1})}{(\frac{t1+t2}{2})^{1.04}(t2-t1)^{2.56}}$$

The practical implications of using equation (5.47) to model correlation, in terms of obtaining a conditional weight gain Z-score, will be explored further in Chapter 7.

We considered whether the Cambridge and Newcastle correlations could essentially be described by the same function. The fortnightly correlations from the Newcastle infancy data were merged with the correlations from the Cambridge study. We

considered the interaction of the Newcastle-Cambridge indicator variable with the Argyle model and the model proposed by Cole (1995). When applying the Argyle model to the combined Cambridge-Newcastle correlations it would appear that the addition of 2 weeks appears to optimise the fit of the Argyle model. Furthermore, the effects of $\log(t1+2)$ and $\log(t2+2)$ are different for Cambridge and Newcastle and that the intercept term for both models is different. The coefficients of the log terms for Newcastle correlations are of smaller magnitude than the Cambridge correlations. Suggesting that the Cambridge weight Z-scores are more correlated than the Newcastle weight Z-scores. When applying the model proposed by Cole (1995) to the combined Cambridge-Newcastle correlations the terms involving the Cambridge-Newcastle indicator were not significant. It would appear that the Cambridge and Newcastle correlations could be described by the same model.

A further consideration, is that the correlations between weight Z-scores from the Cambridge study might not be representative of correlations observed between weight Z-scores obtained routinely. Cole (1998a) does raise this issue in his discussion,

> the Cambridge study infants were screened by midwives before recruitment, and those thought to be at risk were excluded. For this reason the study population is likely to be unusually homogeneous. (Cole 1998a, pp2706)

However the research conditions for the Cambridge study are likely to elevate the correlations in comparison to community weights, but this is partly counteracted by the issue of homogeneity (raised in the previous point) which is likely to reduce the correlations. The plots for the Cambridge correlations are very regular in comparison to those obtained from the Newcastle weekly or fortnightly correlations, which is largely due to the issues discussed above and the same group of individuals contribute to each correlation. The Cambridge study weights were obtained in a research study setting, i.e. the same set of scales, state of undress and observer were used. This is likely to lead to reduced measurement error and may lead to a more regular appearance of the correlation structure than that obtained from the routine Newcastle data (these weights were obtained on different scales, in different states of undress and recorded by different observers).

An additional issue, already raised earlier, is that the Cambridge correlations exhibit an upward trend towards the end of infancy. This seems counter intuitive, as we might expect the correlation to plateau or decrease. However there may be some indication of this feature in the weekly Newcastle correlations between later

weight Z-scores with birth weight Z-score. Otherwise, the correlations obtained from the routine Newcastle weight data don't really extend much beyond 1 year. Boryslawski (1988) presented correlations with birth weight for 100 boys and 100 girls born in Wroclaw, Poland between 1973 and 1975. These samples excluded children that were born at less than 37 weeks gestation or whose birth weight was below 2.3 Kg. The children were monitored at monthly intervals in the first year of life and three month intervals in the second year of life. These are correlations between raw weights, but they show the expected trend that as time elapse increases to beyond 1 year the correlation decreases, see table 5.40. An additional point with regards to the Wroclaw weights is that the correlations for boys decrease to 6 months and then increase to 9 months, the correlations then continue to decrease. A similar pattern is observed for girls until the age of 1 year, but there is again a peak in the correlation at 15 months and a gradual decrease until 2 years. This observation on the raw weight scale may be due to variability observed in raw weights at different ages.

In the statistical literature, explicit modelling of the correlations is rare. In general the correlations tend to be modelled as an afterthought to improve the fit of a specific model or to incorporate non-independent measurements into a reference. Wade and Ades (1998) incorporated the correlation in the construction of references for CD4 counts of uninfected children born to HIV-1 infected woman. Wade and Ades (1998) looked at five models (where $\sigma^2$ represents measurement error), previously proposed by Grady and Helms (1995): zero correlation; constant correlation; time (continuous) dependent correlation ($\rho_{ij} = \sigma^2\rho^{t_j - t_i}$); time (discrete) dependent correlation ($\rho_{ij} = \sigma^2\rho^{(t_j - t_i)^\gamma}$); and age and time (discrete) dependent ($\rho_{ij} = \sigma^2(\rho_1 + \rho_2 \log(t_i))^{(t_j - t_i)^\gamma}$). Wade and Ades (1998) concluded that incorporation of the correlation had little effect on choice of model or on precision of fitted

Table 5.40: Correlation of birth weight with weights up to 2 years for 100 boys and 100 girls from Wroclaw, Poland.

| Age (months) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Boys | 0.87 | 0.76 | 0.62 | 0.52 | 0.47 | 0.45 | 0.41 | 0.43 |
| Girls | 0.82 | 0.58 | 0.53 | 0.49 | 0.44 | 0.43 | 0.44 | 0.44 |
| Age (months) | 9 | 10 | 11 | 12 | 15 | 18 | 21 | 24 |
| Boys | 0.46 | 0.46 | 0.43 | 0.40 | 0.34 | 0.33 | 0.28 | 0.27 |
| Girls | 0.46 | 0.46 | 0.45 | 0.45 | 0.48 | 0.47 | 0.45 | 0.41 |

centiles.

Diggle (1988) discussed modelling correlation in the context of a linear model for repeated measures. Diggle (1988) suggested using the empirical semi-variogram on the residuals from the linear model to suggest a suitable model or check a model for the correlation structure. The $n \times n$ variance matrix $V$ incorporates parameters for measurement error, variation between experimental units and serial correlation within units. The serial correlation component, took on the form $\rho(|t_j - t_i|) = \exp(-\alpha|t_j - t_i|^c)$ where $c$ is 1 (equivalent to AR(1)) or 2. Donnelly et al. (1995) applied the empirical Bayes approach of Laird and Ware (1982) to the longitudinal model proposed by Diggle (1988). Diggle and Verbyla (1998) later estimated the covariance structure non-parametrically by using kernel weighted local linear regression smoothing of the sample variogram ordinates and squared residuals. More recently, Hooper et al. (2002) estimated the covariance structure of estimated foetal weight Z-scores over gestational age using a non-linear regression technique. This method was developed to aid diagnosis of intrauterine growth restriction. Hooper et al. (2002) used a linear combination of logistic basis functions to model the covariance as a function of age. The model for covariance of Z-scores provided a breakdown of the variance into components attributable to measurement error and other factors, that varied with gestational age (Hooper et al. 2002). Hooper et al. (2002) felt that the multi-level approach of Pan and Goldstein (1997), that usually assumes that the error variance is constant, was not appropriate for foetal weights because the error variance decreases substantially over time in pregnancy.

To conclude, it appears that the Cambridge correlations exhibit an upward trend at the end of infancy. The model proposed by Cole (1995) and Cole (1998a) for the correlation structure of weight Z-scores appears to include a quadratic term to model this upturn. The correlations generated from the Newcastle weight Z-scores do not exhibit this trend when age is grouped to the nearest fortnight. Exploratory research suggested that the correlation structure of weight Z-scores was near Markov. Thus suggesting that it may be worth modelling log-correlation. The Argyle model, given by equation (5.39), was found to be a promising model for the correlation structure of weight Z-scores in infancy. The Argyle model has two coefficients and a constant term, and is much simpler than the model proposed by Cole (1995). There is scope for further research on weightings used in weighted least squares to improve the fit of the Argyle model. In Chapter 7 we will explore the adequacy of the Argyle model for calculating conditional weight gain Z-scores.

Figure 5.1: Matrix of scatterplots for weight Z-scores at birth and grouped ages of 6 weeks; 3, 6, 9, 12 and 18 months

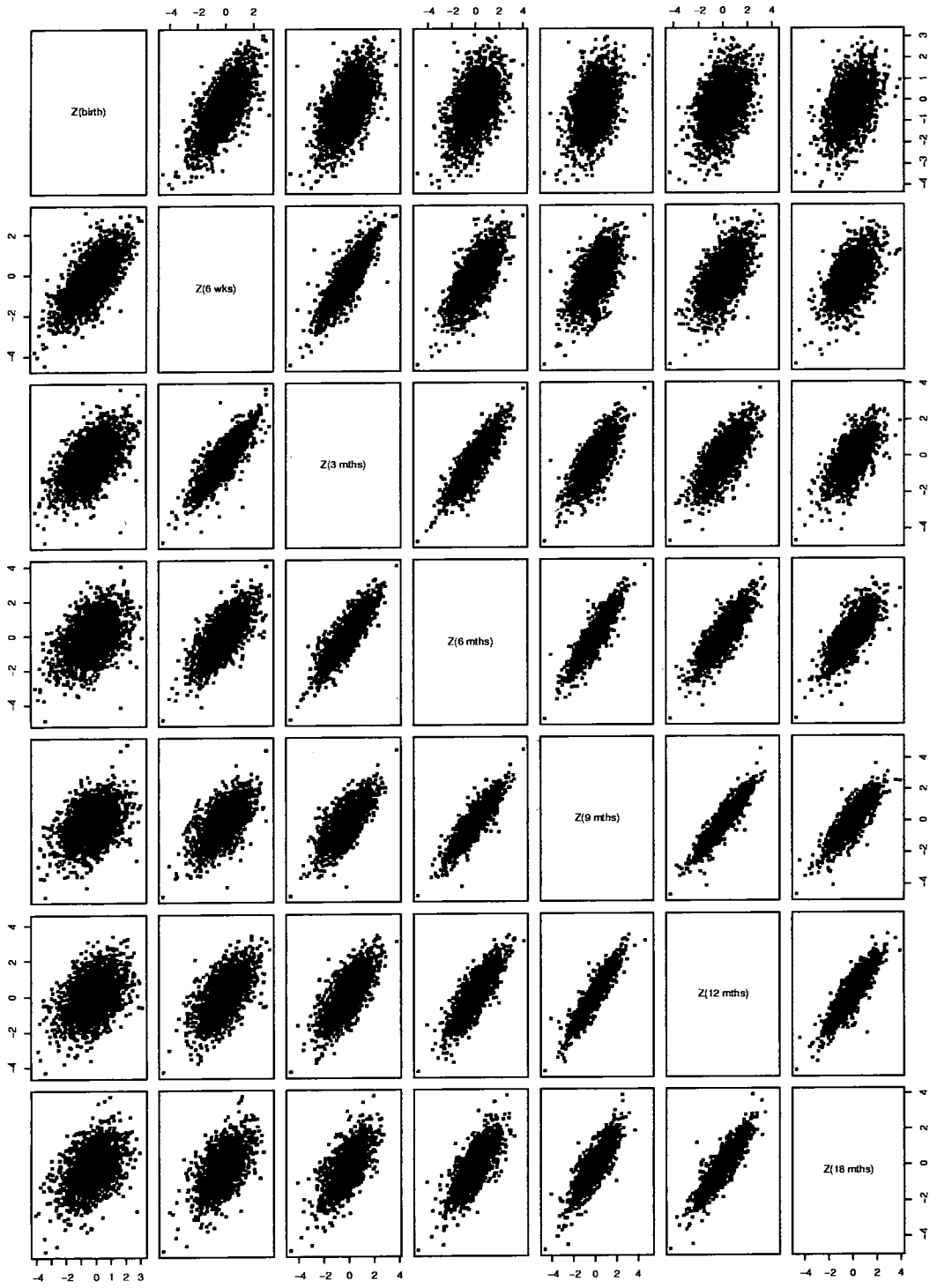


Figure 5.2: Scree plot of variances for principal components

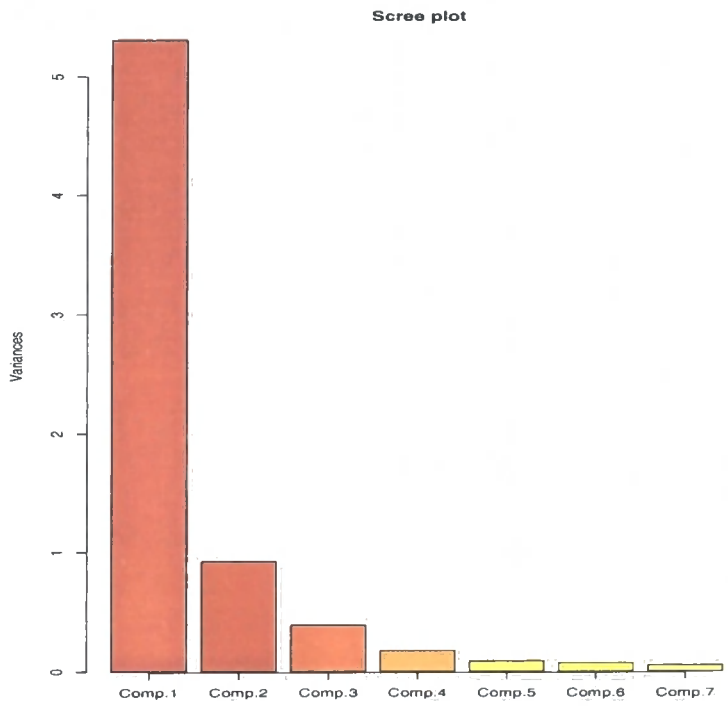


Figure 5.3: Plot of second principal component versus first principal component labelled by: Upper panel case or other Lower panel sex

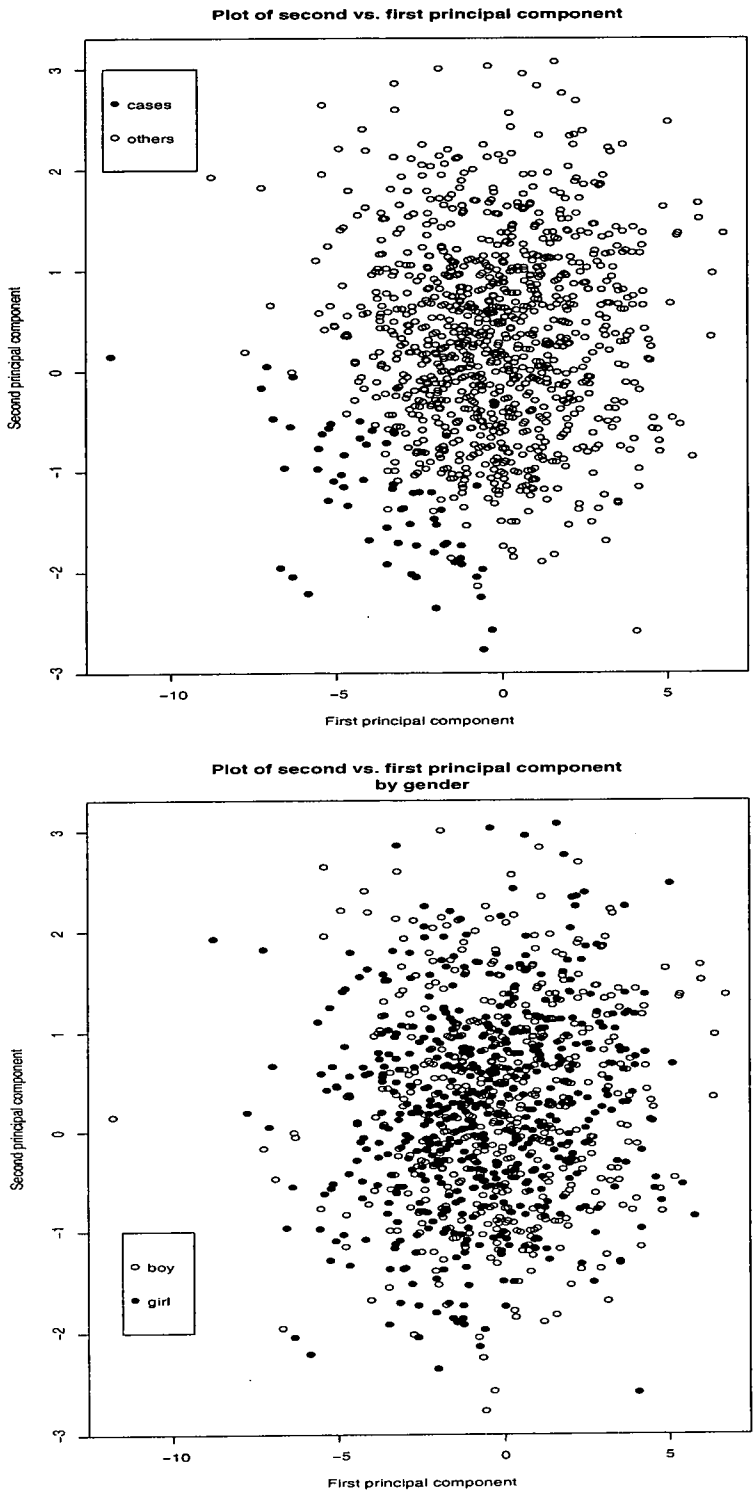


Figure 5.4: Plot of second principal component versus first principal component labelled by: Upper panel level of deprivation Lower panel indicator of gestational age if known

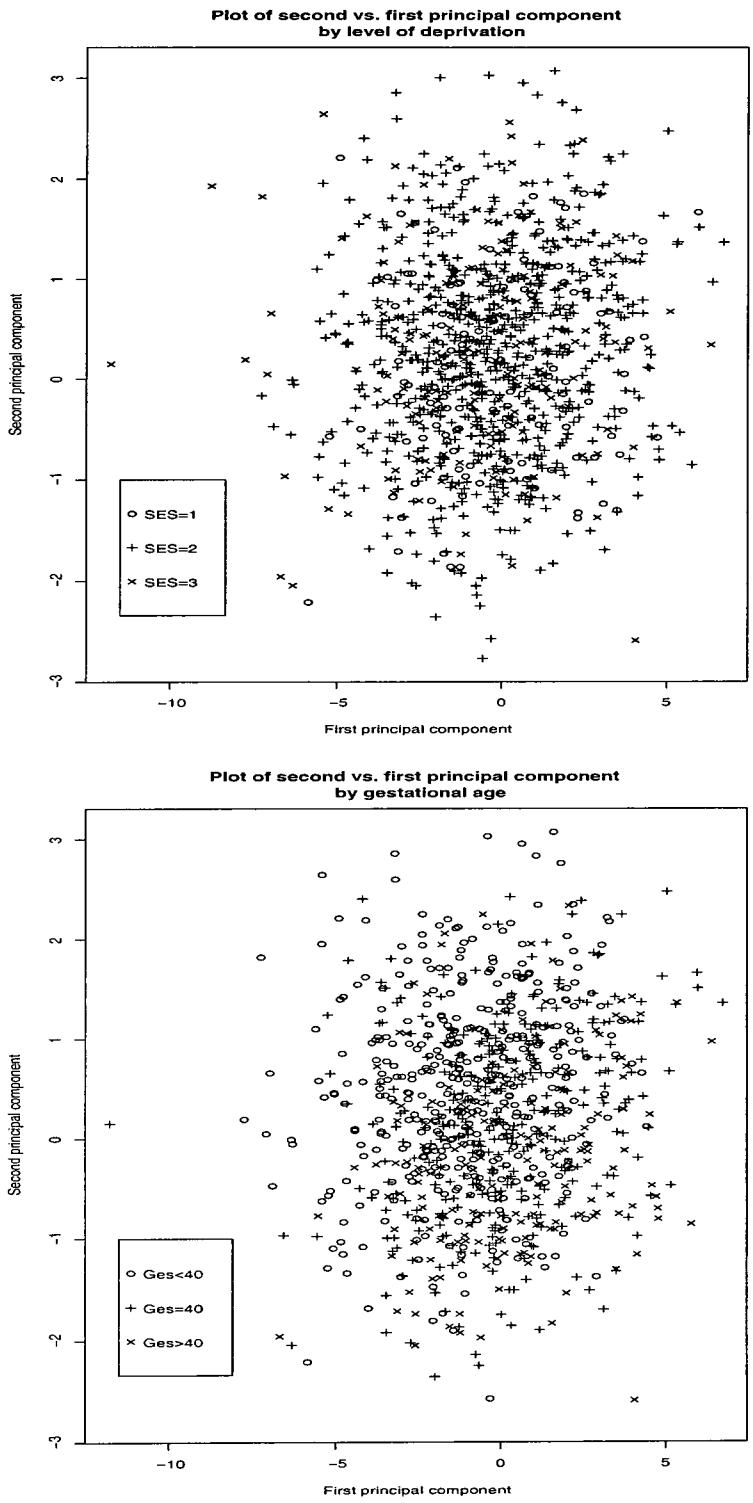


Figure 5.5: Plots of two cubics derived from covariance matrix of weight Z-scores: Left Cubic $4.58\rho^3 - 4.022\rho^2 - 6.426\rho + 5.631$. Right Cubic $2.676\rho^3 - 2.207\rho^2 - 4.454\rho + 3.679$

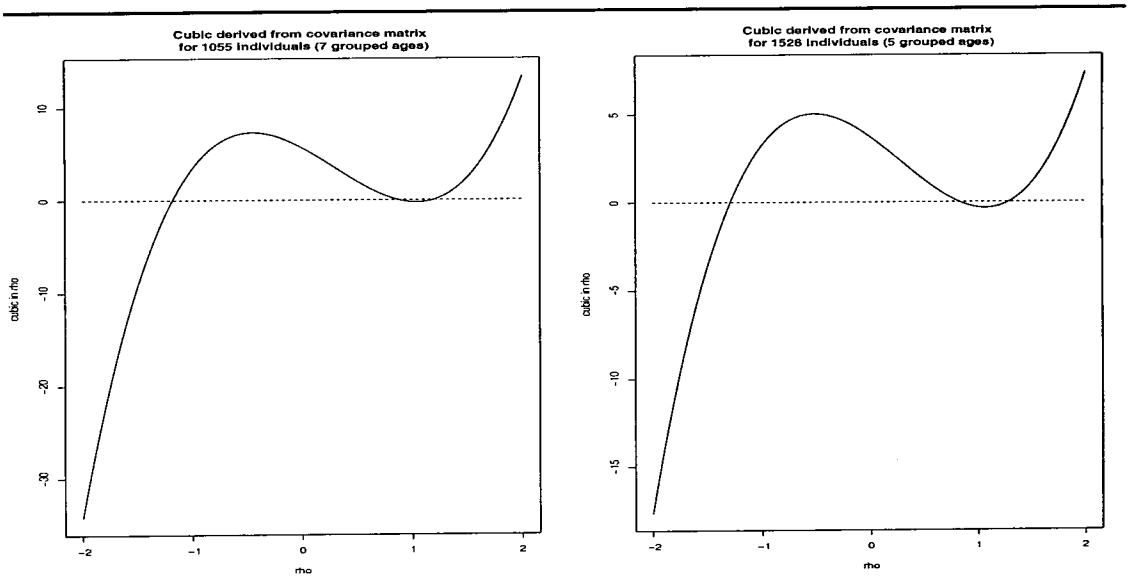


Figure 5.6: Surface plot of correlation at various $Time1$ ($t1$) and $Time2$ ($t2$)

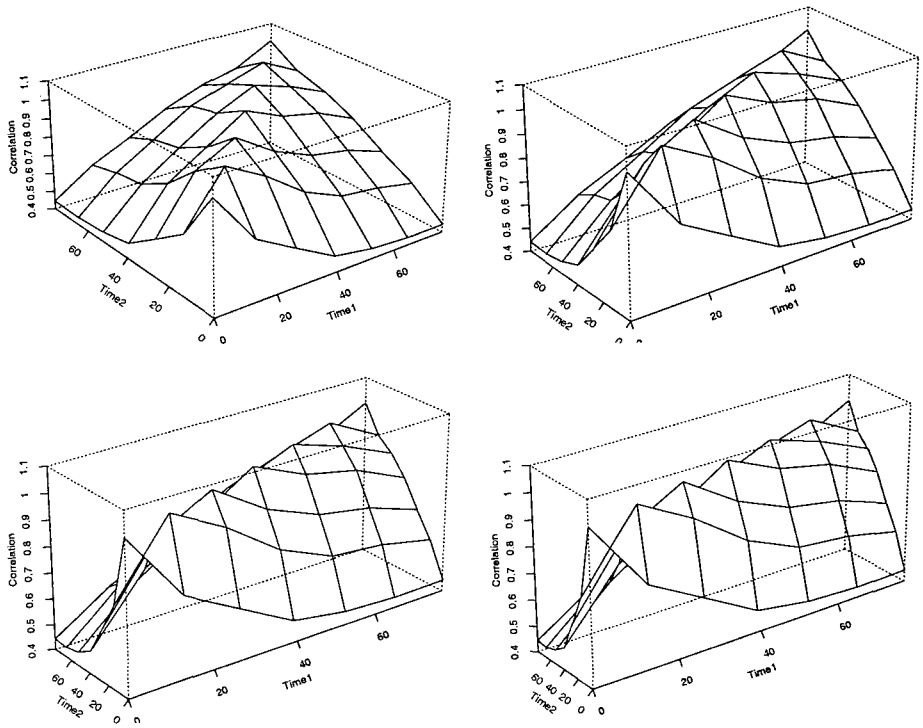


Figure 5.7: **Newcastle correlations for original age grouping:** Scatterplots of correlations versus $t2$ and $t2-t1$ (points that take on the same value of $t1$ are connected)

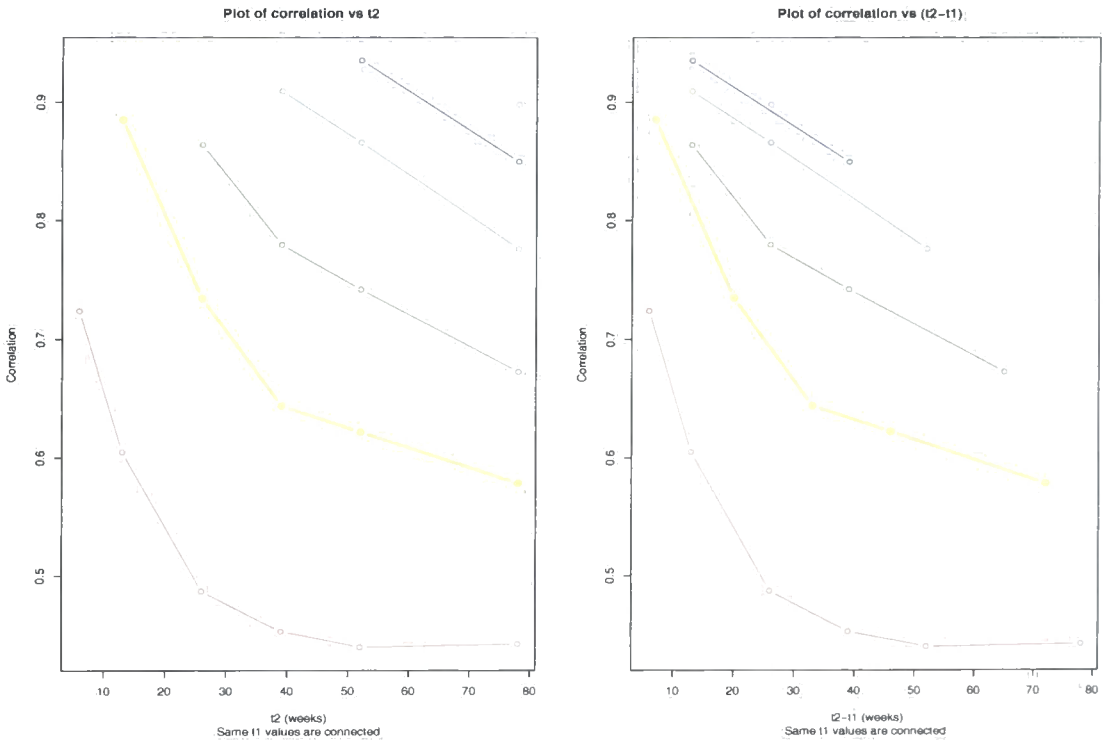


Figure 5.8: **Original Newcastle correlations [N=21]** (based on 1055 individuals with weight Z-scores in 7 age groupings): Plots of standardised residuals versus fitted values for Argyle model with $c = 1$ using ordinary and weighted least squares

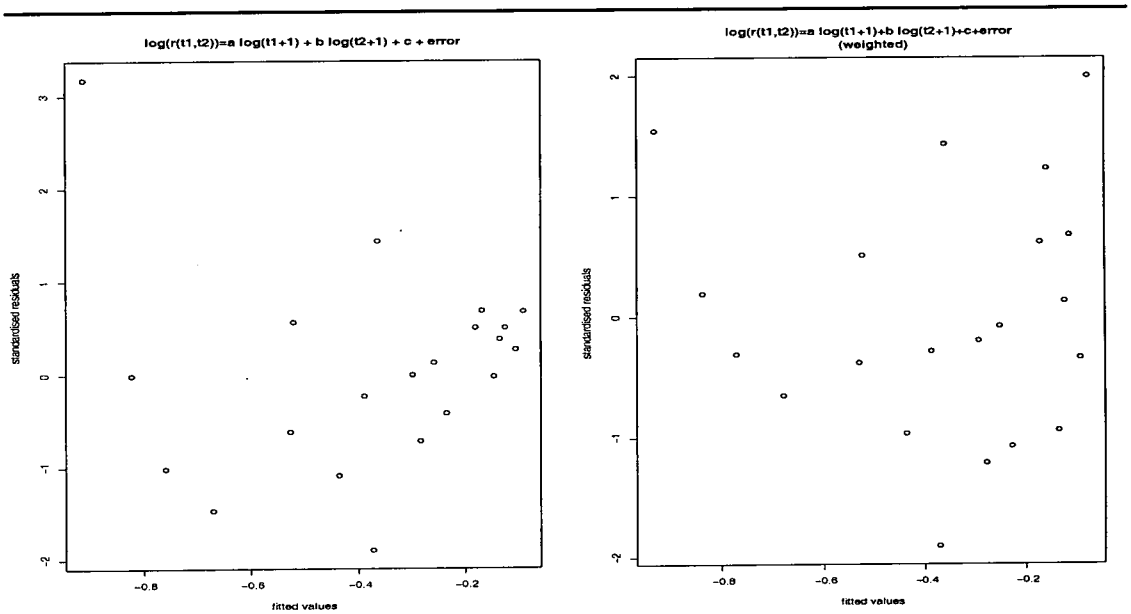


Figure 5.9: **Original Newcastle correlations [N=21]** (based on 1055 individuals with weight Z-scores in 7 age groupings): Plots of standardised residuals versus fitted values for log correlation regressed on $\log(t1+1)$ and $\log(t2-t1)$ using ordinary and weighted least squares

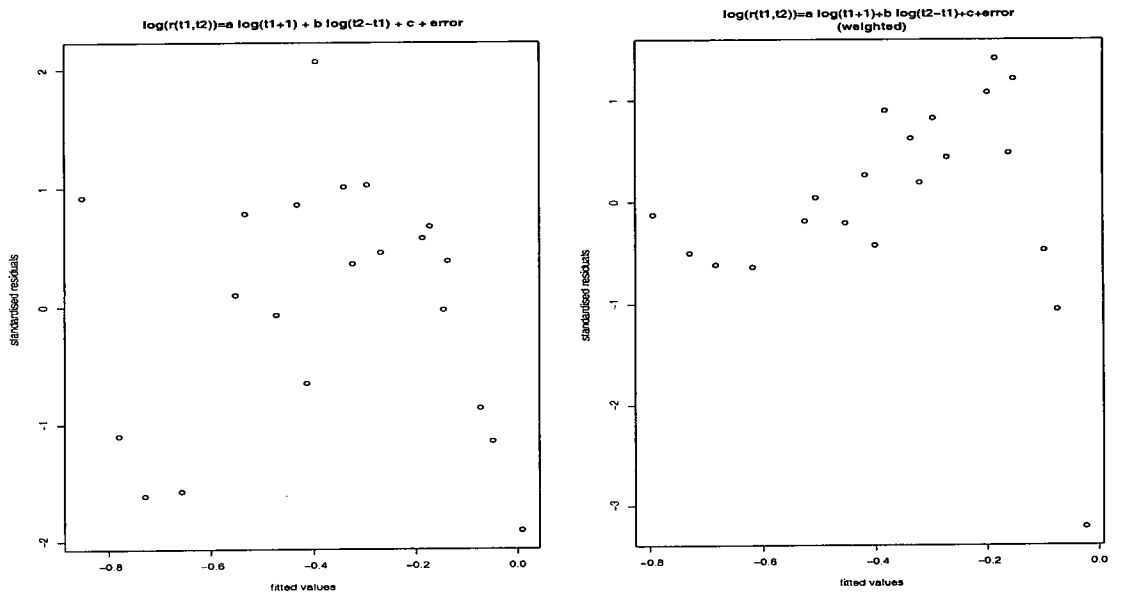


Figure 5.10: **Original Newcastle correlations[N=21]**: Exploratory plots to see how constant ($c$) added in model $\log(r(t1,t2)) = A\log(t1+c) + B\log(t2+c) + C + \epsilon$ effects model fit, term coefficients and intercept (using weighted least squares)

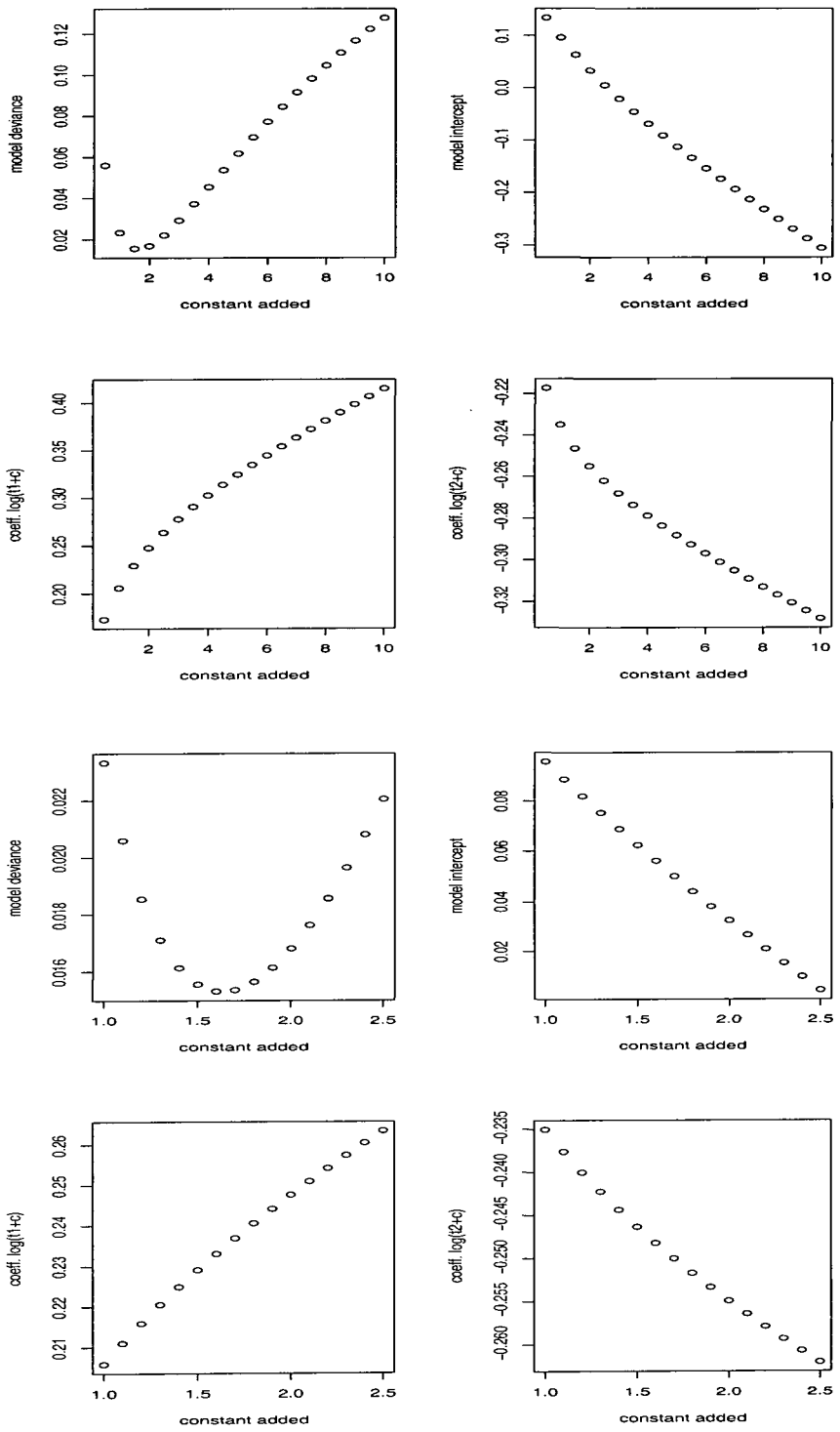


Figure 5.11: Variable width notch box plot of weight Z-scores for birth and six age groupings (6, 13, 26, 39, 52 and 78 weeks). The first and second box-plot for each age group represents Z-scores for 2360 individuals in rest of birth cohort and 1055 individuals with weights in all 7 age groupings, respectively

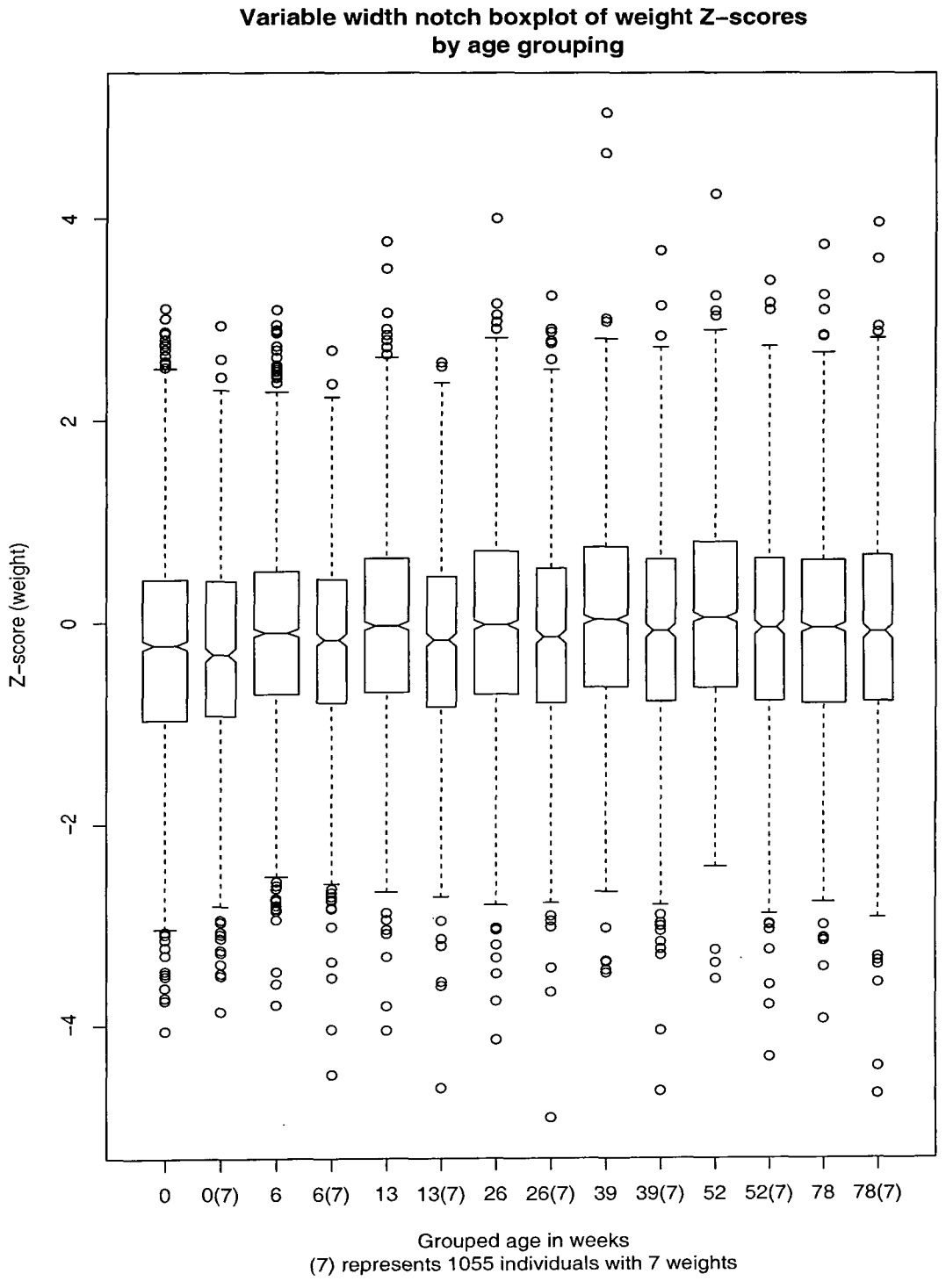


Figure 5.12: Plot of weight Z-score versus age for 1055 individuals with weights in all 7 age groupings (black points correspond to 63 cases and grey points to other children)

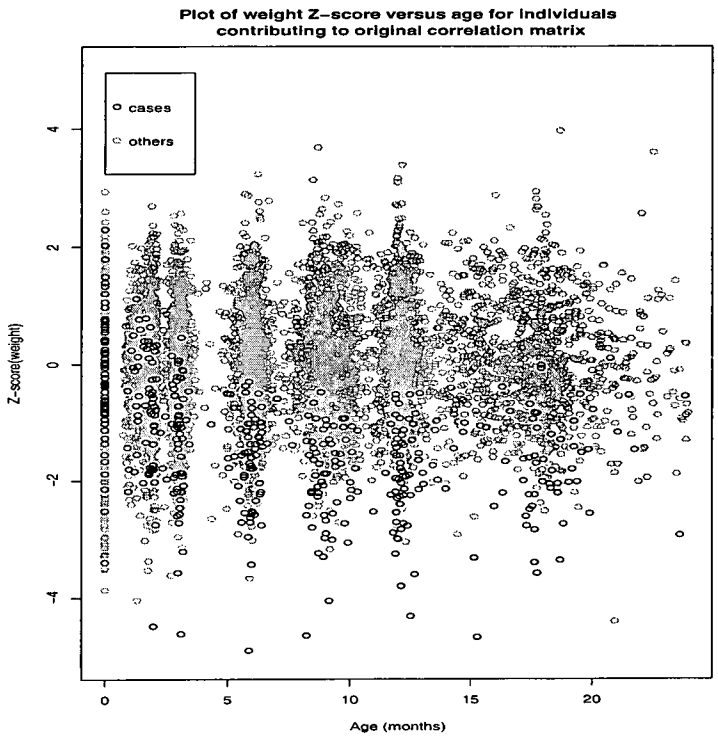


Figure 5.13: Plot of number of individuals contributing to correlation for various age groupings versus $t1$ and $t2$

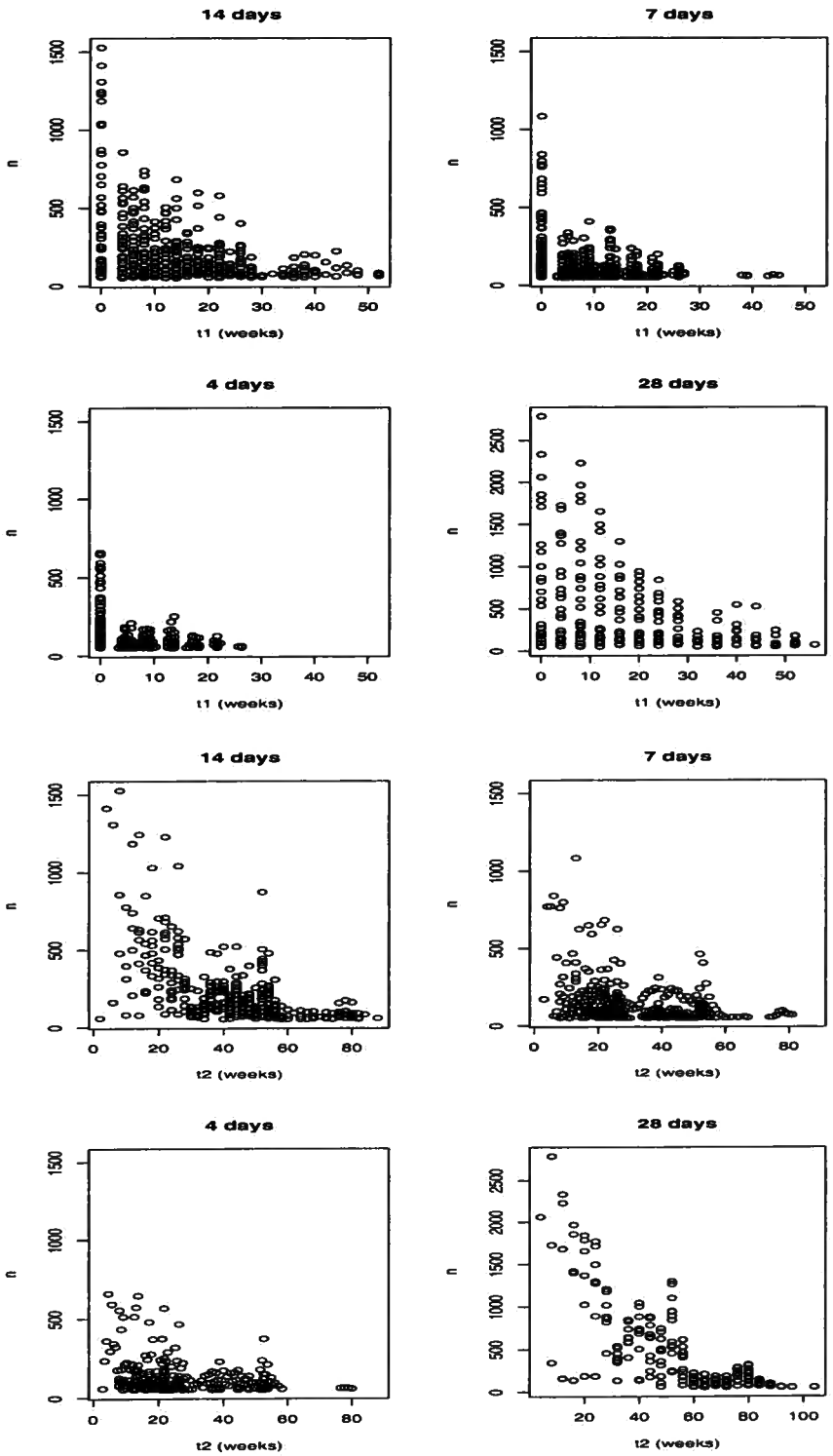


Figure 5.14: Plot of number of individuals contributing to correlation for various age groupings versus $t2 - t1$

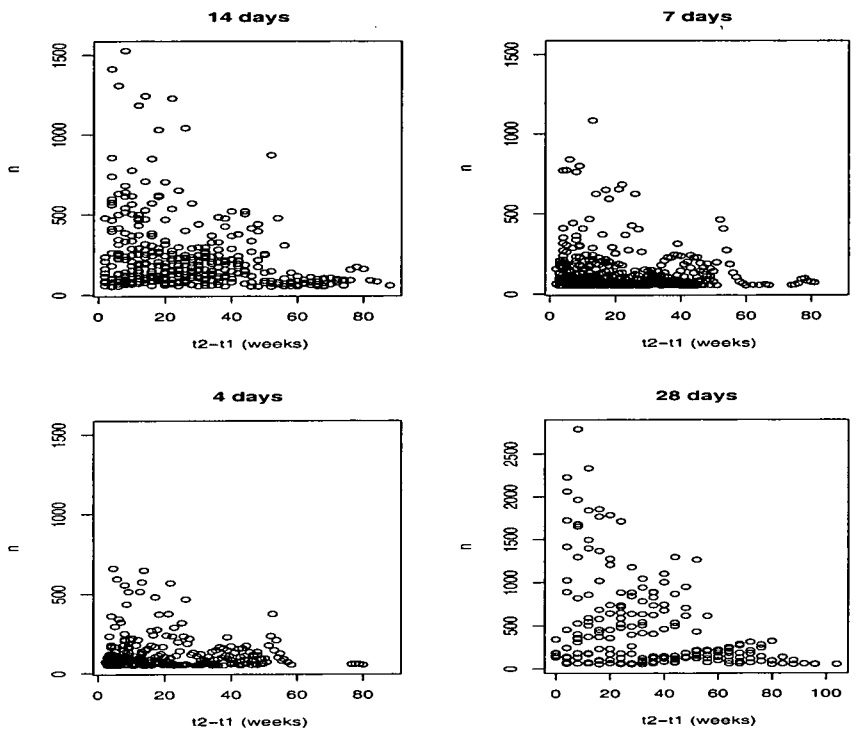


Figure 5.15: **Newcastle correlations with age smoothed to nearest fortnight:** Scatterplots of transformed and untransformed correlations versus $t2$ and $t2 - t1$ (points that have same value of $t1$ are connected)

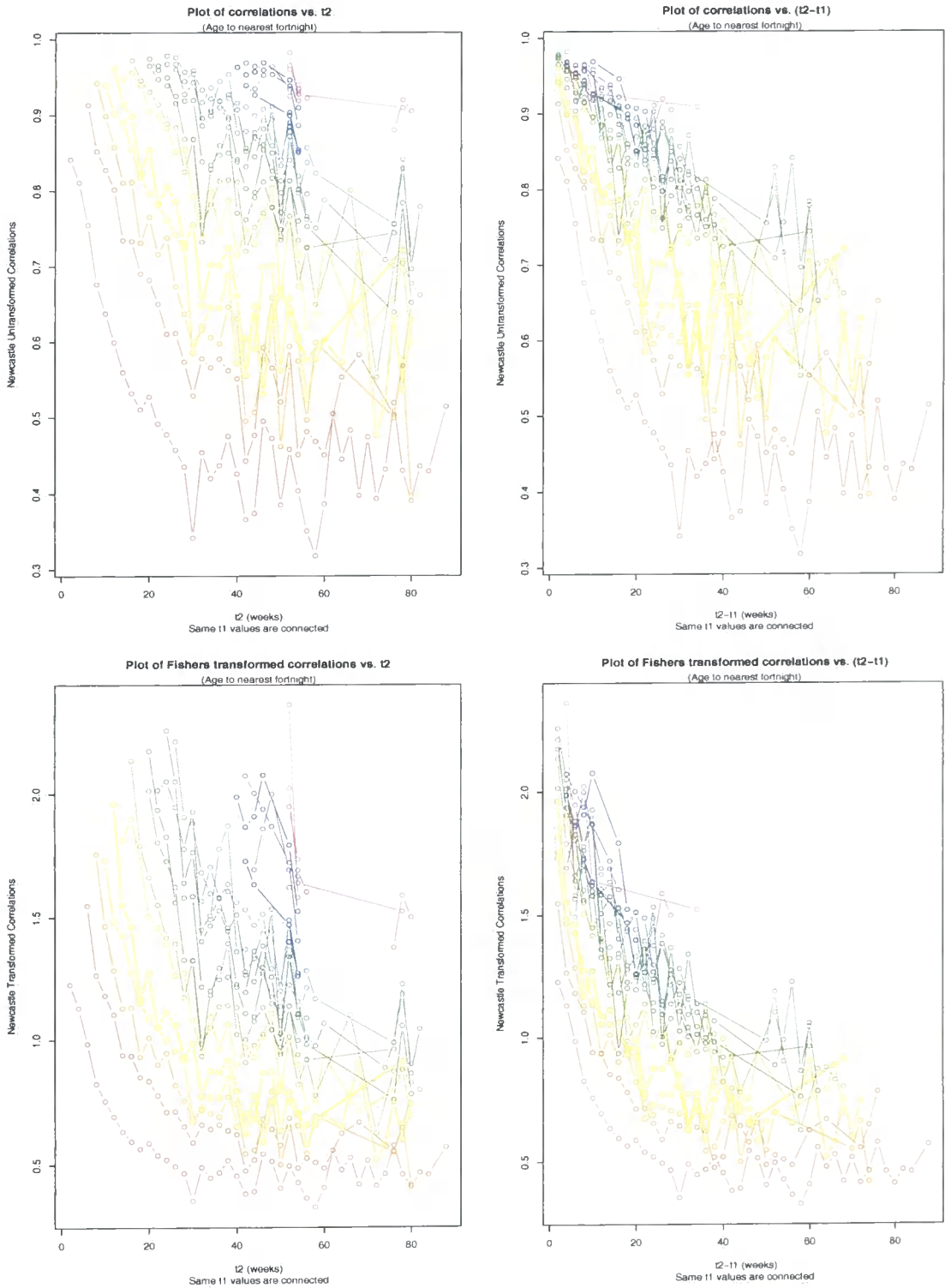


Figure 5.16: **Newcastle correlations with age smoothed to nearest week:** Scatterplots of transformed and untransformed correlations versus $t2$ and $t2 - t1$ (points that take on the same value of $t1$ are of the same colour, the lowess curves for each value of $t1$ are the same colour as the scatter of points that derived them)

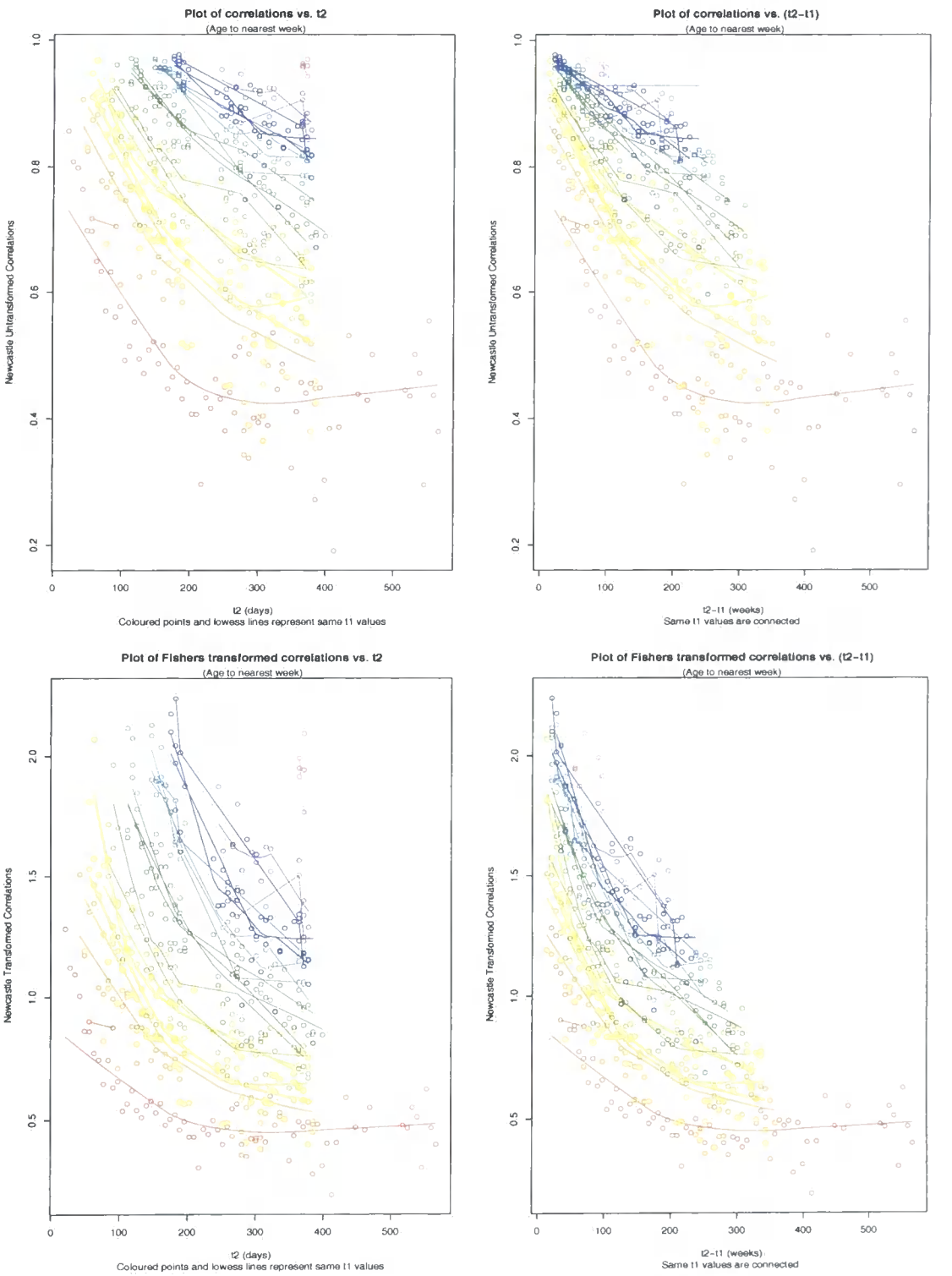


Figure 5.17: **Age to nearest fortnight:** Scatterplots of standardised residuals versus fitted values for Argyle model ($c = 1$). Left Argyle model fitted using ordinary least squares Right Argyle model fitted using weighted least squares

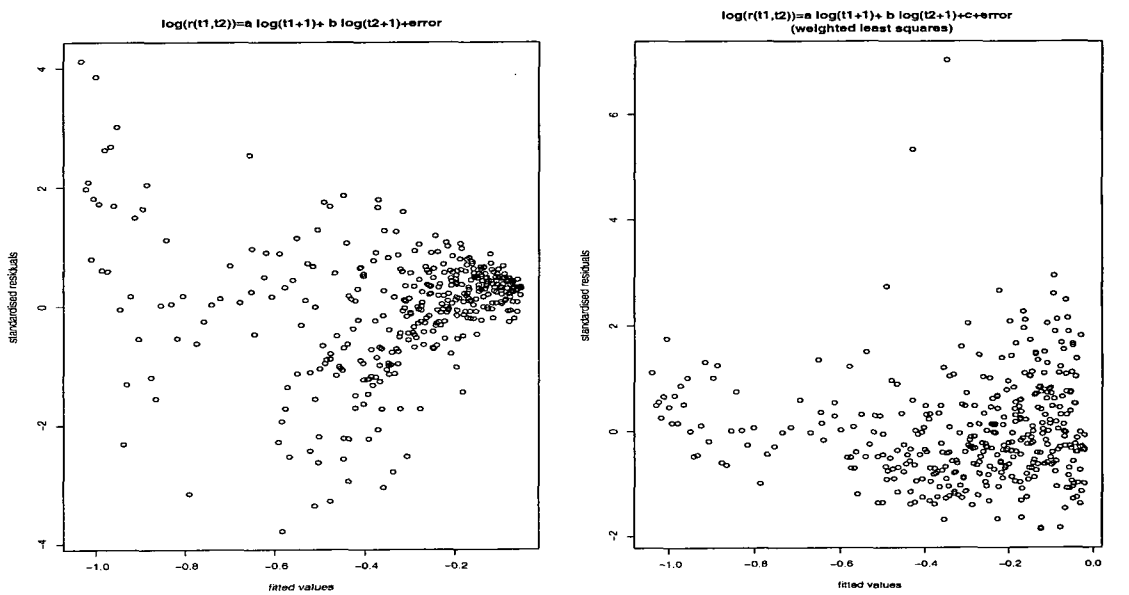


Figure 5.18: **Age to nearest fortnight:** scatterplots of standardised residuals versus fitted values for Argyle model using weighted least squares: Upper panel labelled by size of correlation, red points through to violet points are low correlations (min($r$)=0.32) through to high correlations (max($r$)=0.98) Lower panel labelled by indicator of sample size ($n < 500$ or $n \geq 500$)

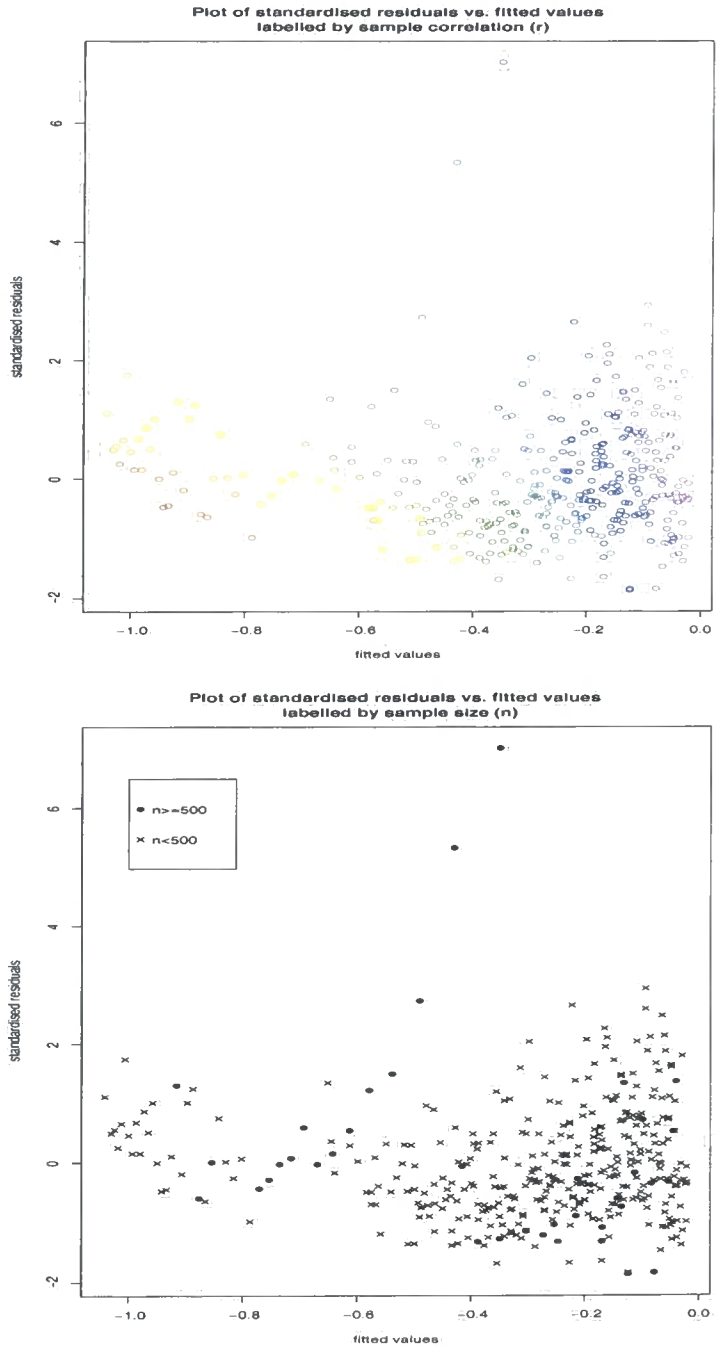


Figure 5.19: **Age to nearest fortnight:** Plot of Cook's distance versus index Left Argyle model (c=1) Right Argyle model (c=2.3). Index runs from 1 to 385, where 1 is the correlation between weight Z-scores at birth and 2 weeks, and 385 is the correlation between weight Z-scores at 52 and 80 weeks. These indexes represent ordering of the correlations by $t1$ values, then ordered by $t2$ values.

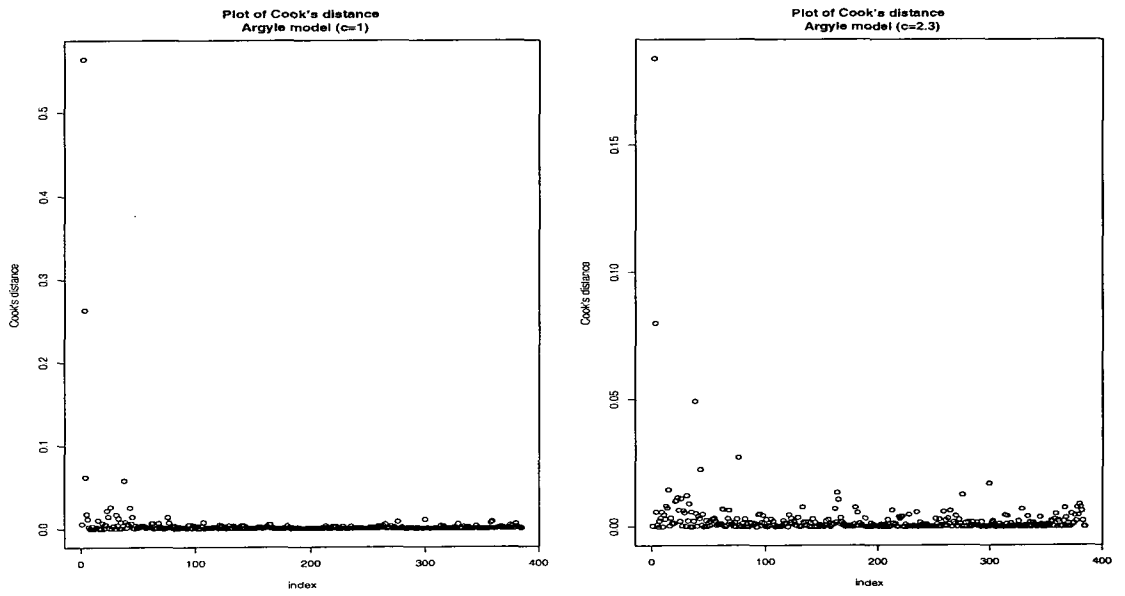


Figure 5.20: **fortnightly correlations**: Exploratory plots to see how constant (c) added in model $\log(r(t1,t2)) = A\log(t1+c) + B\log(t2+c) + C + \epsilon$ effects model fit, term coefficients and intercept

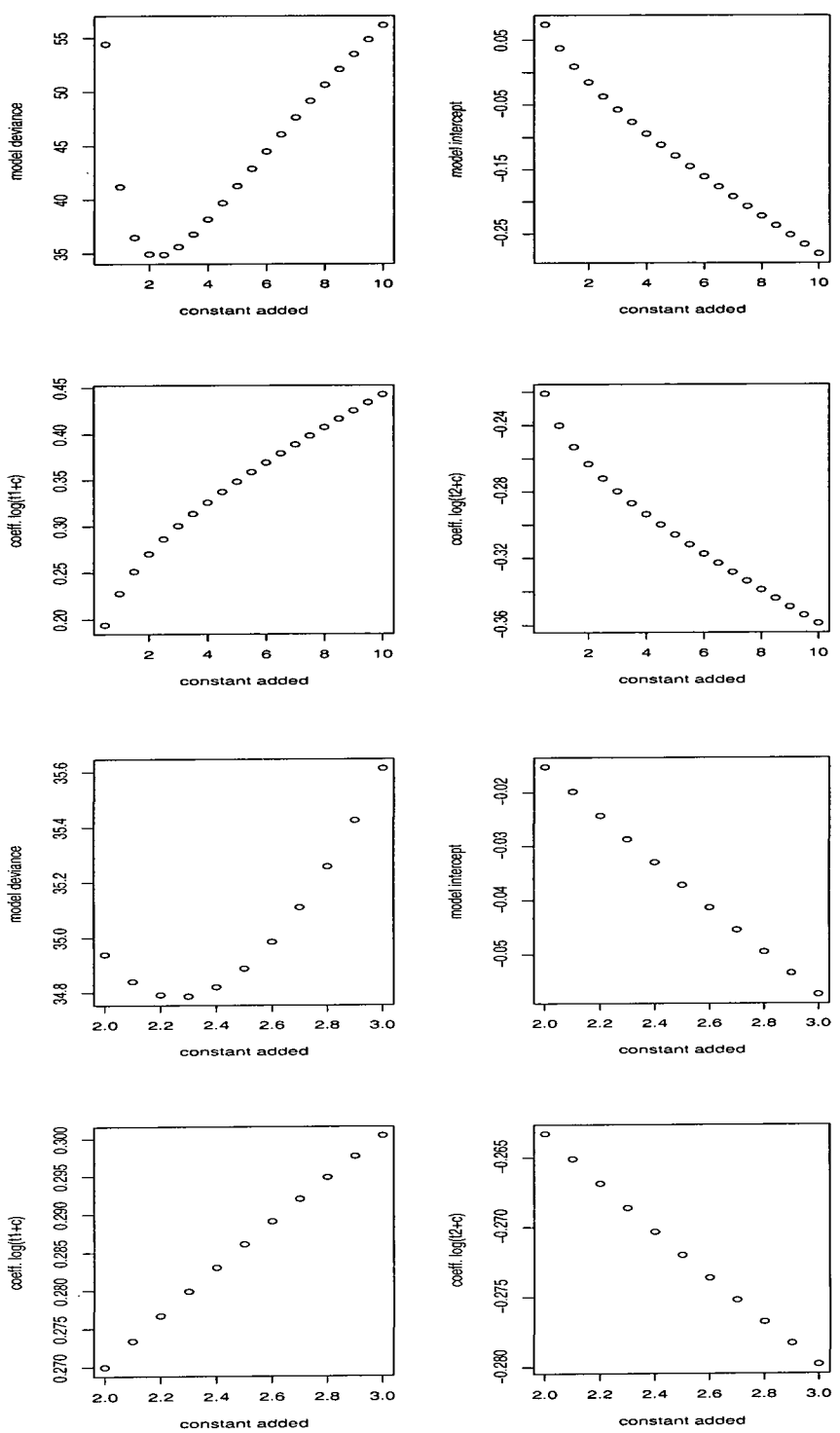


Figure 5.21: **fortnightly correlations (excluding two most influential)**: Exploratory plots to see how constant (c) added in model $\log(r(t1,t2)) = A\log(t1+c) + B\log(t2+c) + C + \epsilon$ effects model fit, term coefficients and intercept

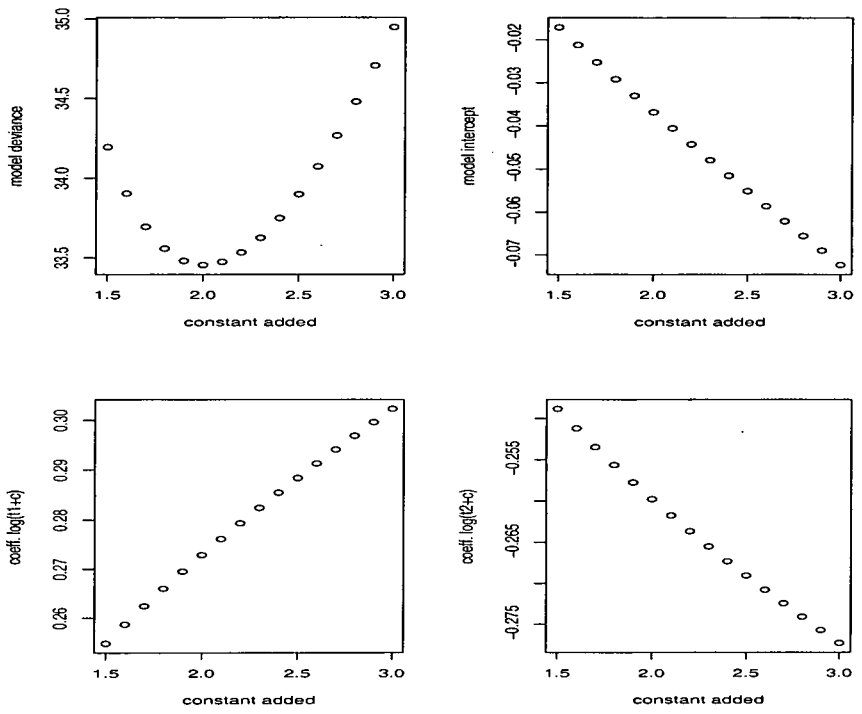


Figure 5.22: **Age to nearest fortnight:** Scatterplots of standardised residuals versus fitted values. Left Argyle model with $c = 2.3$. Right After excluding two most influential observations: Argyle model with $c = 2$.

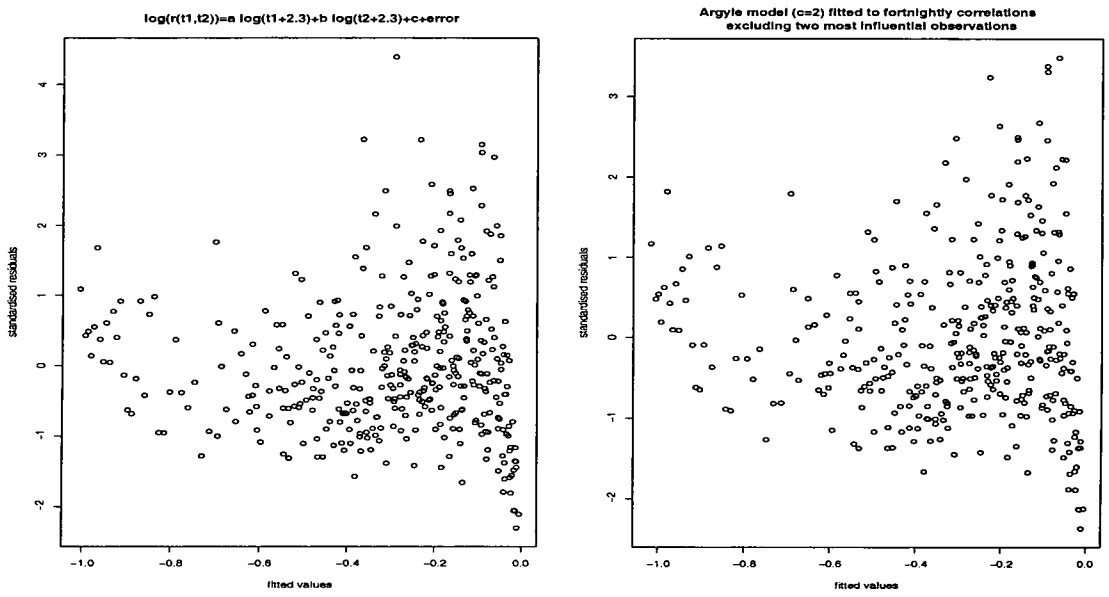



Figure 5.23: **Age to nearest week:** Scatterplots of standardised residuals versus fitted values Left Argyle model with $c = 1$ fitted using ordinary least squares Right Argyle model with $c = 1$ fitted using weighted least squares

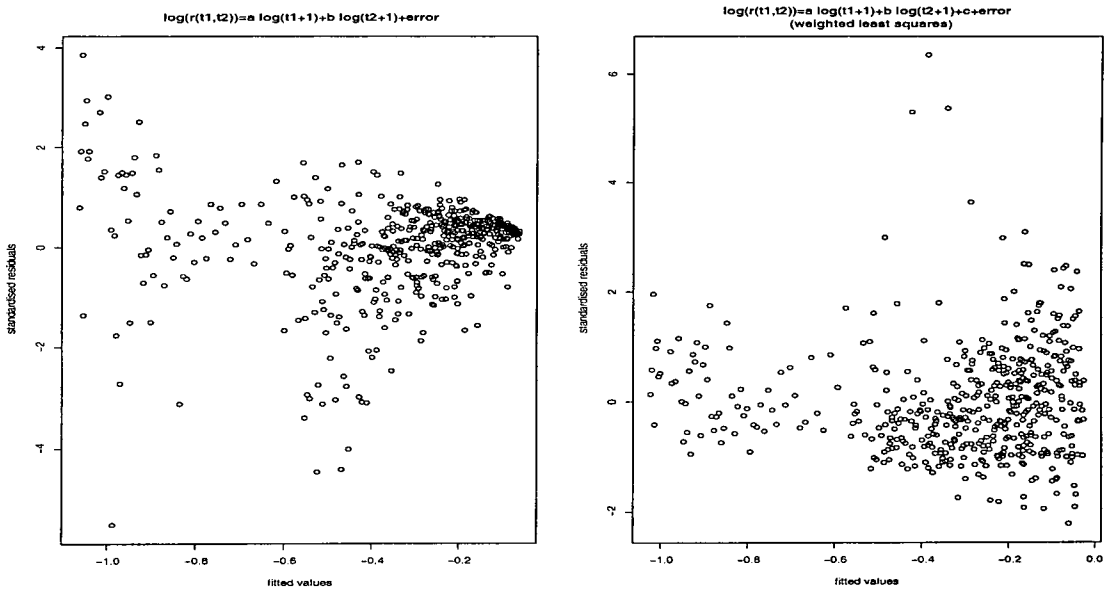


Figure 5.24: **Newcastle correlations derived from age grouped to nearest week**: Plot of Cook's distance versus index Left Argyle model ($c = 1$) Right Argyle model ($c$=2.9)

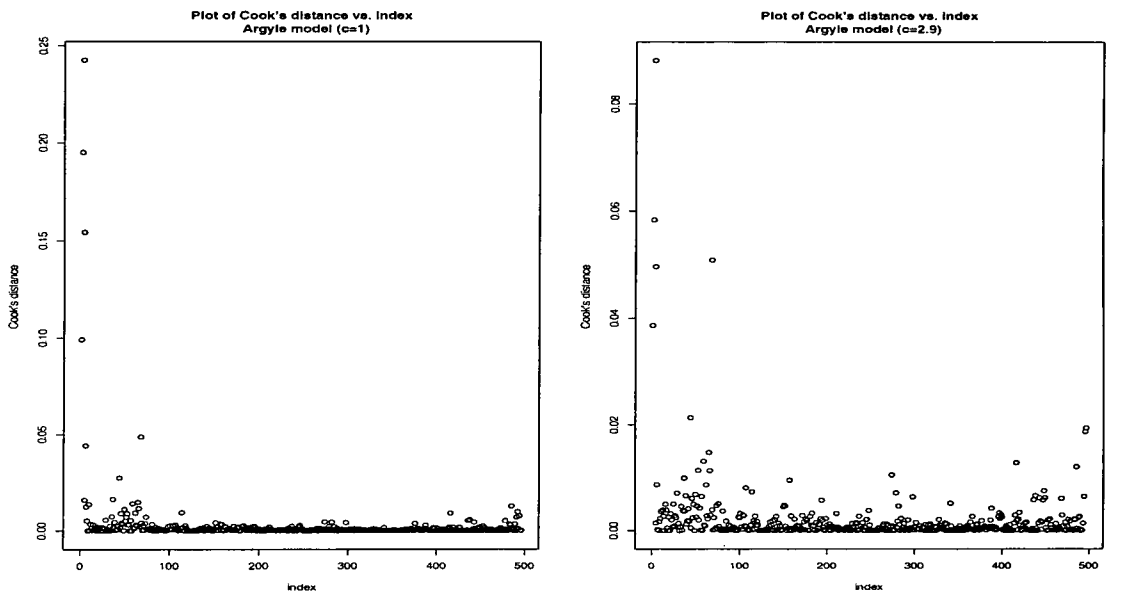


Figure 5.25: **Weekly correlations**: Exploratory plots to see how constant (c) added in model $\log(r(t1,t2)) = A\log(t1+c)+B\log(t2+c)+C+\epsilon$ effects model fit, term coefficients and intercept

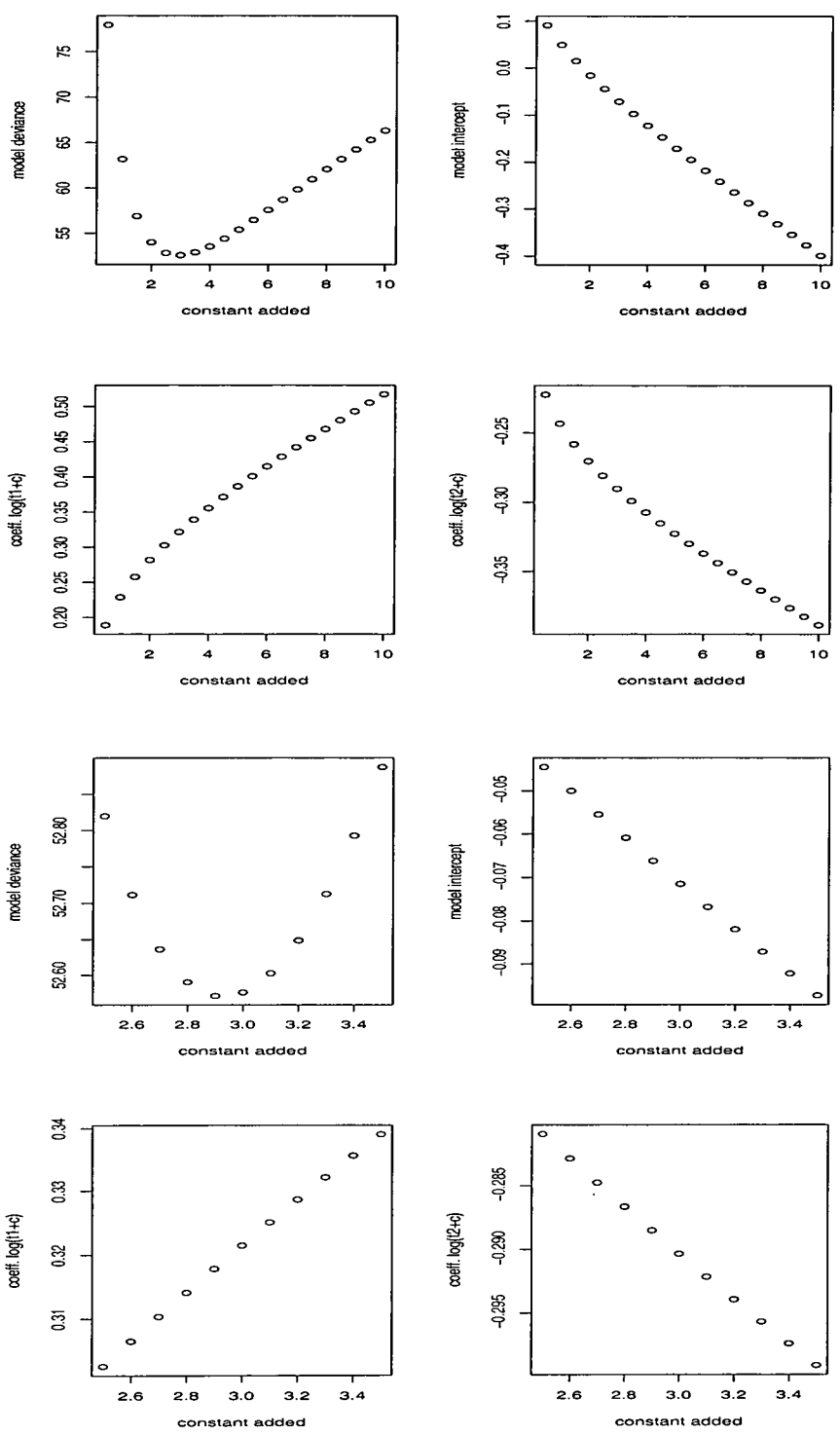


Figure 5.26: **Weekly correlations** (excluding correlation between weight Z-scores at birth and 5 weeks): Exploratory plots to see how constant (c) added in model $\log(r(t1,t2)) = A\log(t1+c) + B\log(t2+c) + C + \epsilon$ effects model fit, term coefficients and intercept

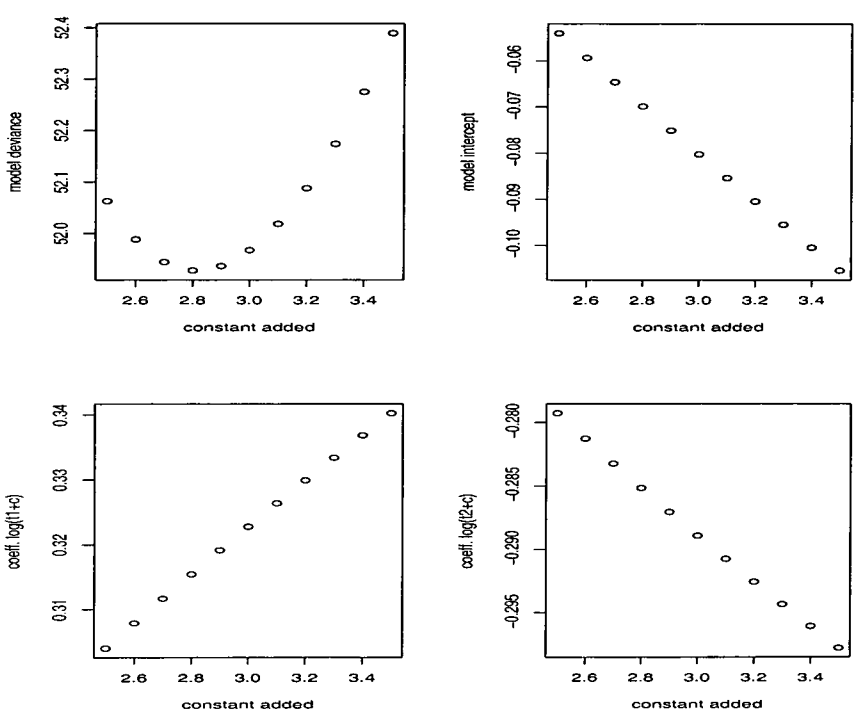


Figure 5.27: **Age to nearest week:** Scatterplots of standardised residuals versus fitted values Left Argyle model fitted with $c = 2.9$ using weighted least squares Right Argyle model fitted with $c = 2.8$ using weighted least squares (after excluding the most influential observation)

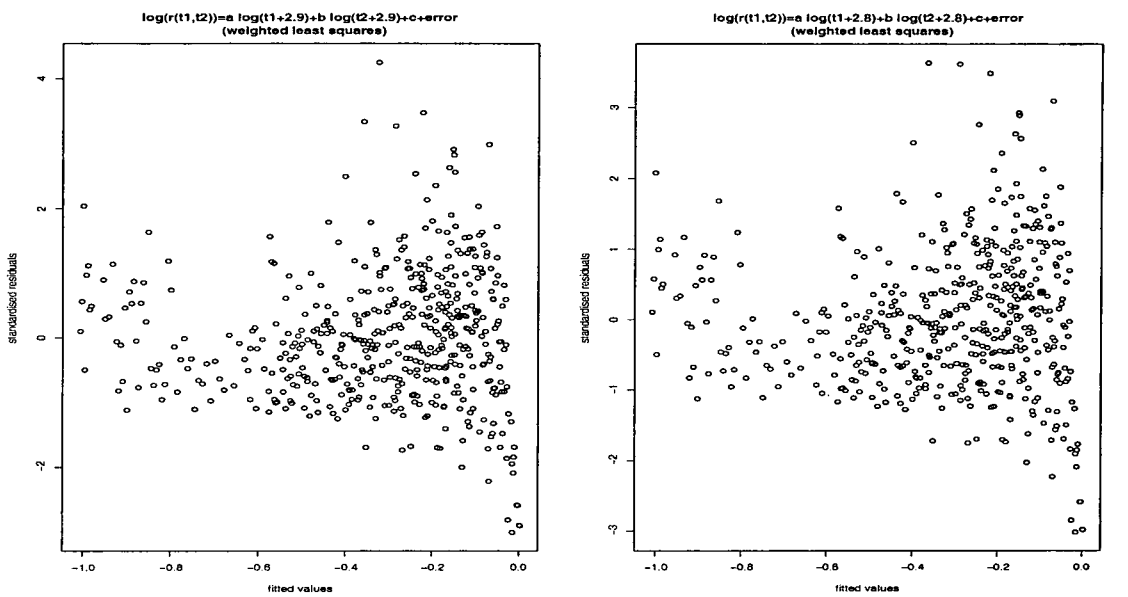


Figure 5.28: **Cambridge correlations (including birth weight):** Scatterplots of transformed and untransformed correlations versus $t2$ and $t2 - t1$ (points that take on the same value of $t1$ are connected)

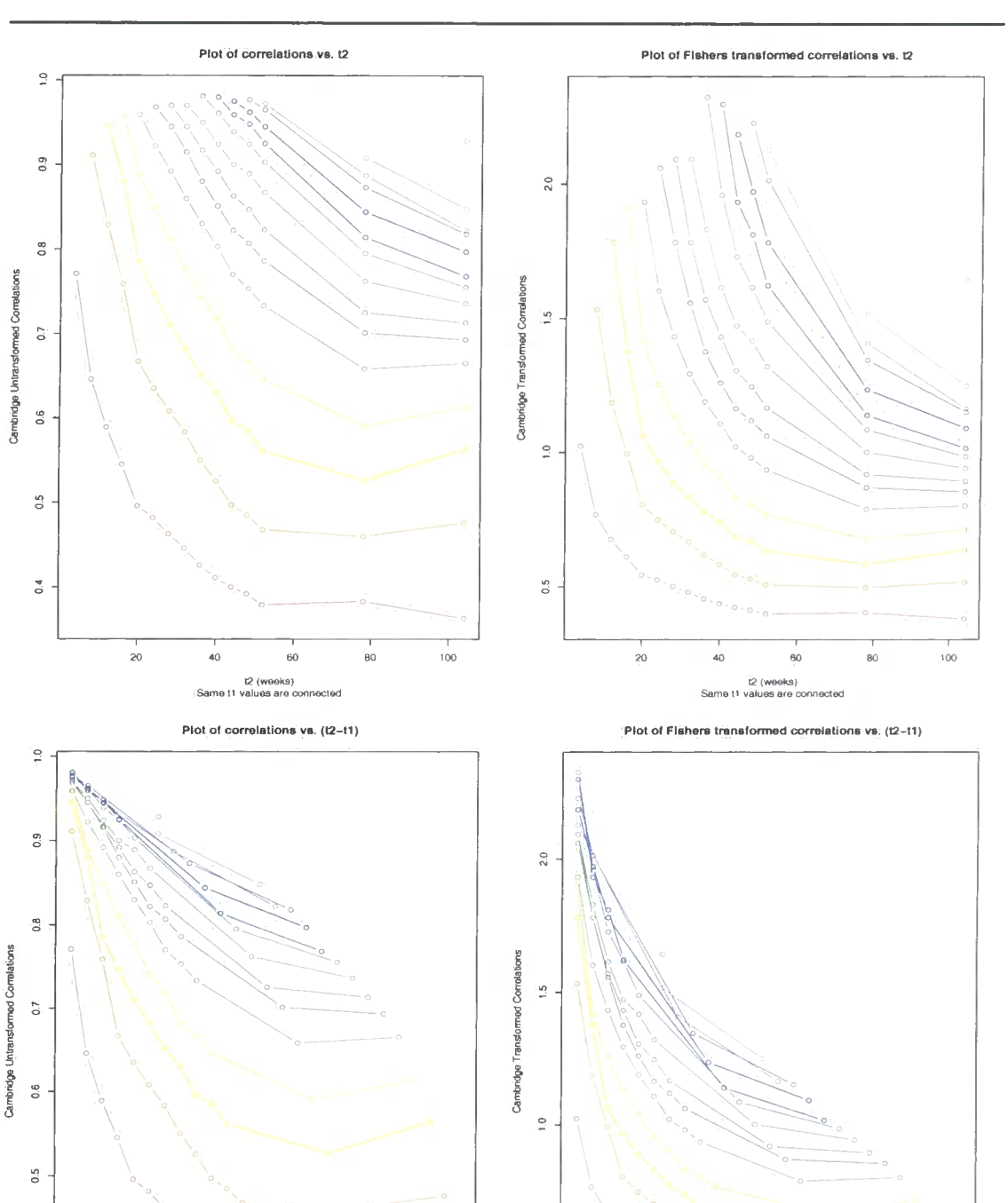


Figure 5.29: **Cambridge correlation matrix (including birth weight)**[N=120,n=221]: Left Plots of standardised residuals versus fitted values for Cole's (1995,1998) model Right Plot of Cook's distance versus index for Cole's (1995,1998) model (Index runs from 1 to 120, where 1 is the correlation between weight Z-scores at birth and 4 weeks, and 120 is the correlation between weight Z-scores at 78 and 104 weeks. This index represent ordering of the correlations by $t1$ values followed by ordering of $t2$ values).

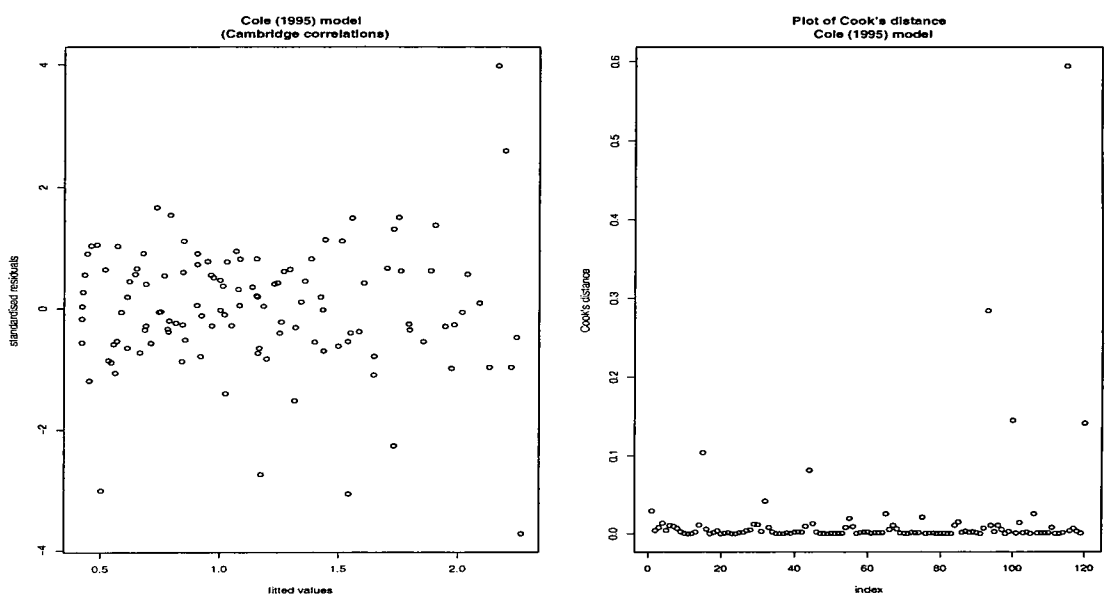


Figure 5.30: **Cambridge Correlation matrix (excluding birthweight)** Top left Box-plots of correlation, Fisher's transformation of correlation and log correlation, Top right Plot of Fisher's transformation of correlation vs. correlation, Bottom left Plot of log correlation vs. correlation, Bottom right Plot of log correlation vs. Fisher's transformation

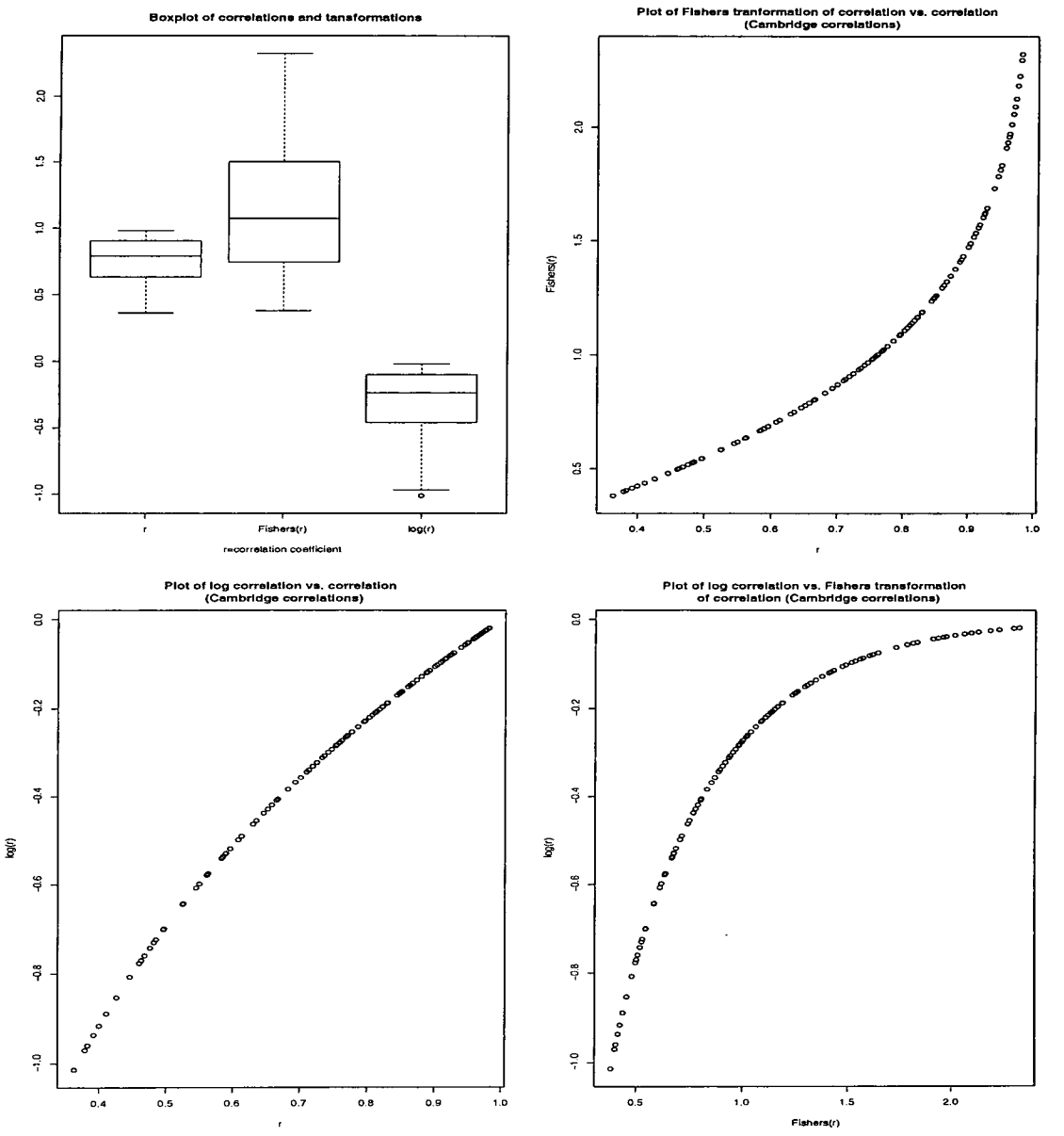


Figure 5.31: **Newcastle correlations:** Plots of standardised residuals versus fitted values for model proposed by Cole (1995)

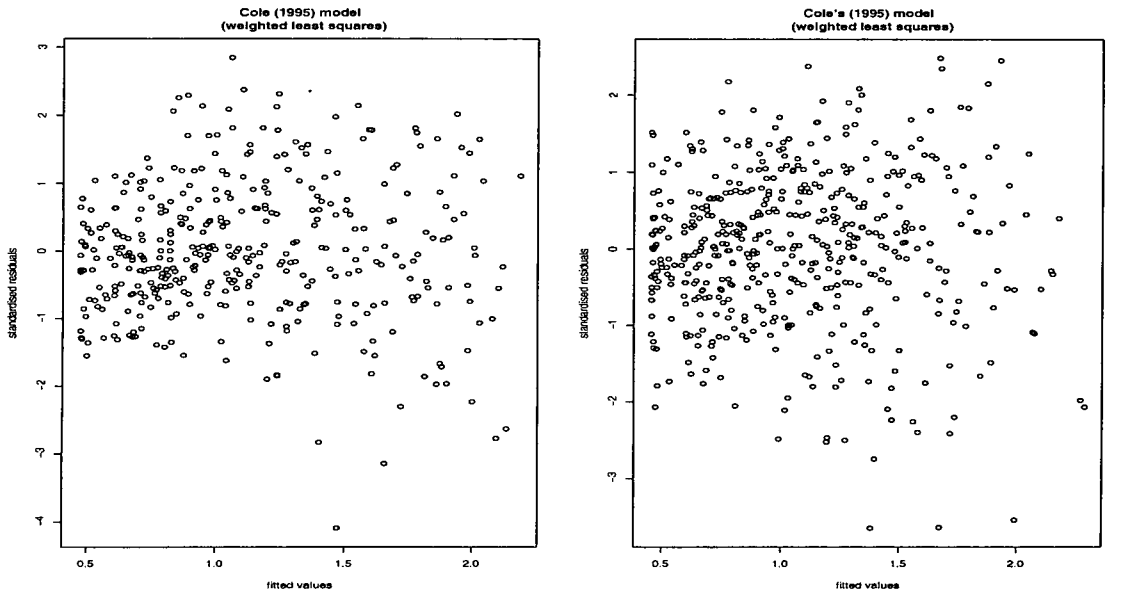



Figure 5.32: **Newcastle correlations**: Plots of Cook's distance versus index for model proposed by Cole (1995) Left Newcastle correlations derived from age grouped to nearest fortnight Right Newcastle correlations derived from age grouped to nearest week

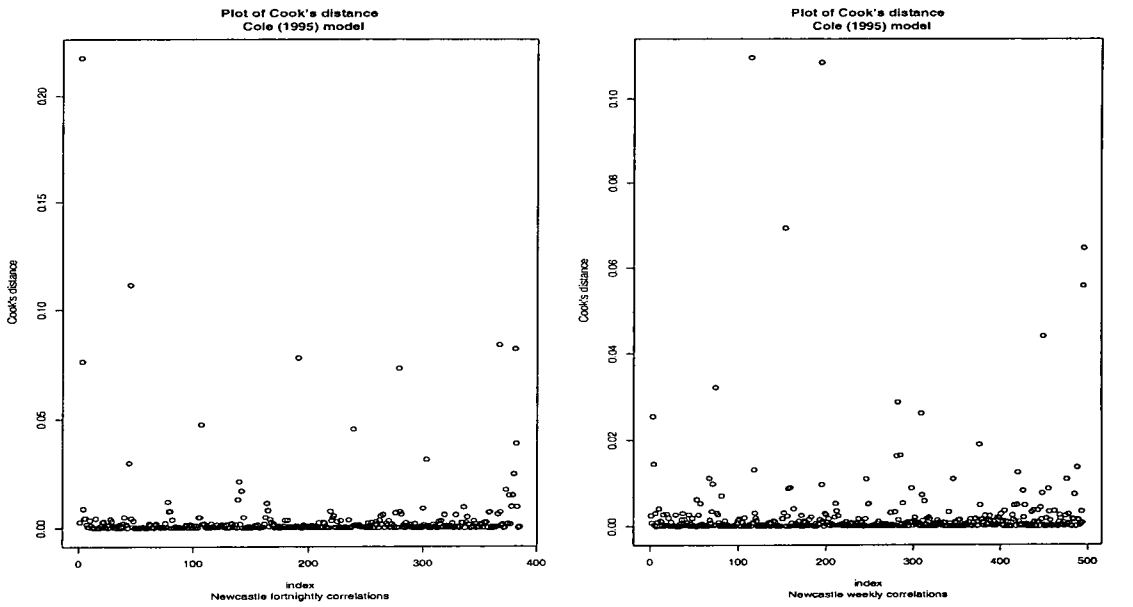


Figure 5.33: **Cambridge correlation matrix (including birth weight)**[N=120,n=221]: Left Plots of standardised residuals versus fitted values for Argyle model Right Plot of Cook's distance versus index for Argyle model ($c = 1$) (Index runs from 1 to 120, where 1 is the correlation between weight Z-scores at birth and 4 weeks, and 120 is the correlation between weight Z-scores at 78 and 104 weeks. This index represent ordering of the correlations by $t1$ values followed by ordering of $t2$ values).

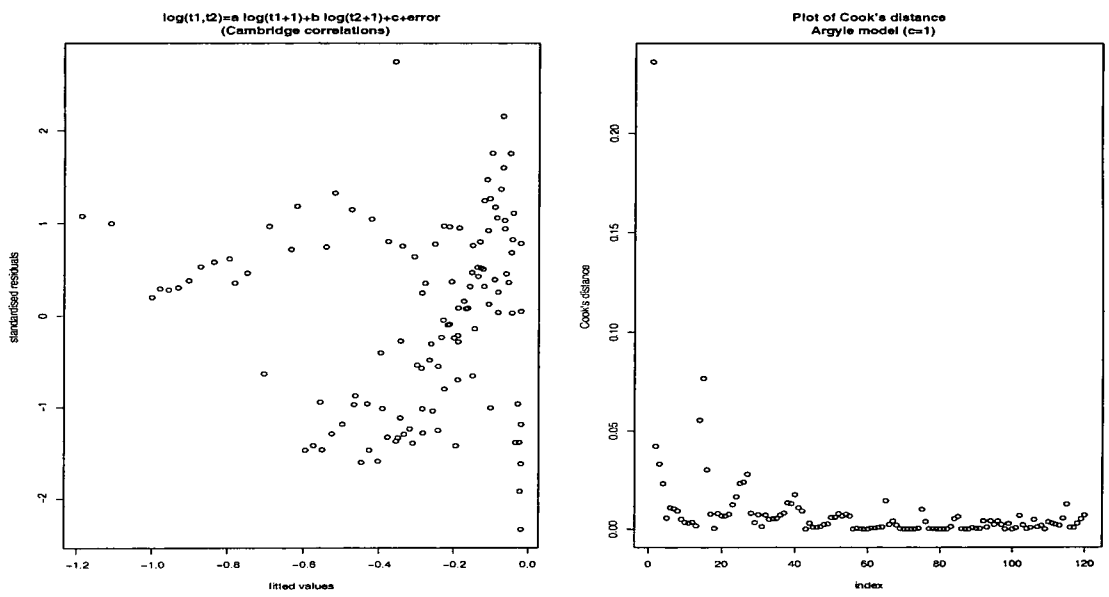


Figure 5.34: **Cambridge correlations (including birth weight)**: Exploratory plots to see how constant (c) added in model $\log(r(t1,t2)) = A\log(t1+c) + B\log(t2+c) + C + \epsilon$ effects model fit, term coefficients and intercept

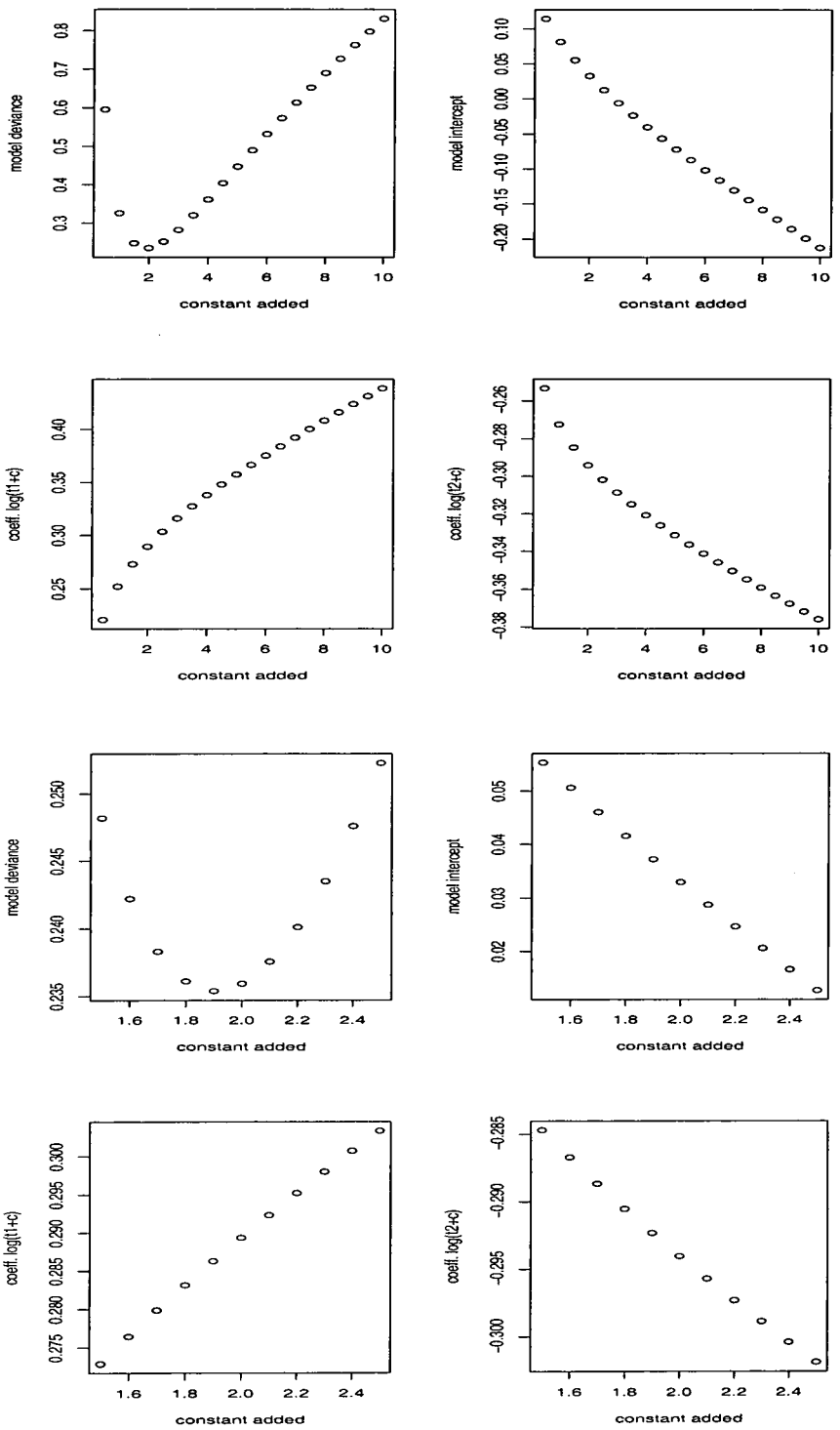


Figure 5.35: **Cambridge correlation matrix (including birth weight)**[N=120,n=221]: Left Plots of standardised residuals versus fitted values for Argyle model with $c$=1.9 Right Plot of Cook's distance versus index for Argyle model with $c$=1.9 (Index runs from 1 to 120, where 1 is the correlation between weight Z-scores at birth and 4 weeks, and 120 is the correlation between weight Z-scores at 78 and 104 weeks. This index represent ordering of the correlations by $t1$ values followed by ordering of $t2$ values).

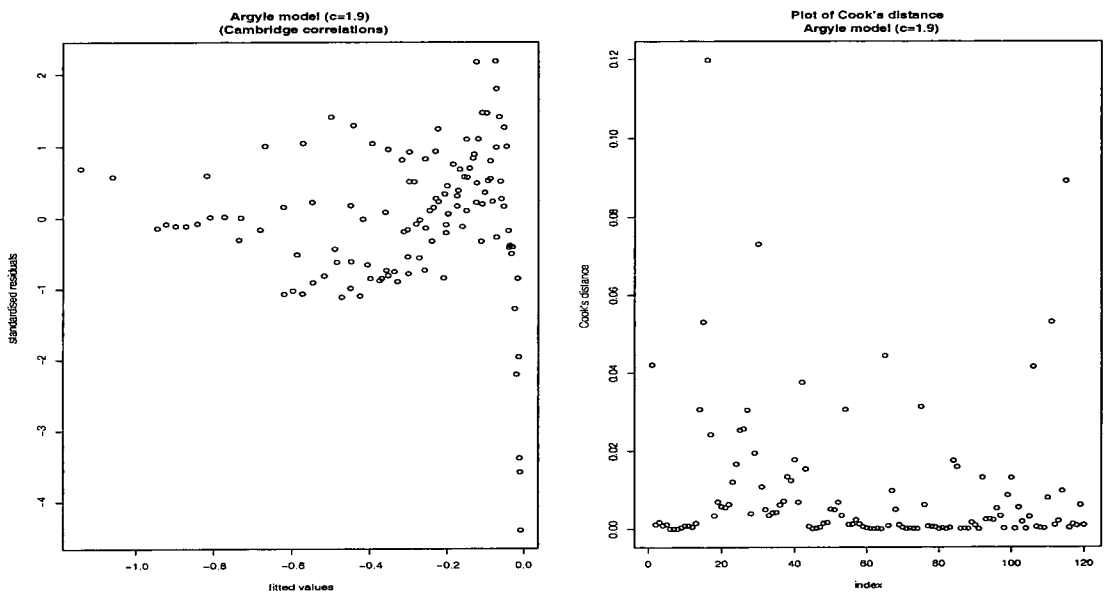


Figure 5.36: **Newcastle & Cambridge correlations**: Argyle model interacted with Newcastle-Cambridge indicator variable Left Plot of Cook's distance versus index for Argyle model with $c = 1$ Right Plot of standardised residuals versus fitted values (after excluding two influential Newcastle correlations (0,4) and (0,6))

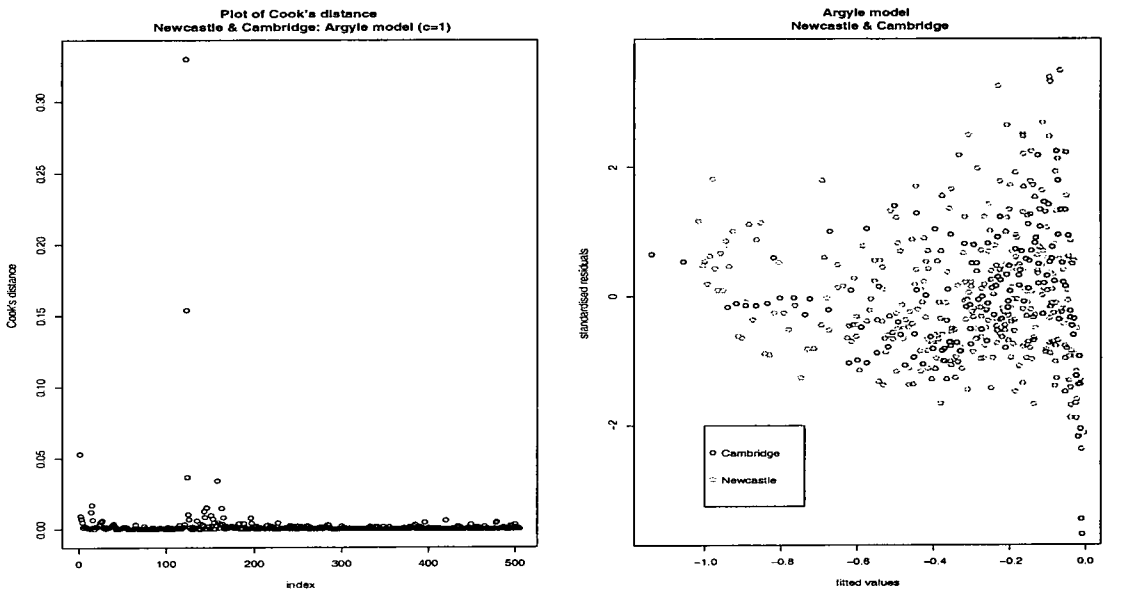


Figure 5.37: **Newcastle & Cambridge correlations [N=503]** Exploratory plots to see how varying constant ($c$) added to the time points effects fit and coefficients in Argyle model interacted with Newcastle-Cambridge indicator variable (using weighted least squares)

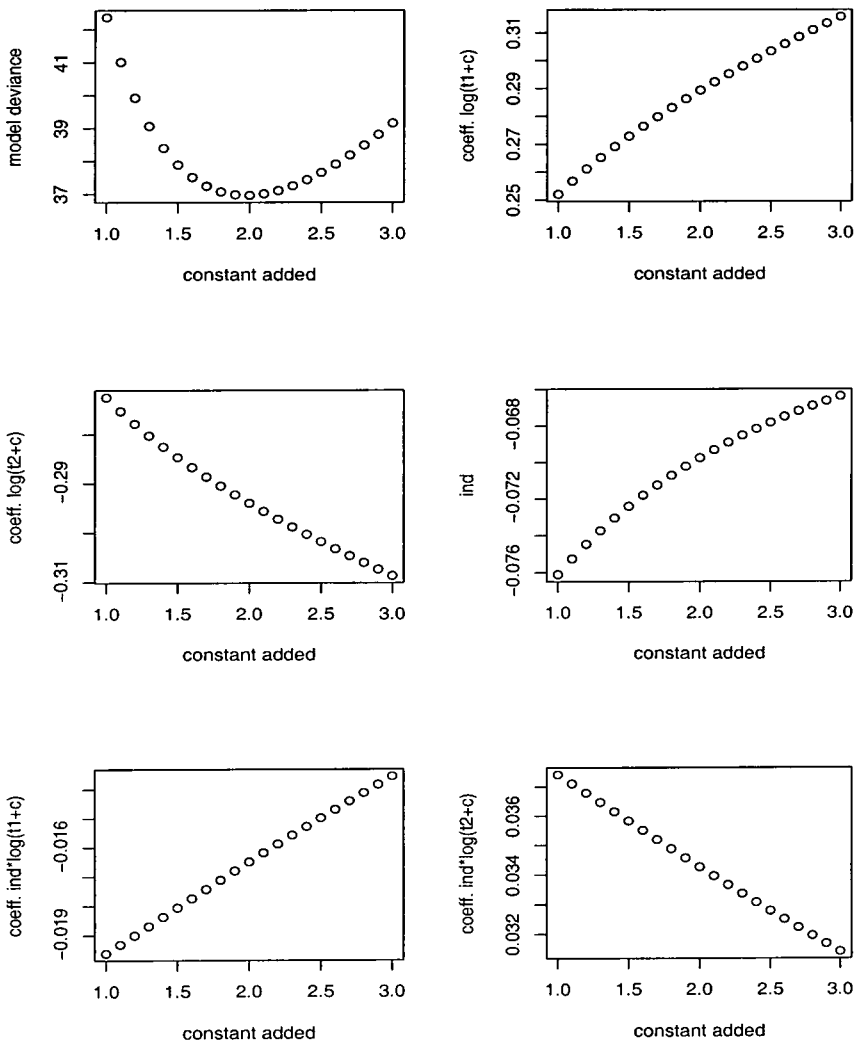
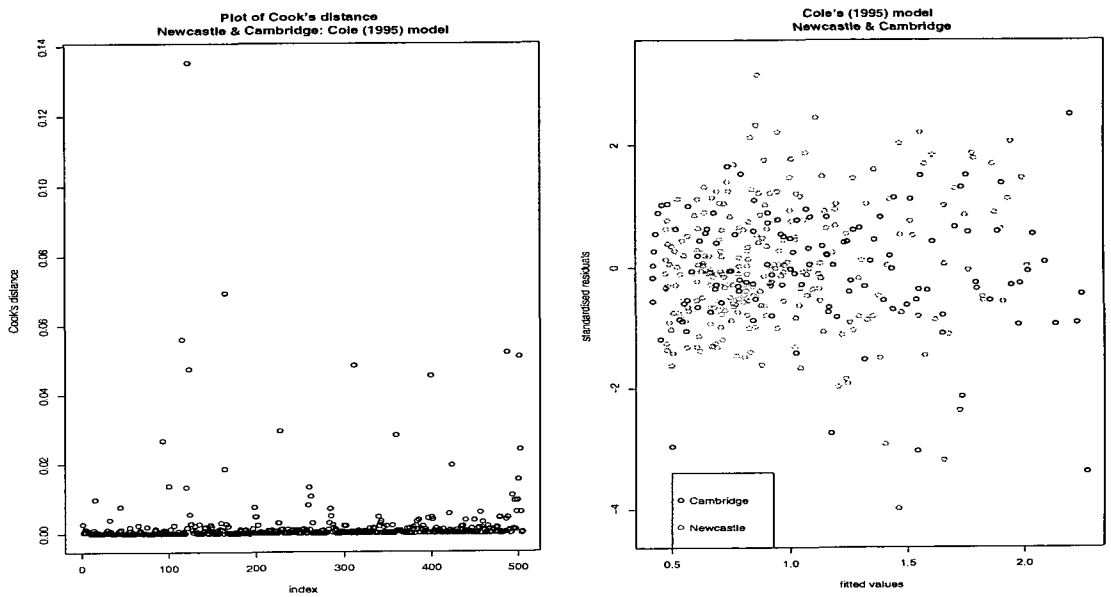


Figure 5.38: **Newcastle & Cambridge correlations**: Model proposed by Cole (1995) interacted with Newcastle-Cambridge indicator variable Left Plot of Cook's distance versus index Right Plot of standardised residuals versus fitted values (after excluding one influential observation)



# Chapter 6

# Follow up study at 7-9 years

In this chapter we introduce data from the Newcastle follow-up study at 7-9 years. In section 6.1 we briefly outline the motivation behind the follow-up study and in section 6.2 we discuss the variables measured in the research study. In section 6.3 we summarise the published research on the case-control study data considered within this chapter. A preliminary analysis of data from the systematic sample and case-control sample can be found in sections 6.5 and 6.4 respectively. We finally conclude this chapter with an analysis of the reported parental height data and obtain mid-parental height Z-scores for children (where possible) within the follow-up study; see section 6.6.

## 6.1 Motives for follow up study

In Chapter 4 we introduced the routine infancy weight data. The routine weights were retrieved in 1989 for a complete birth cohort of 3415 term infants born in Newcastle-upon-Tyne between 1st April 1987 and 31st March 1988. Five years later, in 1996, two subsets of these same children formed part of a follow-up study. At the time these children were aged 7 to 9 years.

The primary motivation for the follow-up study was to look at the long term implications of failure-to-thrive; in terms of attained growth and cognitive outcome. As discussed in chapter 3, infants that experience failure-to-thrive in infancy may later have delayed growth, delayed cognitive development, poor educational attainment and behavioural problems. Early research in the area of failure-to-thrive concentrated on children referred to hospital. However the decision to refer a child is often made on the basis of their psycho-social background and not on their rate of weight

gain (Batchelor and Kerslake (1990) in Drewett et al. (1999)). Therefore hospital samples are unlikely to be representative of all children that fail to thrive in infancy.

All the Newcastle infancy data was retrieved from child health records and primary health care professionals. The children and their parents were contacted for the first time at the start of the follow-up study. Therefore it seems unlikely that there will be any intervention effect on the children that failed to thrive in infancy (Corbett 1998). In carrying out the follow-up study it was discovered that only nine out of 136 cases had been investigated solely for growth problems or failure-to-thrive and two exclusively for feeding problems (Drewett et al. 1999). Similarly, a population study, carried out in an inner city area of London, found that only four out of their 23 cases had been investigated for failure-to-thrive (Dowdney et al. 1987).

## 6.2 Data collected in follow-up study and measurement technique

A total of 810 children were singled out to take part in the follow-up study. One subset of 326 children were chosen to take part in a case-control study to explore the outcomes of failure-to-thrive in infancy. Half of this first subset were identified as cases using the 'thrive index' methodology (Wright et al. 1994). The second subset of 562 children formed a 20% systematic sample of the birth cohort. These children were selected from the 2182 infants that had at least three weights in infancy, with every fifth child followed-up (Wright and Cheetham 1999). This was designed to achieve a representative sample of Newcastle children. These two subsets give a total of 888 children, hence 78 children were selected to be part of the case-control study and the systematic sample. The parents of 27 individuals within the case-control study agreed to height and weight measurement but not to psychological testing (Drewett et al. 1999). So we have anthropometric data on 82% of cases and 91% of controls. Of the 562 eligible children for the systematic sample, 448 (80%) were successfully traced and measured.

A study consent form was sent to parents of all children involved in the follow-up study to gain a report of both their heights. The heights, weights, head circumference and blood pressure of the follow-up study children were measured in school by the same research nurse.

Height was measured to the nearest 0.1 cm using the Leicester Height Measurer. The children removed their shoes and were positioned with their feet on the marked

area. In measuring the height the child was stretched by applying gentle upward pressure to the mastoid process. This is supposed to minimise diurnal variation (i.e. humans are taller in the morning than in the afternoon) (Whitehouse et al. 1974). Although this is a recommended procedure, Voss and Bailey (1997) found that this process didn't reduce the effects of diurnal variation. Each auxologist that uses the stretching procedure has their own measuring error. However, all the height measures were taken by the same research nurse and we know roughly the time of day of the height measure as this was noted when the blood pressure was taken.

Weight was measured to the nearest 100g using portable electronic scales (SECA Scales, Model Number 835). Digital scales were used to obtain weights, as dial scales are subject to more problems with reader error. The children involved in the study were asked to wear light clothing (e.g. shorts and T-shirts or leggings and T-shirt or summer dress) and were weighed without shoes. Some of the children failed to wear these items either because it was cold or they forgot, so a code between 1 and 6 was used for various additional items. Similar items of clothing were subject to test weighing. The light clothing weighed between 0.16 Kg for a summer dress and 0.28 Kg for leggings and a T-shirt, so a correction of 0.2 Kg was suggested for these individuals. The individuals not wearing light clothing were assumed to be wearing a T-shirt, which lead to a correction of 0.1 Kg on the top of the average weight of coded clothing. Table 6.1 contains the results of test weighing along with the suggested average correction. The child's recorded weight was adjusted for the weight of clothes worn, this is a standard approach suggested by Tanner et al. (1966).

Head circumference was measured to the nearest 0.1 cm using a circular Leicester

Table 6.1: Dress coding and suggested weight corrections for various items of clothing

| Type of clothing | DRESS code | Average weight (Kg) | Suggested correction (Kg) |
|---|---|---|---|
| Baseline | NA | 0 | 0.2 |
| Trousers | 1 | 0.25 | 0.35 |
| Jeans/dungarees | 2 | 0.6 | 0.7 |
| Jogging bottoms | 3 | 0.25 | 0.35 |
| Pinafore dress | 4 | 0.25 | 0.35 |
| Corduroy trousers | 5 | 0.4 | 0.5 |
| Sweater/jumper | 6 | 0.35 | 0.45 |

measure positioned above the supra orbital ridges. This may be subject to some error due to random variation, hair thickness and slippage of tape.

Systolic blood pressure was recorded using a mercury sphygmomanometer and a Doppler probe. This was taken three times in succession, by convention the systolic blood pressure and time were both noted on each occasion. Children of this age are not used to having their blood pressure taken, so the initial observation may be slightly elevated and any data analysis would be based on the last recorded blood pressure. The systolic blood pressure at 7-9 is of interest because poor weight gain in infancy has been shown to be linked to ischaemic heart disease (Barker et al. 1989). Analysis of the systolic blood pressure data is deferred to chapter 9.

The follow-up study data set has the following variables for each individual:

1. **ID** Unique identifier from infancy study.
2. **Date.fol** - Date at which height, weight, head circumference and blood pressure were recorded, which was between February and October of 1996.
3. **Sex**
4. **Date of birth** - **Age.fol** age (in years to 2 decimal places) at follow-up was derived from taking the time between the date of birth and date of follow-up.
5. **Height.fol** - Height is to the nearest 0.1 cm and is recorded in centimetres.
6. **Weight.fol** - Weight is to the nearest 0.1 Kg and is recorded in Kilograms.
7. **Head.fol** - Head circumference is taken to the nearest 0.1 cm and is recorded in centimetres.
8. **SYS1, SYS2,** and **SYS3** are the systolic blood pressure taken at times **BPTIME1, BPTIME2,** and **BPTIME3**, respectively.
9. **DRESS** - code of 1 to 6 for various items of clothing additional to 'light clothing'.
10. **BMI.fol** - The body mass index was derived from the weight adjusted for clothing and child's height.

11. **Mother and Fathers reported heights** - These were obtained from the consent form for the follow up study. This provided an estimate of the fathers and mothers height in inches, which were later converted to metric heights.

In using the reported parental heights we should be aware that adults tend to overestimate their heights (Ziebland et al. 1996) and Beyer and Doerr (1998) found that estimated heights of short mothers with children of short stature were not reliable. However Wright and Cheetham (1999) compared the heights of 48 measured mothers with the rest of the reported mothers heights and found the mean height to be similar within these two groups, although the standard deviation was larger for the reported mothers heights.

The heights, weights, head circumference and BMI were converted to Z-scores using the revised UK 1990 growth reference (Freeman et al. 1995; Cole et al. 1995; Cole et al. 1998), to give **ZscoreH.fol**, **ZscoreW.fol**, **ZscoreC.fol** and **ZscoreB.fol**, respectively.

There is also an additional case-control indicator (**CACO**). This variable has 9 codes (including missing) which represent:

- 1 = Cases however sampled
- 2 = Controls for 1
- 3 = Cases identified using original UK 1990 reference
- 4 = Controls for 3
- 5 = Cases identified using revised UK 1990 reference
- 6 = Controls for 5
- 7 = Cases identified using original UK 1990 reference with sex correction
- 8 = Controls for 7
- NA = not within case-control study

As discussed in chapter 4, the Newcastle infancy data indicated that there was a sex bias towards females (Wright et al. 1996) in the original UK 1990 reference (Freeman et al. 1995). Combining individuals with CACO 1 and 3 will give cases

identified using original UK 1990 standard (Freeman et al. 1995), combining 1 and 5 gives cases using revised UK 1990 standard (Preece et al. 1996) and similarly combining 1 and 7 gives cases identified using a sex correction with obsolete standard (this was a temporary measure taken while standards were being revised). The number of children within each of the above codings can be found in table 6.2.

## 6.3 Published results from case-control study at 7-9 years

The children within the case-control study were also studied as part of a project for a PhD thesis within the Psychology department in the University of Durham, the focus here was on the results of psychological and reading ability tests (Corbett 1998). Corbett (1998) overall study aim was to ascertain whether there is an association between failure to thrive (FTT) in infancy and enduring psychological and educational deficits. The children's Intelligence Quotient (IQ) and reading ability were tested in school. The mothers of these children were also given an IQ test and their heights measured. In addition an extensive interview was carried out at home with the mother to collect social, economic, demographic and educational data along with a medical history of the child (Corbett 1998). If a mother reported a medical problem that might have been associated with poor weight gain, then the child's notes were retrieved and reviewed by Dr C.M. Wright (blind to case status) to rate whether the child could have an organic condition that would 'definitely' or 'possibly' result in poor weight gain (Corbett 1998). Fourteen individuals within the case-control study had conditions that could possibly or definitely affect growth. Two cases and one control were identified as having a medical condition that would definitely affect growth. Nine cases and two controls had a medical condition that could possibly affect growth.

The cases within this published research (Drewett et al. 1999) are taken to be

Table 6.2: Number of children taking part in case-control follow-up study

| CACO | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
|---|---|---|---|---|---|---|---|---|---|
| No. of children | 129 | 129 | 19 | 19 | 7 | 7 | 8 | 8 | 326 |

the individuals with CACO indicators 1 and 5, with corresponding controls having CACO indicators 2 and 6. Of the 136 cases (52 boys and 84 girls) and 136 controls, 107 of cases (79%) and 117 of controls (87%) were successfully studied (a slightly higher follow-up rate was observed for controls as the 14 individuals that had moved outside of the area were replaced). The main conclusions from this research were (Drewett et al. 1999; Corbett 1998):

- At eight years old, mothers in the case-group reported more feeding problems in infancy and more organic conditions (i.e. medical conditions that may effect growth and cognition).
- From the interview with the mother, the only statistically significant differences were in the proportion of mothers reporting feeding problems and the reported height of fathers.
- The only further differences observed between case and control groups, were birth weight and gestational age. The control group had a slightly lower birth weight (see graph 6.1) and shorter reported gestational age than cases.
- The cases were significantly shorter, lighter, thinner and had smaller head circumferences than controls (see table 6.3). After adjusting for parental heights (using linear regression) a difference of 4.4 cm in mean height between case and control children (95% CI 2.92 to 5.90 cm).
- There were no statistically significant differences in IQ and reading ability either before or after adjusting for maternal IQ.

Several possible explanations were offered within Corbett (1998) and Drewett et al. (1999) for finding no significant difference in cognitive outcome. These included statistical power (the numbers in this study provide a power of between 75% and 80% to detect a 5-point difference in IQ), exact definition of FTT, that as children aged the effect of failure to thrive in infancy diminished and lastly the chronicity of FTT is a key variable. The chronicity of FTT refers to the duration. Dowdney et al. (1987) and Dowdney et al. (1998) found persisting cognitive differences between their cases and controls, retrospective analysis of clinic weight data revealed that cases could have been discriminated on the basis of their declining trajectory within the first 12 months of life, however their case criterion for FTT required that both height (allowing for parental height) and weight were below the tenth percentile at four years so these cases had enduring poor growth. The overall

Table 6.3: Comparing anthropometric measures of case and control children at 7-9 years (adapted from table in Drewett et. al. (1999))

| | Cases (N=111) | Controls (N=122) | | |
|---|---|---|---|---|
| | Mean | Mean | t | p |
| Height (cm) | 126.0 | 130.7 | 6.3 | <0.01 |
| Head circumference (cm) | 51.9 | 52.8 | 3.86 | <0.01 |
| | Median | Median | $\chi^2$ | p |
| Weight (Kg) | 23.8 | 27.9 | 27.9 | <0.001 |
| BMI ($Kg/m^2$) | 14.9 | 16.3 | 26.3 | <0.001 |

conclusion by Corbett (1998) was that failure to thrive may affect individuals by disrupting their physiological development which predisposes them to health problems in later life, or may be indicative of poor appetite with the potential to lead to eating disorders, or may adversely affect emotional and behavioural development. Future research plans to investigate behaviour problems and psychiatric disorders in this group; in particular, the possibility of continuing problems relating to food and food intake in children who have failed to thrive in infancy (Drewett et al. 1999). A further study has been conducted on just the case-control group, which involves re-measuring individuals and administering various psychological questionnaires.

My research is only on the anthropometric and blood pressure data, so the results from IQ and reading tests will not be discussed in any further detail in this chapter. As already discussed in Chapter 4, the birth weight Z-scores in the original analysis were slightly out because of a numerical error in the FORTRAN code used by Dr C.M. Wright to obtain the weight Z-scores at birth. The identification numbers of children defined as cases can be found in table 4.14.

## 6.4 Preliminary analysis of systematic sample

### 6.4.1 Summary plots and statistics

The systematic sample data frame contains data for 561 individuals (277 boys, 284 girls), because data for ID 1090 was discovered as part of the case-control study to have been born at 32 weeks. Seventy seven of these individuals also belonged to the

case-control study. Obviously we should be aware of the 34 individuals that belong to the case arm of the study because these children have poor attained heights, weights, head circumferences and BMI's at follow-up (Drewett et al. 1999). It was felt that the presence of the case children could unduly influence our assessment of the revised UK 1990 reference. Therefore, we will only consider the systematic sample excluding cases in our data analysis below. This leaves 527 children (263 boys, 264 girls).

Boxplots of the Z-scores for height, weight, BMI and head circumference by sex are produced in figure 6.2. The boxplots of Z-scores for height, weight and BMI look reasonably symmetric about zero. However, the distribution of Z-scores for head circumferences are shifted below zero and there are several females with outlying head circumference Z-scores. ID 3594 had an extremely high head circumference (head circumference = 62.9cm, Z = 5.62); either due to measuring error or because he is also the heaviest child. Therefore this child's head circumference Z-score was excluded from the boxplot for head circumference and the summary statistics presented in table 6.4.

Summary statistics for boys and girls are given in tables 6.4 and 6.5, respectively. We have weight and height data for 428 individuals (214 boys, 214 girls), head circumference data for 426 individuals (213 boys, 213 girls) and systolic blood pressure for 420 individuals (211 boys, 209 girls). We have height and weight data for 7 additional individuals compared to previously published work (Wright and Cheetham 1999), because some individuals that were thought to belong only to the case-control study were discovered to be part of the systematic sample as well. As noted above, the distribution of Z-scores for head circumference are shifted to left of zero and the variance of the head circumference Z-scores for girls appears to be higher than the boys.

### 6.4.2 Adequacy of UK 1990 growth reference for Newcastle children at 7-9 years

As discussed in chapter 4, Z-scores, by construction, are expected to be symmetrically distributed and have a zero mean and variance of one. Quantile-quantile plots were produced for the Z-scores for height, weight, BMI and head circumference by sex. The quantile-quantile plots indicated that there was no reason to doubt the normality of the Z-scores for weight, height, BMI and head circumference (boys only). However, the quantile-quantile plot of the head circumference Z-scores for

Table 6.4: Boys: Summary statistics of measurements made in follow-up study (systematic sample excluding cases)

| | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| Age.fol | 7.720 | 8.340 | 8.565 | 8.559 | 8.788 | 9.220 | 0.301 | 214 |
| Height.fol | 118.5 | 127.5 | 130.9 | 130.9 | 134.1 | 146.5 | 5.3 | 214 |
| ZscoreH.fol | -2.450 | -0.590 | -0.005 | -0.003 | 0.600 | 2.310 | 0.934 | 214 |
| Weight.fol | 19.20 | 24.90 | 27.30 | 28.37 | 30.49 | 50.40 | 5.127 | 214 |
| ZscoreW.fol | -2.560 | -0.635 | 0.080 | 0.088 | 0.710 | 3.020 | 1.059 | 214 |
| BMI.fol | 11.80 | 14.95 | 15.92 | 16.46 | 17.40 | 29.07 | 2.319 | 214 |
| ZscoreB.fol | -3.900 | -0.655 | 0.000 | 0.117 | 0.865 | 3.550 | 1.123 | 214 |
| HeadC | 49.00 | 52.20 | 53.15 | 53.34 | 54.40 | 62.90 | 1.631 | 214 |
| ZscoreC.fol | -3.170 | -1.120 | -0.580 | -0.473 | 0.190 | 2.120 | 0.9486 | 213 |
| SYS BP | 65.00 | 82.00 | 90.00 | 88.55 | 95.00 | 112.00 | 9.615 | 211 |

Table 6.5: Girls: Summary statistics of measurements made in follow-up study (systematic sample excluding cases)

| | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| Age.fol | 7.820 | 8.320 | 8.600 | 8.585 | 8.838 | 9.220 | 0.320 | 214 |
| Height.fol | 113.8 | 126.0 | 129.5 | 129.8 | 133.9 | 145.5 | 6.0 | 214 |
| ZscoreH.fol | -3.120 | -0.798 | -0.160 | -0.130 | 0.505 | 2.430 | 1.009 | 214 |
| Weight.fol | 17.70 | 25.00 | 27.70 | 28.43 | 31.60 | 44.80 | 5.213 | 214 |
| ZscoreW.fol | -2.980 | -0.678 | -0.090 | -0.004 | 0.705 | 2.690 | 1.0482 | 214 |
| BMI.fol | 12.50 | 15.12 | 16.28 | 16.74 | 17.71 | 25.53 | 2.282 | 214 |
| ZscoreB.fol | -2.610 | -0.590 | 0.070 | 0.100 | 0.728 | 2.800 | 1.038 | 214 |
| HeadC | 48.00 | 51.20 | 52.20 | 52.27 | 53.20 | 57.30 | 1.612 | 213 |
| ZscoreC.fol | -4.280 | -1.650 | -0.810 | -0.790 | -0.110 | 3.320 | 1.296 | 213 |
| SYS BP | 64.00 | 80.00 | 88.00 | 87.46 | 96.00 | 116.00 | 9.526 | 209 |

girls indicated that this distribution has heavier tails than the normal distribution. There is no reason to doubt that the mean Z-score for height, weight and BMI is zero. However there is strong evidence to suggest that the mean Z-score for head circumference is less than zero; see table 6.6. There is no reason to doubt that the Z-scores for height, weight, BMI (girls only) and head circumference (boys only) have a variance of 1. There is reason to doubt that the variance of the head circumference Z-scores for girls is 1. However we should be aware that the F-test is sensitive to departures from normality and quantile-quantile plots indicated that the distribution of the Z-scores for head circumference had heavy tails for girls. Furthermore, the

elevated variance is partly a result of the 8 outlying Z-scores for head circumference (see figure 6.2). There is also reason to doubt that the variance of BMI Z-scores for boys is 1. Again, the elevated variance is partly due to 4 outlying Z-scores for BMI (see figure 6.2).

The World Health Organisation (WHO) has produced guidelines for assessing the standard deviation of height-for-age and weight-for-age Z-scores (WHO 1995). The standard deviation (SD) of Z-scores should be close to the expected value of 1. A SD significantly lower than 0.9 describes a distribution that is more homogeneous, or one that has a narrower spread, compared to the reference population (WHO 1995). Similarly if the SD is between 1.1 and 1.2, the distribution is wider spread than the reference (WHO 1995). A SD above 1.3 indicates inaccurate data due to measurement error or incorrect age reporting (WHO 1995). According to WHO guidelines (WHO 1995), the expected range of SD's of the Z-score distribution for the anthropometric indicators are:

1. height-for-age Z-score: 1.1 to 1.3
2. weight-for-age Z-score: 1.0 to 1.2

The standard deviation of weight Z-scores for boys and girls are within the region suggested in point (2). The standard deviation of height Z-scores for girls is within the range suggested in point (1). However the standard deviation of height Z-scores for boys is a little lower than one.

As discussed in chapter 4, we are also need to examine whether there is any linear trend in the Z-scores with age. Plots of the Z-scores for height, weight, BMI and head circumference versus age can be found in figures 6.3 and 6.4. There appears to be some indication of a negative linear trend with age for the height and weight Z-scores of boys. Table 6.7 summarises the results of regressing the Z-score for weight, height, BMI and head circumference on age by sex. The results from table 6.7 indicate that there is no reason to doubt the null hypothesis of zero slope for BMI and head circumference, indicating that the BMI and head circumference data at 7-9 years may be appropriately adjusted for age by the revised UK 1990 growth reference. In table 6.7, for girls, there is no reason to doubt the null hypothesis of zero slope for height and weight. However, the same isn't true for the boys' height and weight Z-scores. Thus indicating there may be some trend in the Z-scores for weight and height with age for boys.

The main conclusion from assessing the adequacy of the revised UK 1990 growth reference is that children in Newcastle tend to have smaller head circumferences at

Table 6.6: Systematic sample (excluding cases): Results of testing that (a) the mean Z-score for height, weight, BMI and head circumference is zero (by sex) and (b) the variance of the Z-score for height, weight, BMI and head circumference is one (by sex)

| Z-score | sex | t | p | 95% CI |
|---|---|---|---|---|
| Height | Boys | -0.0424 | 0.9662 | [-0.1286, 0.1232] |
| | Girls | -1.8883 | 0.0604 | [-0.2662, 0.0057] |
| Weight | Boys | 1.2126 | 0.2266 | [-0.0549, 0.2305] |
| | Girls | -0.0515 | 0.959 | [-0.1449, 0.1375] |
| BMI | Boys | 1.5191 | 0.1302 | [-0.0347, 0.2680] |
| | Girls | 1.4133 | 0.159 | [-0.0396, 0.2400] |
| Head Circumference | Boys | -7.277 | $6.539 \times 10^{-12}$ | [-0.6011, -0.3449] |
| | Girls | -8.8925 | $2.647 \times 10^{-16}$ | [-0.9647, -0.6146] |
| Z-score | sex | $\frac{(n-1)s^2}{\sigma^2}$ | approx. p-value | 95% CI |
| Height | Boys | 185.95 | 0.1900 | [0.7283, 1.0658] |
| | Girls | 216.82 | 0.8531 | [0.8492, 1.2427] |
| Weight | Boys | 238.9907 | 0.2079 | [0.9361, 1.3698] |
| | Girls | 234.007 | 0.3088 | [0.9165, 1.3412] |
| BMI | Boys | 268.6928 | 0.0070 | [1.0524, 1.5400] |
| | Girls | 229.2591 | 0.4308 | [0.8980, 1.3140] |
| Head Circumference | Boys | 190.7841 | 0.3029 | [0.7505, 1.0992] |
| | Girls | 356.0492 | $2.6403 \times 10^{-12}$ | [1.4006, 2.0514] |

7-9 years than those children that contribute to the UK reference. Furthermore, it is not entirely unreasonable to use the revised UK 1990 reference to convert weight, height and BMI of Newcastle children to Z-scores at 7-9 years.

Table 6.7: Slope coefficients from regression of Z-scores for height, weight, BMI and head circumference on age of follow-up assessment (by sex)

| Sex | Z-score | Estimate | Std. Error | t value | Pr(> \|t\|) |
|---|---|---|---|---|---|
| Boys | Height | -0.4643 | 0.2104 | -2.206 | 0.0284 |
| | Weight | -0.5668 | 0.2381 | -2.380 | 0.0182 |
| | BMI | -0.3842 | 0.2545 | -1.510 | 0.133 |
| | HeadC | -0.1718 | 0.2158 | -0.796 | 0.427 |
| Girls | Height | 0.1192 | 0.2166 | 0.55 | 0.583 |
| | Weight | 0.0038 | 0.2252 | 0.017 | 0.987 |
| | BMI | -0.1050 | 0.2228 | -0.471 | 0.638 |
| | HeadC | 0.3760 | 0.2778 | 1.353 | 0.1774 |

### 6.4.3 Impact of socio-economic status on attained height, weight, BMI and head circumference

Historically it has been found that children of a lower socio-economic status end up smaller adults. For example, in the 1872-3 survey, urban children who worked in factories and had parents that did the same, were 4 centimetres shorter and 2 Kg lighter by the age of 12 than children who lived in non-factory rural or suburban districts (Tanner 1981). More recently in the UK, the difference in height between children of professional and unskilled manual fathers, is:

1. For boys, nearly 2 cm at age 3 rising to 4 cm at adolescence and falling again to 3 cm in young adulthood.

2. For girls, it is about 1.5 cm at three rising to 3 cm at adolescence and falling to 2 cm in adults (Rosenbaum et al. (1985) in Tanner (1989)).

Wenlock et al. (1986) (in Tanner (1989)) found that the weight differences tended to be somewhat less than height differences, because the worse-off children have a higher weight for height.

In 1989, each child was matched to a neighbourhood using their postcode, the levels of deprivation for these areas were then assessed using the Townsend score, which summarises census data on car and home ownership, overcrowding and unemployment rates (Wright et al. 1994). The child's socio-economic status may have changed since this mapping, e.g. moved house, father becoming unemployed, divorce etc. However, Lasker and Mascie-Taylor (1989) found that growth in stature and weight after age seven is little affected by social class as most differences are already established by this age. Thus it seems reasonable to consider the impact of level of deprivation (in infancy) on anthropometric measures at 7 to 9 years.

Table 6.8: Summary of the number of children that have anthropometric data by sex and level of deprivation

| SES | level | Boys | Girls | Total |
|---|---|---|---|---|
| affluent | 1 | 21 | 17 | 38 |
| intermediate | 2 | 147 | 151 | 298 |
| deprived | 3 | 46 | 46 | 92 |

In table 6.8 we summarise the number of children with anthropometric data at 7 to 9 years by sex and level of deprivation. The sample sizes under consideration for the deprived and affluent children are small. Thus, the data analysis is purely exploratory and the discussion that follows only provides tentative conclusions. We would require much larger sample sizes to explore the impact of level of deprivation on attained height, weight and BMI at 7 to 9 years. In figures 6.5 and 6.6 we produce variable width notch box-plots for the systematic sample (excluding cases) of the Z-scores for height, weight, BMI and head circumference by sex and level of deprivation. The boxplots for BMI and head circumference, by sex, indicate that there is no difference in medians for different levels of deprivation (see figure 6.6), thus suggesting that level of deprivation has little impact on BMI and head circumference. If we consider the boxplots for height and weight Z-scores by level of deprivation for girls, see figure 6.5, there is a significant difference between the medians for deprivation levels 2 and 3 ($W = 4513$, $p = 0.0021$ for height Z-score and $W = 4251$, $p = 0.0216$ for weight Z-score). If we now consider the boxplot for height Z-scores by level of deprivation for boys, see figure 6.5, there is a significant difference between the medians for deprivation levels 1 and 2 ($W = 2159.5$, $p = 0.0032$). Although the notches in boxplot for weight Z-scores by level of deprivation for boys in figure 6.5 overlap, there is some evidence to suggest that the median weight Z-scores for deprivation levels 1 and 2 are different ($W = 1954$, $p = 0.0493$). Thus, in agreement with previous studies, there is some indication that level of deprivation may influence height and weight Z-scores at 7-9 years. Furthermore, the impact of level deprivation appears to be different for girls and boys suggesting that the intermediate level of deprivation may not be well discriminating.

## 6.5 Preliminary analysis of case-control sample

There have been very few long term studies on the consequences of failure to thrive (Wright 2000). In the pre-school years there tends to be a gradual improvement; i.e. partial catch up growth (Kristiansson and Failstrom 1987; Wright et al. 1998) but it is not known if there is any lasting deficit (Wright 2000). As discussed in the introduction to this chapter, the role of the case-control study was to establish whether failure to thrive in infancy had any effect on later IQ and reading ability at 7-9 years. However, we are also interested in the impact that failing to thrive in infancy has on future growth outcome. Dowdney, Skuse, Heptinstall, Puckering, and Zur-Szpiro (1987) in 1984 identified 25 cases of growth retardation at the age of 4 years from health clinic records in a socioeconomically disadvantaged inner-city

area of London from 1868 children. The cases identified were significantly, shorter, lighter and leaner than a comparison group at 11 years (Dowdney et al. 1998). A group of 42 cases, that had weights below the third centile for at least 3 months, were identified from a one year cohort in an inner-city area of south London (Boddy et al. 2000). At 6 years, these case children were considerably smaller than matched comparisons (Boddy et al. 2000).

There are 136 individuals (52 boys, 84 girls) in the case data frame and 133 individuals (54 boys, 79 girls) within the control data frame. Summary statistics were produced for the case and control groups by sex and these can be found in tables 6.9 and 6.10. It would appear that distribution of the Z-scores for case children is shifted to the left of control children.

In performing a t-test we assume that the cases and controls comprise two independent normal random samples. Quantile-quantile plots for the Z-scores for height, weight, BMI and head circumference were produced for the case-control study boys and girls. The quantile-quantile plots suggested that the assumption of normality

Table 6.9: Case-control study: Summary statistics of measurements made in follow up study for boys

| | group | Min | LQ | Median | Mean | UQ | Max | SD | no. |
|---|---|---|---|---|---|---|---|---|---|
| Age | Case | 7.870 | 8.305 | 8.480 | 8.501 | 8.740 | 9.070 | 0.309 | 43 |
| | Control | 8.130 | 8.370 | 8.600 | 8.556 | 8.710 | 9.210 | 0.246 | 45 |
| Height | Case | 115.0 | 122.0 | 125.4 | 125.7 | 128.9 | 136.7 | 5.196 | 42 |
| | Control | 119.9 | 128.0 | 131.5 | 132.1 | 134.9 | 148.6 | 5.850 | 45 |
| ZscoreH | Case | -2.770 | -1.595 | -0.870 | -0.878 | -0.195 | 1.190 | 0.930 | 42 |
| | Control | -1.750 | -0.560 | 0.150 | 0.205 | 0.590 | 2.960 | 0.984 | 45 |
| Weight | Case | 19.45 | 21.30 | 23.28 | 23.57 | 25.06 | 29.05 | 2.609 | 42 |
| | Control | 20.85 | 26.10 | 28.80 | 29.33 | 31.80 | 47.90 | 5.561 | 45 |
| ZscoreW | Case | -2.620 | -1.728 | -1.110 | -1.086 | -0.543 | 0.490 | 0.876 | 42 |
| | Control | -1.850 | -0.370 | 0.380 | 0.286 | 1.030 | 2.710 | 1.0419 | 45 |
| BMI | Case | 12.96 | 14.10 | 14.72 | 14.86 | 15.44 | 17.34 | 1.038 | 42 |
| | Control | 12.95 | 15.34 | 16.25 | 16.73 | 17.29 | 29.07 | 2.757 | 45 |
| ZscoreB | Case | -2.470 | -1.360 | -0.770 | -0.779 | -0.293 | 0.770 | 0.785 | 42 |
| | Control | -2.420 | -0.420 | 0.200 | 0.222 | 0.730 | 3.550 | 1.177 | 45 |
| Head | Case | 49.10 | 51.35 | 52.20 | 52.41 | 53.25 | 57.30 | 1.710 | 43 |
| | Control | 50.70 | 52.30 | 53.40 | 53.44 | 54.40 | 56.40 | 1.367 | 45 |
| ZscoreC | Case | -3.090 | -1.750 | -1.120 | -1.019 | -0.475 | 2.080 | 1.096 | 43 |
| | Control | -2.120 | -1.110 | -0.410 | -0.383 | 0.190 | 1.460 | 0.863 | 45 |
| SYS BP | Case | 68.00 | 78.00 | 86.00 | 83.72 | 89.00 | 100.00 | 8.172 | 43 |
| | Control | 68.00 | 84.00 | 92.00 | 90.22 | 100.00 | 110.00 | 11.129 | 45 |

Table 6.10: Case-control study: Summary statistics of measurements made in follow up study for girls

| | group | Min | LQ | Median | Mean | UQ | Max | SD | no. |
|---|---|---|---|---|---|---|---|---|---|
| Age | Case | 7.910 | 8.360 | 8.600 | 8.571 | 8.840 | 9.270 | 0.325 | 69 |
| | Control | 7.780 | 8.320 | 8.540 | 8.558 | 8.790 | 9.150 | 0.301 | 77 |
| Height | Case | 108.5 | 122.5 | 126.5 | 126.3 | 130.0 | 144.0 | 6.110 | 69 |
| | Control | 113.8 | 125.6 | 129.9 | 129.6 | 132.9 | 141.9 | 5.838 | 77 |
| ZscoreH | Case | -3.780 | -1.400 | -0.650 | -0.738 | -0.050 | 2.110 | 1.060 | 69 |
| | Control | -2.920 | -0.790 | -0.170 | -0.154 | 0.450 | 1.900 | 0.987 | 77 |
| Weight | Case | 17.30 | 21.75 | 24.83 | 24.77 | 27.40 | 33.80 | 3.968 | 68 |
| | Control | 20.55 | 24.95 | 26.90 | 28.51 | 31.10 | 43.40 | 5.075 | 77 |
| ZscoreW | Case | -3.690 | -1.705 | -0.610 | -0.825 | -0.030 | 1.000 | 1.055 | 68 |
| | Control | -1.740 | -0.670 | -0.120 | 0.049 | 0.700 | 2.000 | 0.936 | 77 |
| BMI | Case | 12.76 | 14.09 | 15.17 | 15.41 | 16.55 | 20.60 | 1.661 | 68 |
| | Control | 13.51 | 15.14 | 16.23 | 16.86 | 17.95 | 22.76 | 2.104 | 77 |
| ZscoreB | Case | -2.390 | -1.330 | -0.625 | -0.561 | 0.078 | 1.750 | 0.925 | 68 |
| | Control | -1.660 | -0.570 | -0.020 | 0.191 | 0.800 | 2.170 | 0.913 | 77 |
| Head | Case | 47.30 | 50.60 | 51.70 | 51.49 | 52.50 | 56.00 | 1.735 | 65 |
| | Control | 47.00 | 51.40 | 52.50 | 52.34 | 53.30 | 56.50 | 1.830 | 77 |
| ZscoreC | Case | -4.900 | -2.090 | -1.230 | -1.407 | -0.550 | 2.120 | 1.414 | 65 |
| | Control | -5.060 | -1.500 | -0.600 | -0.722 | 0.070 | 2.520 | 1.470 | 77 |
| SYS BP | Case | 60.00 | 78.00 | 86.00 | 84.94 | 94.00 | 108.00 | 10.303 | 64 |
| | Control | 70.00 | 82.00 | 88.00 | 88.17 | 94.00 | 128.00 | 10.762 | 76 |

was not entirely unreasonable. Results of the two-sample t-test on case and control samples (assuming unequal variance) can be found in table 6.11. It appears that case children are significantly shorter, lighter, leaner and have smaller head circumferences than control children.

In figure 6.7 we have produced a variable width notch boxplot for the Z-scores for weight, height, BMI and head circumference by case-control status and sex. For height, weight and BMI the difference in medians between case and control boys is greater than the difference in medians for case and control girls. In table 6.12 we summarise results from testing that the median of the Z-scores for height, weight, BMI and head circumference are the same in case and control children by sex. Again, it would appear that the case children are significantly shorter, lighter, leaner and have smaller head circumferences than control children.

With the exception of control boys, the boxplots for height Z-score are reasonably symmetric (see figure 6.7). There is a case girl (ID 1813) with an extremely low height Z-score (ZscoreH.fol = -3.78); she is an organic case with a condition

Table 6.11: Case-control study: Results of testing that the mean Z-score for height, weight, BMI and head circumference is the same in cases and controls (by sex)

| Z-score | sex | t | p | 95% CI |
|---|---|---|---|---|
| ZscoreH.fol | Boys | -5.2813 | $9.718 \times 10^{-7}$ | [-1.4916, -0.6757] |
| | Girls | -3.436 | 0.0008 | [-0.9206, -0.2481] |
| ZscoreW.fol | Boys | -6.6589 | $2.670 \times 10^{-9}$ | [-1.7808, -0.9617] |
| | Girls | -5.2505 | $5.751 \times 10^{-7}$ | [-1.2041, -0.5452] |
| ZscoreB.fol | Boys | -4.6978 | $1.125 \times 10^{-5}$ | [-1.4260, -0.5770] |
| | Girls | -4.9186 | $2.406 \times 10^{-6}$ | [-1.0549, -0.4500] |
| ZscoreC.fol | Boys | -3.0137 | 0.0035 | [-1.0551, -0.2158] |
| | Girls | -2.8253 | 0.0054 | [-1.1647, -0.2056] |

Table 6.12: Case-Control study: Results of testing that the median Z-score for height, weight, BMI and head circumference is the same in cases and controls (by sex)

| | sex | Case notches | Control notches | W | p |
|---|---|---|---|---|---|
| ZscoreH.fol | Boys | (-1.179, -0.561) | (-0.097, 0.397) | 416 | $7.144 \times 10^{-6}$ |
| | Girls | (-0.883, -0.417) | (-0.390, 0.050) | 1798 | $7.702 \times 10^{-4}$ |
| ZscoreW.fol | Boys | (-1.371, -0.849) | (0.079, 0.681) | 294 | $3.277 \times 10^{-8}$ |
| | Girls | (-0.900, -0.320) | (-0.364, 0.124) | 1526 | $1.528 \times 10^{-5}$ |
| ZscoreB.fol | Boys | (-1.029, -0.511) | (-0.047, 0.4470 | 442.5 | $2.006 \times 10^{-5}$ |
| | Girls | (-0.869, -0.381) | (-0.264, 0.224) | 1504 | $1.024 \times 10^{-5}$ |
| ZscoreC.fol | Boys | (-1.399, -0.841) | (-0.690, -0.130) | 611.5 | 0.003 |
| | Girls | (-1.495, -0.965) | (-0.879, -0.321) | 1781.5 | 0.003 |

that would definitely affect growth. The boxplots for weight Z-scores are reasonably symmetric for the boys but not for the girls (see figure 6.7). There are three control boys with extreme BMI Z-scores (see figure 6.7). If we now consider the boxplot for head circumference Z-scores, there are four girls with extremely low head circumference Z-scores (see figure 6.7): these are two cases (IDs 1766 and 2188: both have a medical condition that could affect growth) and two controls (IDs 342 and 639; ID 639 also has a outlying low height Z-score).

## 6.6 Reported parental height data

In section 2.6 we discussed the role of parental height data in the assessment of a child's attained height. Here we carry out a preliminary analysis of the reported parental height data. We then go on to obtain mid-parental height Z-scores and use the conditional approach to contrast a child's height Z-score at 7-9 years with their reported mid-parental height Z-score. In the follow-up study parents were asked to report their heights on the consent form for the follow-up study in feet and inches. In addition, mothers of children that participated in the case-control study were measured by Dr S.S. Corbett and were also asked to report the fathers height. The reported heights were then converted to centimetres and Z-scores were obtained for all parental heights using the LMS values at the age of 23 from the revised UK 1990 growth reference (Freeman et al. 1995; Preece et al. 1996).[1]

### 6.6.1 Parental height data - systematic sample

The summary statistics for the mothers and fathers reported heights within the systematic sample can be found in table 6.13. We have reported heights for 397 fathers and 406 mothers with 396 children having both. There is slightly more variability in the reported parental heights than the adult heights that contribute to the UK 1990 reference. The mean of the reported fathers heights was just over 5'9", whereas the mode and median was 5'10". The mean of the reported mothers heights was 5'4", whereas the mode and median were 5'2" and 5'4", respectively. There appears to be an outlying reported height for a mother of 6'5.5" (her partner is 6'2.5"). As can be seen from table 6.14, the majority of parents reported their heights to the nearest whole inch, with mothers giving their heights in fractions of an inch more frequently than men (Chi-square test on table 6.14 after combining height measures given to quarter and half inch in one column: $\chi^2$=4.787, $p$=0.029). The mean paternal height is 176.3 cm and the mean maternal height is 162.3 cm, giving a sex difference of 14 cm which is the same as the recommended sex correction. A histogram of the reported imperial parental heights can be found in figure 6.8, this indicates that the reported heights of the mothers are fairly normally distributed with one outlier as mentioned above. However, the histogram of the reported fathers heights indicates a much higher proportion of fathers reporting heights of 5'10"or 5'11", this could

[1] Cole (2000b) suggested using LMS values at the age of 22 years to convert parental heights to Z-scores because the macro to calculate Z-scores in Excel does not allow a Z-score to be calculated at the age of 23. This will not be of practical importance as the LMS values change little at this age.

reflect the tendency for adults to overestimate their heights (Ziebland et al. 1996). Therefore the distribution of reported fathers heights is skewed with tail to left; indicating that the distribution of fathers reported height measures are not normal.

If we carry out a t-test on the null hypothesis that the mean of the Z-scores for the reported heights is zero, then t-values are -4.31 ($p = 2.06 \times 10^{-5}$) and -4.86 ($p = 1.71 \times 10^{-6}$) for mothers and fathers, respectively. However this result is not exactly surprising, as the current UK 1990 growth reference is based on modern data and in the UK there is a continuing secular trend in height. The implication of the secular trend in height means that the average height of adults today is greater than it was say 30 years ago. If we consider the obsolete Tanner-Whitehouse standards for height (Tanner et al. 1966), the 50th centile for boys is at 174.7 cm at 18 years and the 50th centile for girls is 162.2 cm at 16 years, these values are slightly lower than the median reported parental heights.

A box-plot of the reported parental heights (converted to metric units) by socio-economic status can be found in figure 6.9. The sample sizes under consideration for the parents of deprived and affluent children are small. Thus, the data analysis

Table 6.13: Summary statistics for reported heights in systematic sample

| Height | Min. | LQ | Median | Mean | UQ | Max. | SD |
|---|---|---|---|---|---|---|---|
| Fathers (n=397) | | | | | | | |
| (feet) | 5.000 | 5.583 | 5.833 | 5.783 | 6.000 | 6.500 | 0.262 |
| (cm's) | 152.4 | 170.2 | 177.8 | 176.3 | 182.9 | 198.1 | 7.998 |
| Z(Father) | -3.7810 | -1.1780 | -0.0629 | -0.2855 | 0.6807 | 2.9120 | 1.171 |
| Mothers (n=406) | | | | | | | |
| (feet) | 4.833 | 5.167 | 5.333 | 5.325 | 5.500 | 6.458 | 0.231 |
| (cm's) | 147.3 | 157.5 | 162.6 | 162.3 | 167.6 | 196.9 | 7.036 |
| Z(mother) | -2.7370 | -1.0520 | -0.2090 | -0.2496 | 0.6336 | 5.4790 | 1.167 |

Table 6.14: Pattern of reporting in mothers and fathers

| parent | inch | 1/2 inch | 1/4 inch |
|---|---|---|---|
| father | 369 | 26 | 2 |
| mother | 358 | 41 | 7 |

is purely exploratory and the discussion that follows only provides tentative conclusions. We would require much larger sample sizes to explore the impact of level of deprivation on reported parental height. The median reported heights of fathers for classes 1 and 2 are very similar, however the median height of fathers from class 3 is markedly shorter. The notches for affluent are at (174.91,180.69) and for deprived at (171.27, 174.17); these notches do not overlap indicating there is some evidence that the median heights of affluent and deprived are different (W = 1640, p = 0.0484). A similar pattern is observed for the reported mothers heights. The notches for affluent are at (162.76,167.44) and for deprived at (158.57, 161.47); these notches do not overlap but the difference in medians fails to reach 5% significance level when we apply the Wilcoxon (Mann-Whitney) test (W = 1698, p = 0.0899).

The mid-parental height was calculated for 396 children using equation (2.19). The target centile range for each individual was obtained by adding and subtracting 10 cm to the mid-parental height and then obtaining the corresponding Z-score. Three hundred and thirty one children had heights at 7-9 years within their target centile range, fifty six were outside their target centile range (38 children were taller, 18 were shorter).

### 6.6.2 Parental height data - case-control study

Table 6.16 contains the summary statistics for the parental height data from the case-control study. A two sample t-test (for independent samples) was applied to the mothers measured heights because the assumption of normality (see figure 6.10) was reasonable. The t-value from the two-sample t-test was -1.547 (p=0.123) (Wilcoxon Mann-Whitney test W=5029.5, p=0.084); which indicates there's no reason to doubt the null hypothesis that the mean height of the case mothers is the same as the mean height of the control mothers. A t-test was not applied to the reported fathers heights (by mothers) as the box-plot for the case fathers showed evidence of

Table 6.15: Summary statistics for reported mid-parental heights in systematic sample

| (n=396) | Min. | LQ | Median | Mean | UQ | Max. | SD |
|---|---|---|---|---|---|---|---|
| MPH | 147.9 | 162.2 | 170.8 | 169.5 | 176.7 | 191.1 | 8.98 |

skewness. If instead we use the non-parametric Wilcoxon (Mann-Whitney) test, we find that W = 4598.5 (p = 0.0054). Therefore there is strong evidence against the null hypothesis that the median reported heights of fathers in the case and control arms of the study are the same. A box-plot of the measured mothers heights and reported fathers heights can be found in figure 6.11, with the exception of the case fathers the distribution of measured and reported heights are reasonably symmetric. As expected; the notches of the box-plots for case and control fathers do not overlap indicating the medians are significantly different at the 5 percent level.

We only have measured and reported heights for 24 mothers and as can be seen from figure 6.12 there appears to be no systematic pattern in terms of whether mothers over or under estimates in relation to actual height. The majority of mothers actually underestimate their heights, which is at odds with previous reports (Ziebland et al. 1996; Himes and Roche 1982), although these authors results are based on much larger samples. There are 25 fathers with two reported heights (on consent form and reported by mother), 10 of these agree, 6 mothers report lower heights and 9 mothers report higher heights.

Table 6.16: Summary statistics for measured and reported parental heights, and mid-parental heights in case-control study (reported fathers height is taken from mother).

| | Group | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|---|
| Measured mothers | case | 142.5 | 157.0 | 161.1 | 161.4 | 165.0 | 179.0 | 6.22 | 104 |
| | control | 142.5 | 158.9 | 162.4 | 162.7 | 167.0 | 181.6 | 6.52 | 112 |
| Z(Mother) | case | -3.536 | -1.135 | -0.443 | -0.403 | 0.196 | 2.518 | 1.031 | 104 |
| | control | -3.536 | -0.812 | -0.244 | -0.181 | 0.532 | 2.949 | 1.082 | 112 |
| Reported fathers | case | 157.5 | 170.2 | 172.7 | 174.6 | 180.3 | 195.6 | 7.81 | 104 |
| | control | 154.9 | 172.7 | 177.8 | 177.0 | 182.9 | 193.0 | 7.24 | 113 |
| Z(Father) | case | -3.037 | -1.178 | -0.807 | -0.535 | 0.309 | 2.540 | 1.144 | 104 |
| | control | -3.409 | -0.807 | -0.063 | -0.181 | 0.681 | 2.168 | 1.060 | 113 |
| MPH | case | 149.8 | 159.3 | 166.1 | 166.6 | 172.7 | 188.3 | 8.83 | 104 |
| | control | 153.7 | 160.9 | 166.3 | 168.2 | 175.6 | 189.0 | 8.89 | 112 |

### 6.6.3 Obtaining Z-scores for height conditional on mid-parental height

In the above analysis for the systematic sample we just considered the reported height data. If we now consider all the parental data for the systematic sample, then it seems reasonable to take the measured mothers height (over reported) if the child was also part of the control arm of the follow-up study. However for the fathers height data, we will only take the mothers report of the fathers height, if there was no fathers height reported on the consent form. One of the reasons for doing this, is that whoever completed the consent form may have consulted their partner.

These data can be used to obtain estimates of the correlation between reported parental heights and child's height. There are 417 children with both parents heights and 10 single mothers within the systematic sample. The correlation between the mothers and fathers reported heights is 0.2209 (95% CI [0.1276, 0.3104]), which is lower than quoted by Cole (1996). However, the 95% confidence interval for this correlation contains 0.3. Taking $r(m, f)$ to be 0.3 in equation (2.21), the Z-score for mid-parental height was obtained from the reported/measured mothers height Z-score and reported fathers height Z-score. Summary statistics of the Z-scores for mothers, fathers and mid-parental heights can be found in table 6.17.

The correlations between parental height Z-scores and child's height Z-score at follow-up can be found in table 6.18. It appears that there is a tendency for girls heights to be more correlated with mothers, fathers and mid-parental heights than boys heights. This contradicts Tanner et al. (1970) and Byrad et al. (1983) observations, but we are dealing with reported heights. The use of a correlation of 0.4 by Cole (2000b) seems reasonable because this value is within the confidence intervals for the correlation of mid-parental height with child height.

Table 6.17: Summary statistics of Z-score for mothers, fathers and mid-parental heights

| | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| Z(Mother) | -2.7370 | -1.0520 | -0.2090 | -0.2299 | 0.6336 | 5.4790 | 1.1488 | 427 |
| Z(Father) | -3.7810 | -1.1780 | -0.0629 | -0.2957 | 0.6807 | 2.9110 | 1.1651 | 417 |
| Z(MPH) | -3.8120 | -1.1370 | -0.3378 | -0.3238 | 0.3846 | 4.3960 | 1.1250 | 417 |

Table 6.18: **Correlation between Z-score for height at follow-up and parental height Z-scores**: correlation coefficients and approximate 95% CI's

| Z-score | group | correlation | n | 95% CI |
|---|---|---|---|---|
| mothers height | all | 0.3447 | 416 | [0.2571, 0.4267] |
| | boys | 0.3168 | 210 | [0.1895, 0.4336] |
| | girls | 0.3819 | 206 | [0.2587, 0.4928] |
| fathers height | all | 0.3677 | 406 | [0.2804, 0.4490] |
| | boys | 0.2511 | 202 | [0.1172, 0.3762] |
| | girls | 0.4659 | 204 | [0.3509, 0.5669] |
| mid-parental height | all | 0.4572 | 406 | [0.3766, 0.5309] |
| | boys | 0.3721 | 202 | [0.2467, 0.4852] |
| | girls | 0.5322 | 204 | [0.4260, 0.6240] |

In order to assess whether a child is an appropriate height for their parents heights we need to condition the child's current height Z-score on their mid-parental height Z-score. Substituting a child's height Z-score and their mid-parental height Z-score into equation (3.11) leads to the following equation:

$$Z_{h|mph} = \frac{Z_h - rZ_{mph}}{\sqrt{1-r^2}} \tag{6.1}$$

where $r$ is the correlation between $Z_h$, the child's height Z-score and $Z_{mph}$, their mid-parental height Z-score. In the follow-up study the correlation between height and mid-parental height is about 0.46 (see scatterplot on left of figure 6.13).

Summary statistics of the Z-score for height conditional on mid-parental height can be found in table 6.19. It would appear that there is no reason to doubt that the variance of the Z-score for height conditional on mid-parental height is one (see table 6.20). There is no reason to doubt that the mean of the Z-scores for height conditional on mid-parental height is zero for girls. However, there is reason to doubt that the mean of the Z-scores for height conditional on mid-parental height is zero for boys. Therefore, it would be appropriate to use conventional cut-offs ($\pm 2.67$) for detecting girls with unusually low or high heights given mid-parental height. However, if the same cut-offs were used with the Z-score for height conditional on mid-parent height for boys we would detect too few boys with unusually short heights and too many boys with unusually high heights. This may indicate the need to use a sex-specific correlation between child's height Z-score and mid-parental height Z-score. It is also desirable that the Z-score for height conditional on mid-parental height is uncorrelated with the mid-parental height Z-score, in fact the correlation

is small ($r = -0.0841$).

Boxplots of the Z-score for height conditional on mid-parental height at follow-up by sex can be found on the right of figure 6.13. The median Z-score for height conditional on mid-parental height exhibits the same pattern as the mean, namely that there is reason to doubt that the median is zero for boys (V = 12611, p = 0.0046). In figure 6.14 we produce a variable width notch boxplot of the Z-score for height conditional on mid-parental height by level of deprivation; the noticeable feature is that there is less variability in the conditional Z-scores within the affluent group.

If we now consider the parental height data from the case-control study. In figure 6.15 we produce a scatter plot of the Z-score for child's height at follow up versus

Table 6.19: **Systematic sample (excluding cases)**: Summary statistics of Z-score for height at follow-up conditional on mid-parental height

| Z(h.fol\|mph) | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| All | -2.5270 | -0.5486 | 0.1246 | 0.1311 | 0.7900 | 2.7820 | 0.9741 | 406 |
| Boys | -2.2400 | -0.4686 | 0.2092 | 0.2121 | 0.8318 | 2.7820 | 0.9949 | 202 |
| Girls | -2.5270 | -0.6430 | 0.0081 | 0.0508 | 0.7723 | 2.4180 | 0.9488 | 204 |

Table 6.20: **Systematic sample (excluding cases)**: Results of testing that (a) the mean Z-score for height conditional on mid-parental height is zero (by sex) and (b) the variance of the Z-score for height conditional on mid-parental height is one (by sex)

| | t | p | 95% CI |
|---|---|---|---|
| All | 2.7113 | 0.0070 | [0.0360, 0.2261] |
| Boys | 3.0304 | 0.0028 | [0.0741, 0.3502] |
| Girls | 0.7652 | 0.445 | [-0.0801, 0.1818] |
| | $\frac{(n-1)s^2}{\sigma^2}$ | approx p-value | 95% CI |
| All | 384.31 | 0.4673 | [0.8307, 1.0945] |
| Boys | 198.95 | 0.9184 | [0.8216, 1.2158] |
| Girls | 182.73 | 0.3144 | [0.7478, 1.1045] |

the Z-score for mid-parental height labelled by case-control status. The scatterplot indicates that the case and control group form two overlapping clusters. If we now initially just consider the mid-parental height Z-score for the case and control groups. Summary statistics for the mid-parental height Z-score by case-control status can be found in table 6.21. Quantile-quantile plots indicate that there seems no reason to doubt normality of mid-parental height Z-scores. If we compare the means of the mid-parental height Z-scores in the case control study, we find that they are significantly different (t = -2.4793, p = 0.0139). On the left of figure 6.16 we produce a notch boxplot of the mid-parental height Z-scores by case control status, the notches don't overlap indicating that the median mid-parental height Z-score is significantly different in the case and control groups (W = 4593.5, p = 0.0074).

The Z-scores for height conditional on mid-parental height for individuals within the case-control study were calculated using equation (6.1) with $r = 0.46$. Summary statistics for the Z-score of height at follow up conditional on mid-parental height can be found in table 6.22. In contrast to the cases, the Z-scores for height conditional on mid-parental height of controls cover a larger range. In figure 6.16 we produce a notch boxplot of the Z-scores for height conditional on mid-parental height by sex. At follow-up the notches on the case and control boxplots by sex do not overlap indicating that the medians within the case and control groups are significantly different (W = 413, p = $1.449 \times 10^{-4}$ for boys and W = 1481, p = 0.0024 for girls).

The quantile-quantile plots of Z-scores for height conditional on mid-parental height (see figure 6.17) indicate that there may be some reason to doubt normality (one outlying individual in plot for case girls and evidence of skewness in plot for controls). Thus tests to compare means in case and control groups were not pursued further.

Table 6.21: Case-Control study: Summary statistics for mid-parental height Z-scores

| Group | Min | LQ | Median | Mean | UQ | Max | SD | no. |
|---|---|---|---|---|---|---|---|---|
| case | -2.7050 | -1.2230 | -0.7655 | -0.58160 | 0.0500 | 1.9120 | 1.0263 | 104 |
| control | -3.6160 | -0.8396 | -0.2571 | -0.2292 | 0.4332 | 2.4330 | 1.0622 | 112 |

Table 6.22: Case-Control study: Summary statistics of Z-score for height conditional on mid-parental height at follow-up

| Sex | Group | Min | LQ | Median | Mean | UQ | Max | SD | no. |
|---|---|---|---|---|---|---|---|---|---|
| Boys | case | -2.7900 | -1.6670 | -0.6907 | -0.7259 | 0.0851 | 0.9009 | 0.9823 | 39 |
| | control | -1.3280 | -0.5965 | 0.0813 | 0.3046 | 0.7817 | 2.8180 | 1.0844 | 41 |
| Girls | case | -4.9780 | -1.2970 | -0.4808 | -0.5024 | 0.2749 | 2.3210 | 1.1905 | 62 |
| | control | -1.8630 | -0.5318 | 0.0598 | 0.0742 | 0.6391 | 2.4360 | 0.8408 | 69 |

## 6.7 Discussion and Conclusions

In this chapter we have considered the growth outcomes of the Newcastle sample at 7-9 years. It is assumed that the systematic sample (excluding cases) is representative of children in Newcastle at 7 to 9 years. Anthropometric data from the systematic sample was used to assess the adequacy of the revised UK 1990 growth reference for converting height, weight, BMI and head circumference measures to Z-scores. There seems to be no reason to question the normality of the Z-scores for height, weight, BMI and head circumference (boys only). There is no reason to doubt that the mean of weight, height and BMI Z-scores is zero. There is no reason to doubt that variance of Z-scores is one for height, weight and BMI (girls only). However, there is reason to doubt that the variance of the BMI Z-scores for boys is one. It would appear that the Z-scores for BMI, height (girls), weight (girls) are appropriately adjusted for age. However, there is some indication that the Z-scores for weight and height for boys may not adjust completely for age. The main conclusion, is that it not entirely unreasonable to use the revised UK 1990 reference to convert weight, height and BMI to Z-scores at 7-9 years.

The conclusion that the UK 1990 reference is suitable for converting BMI to Z-scores at 7 to 9 years is at odds with the observations made in the Wessex growth study (Voss and Mulligan 1999b). Voss and Mulligan (1999b) observed the change in BMI of 120 children of average height from school entry until 16 to 17 years, they found at five that the distribution of BMI corresponds reasonably closely to the UK 1990 reference (mean BMI close to 45th centile), however by 16-17 years the mean BMI was close to the 65th centile. Rudolf et al. (2000) also noted a similar trend of increasing BMI Z-score with age in their study of children aged 7 to 10 years from schools in the Leeds area. Reilly et al. (1999) also noted a tendency for an

increasing prevalence of overweight and obesity with age, from 6 until 15 years, in a nationally representative sample of 2360 children.

There is strong evidence to suggest that the mean Z-score for head circumference is less than zero. Thus indicating that children in Newcastle have smaller head circumferences than those that contribute to the UK 1990 growth reference. There is no reason to doubt that the variance of head circumference Z-scores is one for boys. However, the head circumference data for girls has several outlying measurements which results in a variance for the head circumference Z-scores that is greater than one. It would appear that the Z-scores for head circumference are appropriately adjusted for age. One possible reason for the poor match between the Newcastle children and the revised UK 1990 reference is that Newcastle children may have a higher level of deprivation resulting in smaller head circumferences. Wright et al. (1992) found that the mean head circumference of primary school children was lower in a sample of 219 children from a deprived area of Newcastle compared to a sample of 1016 children from Oxfordshire (with few areas of deprivation). Another possibility is that the revised UK 1990 reference for head circumference may not be nationally representative because this is derived from children within the Edinburgh growth study (Ratcliffe et al. 1994) beyond infancy (Cole et al. 1998). It should also be noted that in the creation of the head circumference references (Cole et al. 1998), slight evidence of skewness was observed for the girls and the tails were relatively heavy. Recent consensus also suggests that there is a need for revision of the UK 1990 reference for head circumference beyond infancy (Wright et al. 2002).

Rudolf et al. (2000) assessed the adequacy of the revised UK 1990 growth reference using weight and height measures from children aged 7 to 10 years from 10 primary schools in the Leeds area. Rudolf et al. (2000) found that children in Leeds were marginally heavier and taller than those children the contributed to the UK 1990 growth reference but concluded this could reflect participation in the APPLES project. Rudolf et al. (2000) also noted a tendency for an increasing mean Z-score with age for weight in boys and both weight and height for girls. Rudolf et al. (2000) supported the use of the revised UK 1990 growth reference for converting height, weight and BMI to Z-scores. Reilly et al. (1999) assessed the suitability of the revised UK 1990 reference for converting heights and weights to Z-scores at 7.5 years. Reilly et al. (1999) found that there was no reason to doubt that the mean Z-score for height and weight was zero. However Reilly et al. (1999) didn't consider the variance of the Z-scores but did note an excess of girls below the 10th centile for weight.

The effect of level of deprivation (in infancy) on each of the anthropometric

measures at 7-9 years was tentatively explored. It would appear that level of deprivation has little impact on BMI and head circumference at 7-9 years. Other authors have also noted that level of deprivation has little impact on BMI because the more deprived tend to have an appropriate weight for their lower height (Tanner 1989). There may be some indication that level of deprivation may influence the height and weight Z-score at 7 to 9 years. There may be a tendency for the more deprived to be shorter and lighter but sample sizes involved are too small to be conclusive. A similar pattern is observed for parental heights, with the more deprived fathers being shorter.

The anthropometric data at 7 to 9 years from the case-control study was converted to Z-scores. The case children were found to be significantly shorter, lighter, leaner and to have smaller head circumferences than control children. These observations are in general agreement with the follow up study of Dowdney et al. (1998) on infants with prolonged growth retardation (weight gain faltered in first year of life and below tenth centile for weight and height at 4 years) followed up at 11 years. A rough comparison on the Z-score scale suggests that the case children participating in the Dowdney et al. (1998) study tend to be shorter, lighter, leaner and have smaller head circumferences than the cases within the Newcastle study. It would also appear that the observed differences in height, weight, BMI and head circumference Z-scores are greater between case and control boys than case and control girls. Research in Peru arrived at a similar conclusion, namely that girls that had FTT demonstrated better catch up than boys that had FTT when followed up for 15 years (Rudolf and Hochberg 1990).

The measured mothers heights and reported fathers heights from the case-control study were considered. It was found that there was no difference in the mean heights of mothers from the case and control study. However, it was found that the median of the reported heights of case fathers was significantly lower than that of control fathers. We should be aware that this difference is related to mothers reporting of fathers heights, although Himes and Roche (1982) found that mothers tended to overestimate their husbands heights by about 1.3 cm. Hypothetically, the shorter reported heights of case fathers could indicate that some of the children with faltering weight gain were just homing in towards their genetic potential. Smith et al. (1976) found that birth length related predominantly to maternal size whereas by two years of age the length correlated best to mean parental height, reflecting genetic growth factors of both parents. Smith et al. (1976) also observed that those shifting downward did not decelerate until after the first 3-6 months, indicating those that were 'lagging down' did so in mid-infancy. A further possibility is that the case

fathers may not have achieved their genetic potential because of deprivation in their own childhood. A literature research revealed that the observation that fathers of FTT infants tend to be shorter has not been noted elsewhere.

Research on adult outcome of 'normal' short[2] and underweight children at 7 years suggests that one in three of these children became short or underweight adults (Greco et al. 1995). The long term implications of failure-to-thrive in infancy are unknown, these results in mid-childhood indicate that the growth status of these children is affected, which in turn may have future implications, for example: delayed puberty, failure to achieve genetic potential in height and may effect peak bone density. There are also psychological consequences for children that have short stature (Skuse 1987), although recent research by Voss and Mulligan (1994) (in Stratford et al. (1999)) found that there were no differences between short and normal height children on any measures of self-esteem, intelligence or behaviour. However teachers ratings of attainment were lower for the shorter group but this difference was removed when an allowance was made for the difference in socio-economic distribution between the groups. Many of the children within the Wessex growth study are aware of the group to which they belong (Stratford et al. 1999).

Data collected on reported parental heights was converted to Z-scores using the revised UK 1990 growth reference values for height at 23 years. It was found that a child's height Z-score was reasonably correlated with their mid-parental height Z-score ($r = 0.46$). There was also a slight tendency for girls height Z-scores to be more correlated with parental height Z-scores than boys. This observation contradicts Tanner et al. (1970) although we are dealing with reported heights here rather than measured heights. The Z-score for height conditional on mid-parental height was calculated using a correlation of 0.46. It was found that the Z-score for height conditional on mid-parental height exhibited the desirable properties: uncorrelated with mid-parental height Z-score, no reason to doubt variance of Z-scores is one and no reason to doubt mean of Z-scores is zero (girls only). However, there is reason to doubt the mean of the Z-score for height conditional on mid-parental height is zero for boys. This may indicate the need for a sex-specific correlation in the calculation of the conditional gain Z-score. So we should be cautious when assessing a boys height given his mid-parental height.

The Z-score for height conditional on mid-parental height was calculated for children within the case and control study. It was found that the median of the

[2]Only children with organic causes of short stature were excluded. Greco et al. (1995) acknowledge that short stature may be a consequence of psychosocial causes such as failure-to-thrive.

Z-scores for height conditional on mid-parental height in the case and control group were significantly different. This suggests that even after accounting for mid-parental height, case children were shorter than control children.

To conclude, the use of the revised UK 1990 reference to convert weight, height and BMI to Z-scores is not entirely inappropriate. However, some caution may be needed when interpreting the Z-scores for BMI of boys. It was found that girls height Z-scores were more correlated with reported parental height Z-scores than boys. The use of the Z-score for height conditional on mid-parental height to assess a child's current height looks promising. However, this conditional Z-score should be interpreted cautiously at the extremes for boys. As reported by Drewett et al. (1999), case children were found to be significantly shorter (even after adjusting for mid-parental height), leaner and lighter than control children. Furthermore, it appears that boys are more sensitive than girls to failure-to-thrive in infancy. The significant difference between reported fathers' heights of case and control children suggest that there may be some genetic component to the growth faltering observed in infancy.

Figure 6.1: Mean weight Z-scores for cases and controls, with 95% confidence intervals (after correction to birth weight Z-score)

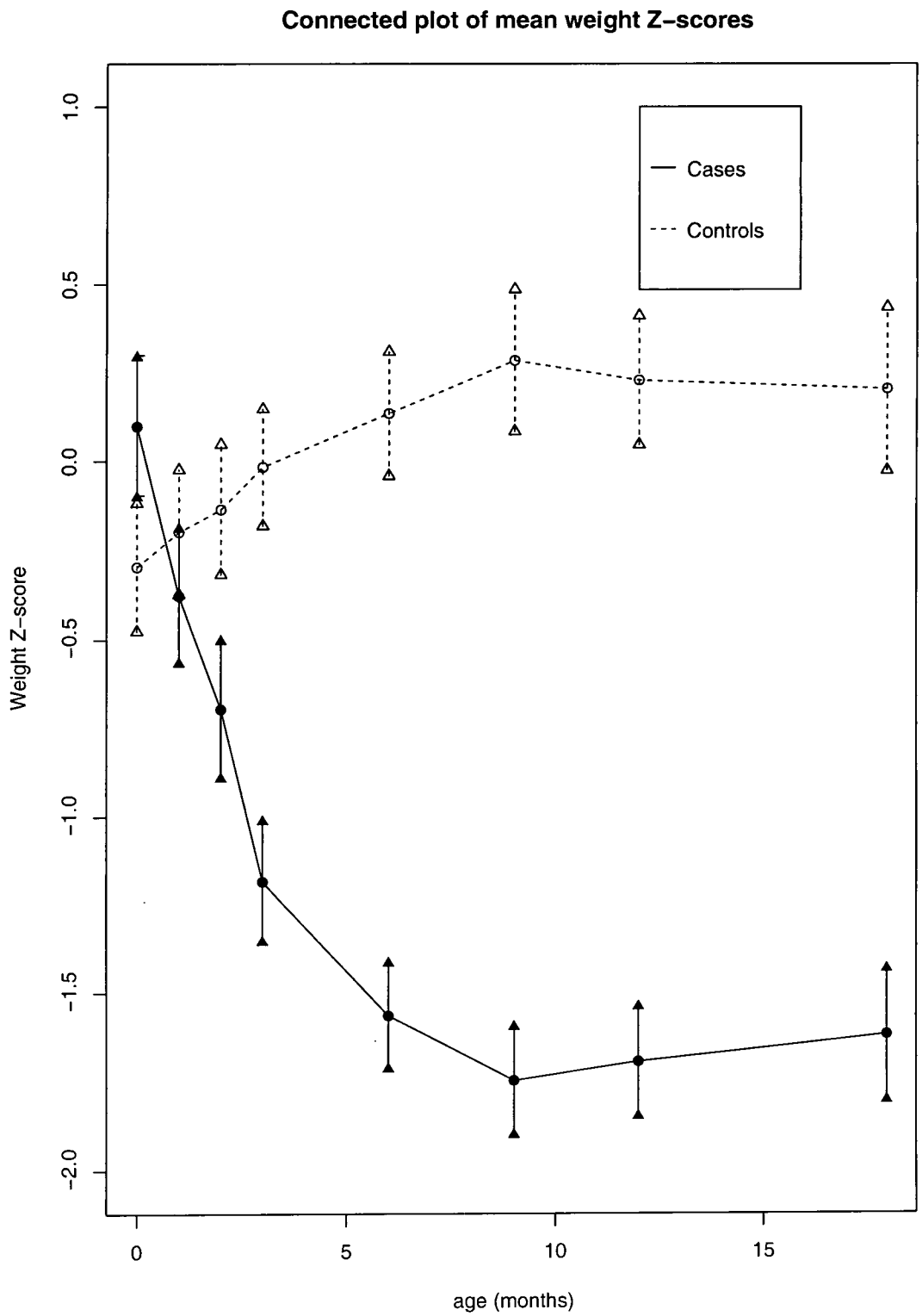


Figure 6.2: **Systematic Sample (excluding cases)**: Box-plots of Z-scores for heights, weights, BMI's and head circumference by sex

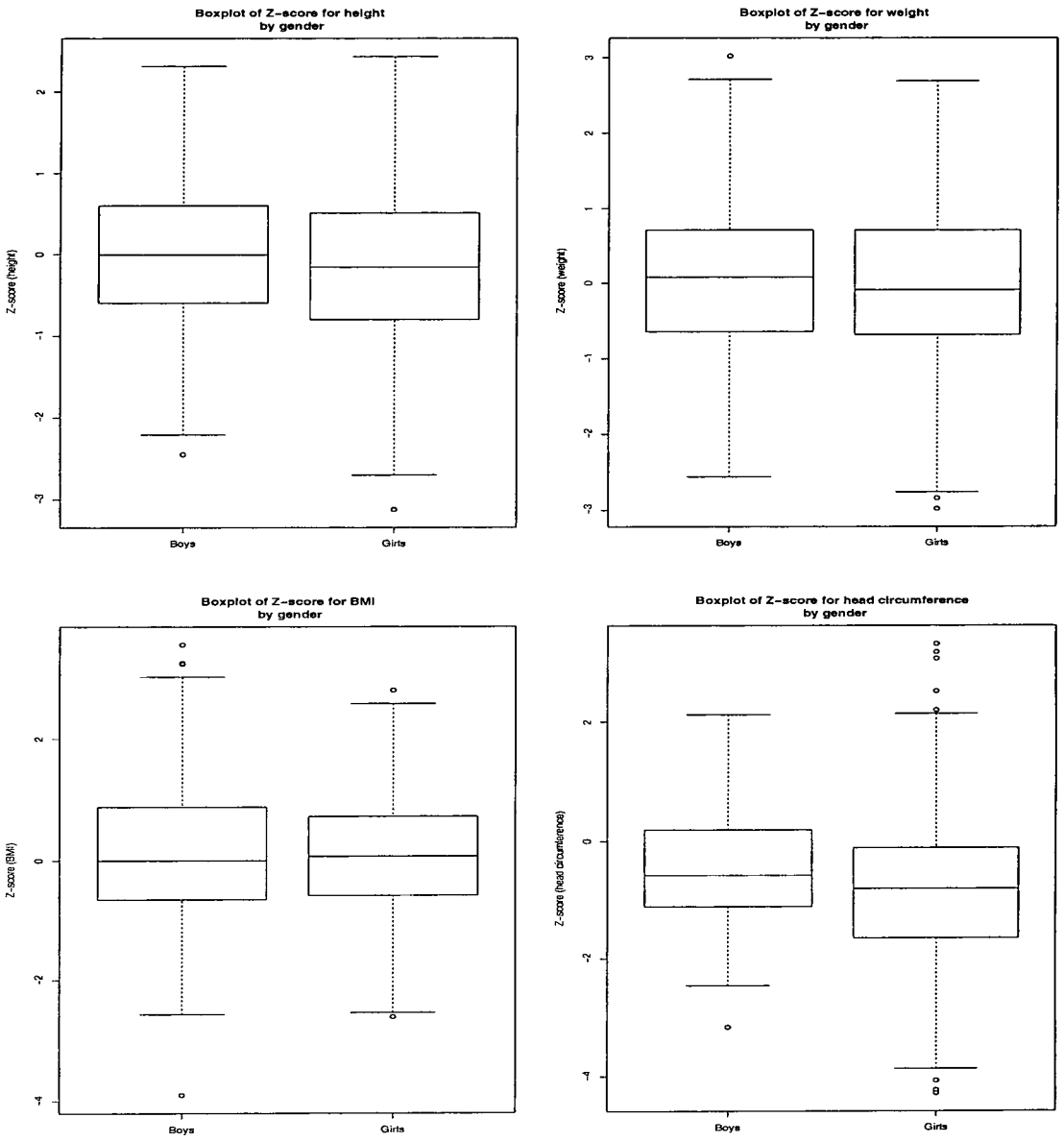


Figure 6.3: **Systematic Sample (excluding cases)**: Plot of Z-scores for height, weight, BMI and head circumference versus age for boys

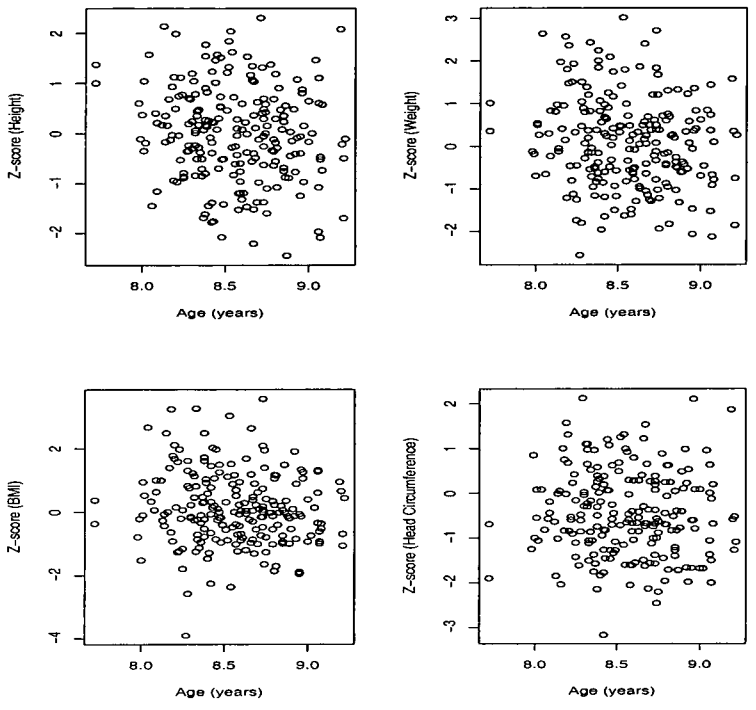


Figure 6.4: **Systematic Sample (excluding cases)**: Plot of Z-scores for height, weight, BMI and head circumference versus age for girls

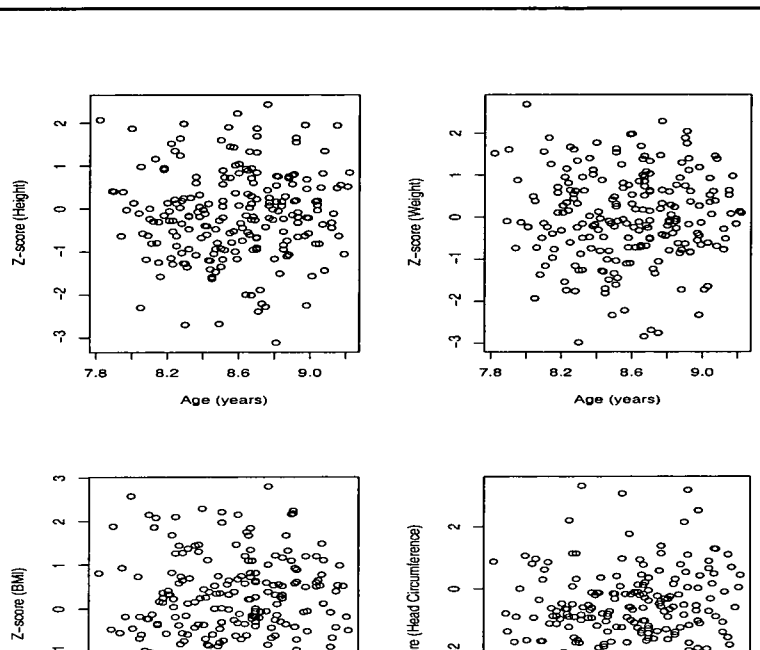


Figure 6.5: **Systematic Sample (excluding cases)**: Variable width notch box-plots of Z-scores for heights and weights by sex and level of deprivation

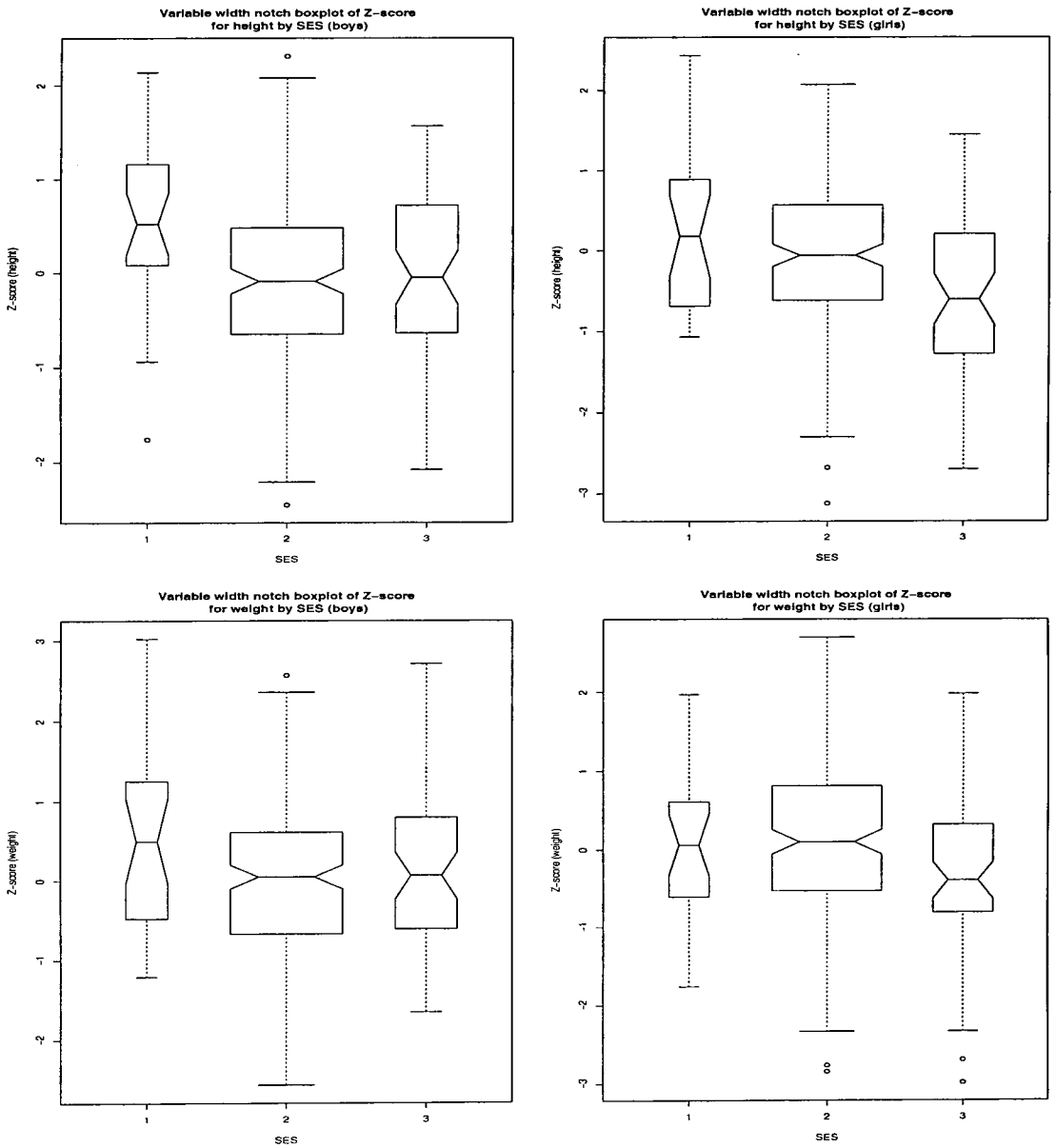


Figure 6.6: **Systematic Sample (excluding cases)**: Variable width notch box-plots of Z-scores for BMIs and head circumferences by sex and level of deprivation

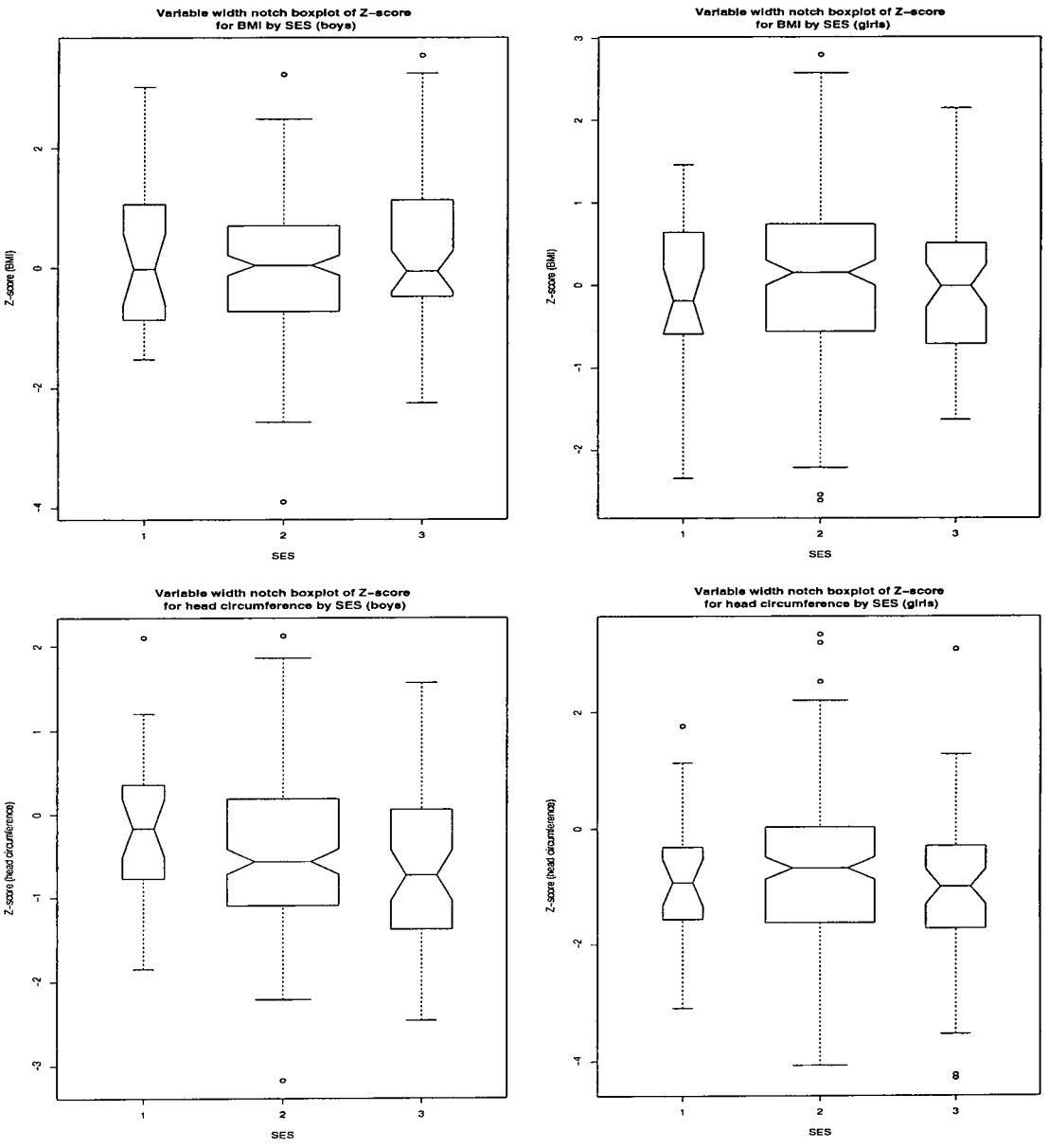


Figure 6.7: Variable width notch boxplots of Z-scores for height, weight, BMI and head circumference at follow up assessment (grouped by case-control status and sex)

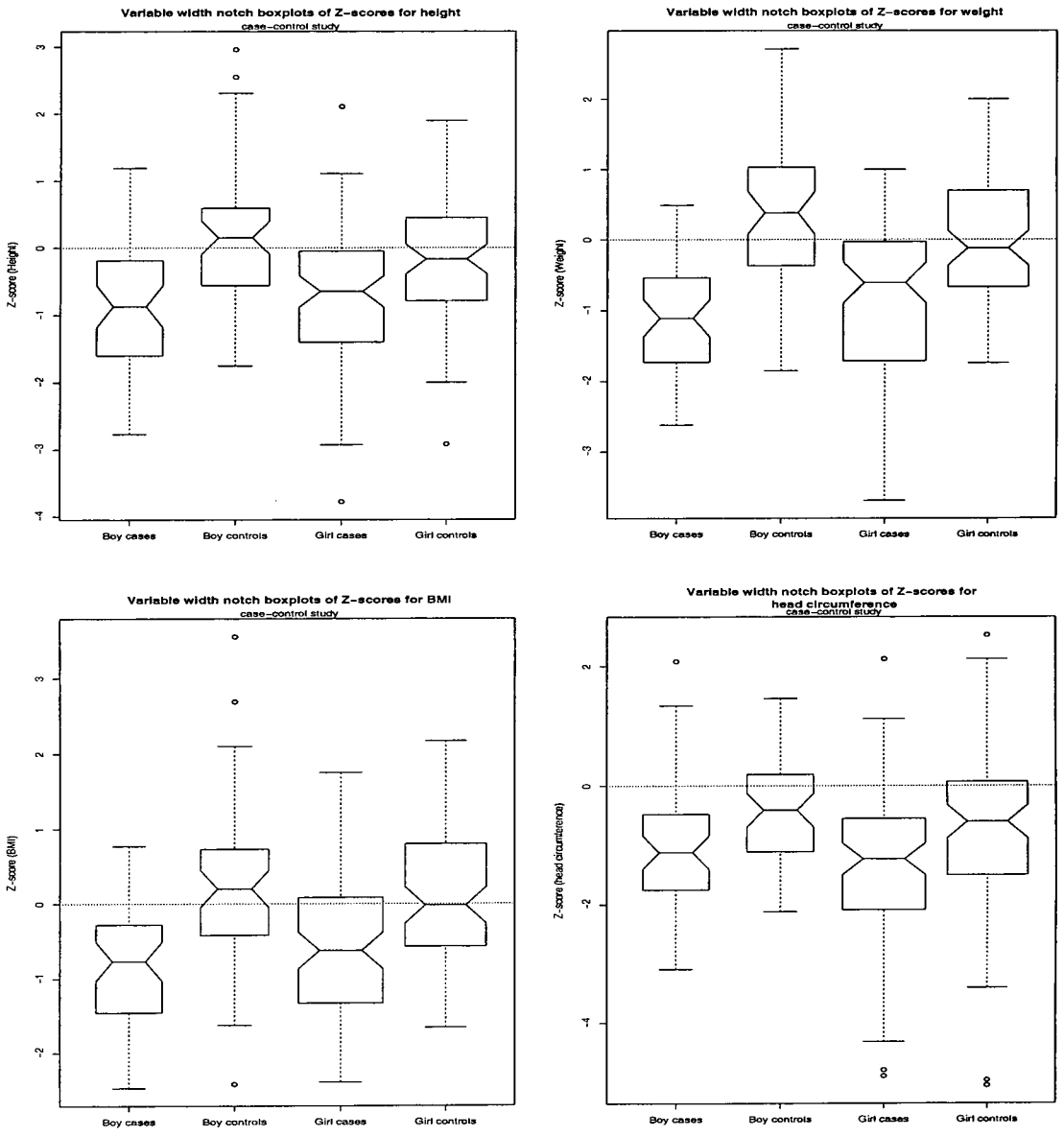


Figure 6.8: **Systematic sample**: Histograms of reported parental heights (feet)

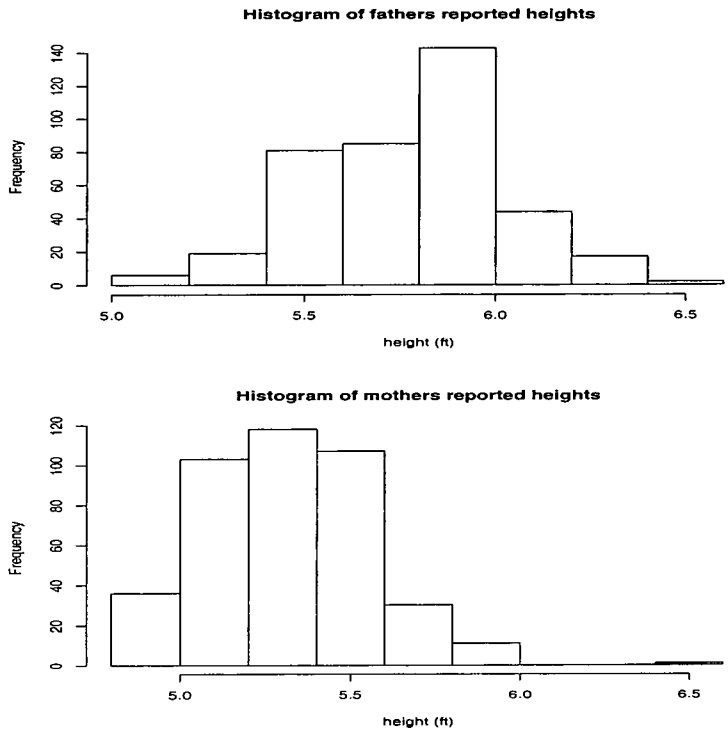


Figure 6.9: Boxplots of reported parental heights (metric) grouped by SES

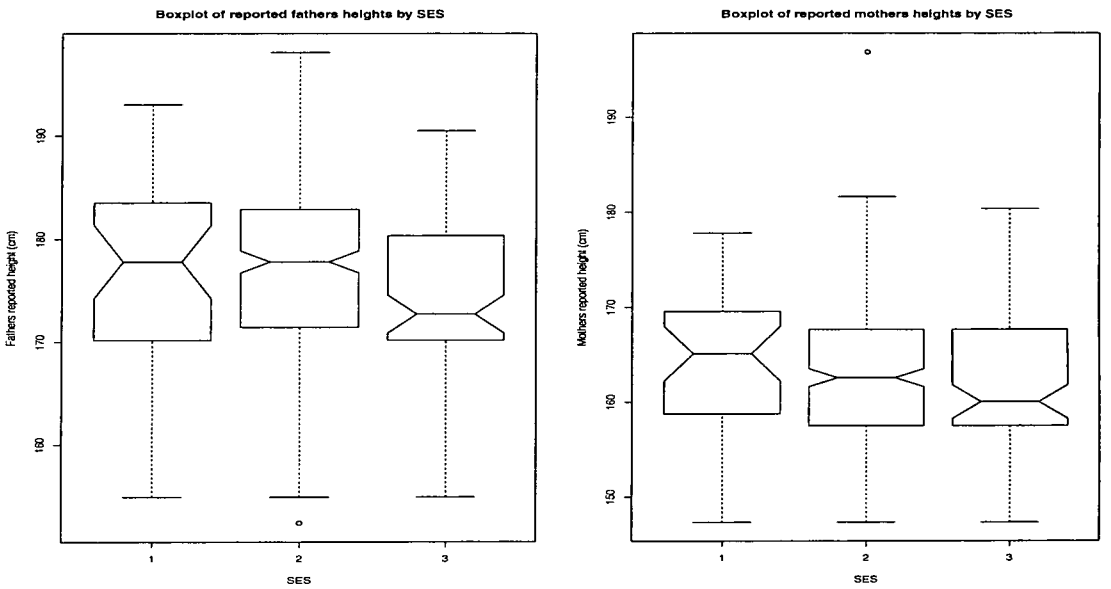



Figure 6.10: Quantile-quantile plots of measured mothers heights (case-control study)

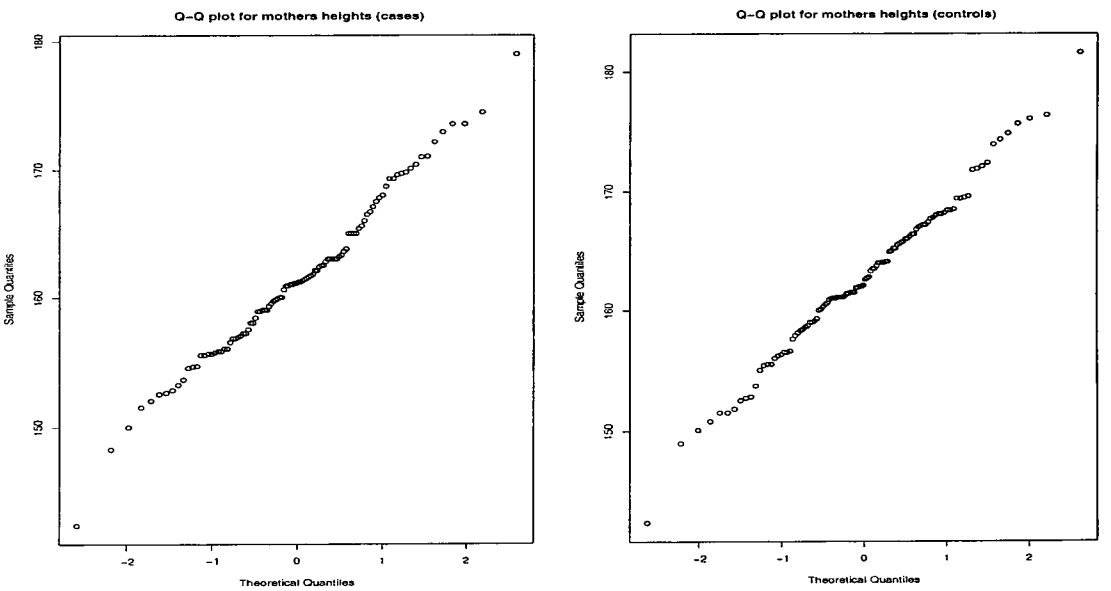


Figure 6.11: Notch boxplots of Z-scores for measured and reported parental heights grouped by case-control status

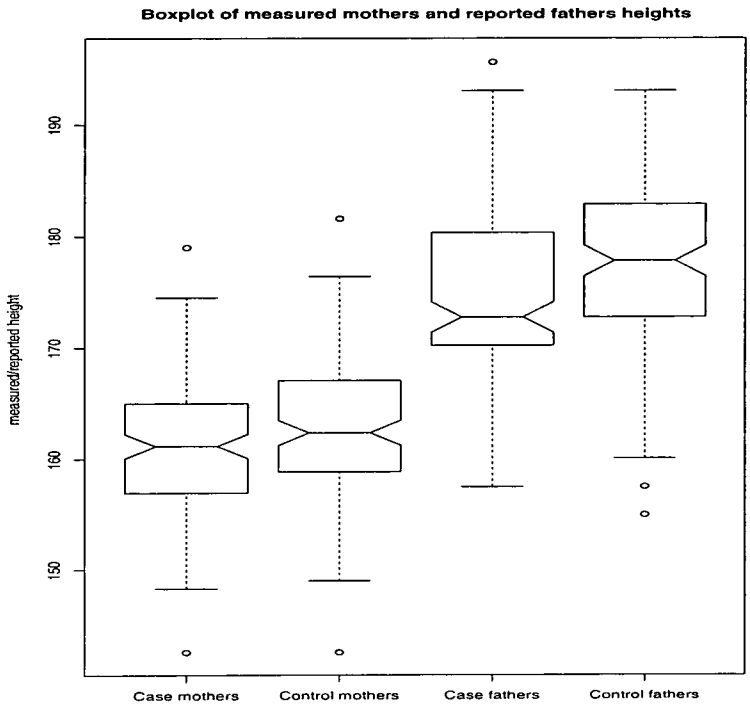


Figure 6.12: Plot of discrepancy between measured and reported mothers heights versus measured mothers heights

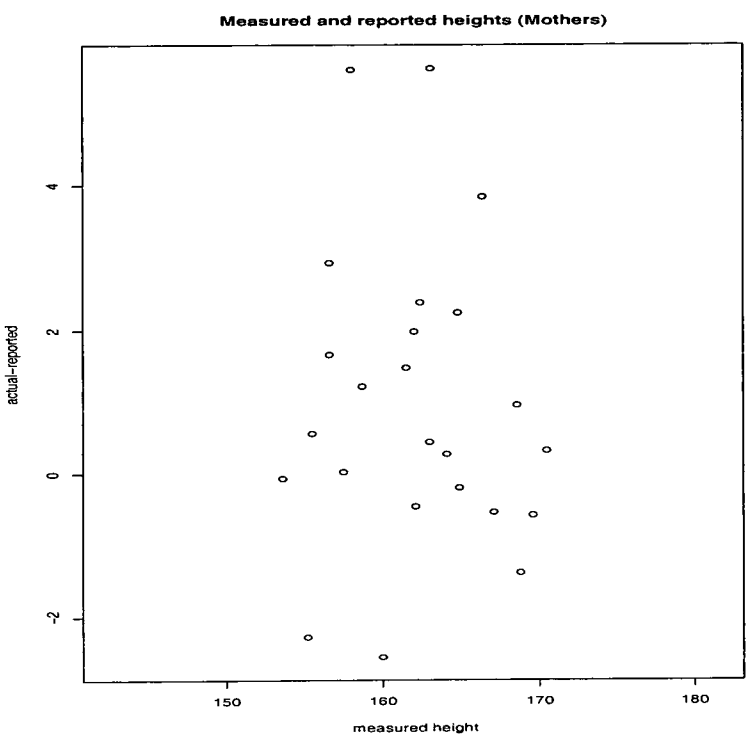


Figure 6.13: **Systematic sample (excluding cases)**: Left: Plot of Z-score for height at follow-up versus mid-parental height Z-score Right: Boxplot of Z-score for height conditional on mid-parental height by sex

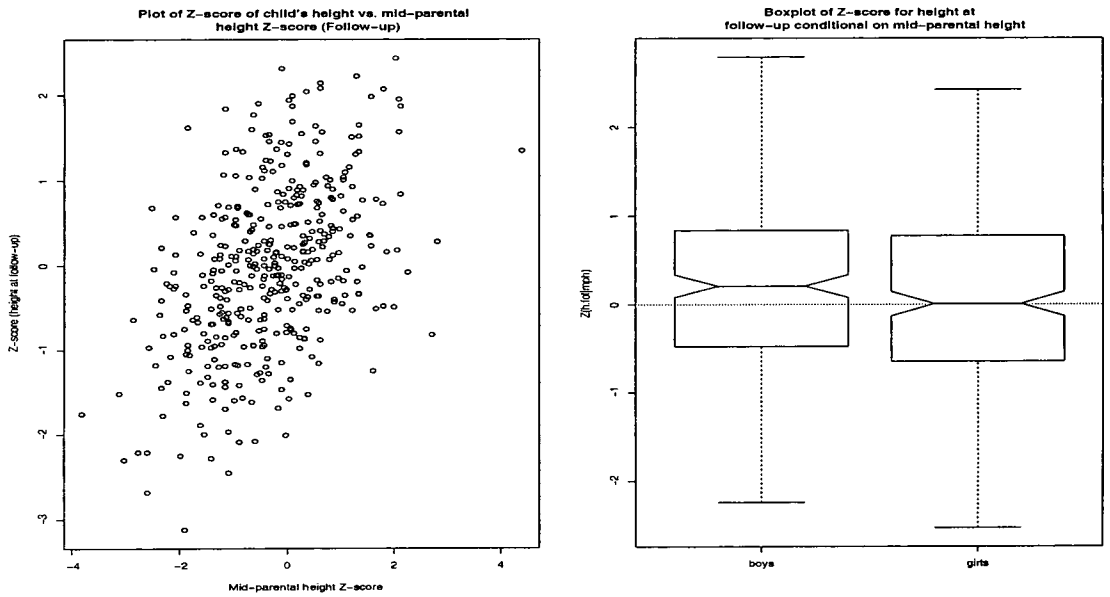



Figure 6.14: Systematic sample (excluding cases): Boxplots of Z-score for height at follow-up conditional on mid-parental height by level of deprivation

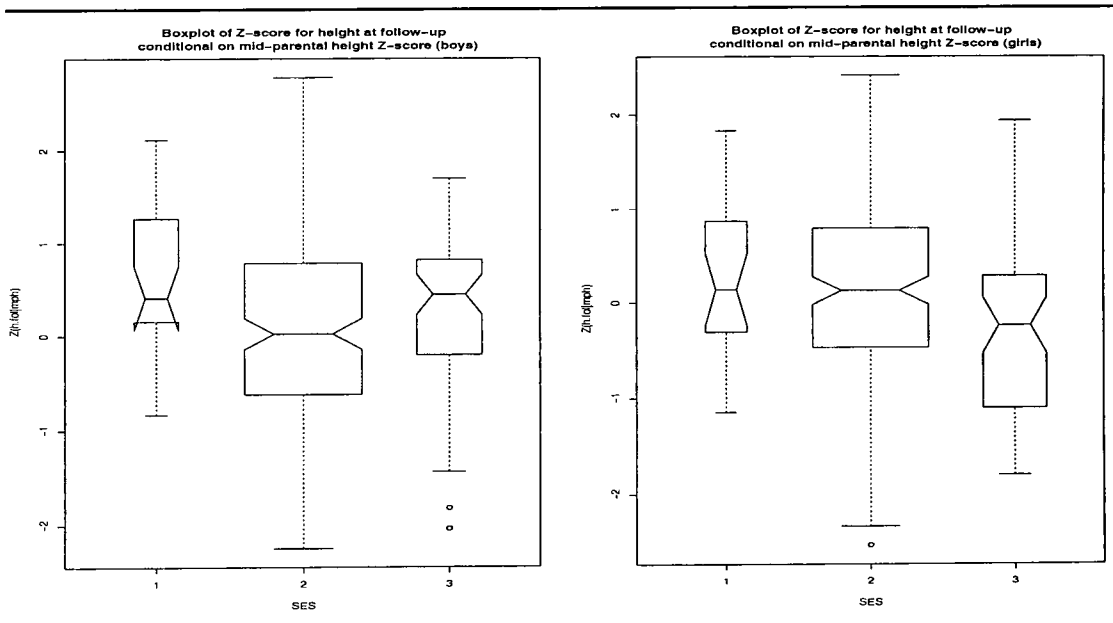


Figure 6.15: **Case-control study**: Plot of Z-score for height at follow-up versus mid-parental height Z-score

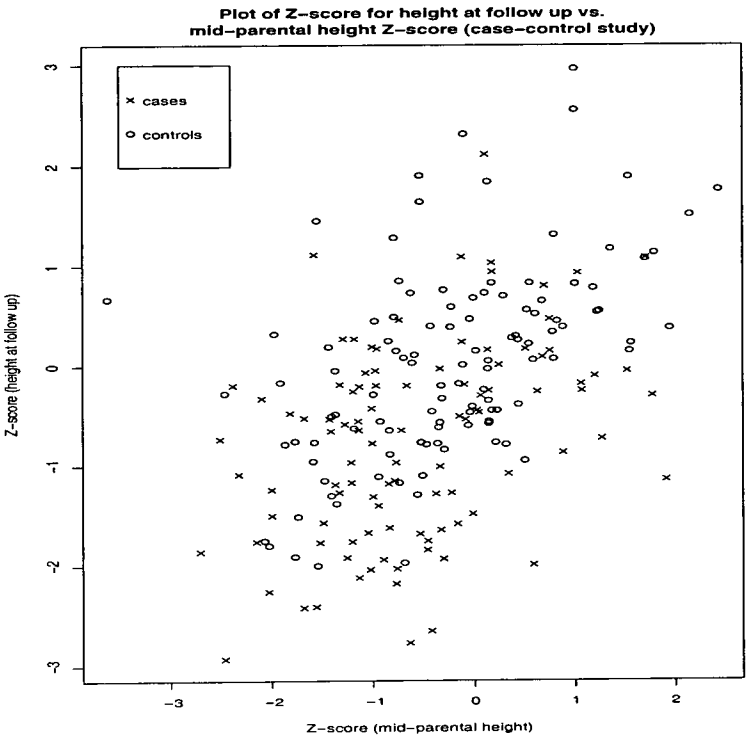


Figure 6.16: **Case-control study**:Left: Notch boxplot of mid-parental height Z-score by case-control status Right: Notch boxplot of Z-score for height conditional on mid-parental height by case-control status and sex at follow-up

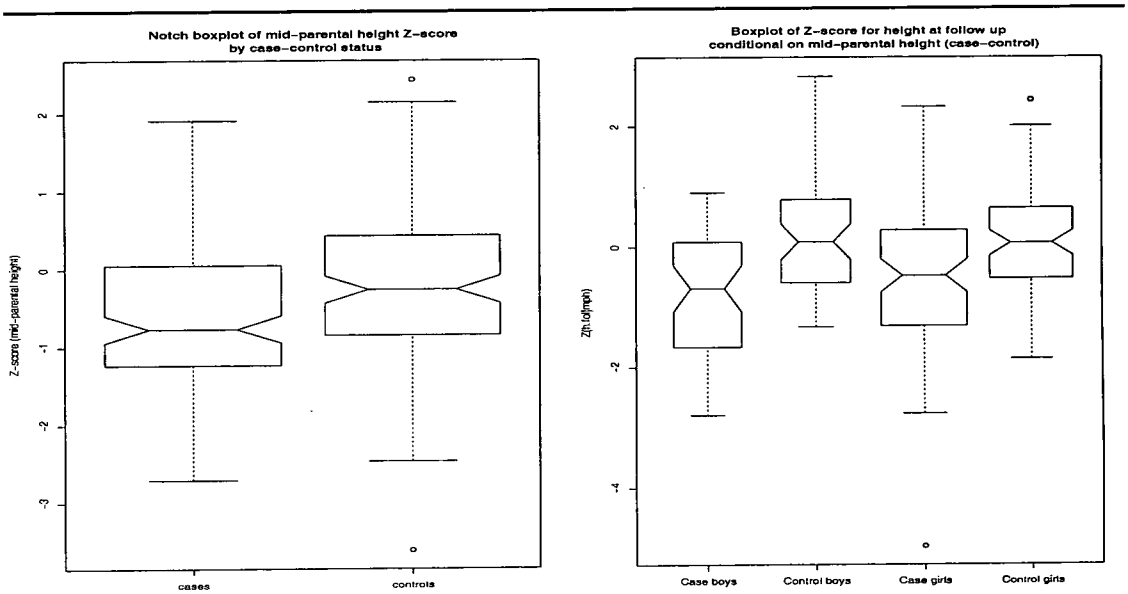
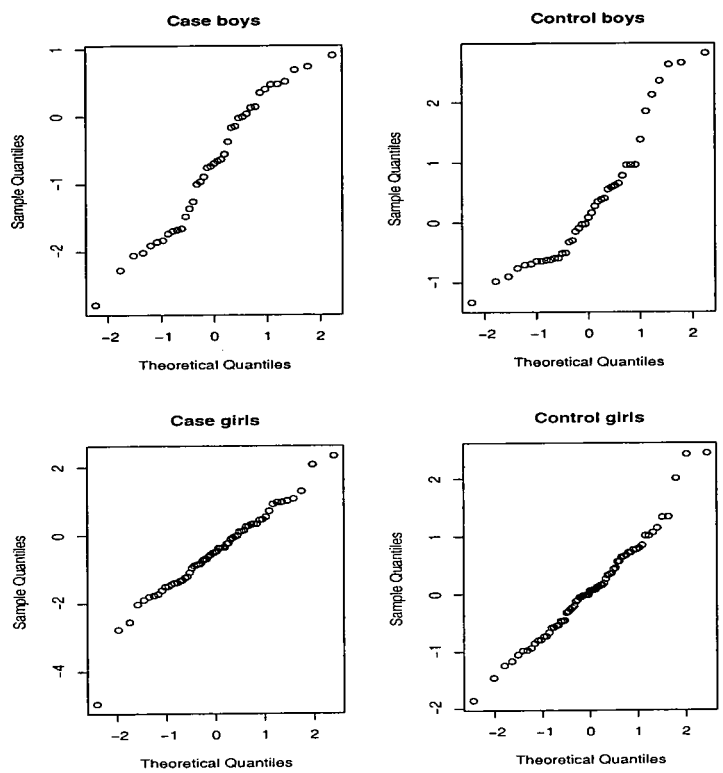


Figure 6.17: Quantile-quantile plots of Z-score for height at follow up conditional on mid-parental height (case-control study)



# Chapter 7

# The utilisation of conditional weight gain Z-scores

## 7.1 Introduction

In chapter 5 we developed a model for the correlation structure of weight Z-scores in infancy. Given a model for the correlation between any two time-points we can now calculate a conditional weight gain Z-score. In this chapter we explore possible ways of using and interpreting conditional weight gain Z-scores when we have more than two weights.

In section 7.2 we develop a one variable version of the Argyle model for the correlation structure. This allows us to calculate all possible conditional weight gain Z-scores for each child. In section 7.3 we consider the summary statistics of the conditional weight gain Z-scores generated when age is grouped to the nearest fortnight.

A conditional weight gain Z-score can be used to assess a child's weight gain between two measurement ages. However, infants are rarely weighed just twice in infancy; in the Newcastle infancy data frame only 133 (3.9%) infants have just two routine weights. Even then, there may only be two weights in the data frame because some of the clinic weights may have gone astray. The conditional weight gain Z-score methodology allows one to contrast two weights. If we stick to using pair-wise comparisons, then what is a meaningful way to interpret these? For example, an infant weighed 7 times in infancy would generate 21 pair-wise conditional weight gain Z-scores.

An additional issue is the short term variability of conditional weight gain Z-scores. The problem with a short interval is that the amount of growth will be small compared to the variability from measuring error and other causes (for example, biological variation, a saltatory pattern (Lampl et al. 1992) or pulsatile pattern (Greco et al. 1990) of growth). Cole (1995) suggests that conditional weight gain Z-scores calculated over short time periods should be interpreted cautiously, because even normally growing children are likely to throw up occasional significant weight gains simply by chance owing to the large number of possible measurement pairs. Cole (1995) demonstrates the variability of conditional weight gains using the routine weights of one infant (ID 2149) from the Newcastle dataset. Cole (1995) advocates using the longest available time interval to calculate a conditional weight gain Z-score so that the effect of measurement error is reduced. However, if a longer interval is used, a slowing-down in growth may persist for some time before a measurement occasion makes it possible to detect it. Moreover, a clinician who is concerned about a child's weight gain doesn't want to wait for more time to elapse before they act. If there are more than two weight measures then this increases the number of conditional weight gain Z-scores that can be obtained, and it raises the question of how to interpret and utilise these as they evolve.

In clinical situations, often a sequence of measurements is available. It would be desirable to use these measurements efficiently to judge the 'typicality' of the growth pattern rather than simply the growth at one point in time. This leads to the issue of how we characterise growth in such circumstances. An additional issue is the number of weight measures that are useful to condition on. In using the conditional weight gain approach contrasting a late weight with an early weight or a combination of weights, there is the implicit assumption that this early weight is in itself not 'abnormal' (Healy 1986).

The routine infancy data set has up to 11 weights for any individual leading to a maximum of 55 conditional weight gain Z-scores for each individual. This is a lot of weight Z-scores and conditional weight gain Z-scores to throw away in order to just contrast two weight measurements and derive one gain. The patterns exhibited by the conditional weight gain Z-scores may possibly tell us something about an infants growth and provide some indication of the variability. We may expect an infant that fails to thrive in infancy to have a high proportion of conditional weight gain Z-scores that are negative. Whereas for an infant growing 'normally' we may anticipate a set of conditional weight gains Z-scores that are 'close' to zero, perhaps fluctuating either side of zero.

It seems sensible to look at patterns in 'runs' of these conditional weight gain

Z-scores. The list below gives some sequences of conditional weight gain Z-scores that may be of interest for a child that has $n$ weights at times $t_1, \ldots, t_n$:

1. $Z(t_2|t_1)\ Z(t_3|t_1)\ \ldots\ Z(t_n|t_1)$

   In the majority of children, the first weight Z-score ($Z(t_1)$) will be the Z-score that corresponds to birth weight. However, birth weight is strongly influenced by maternal factors.

2. $Z(t_3|t_2)\ Z(t_4|t_2)\ \ldots\ Z(t_n|t_2)$

   We may view the first weight Z-score after the birth weight Z-score to be a more useful indicator of expected growth.

3. $Z(t_2|t_1)\ Z(t_3|t_2)\ \ldots\ Z(t_n|t_{n-1})$

   Here we compare a child's current weight Z-score to their last weight Z-score; this enables us to say something about recent growth. This pattern of conditional weight gain Z-scores is likely to identify unusual conditional weight gain Z-scores sooner. However this may be at a cost, owing to increased variability due to short term variation in growth.

4. $Z(t_2|t_1)\ Z(t_3|t_1, t_2)\ \ldots\ Z(t_n|t_1, \ldots, t_{n-1})$

   This pattern assumes that the growth trajectory the infant is expected to follow is continually updated at each time point from birth.

5. Some weighted combination that utilises possibly all of the conditional weight gain Z-scores, but also takes into account the short term variability of conditional weight gain Z-scores

In section 7.4 we consider the implications of using the model developed in section 7.2 to calculate conditional weight gains in situations (3) and (4) above. We will find that the model developed for correlation between weight Z-scores in infancy has an effect on how we assess weight gain patterns in infancy.

In section 7.5 we explore trends in the sequence of conditional weight gain Z-scores suggested above. We also consider the impact of level of deprivation and sex on the conditional weight gain Z-scores. In section 7.6 we explore patterns in runs of conditional weight gain Z-scores. We also suggest other possible strategies for detecting infants that are experiencing growth faltering. We also propose a possible cost mechanism to aid the decision of arriving at a suitable threshold for 'no change' in weight Z-score.

We conclude by discussing the graphical display of conditional weight gain Z-scores, see section 7.7. In this section we present the expected pattern of weight gain derived from the model proposed for the correlation between weight Z-scores. We conclude this section by presenting the conditional fifth centile for this same model.

## 7.2 Developing a practical model for the correlation structure of weight Z-scores in infancy

In chapter 5, we developed the Argyle model and arrived at the 'best' model, in the least squares sense for the correlation structure. The final model from table 5.25 was:

$$\log(r(t1,t2)) = -0.0368 + 0.2729\log(t1+2) - 0.2598\log(t2+2) \qquad (5.47)$$

where $t1$ and $t2$ are the initial age and later age (given in weeks).

A desirable feature of any correlation model is that it predicts a correlation in the range -1 to 1, inclusive and equal to one when $t1 = t2$. If equation (5.47) is used to predict correlation in the age range birth to 2 years, then this model falls down in late infancy. For example if t1=404 and t2=425 days, then the predicted correlation is 1.0039. It is suspected that in late infancy the correlation may be close to 1. In figure 7.1 we produce a contour plot for the correlation generated from equation (5.47); the correlation is greater than one in the area between the dashed line (representing $t1 = t2$) and the contour representing $r = 1$.

An alternative approach is to use a one variable version of the Argyle model, developed below. To arrive at a one variable model, the constant added to the time-point is varied, in the same way as described in Chapter 5. However, instead of trying to minimise the deviance, we now want the coefficients of both log terms to be of similar magnitude. The addition of 1.6 weeks gave coefficients that were of the same magnitude to three decimal places, see upper table in table 7.1 for summary of fit. The lower table provides a summary of the fit if the Argyle model has one coefficient and a constant. The accompanying residual plots for both models in table 7.1 can be found in figure 7.2, these illustrate that the one coefficient model gives as good a fit as the two coefficient model. A plot of the Cook's distance for the one variable Argyle model versus index, indicates that there are three influential points (see figure 7.3); corresponding to the correlation between birth weight Z-score and weight Z-score at: 76, 4 and 6 weeks. The affect of excluding these three influential points on the constant added to age was explored. If all three correlations were

excluded then the optimum constant was 1.1 weeks. A summary of the fit of this model can be found in the uppermost table of table 7.2. The middle table in table 7.2 summarises the fit of the one variable Argyle model with a constant. Notice that the constant term is not significantly different from zero. In general, assuming a zero intercept is a very strong assumption and is in general not justifiable (Draper and Smith 1998). Possible exceptions are when there is data in the region of the origin or when the model is 'known' to pass through the origin (Draper and Smith 1998). If we consider excluding the constant term, which although this is a strong assumption does give a correlation of 1 when $t1 = t2$. Therefore for practical purposes it seems appropriate to use the model in the bottom table of table 7.2 (note that the summary statistics are excluded here, because the multiple correlation coefficient provides a measure of the usefulness of all the terms in the model with the exception of the intercept). In figure 7.4 we plot the standardised residuals versus fitted values for the one coefficient models in table 7.2; the impact of excluding the intercept term appears to be negligible. A contour plot of the correlation derived from equation (7.2) can be found in figure 7.5. Therefore, from this point onwards, the correlation between weight Z-scores is obtained using equation (7.2).

$$\log(r(t1, t2)) = 0.24[\log(t1 + 1.1) - \log(t2 + 1.1)] \tag{7.1}$$

$$r(t1, t2) = \left(\frac{t1 + 1.1}{t2 + 1.1}\right)^{0.24} \tag{7.2}$$

where $t1$ and $t2$ are the initial and later age in weeks.

In figure 7.6 we plot the fitted curve for the Argyle model ($c = 1.1$), along with the confidence and prediction intervals on a log-scale and the original scale. The Newcastle fortnightly correlations are well within the prediction intervals.

## 7.3 Properties of conditional weight gain Z-scores

Conditional weight gain Z-scores should (by definition) be normally distributed, have zero mean and variance one and be uncorrelated with initial weight Z-score (Cole 1995). However, we should be most concerned with the standard deviation of the distribution of the conditional weight gain Z-scores, because an inflated standard deviation would result in more individuals than expected having extreme gains (Cole 1995).

It is only possible to consider the properties of conditional weight gain Z-scores generated from individuals that contributed to the correlations modelled, otherwise

Table 7.1: **Newcastle correlations based on data grouped to nearest fortnight [N=385]**: Regression of log correlation on $\log(t1 + 1.6)$ and $\log(t2 + 1.6)$, and $\log(t1 + 1.6) - \log(t2 + 1.6)$ using weighted least squares.

$$\log(r(t1, t2)) = a\log(t1 + 1.6) + b\log(t2 + 1.6) + c + \epsilon$$

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.003906 | 0.015972 | 0.245 | 0.807 |
| $\log(t1 + 1.6)$ | 0.255243 | 0.003912 | 65.251 | $< 2 \times 10^{-16}$ |
| $\log(t2 + 1.6)$ | -0.255246 | 0.005519 | -46.247 | $< 2 \times 10^{-16}$ |

$R^2$=0.9198, $R^2$(adj)=0.9194, residual SE=0.307 on 382 df

$$\log(r(t1, t2)) = a[\log(t1 + 1.6) - \log(t2 + 1.6)] + b + \epsilon$$

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.003896 | 0.003252 | 1.198 | 0.232 |
| $\log(t1 + 1.6) - \log(t2 + 1.6)$ | 0.255243 | 0.003850 | 66.293 | $< 2 \times 10^{-16}$ |

$R^2$=0.9198, $R^2$(adj)=0.9196, residual SE=0.3066 on 383 df

Table 7.2: **Newcastle correlations based on data grouped to nearest fortnight [N=382]**: Regression of log correlation on $\log(t1 + 1.1)$ and $\log(t2 + 1.1)$, and $\log(t1 + 1.1) - \log(t2 + 1.1)$ using weighted least squares (after excluding three most influential points: correlations at(0,4), (0,6) and (0,76))

$$\log(r(t1, t2)) = a\log(t1 + 1.1) + b\log(t2 + 1.1) + c + \epsilon$$

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.002110 | 0.016334 | 0.129 | 0.897 |
| $\log(t1 + 1.1)$ | 0.238981 | 0.003675 | 65.037 | $< 2 \times 10^{-16}$ |
| $\log(t2 + 1.1)$ | -0.239950 | 0.005420 | -44.272 | $< 2 \times 10^{-16}$ |

$R^2$=0.9205, $R^2$(adj)=0.9201, residual SE=0.3056 on 379 df

$$\log(r(t1, t2)) = a[\log(t1 + 1.1) - \log(t2 + 1.1)] + b + \epsilon$$

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.001170 | 0.003187 | -0.367 | 0.714 |
| $\log(t1 + 1.1) - \log(t2 + 1.1)$ | 0.239123 | 0.003604 | 66.341 | $< 2 \times 10^{-16}$ |

$R^2$=0.9205, $R^2$(adj)=0.9203, residual SE=0.3052 on 380 df

$$\log(r(t1, t2)) = a[\log(t1 + 1.1) - \log(t2 + 1.1)] + \epsilon$$

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| $\log(t1 + 1.1) - \log(t2 + 1.1)$ | 0.240031 | 0.002617 | 91.72 | $< 2 \times 10^{-16}$ |

we would be considering the properties of conditional weight gain Z-scores generated from small sample sizes. Even then there are still 385 groups of conditional weight gain Z-scores. Plots of the mean, standard deviation, minimum and maximum for each group of conditional weight gain Z-scores are presented in figures 7.7 to 7.10. These summary statistics were plotted against the time-elapsed between measurements and the mean of two measurement ages. This approach was used because plotting summary statistics against initial and later age was found to be uninformative. However, preliminary work on individuals with extreme conditional weight gain Z-scores suggested that gains for weight measurements made closer in time were more variable.

In figure 7.7 we plot the mean of the conditional weight gain Z-scores versus time elapsed and average age. There appears to be a tendency for the mean conditional weight gain Z-score to be positive and more variable for weight measures made close in time. However there are more extreme positive conditional weight gain Z-scores than extreme negative conditional weight gain Z-scores: which are likely to be influential on the mean. There is no obvious trend in the plot of mean conditional weight gain Z-score against average age. The resulting t-values from testing that the mean conditional weight gain Z-score is zero were also plotted against the time elapsed and average age; see figure 7.11. Informally using a cut-off value of 2 for the t-values (the sample size is greater than 50), there are 125 (32.5%) sets of conditional weight gain Z-scores that have a mean significantly different from zero. The two most extreme t-values correspond to groups of conditional weight gain Z-scores calculated for weight Z-scores close in time: a t-value of 6.34 when $t1$ and $t2$ are 14 and 18 weeks, respectively and a t-value of -3.79 when $t1$ and $t2$ are 0 and 4 weeks, respectively. Recall that the correlation between birth weight Z-score and weight Z-score at 4 weeks was identified as an influential point when fitting models to the Newcastle correlations. The t-values appear to have a linear trend with average age with higher values in the early weeks of infancy which gradually decrease in late infancy.

In figure 7.8 we plot the standard deviations of the conditional weight gain Z-scores against time elapsed and average age. These plots illustrate that for weight measures close in time the standard deviation of the conditional weight gain Z-score is more variable. The two most extreme standard deviations (SD $\approx$ 1.8) correspond to conditional weight gain Z-scores obtained between 48 and 54 weeks, and 40 and 44 weeks. An unexpected feature of the plot of the standard deviation of the conditional weight gain Z-score versus average age; is the upward trend in the magnitude of the standard deviation as average age increases. In figures 7.9 and 7.10

we plot the minimum and maximum of the conditional weight gain Z-scores versus time elapsed and average age, respectively. These show the expected trend, that when the two measurement occasion are close the minimum and maximum of the conditional weight gain Z-scores are of greater magnitude. In figure 7.12 we plot the approximate p-values from testing that the variance of the conditional weight gain Z-scores is one, against the time elapsed and average age. There is no apparent trend with age in the resulting approximate p-values from testing that the variance is one. However, we must remember that the F-test is sensitive to non-normality. There is reason to doubt that the variance is one for 241 (62.6%) groups of conditional weight gain Z-scores. Further work is needed to establish the cause of the elevated variance of conditional weight gain Z-scores. The apparent trend in variance with age could be related to the smaller sample sizes as age increases. In figure 7.13 we plot the approximate p-value, from testing that the variance is one, against sample size to explore if there was any relationship between sample size and resulting p-values. Again there doesn't appear to be any strong trend in relation to sample size. The elevated variance of the conditional weight gain Z-scores may be a result of elevated variances in the original weight Z-scores or indicative of some deficiency in the model for correlation between weight Z-scores. Recall from chapter 4, that the variance of the weight Z-scores for the full sample from 6 to 9 months onwards were elevated. Thus the elevated variance on the original weight Z-score scale will contribute to the elevated variance of the conditional weight gain Z-scores. We also explored whether there was a relationship between t-values and approximate p-values from testing if variance is one; see figure 7.13. It would appear that there is no apparent trend between significant t-tests and significant variance tests.

In theory, the conditional weight gain Z-score should be uncorrelated with the initial weight Z-score. We calculated the correlation between the initial weight Z-score and the conditional weight gain Z-score. In figure 7.14 we plot this correlation against average age and time elapsed. There appears to be no strong trend in the correlations with relation to age, with the majority of correlations between -0.3 and 0.3. Two correlations are close to -0.4, these are from conditional weight gain Z-scores calculated, close in time, towards the end of the first year ($t1 = 44$ and $t2 = 54$ weeks, and $t1 = 40$ and $t2 = 44$ weeks).

In theory, the conditional weight gain Z-scores are expected to be normally distributed. In figure 7.15 we plot the p-values from the Shapiro-Wilk normality test against average age and time elapsed. There is no apparent trend with average age or time elapsed. However, in first few weeks of infancy (average age less than 5 weeks) there appears to be reason to doubt the normality of the conditional weight

gain Z-scores. There is no reason to doubt the normality of the conditional weight gain Z-scores for 238 (61.8%) groups of conditional weight gain Z-scores.

## 7.4 Implications of correlation model

In assessing an infants pattern of weight gain we may be interested in following two sequences of conditional weight gain Z-scores: conditioning on previous weight Z-score and conditioning on all previous weight Z-scores. In this section we consider the implications of using the one variable Argyle model, given by equation (7.2), for modelling the correlation between weight Z-scores on these two sequences of conditional weight gain Z-scores.

When conditioning on all previous weight Z-scores, we wish to obtain the following conditional weight gain Z-score:

$$Z_{t+1|1,\ldots,t} = \frac{Z_{t+1} - E(Z_{t+1}|\underline{Z})}{\sqrt{Var(Z_{t+1}|\underline{Z})}} \tag{7.3}$$

Therefore we need to obtain the conditional expectation and conditional variance. Assuming that the Z-scores are multivariate Normal then the conditional expectation and variance are given by equations (7.4) and (7.5), respectively.

$$\begin{aligned} E(Z_{t+1}|\underline{Z}) &= E(Z_{t+1}) + Cov(Z_{t+1}, \underline{Z})Var^{-1}(\underline{Z})(\underline{Z} - E(\underline{Z})) \\ &= Cov(Z_{t+1}, \underline{Z})Var^{-1}(\underline{Z})\underline{Z} \text{ as } E(Z_j) = 0 \text{ for } j = 1, \ldots, t+1 \\ &= \underline{r}^T R^{-1} \underline{Z} \end{aligned} \tag{7.4}$$

where $Var(Z_j)$ is assumed to be 1 for $j = 1, \ldots, t+1$, $\underline{r} = (\rho_{1,t+1}, \rho_{2,t+1}, \cdots, \rho_{t,t+1})^T$ and

$$R = \begin{pmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,t} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,t} \\ \vdots & & \ddots & \vdots \\ \rho_{t,1} & \cdots & \cdots & 1 \end{pmatrix}$$

Similarly,

$$\begin{aligned} Var(Z_{t+1}|\underline{Z}) &= Var(Z_{t+1}) - Cov(Z_{t+1}, \underline{Z})Var^{-1}(\underline{Z})Cov(\underline{Z}, Z_{t+1}) \\ &= 1 - \underline{r}^T R^{-1} \underline{r} \end{aligned} \tag{7.5}$$

Equation (7.2) is used to determine $\rho_{i,j}$, so $\rho_{i,j}$ can be written as:

$$\rho_{i,j} = \frac{v_i}{v_j} \tag{7.6}$$

where $v_i$ and $v_j$ are functions of $t_i$ and $t_j$, respectively and $v_i < v_j$ for $t_i < t_j$. Exploratory research suggested that conditional weight gain Z-scores obtained by conditioning on previous weight Z-score were the same as those obtained by conditioning on all previous weight Z-scores.

Theorem

$$Z(t_n|t_1,\ldots,t_{n-1}) = Z(t_n|t_{n-1}) \quad \text{if } \rho_{i,j} = \frac{v_i}{v_j} \text{ where } t_i < tj \tag{7.7}$$

Thus suggesting that:

$$E(Z_n|Z_{n-1}\ldots Z_1) = E(Z_n|Z_{n-1}) \tag{7.8}$$

$$Var(Z_n|Z_{n-1}\ldots Z_1) = Var(Z_n|Z_{n-1}) \quad \text{if } \rho_{i,j} = \frac{v_i}{v_j} \text{ where } t_i < tj \tag{7.9}$$

Proof

We need to find:

$$E(Z_{m+1}|Z_m\ldots Z_1) = \underline{r}^T R^{-1}\underline{z} = \underline{z}^T R^{-1}\underline{r} \tag{7.10}$$

$$Var(Z_{m+1}|Z_m\ldots Z_1) = 1 - \underline{r}^T R^{-1}\underline{r} \tag{7.11}$$

If we let $R$ be the $m \times m$ symmetric matrix given by:

$$R = \begin{bmatrix} 1 & \frac{v_1}{v_2} & \frac{v_1}{v_2} & \cdots & \frac{v_1}{v_m} \\ \frac{v_1}{v_2} & 1 & \frac{v_2}{v_3} & \cdots & \frac{v_2}{v_m} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \frac{v_1}{v_m} & \cdots & \cdots & \cdots & 1 \end{bmatrix}$$

and let $\underline{r}$ (the correlation between $(Z_1,\ldots,Z_m)$ and $Z_{m+1}$) and $\underline{e}$ be both $m \times 1$ vectors:

$$\underline{r} = \begin{bmatrix} \frac{v_1}{v_{m+1}} \\ \frac{v_2}{v_{m+1}} \\ \vdots \\ \frac{v_m}{v_{m+1}} \end{bmatrix} \quad \text{and} \quad \underline{e} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

Then

$$R\underline{e} = \begin{bmatrix} \frac{v_1}{v_m} \\ \frac{v_2}{v_m} \\ \vdots \\ \frac{v_m}{v_m} \end{bmatrix} = \frac{v_{m+1}}{v_m} \begin{bmatrix} \frac{v_1}{v_{m+1}} \\ \frac{v_2}{v_{m+1}} \\ \vdots \\ \frac{v_m}{v_{m+1}} \end{bmatrix} = \frac{v_{m+1}}{v_m}\underline{r} \tag{7.12}$$

We need to evaluate $R^{-1}\underline{r}$. As $R\underline{e} = \frac{v_{m+1}}{v_m}\underline{r}$, then:

$$R^{-1}\underline{r} = \frac{v_m}{v_{m+1}}R^{-1}R\underline{e} = \frac{v_m}{v_{m+1}}\underline{e} \tag{7.13}$$

So if we substitute the result given by equation (7.13) into equations (7.10) and (7.11), we find:

$$\begin{aligned} E(Z_{m+1}|Z_m \dots Z_1) &= \underline{z}^T R^{-1}\underline{r} \\ &= \frac{v_m}{v_{m+1}}Z_m \\ &= E(Z_{m+1}|Z_m) \end{aligned} \qquad \begin{aligned} Var(Z_{m+1}|Z_m \dots Z_1) &= 1 - \underline{r}^T R^{-1}\underline{r} \\ &= 1 - \frac{v_m}{v_{m+1}}\underline{r}^T\underline{e} \\ &= 1 - \frac{v_m^2}{v_{m+1}^2} \\ &= Var(Z_{m+1}|Z_m) \end{aligned}$$

Alternatively:

$$\begin{aligned} \rho_{t_i t_j} &= \left(\frac{t_i + 1.1}{t_j + 1.1}\right)^{0.24} \\ &= \exp(0.24\log(t_i + 1.1) - 0.24\log(t_j + 1.1)) \\ &= \exp|\tau_i - \tau_j| \end{aligned} \tag{7.14}$$

Equation (7.14) is Markov with a rescaling of the time axis. Therefore the data imply a model which has the Markov property. In assuming that the weight Z-scores are normally distributed and that the model for the correlation between weight Z-scores in infancy is given by equation (7.2) we arrive at a Markov property. Therefore, in order to assess a child's progress, we only need to consider the current weight Z-score conditioned on the last weight Z-score. Implicitly the child's previous weight Z-score contains all the information of all previous weight Z-scores. It also answers the question of how many previous observations is it useful to condition on, namely the last available (or one).

## 7.5 Exploring trends in conditional weight gain Z-scores

In this section we aim to explore whether infants with failure-to-thrive exhibit different growth patterns to 'normally' growing infants. We will consider three different ways of using conditional weight gain Z-scores; namely conditioning on first weight Z-score, conditioning on first weight Z-score after birth and conditioning on previous

weight Z-score. As discussed in section 7.4, using equation (7.2) to model correlation, conditioning on all previous weight Z-scores is the same as conditioning on just the previous weight Z-score.

In figures 7.16, 7.17 and 7.18 we plot the conditional weight gain Z-scores versus current age for conditioning on: first weight Z-score, first weight Z-score after birth and previous weight Z-score, respectively. Plots with lowess trend curves overlaid were produced for the full sample, cases, controls and others. Where 'others' refers to all the Newcastle infants excluding the infants identified as cases.

The lowess curves for the three sequences indicate that:

- Conditioning on first weight Z-score (see figure 7.16) The case group, have conditional weight gain Z-scores that drop away from zero at birth to about -2, they then experience a slight increase in conditional weight gain Z-score towards the end of infancy. In general, the case children still have negative conditional weight gain Z-scores at the end of infancy but their rate of deceleration in weight gain is slowing down. Indicating that individuals within the case group drop from relatively normal birth weights to subnormal weights mid-infancy. Whereas the control and others groups experience a slight increase in the time period soon after birth then level around the zero line.

- Conditioning on first weight Z-score after birth (see figure 7.17) The case group, have conditional weight gain Z-scores that drop away from first weight Z-score after birth to just above -2, they then experience a gradual increase towards the end of infancy (approaching a conditional weight gain Z-score of about -1). Indicating that individuals within the case group drop from their first weight Z-score after birth that is just below zero (median of -0.4 for first weight Z-score after birth) to subnormal weights mid-infancy. The control group, have conditional weight gain Z-scores that rise from first weight Z-score after birth to around 0.5 towards the end of the first year, there is then a slight decrease in value to around 0.25 at the end of infancy. Whereas the others group has conditional weight gain Z-scores around the zero line. It is likely that the different conditional weight gain pattern of control infants is a consequence of selecting individuals that are above the tenth centile TI cut-off.

- Conditioning on previous weight Z-score (see figure 7.18) The case group

drop away sharply from birth until about 6 months (conditional weight gain Z-score about -0.8), they then experience a gradual increase in their conditional weight gain Z-score relative to their attained low weight as they approach the end of infancy (the trend curve crosses the zero line at about 18 months). Indicating that the growth faltering of the infant relative to their previous position is 'slowing up'; representing a deceleration in the lack of growth. Whereas the others and control groups tend to fluctuate from just above the zero line to just below the zero line at about 1 year.

One interesting feature of these plots is the presence of a few extreme conditional weight gain Z-scores when conditioning on previous weight Z-score in early infancy. Cole (1995) has discussed the short term variability in weight gain, it is well known that the first year represents a period of rapid growth and it is possible that some kind of individual error is also incorporated in these plots. To explore the issue of variability in conditional weight gain Z-scores further, a random sample of 20 infants was taken from the 'Case' and 'Other' groups. The connected plots for the case and other samples were produced on the same scale to aid comparisons. A connected plot of weight Z-score versus ages for the case and other samples are produced in figure 7.19. The case sample's weight Z-scores have a downward trend in early infancy but there are some fluctuations in weight Z-score towards the end of infancy. Whereas the others sample's weight Z-scores fluctuate more in early infancy where weight measures are closer together in time than in late infancy. A connected plot of the conditional weight gain Z-score when conditioning on birth weight Z-score (or first weight Z-score) versus age for case and other samples are produced in figure 7.20. The majority of the case sample have negative conditional weight gain Z-scores from birth, with a few not experiencing negative conditional weight gain Z-scores until the third or fourth measurement occasion. In general, when conditioning on first weight Z-score, the case sample have a downward trend in early infancy, experience some fluctuations mid-infancy and then an upward trend towards the end of infancy. Whereas the connected plots for the others sample don't really have a general pattern and a few infants experience conditional weight gain patterns similar to those of the case children. A connected plot of the conditional weight gain Z-score when conditioning on first weight Z-score after birth versus age for case and other samples are produced in figure 7.21. The noticeable feature for the case sample is that the majority of the conditional weight gain Z-scores are negative and the pattern in general is similar to that observed when conditioning on birth (or first) weight Z-score. Again the plot for other sample has no general pattern. A

connected plot of the conditional weight gain Z-score when conditioning on previous weight Z-score versus age for case and other samples are produced in figure 7.22. In both of these plots there are more fluctuations in the conditional weight gain Z-score when conditioning on previous weight Z-score than in figures 7.20 and 7.21 where we were conditioning on birth weight Z-score or first weight Z-score after birth, respectively. The conditional weight gain Z-scores for the case sample tend to oscillate below zero whereas for the other sample the majority oscillate about zero.

A further issue when interpreting conditional weight gain Z-scores is the influence of one 'unusual' weight Z-score. In figure 7.23 we produce connected plots for weight Z-scores and conditional weight gain Z-scores of four individuals, from the others sample, to illustrate the impact of an unusually low or high weight Z-score (relative to rest of weight Z-scores) in early infancy. Conditioning on birth weight Z-score mirrors the weight Z-score pattern to some extent. Whereas conditioning on previous weight Z-score 'exaggerates' any fluctuation. These examples illustrate the caution needed when interpreting conditional weight gain Z-scores in early infancy. Furthermore, they also show the impact of conditioning on an 'unusual' weight Z-score.

It would be of interest to see if other factors, such as sex and level of deprivation exert any influence on the conditional weight gain Z-scores in infancy. In figures 7.24, 7.25 and 7.26 we have produced lowess trend curves for the three sequences of conditional weight gain Z-scores by sex and SES for the full infancy sample excluding cases (others). It would appear that female infants in general tend to have slightly lower conditional weight gain Z-scores than boys but both sexes have trend curves of a similar shape. A similar trend was observed on the original weight Z-score scale; see Chapter 4. An interesting trend is observed when considering conditioning on the previous weight Z-score; the most deprived children tend to have lower conditional weight gain Z-scores. Thus suggesting that level of deprivation may exert some influence on rate of weight gain between consecutive clinic weighing. When conditioning on the first weight Z-score after birth it would appear that the conditional weight gain Z-scores drop away more rapidly in late infancy for affluent infants. When conditioning on first weight Z-score both the affluent and deprived have similar conditional weight gain Z-scores in late infancy. The negative trend observed at the end of infancy in all of the plots produced in figures 7.24, 7.25 and 7.26 possibly reflects that children growing less well are monitored for longer.

## 7.6 Interpretation of conditional weight gain Z-score patterns

It was thought that patterns in conditional weight gain Z-scores may be informative about a child's weight gain and also provide some indication of the variability. We may expect an infant growing 'normally' to have conditional weight gain Z-scores that are 'close' to zero, perhaps fluctuating either side of zero. Whereas an infant with slow weight gain may have a high proportion of 'negative' conditional weight gain Z-scores. Similarly an infant with excessive weight gain may be expected to have a high proportion of conditional weight gain Z-scores that are 'positive'. Thus it seemed sensible to look at patterns in 'runs' of these sequences of conditional weight gain Z-scores: conditioning on first weight Z-score, conditioning on first weight Z-score after birth, and conditioning on previous weight Z-score.

### 7.6.1 Preliminary work on patterns of conditional weight gain Z-scores

Starting with the conventional approach used when applying the 'runs' test to a series of data, a child was said to have a '+' if their conditional weight gain Z-score was positive and a '−' if their conditional weight gain Z-score was negative. Considering a subset of the data which had weights in the seven grouped ages allocated by Dr C.M. Wright. Tabulation of the individual patterns for the runs from birth revealed that the series of conditional weight gain Z-scores were unlikely to come from a completely random process. When conditioning on previous weight Z-score, the run of signs appear to be more random (possibly because the sample size is smaller) than when conditioning on first weight Z-score. We found that in general the 'number of negatives' was more informative for detecting poor weight gain than the run length of a series of conditional weight gain Z-scores. Preliminary analysis also indicated that control children exhibited similar growth patterns to the rest of the birth cohort. So here we only consider two groups of children, the cases and others, where others refers to children that are not cases.

Counting the number of negatives is not an effective way of comparing the case and others groups, because a child can have between 1 and 10 conditional weight gain Z-scores when conditioning on first weight Z-score or previous weight Z-score. Instead, the approach used here is to calculate the proportion of negatives. However this still has the downside that a child with only one conditional weight gain Z-score

that is negative is treated the same has a child that has all 10 conditional weight gain Z-scores that are negative.

The uppermost histograms in figure 7.27 summarise the distribution of the proportion of negatives in the case and other groups when conditioning on first weight Z-score. The cases have a high proportion of negatives (all above 0.5) with the majority of cases having all negative conditional weight gain Z-scores. This is not a surprising result because the case children were identified as an at risk group by using the 'thrive index' approach (see section 3.1), this contrasts late weight Z-scores with a weighted average of weight Z-scores at birth and grouped ages of 1 and 2 months. The others group have a both a high and low proportion of negatives, with a flattening in the distribution between proportions of 0.1 and 0.9.

The uppermost histograms in figure 7.28 summarise the distribution of the proportion of negatives in the case and other groups when conditioning on previous weight Z-score. Again the cases have a high proportion of negatives (all above 0.3). However the highest proportion of negatives in the others group is around 0.5, with the proportion of negatives forming a bell shape distribution but with slightly more children having a lower proportion of negative weight Z-scores.

### 7.6.2 Introduction of a threshold

In the standard 'runs' approach, if 0 would be taken as the median then anything above zero would be classed as positive and anything below as negative. In doing this we are saying that Z-scores, such as -0.01 and 0.01, are distinguished as being of different sign when in fact they are not very different. Therefore various thresholds were explored between 0 and 1 in increments of 0.2. A threshold of 0.2 is taken to mean that the conditional weight gain Z-score is:

- positive (+) if it's greater than 0.2
- negative (−) if it's less than -0.2
- the same (0) if it's between -0.2 and 0.2

In introducing a threshold we were hoping to exclude individuals that may have some negative conditional weight gain Z-scores but are growing 'normally'; that is those children that exhibit the usual weight gain fluctuations seen in infancy. The effects of introducing a threshold are as expected; as the size of the threshold increases the proportion of negatives decreases for both the cases and others.

Histograms of the distribution of the conditional weight gain Z-scores (when conditioning on first weight Z-score) as the threshold increases, for both cases and others, are displayed in figure 7.27. At around a threshold of 0.4 to 0.6, the proportion of negatives is low for the others group but still quite high for cases. Histograms of the distribution for the conditional weight gain Z-scores (when conditioning on previous weight Z-score) as the threshold increases, for both cases and others, are displayed in figure 7.28. At around the same threshold value, 0.4 to 0.6, the distribution of the conditional weight gain Z-scores of the others group is shifted towards the left while the distribution of the case group is more symmetric. Thus suggesting that a thresholds of around 0.5 and -0.5 may be appropriate for determining whether a conditional weight gain Z-score is 'positive' and 'negative', respectively.

### 7.6.3 An alternative approach for detecting growth faltering

It would be ideal to identify infants at risk of failing-to-thrive early so that remedial action can be implemented. It was with this philosophy we decided to explore the use of the criteria that the first two consecutive conditional weight gain Z-scores are 'negative' as an indicator of failure to thrive, i.e. the conditional weight gain Z-score between first and second weight, and second and third weight are both below the lower threshold. This approach may pick up infants at risk of failing-to-thrive earlier as their weight gain is slowing up rather than waiting for the infant's weight to fall below say the third centile or below the 'thrive index' cut-off in late infancy. A cautionary note at this point is that we may arrive at a higher frequency than expected of negative conditional weight gain Z-scores immediately after birth because in this time period there are often problems adjusting to feeding, resulting in a slight loss in weight or delay in weight gain. Inevitably, the large number of others and the small number of cases will lead to a large number of 'false positives' (i.e. others detected as FTT).

The use of this mechanism for identifying children at risk of failing to thrive was explored in both the case and other groups. In figure 7.29 we present bar charts of the proportion of cases and others that have the first two conditional weight gain Z-scores negative, again the threshold is varied between 0 and 1 in steps of 0.2. As the threshold is increased the percentage of cases and others satisfying the first two conditional weight gain Z-scores negative decreases. Even with a zero threshold, this criterion identifies a high proportion of cases. At the same time this criterion also identifies just under one fifth of the others group as an at risk group. In figure 7.30 we present the conditional weight gain Z-scores when conditioning on previous

weight Z-score for cases (upper panel) and others (lower panel). The scatter within these plots are coloured grey if the criterion is true and in pink if the criterion is false, using the zero threshold as an example. A black and red lowess curve are overlaid for the group of children were the criterion is true and false, respectively. The upper panel in figure 7.30 for the case infants indicates that the children that satisfy the criterion, falter earlier in infancy in comparison to the group that don't satisfy the criterion. The lower panel in figure 7.30, for the other infants, indicates that the children that satisfy the criterion experience a period of temporary growth faltering immediately after birth. These may be the infants that are experiencing problems with initial feeding, but this group of children then go on to improve.

The first three weights (hence the first two conditional weight gain Z-scores) are often in the first two months of life. In figure 7.31 we explore whether taking the first three consecutive conditional weight gain Z-scores to be negative is a viable alternative. It would appear that this is also a promising criterion, identifying a similar percentage of cases but a lower percentage of others with zero threshold. In figure 7.32 we consider taking the first four consecutive conditional weight gain Z-scores to be negative. This approach still identifies a large proportion of cases and very few others. However, this criterion could be too stringent and growth retardation may have persisted for a long time by the fifth weight measure in infancy. Finally, we considered the impact of ignoring birth weight and explored the impact of a criterion requiring that the second and third consecutive conditional weight gain Z-scores are negative, see figure 7.33. It would appear that with a zero threshold this criterion identifies 80% of case children and a similar percentage of 'others' as conditioning on first two weight Z-scores negative. Here we have presented some suggestions of how conditional weight gain Z-scores could be used to detect growth faltering. This is a potential area for future work, in collaboration with child growth experts.

### 7.6.4 Introducing a cost mechanism

So far we have not really arrived at any clear cut answers with regards to deciding an appropriate threshold, assuming for the moment that the case group were a definite diagnostic group, then a high proportion of the cases are identified as having their first two conditional weight gain Z-scores negative but at a cost that some of the others group are also identified (these may be vulnerable infants). As the threshold increases the percentage of children identified using the 'two negatives' criterion is reduced, but the same is true in the case group.

It is suspected that the distinction between the two groups, cases and others is not as clear cut in early infancy, with the case group experiencing a prolonged period of growth faltering whilst infants in the others group may experience temporary growth faltering.

> Cost benefit analysis makes it possible to determine, firstly, whether an individual intervention offers an overall net welfare gain and, secondly, how the welfare gain from the intervention compares with that from alternative interventions (Palmer et al. 1999, pp1349)

Cost-benefit analysis is an approach adopted in many disciplines. However, its use in health care has been limited because of objections to valuing health benefits in monetary terms and practical measurement difficulties (Palmer et al. 1999). Kelnar (2000) suggested that a cost-benefit approach may be the way forward in deciding when to use growth hormone therapy with children of short-stature. Here the issue is to discriminate between children with idiopathic short stature, growth hormone deficiency and other potential causes of short stature. Furthermore, the benefits of growth hormone treatment for these different groups of children is debatable.

One possible approach to determining a suitable threshold is to use a cost mechanism. Given an infants weight, there are 4 possibilities:

1. FTT detected, leading to cost of dealing with.
2. FTT not detected, resulting in no cost of dealing with but possibly big costs later in terms of treatment and ethically.
3. Non-FTT classified as non-FTT resulting in no costs.
4. Non-FTT incorrectly classified as FTT which leads to costs of dealing with and possible inconvenience for parents.

Therefore we are mainly choosing a classification rule to balance (2) and (4). If we for example take the following costs:

1. FTT and classified FTT = treatment costs = $k$
2. FTT and classified Non-FTT = future treatment and ethical costs = $\alpha k$

3. Non-FTT and classified non-FTT = zero cost

4. Non-FTT and classified as FTT = treatment and inconvenience costs $= k + hk$

This is a hypothetical set up as the true costs are not available, so for arguments sake we take $k$=1 and $h$=0.

Then if we consider those infants whose first two conditional weight gain Z-scores are 'negative', then for a particular threshold we have the following costs: E(cost)=p(FTT and classified FTT) + p(Non-FTT and classified FTT) + $\alpha$ p(FTT and not classified FTT)

The criterion that the first two conditional weight gain Z-scores are negative was used and the costs tabulated using the cost mechanism described above (the threshold varied between 0 and 1 in steps of 0.1 and $\alpha$ took on values between 1 and 10 in steps of 1), see table 7.3. The overall aim is to minimise the cost. For low values of $\alpha$ the tabulated costs decrease as the threshold increases. As soon as the value of $\alpha$ is about 5-6 then the threshold is that minimises the cost is about 0.8, as $\alpha$ increases the threshold that minimises the cost decreases to about 0.5. However when $\alpha$ reaches 9 or 10, there are several thresholds which give local minima.

Table 7.3: Contrasting others with cases: Tabulated costs for various thresholds and different values of $\alpha$

| Threshold | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.262 | 0.275 | 0.288 | 0.300 | 0.313 | 0.326 | 0.338 | 0.351 | 0.364 | 0.376 |
| 0.1 | 0.221 | 0.236 | 0.251 | 0.266 | 0.281 | 0.296 | 0.311 | 0.326 | 0.341 | 0.356 |
| 0.2 | 0.180 | 0.198 | 0.217 | 0.235 | 0.254 | 0.272 | 0.291 | 0.309 | 0.328 | 0.346 |
| 0.3 | 0.155 | 0.177 | 0.199 | 0.221 | 0.243 | 0.265 | 0.287 | 0.309 | 0.331 | 0.353 |
| 0.4 | 0.131 | 0.155 | 0.179 | 0.204 | 0.228 | 0.252 | 0.276 | 0.301 | 0.325 | 0.349 |
| 0.5 | 0.112 | 0.138 | 0.165 | 0.192 | 0.219 | 0.245 | 0.272 | 0.299 | 0.326 | 0.352 |
| 0.6 | 0.096 | 0.125 | 0.155 | 0.184 | 0.213 | 0.243 | 0.272 | 0.301 | 0.330 | 0.360 |
| 0.7 | 0.082 | 0.114 | 0.146 | 0.178 | 0.210 | 0.241 | 0.273 | 0.305 | 0.337 | 0.369 |
| 0.8 | 0.072 | 0.105 | 0.139 | 0.173 | 0.207 | 0.240 | 0.274 | 0.308 | 0.341 | 0.375 |
| 0.9 | 0.064 | 0.100 | 0.137 | 0.173 | 0.209 | 0.245 | 0.282 | 0.318 | 0.354 | 0.390 |
| 1 | 0.056 | 0.094 | 0.131 | 0.169 | 0.206 | 0.244 | 0.281 | 0.319 | 0.356 | 0.394 |

## 7.7 Graphical display of expected pattern of weight gain

The graphical display of conditional weight gain Z-scores is an area where there may be scope for further research. In this chapter we have adopted the approach of plotting the conditional weight gain Z-score at the time of the current weight measure because it is at this time point the clinician would make an evaluation of the child's weight gain. We have also assumed that the conditional weight gain Z-score was zero at the first measurement occasion. It could be argued that the conditional weight gain Z-score is a measure over the time period between the two weight measures, so it may be appropriate to plot the conditional weight gain Z-score at the average time (in a similar fashion to velocity charts) or use the approach adopted by Cole (1995) discussed below.

The conditional $100\alpha$ centile is given by (Cole 1998a):

$$Z_{t2|t1,100\alpha} = rZ_{t1} + z_\alpha\sqrt{1-r^2} \tag{7.15}$$

where $r$ is the correlation between weight Z-scores at $t1$ and $t2$, and $z_\alpha$ is the normal equivalent deviate.

Cole (1995) presented the median pattern of weight gain derived from the Cambridge correlation matrix (excluding birth weight) for 223 infants, see table 3.1 in chapter 3. Cole (1995) considered a set of hypothetical infants; these had weight Z-scores from -4 to 4 in steps of two-thirds at 4 weeks (i.e. at 4 weeks Z-score $= (\pm 4, \pm 3\frac{1}{3}, \pm 2\frac{2}{3}, \pm 2, \pm 1\frac{1}{3}, \pm\frac{2}{3}, 0)$). The predicted Z-score at 8 weeks was then obtained by multiplying the Z-score at 4 weeks by the correlation between weight Z-scores at 4 and 8 weeks. The correlations on the leading off-diagonal of the Cambridge correlation matrix were used in a similar fashion to obtain the median pattern of weight gain for these hypothetical infants up to the age of 2 years. This gave a Z-score chart of the expected pattern of weight gain over a 4 week interval. Cole (1995) overlaid isoponds on these weight gain Z-score charts; with these contours corresponding to constant weights in kilograms. Thus enabling a child's actual weight to be plotted directly on the weight gain chart. Cole (1995) also overlaid the median pattern of weight gain over the UK 1990 weight reference.

Cole (1998a) presented the conditional 5th and 95th weight gain centiles using the model for correlation derived from the Cambridge correlations including birth weight. The thrive lines in the 3-in-1 chart (Cole 1997b) correspond to the conditional 5th weight gain centiles. Cole (1998a) chose the 5th and 95th centiles as it was

only feasible to present these two extremes without the chart becoming cluttered. In a similar fashion to the median pattern of weight gain, these conditional centiles were then overlaid on the original weight growth chart.

We now produce similar charts, but starting with Z-scores between -4 and 4 in steps of thirds at birth. The correlation at 4 week intervals is obtained by using equation (7.2). The expected pattern for weight gain Z-scores, using the one variable Argyle model, is represented by the black lines on the left plot of figure 7.34. The conditional fifth centile of weight gain is also obtained for these same starting Z-score values at birth and these can be found on the right of figure 7.34. The red lines within both plots in figure 7.34 represent the expected pattern of weight gain and the conditional fifth centile of weight gain generated by using the model proposed by Cole (1998a). The Argyle model generates expected weight gain Z-scores that regress towards the mean sooner than those generated by the model proposed by Cole (1998a) model. This is probably due to the weaker correlations observed between the routine Newcastle weight Z-scores. The conditional fifth centile using the Argyle model drops away from the starting Z-score value earlier than the model proposed by Cole (1998a). However for birth weight Z-scores that are -2 or less the conditional fifth centiles are similar for both models. Thus suggesting that children with initially high weight Z-scores would have to experience a more rapid deceleration in weight gain to fall below the conditional fifth centile generated by the Argyle model than with the model proposed by Cole (1998a).

Cole (1995) presented ID 2149's conditional weight gain Z-scores as an example of chart usage and variability of conditional weight gain Z-scores. These are presented here using the one variable Argyle model to obtain the conditional weight gain Z-scores. Cole (1995) presented conditioning on previous weight Z-score and all possible conditional weight gain Z-scores. Cole (1995) chose to present the conditional weight gain Z-scores as dotted lines from age of first weight to age of second weight and these were overlaid on the expected pattern of weight gain chart. In figure 7.35 we use the same approach as Cole (1995), but the expected pattern of weight gain is represented by grey lines and two charts are produced to avoid information overload. The chart on the left of figure 7.35 presents conditional weight gains Z-scores when conditioning on previous weight Z-score. The chart on the right of figure 7.35 presents conditional weight gains Z-scores when conditioning on birth weight Z-score. Similar charts are also produced for a case child with ID 3662 for comparison purposes; see figure 7.36.

As discussed in chapter 3, velocity charts are restricted to height measurements made one year apart. Similarly, the charts produced in figure 7.34 have limited

use because they are only suitable for conditional weight gain Z-scores over a four week interval. Furthermore, Cole (1995) acknowledges that the chart for expected pattern of weight gain fails to identifying growth that departs from the median. An approach favoured here is the use of the conditional weight gain Z-score, because this index can be applied over any time interval and in theory should be distributed with zero mean and variance one. The only caution is that this conditional weight gain Z-score is likely to be more variable over shorter time intervals.

## 7.8 Discussion and Conclusions

At the start of this chapter we developed the one variable version of the Argyle model; this achieved the desirable property of a correlation of one when $t1 = t2$. Furthermore, the one variable Argyle model developed has an interesting and useful result; it has a Markov property. The properties of conditional weight gain Z-scores generated using the Argyle model given by equation (7.2) were considered. It was only possible to consider 385 groups of conditional weight gain Z-scores; these were the same infants that derived the model for correlation. Therefore we were unable to assess conditional weight gain Z-scores of individuals that have out of sample correlations because the sample sizes of these groups would be too small ($n < 50$). In considering the properties of these 385 sets of conditional weight gain Z-scores, grouped by $t1$ and $t2$, we found that the conditional weight gain Z-scores generated using equation (7.2) tended to have a mean that was greater than zero. The magnitude of the standard deviation and mean was greater for weight measurements made close in time. Thus suggesting that we should be cautious about using cut-off points with conditional weight gain Z-scores.

Cole (1995) verified his model for correlation on all data from Cambridge infant study and a sub-sample of the Newcastle data considered here. Cole (1995) grouped each set of conditional weight gains according to mean and gap ages, because these were the two variables within his model for correlation. These groupings corresponded to the mean and gap ages' generated by the Cambridge correlation matrix. In the set of conditional gains from Newcastle; Cole (1995) discarded any groups with fewer than 50 individuals. Cole (1995) found that the Newcastle data generated more extreme gains than the Cambridge data. Cole (1995) advocated using his model for time periods greater than 4 weeks, but my research suggests that there is still the odd extreme group of conditional weight gain Z-scores at age intervals greater than 4 weeks.

An additional issue arises when we need to calculate a conditional weight gain Z-score by extrapolating outside the age range that the Newcastle correlations were modelled. This applies to infants that have late weights at an age greater than the grouped age of 88 weeks, and an early weight at a grouped age greater than 52 weeks. In total there are 3111 (3.1%) conditional weight gain Z-score pairs that would involve extrapolation. This is only a small percentage, but an alternative would be to use the model proposed by Cole (1998a) because this models the correlation between weight Z-scores up to the age of 2 years. In using the model proposed by Cole (1998a) we would be extrapolating for 1052 (1.1%) conditional weight gain Z-score pairs.

Exploring trends in sequences of conditional weight gain Z-scores verifies that case infants experience a different pattern of weight gain in contrast with the rest of the Newcastle infancy data. The trend is for case infants to drop to conditional weight gain Z-scores around -2 mid-infancy with a slowing up in weight faltering towards the end of infancy. It appears that the majority of case infants experience decelerating weight gain from birth.

We considered several possible criteria for detecting growth faltering in early infancy. However, consultation with an expert on child growth is required to draw any firm conclusions. The use of the criterion that the first two (or three) consecutive conditional weight gain Z-scores are 'negative' to detect growth faltering seemed to be sensible. The 'two negatives' criterion selected a large percentage of the children identified as failing to thrive using Dr C.M. Wright's 'thrive index' approach. However there were still a small proportion of children identified from the others group. However preliminary analysis suggested that this group of children exhibited a period of temporary growth faltering immediately after birth. Furthermore, it is not unusual for infants growth to slow in the early weeks of life due to adjusting to feeding (Tanner 1989). A further consideration is that the baseline Z-score, see Chapter 4, is an average of Z-scores for weights at birth and grouped ages of 1 and 2 months. Thus a child whose first two conditional gain Z-scores are 'negative' may have a much lower baseline Z-score and hence have to drop even further before being detected as a case.

A similar approach has been used in the past for raw length increments by Healy et al. (1988), as discussed in chapter 3. Healy et al. (1988) suggested using a 25th percentile 'warning' limit with their increment chart for length, i.e. if the infants length increment was below the 25th percentile on two consecutive occasions then action was taken.

Quantifying the measuring error of weights in infancy could be of potential value in the use and interpretation of conditional weight gain Z-scores. However quantifying the measurement error is not a straightforward task, because very little has been done on the measurement errors of weight in infancy, Alsop-Shields and Alexander (1997) being the exception[1]. When we explored patterns in conditional weight gain Z-scores we introduced a threshold to allow for normal variation in the growth process along with a cost mechanism, this threshold could be based on measuring error. However, if we could quantify the measurement error in infancy in general, then on a Z-score scale we would arrive at a threshold that varies with age. It would also be useful to determine an optimal measurement interval for contrasting two weights using the conditional weight gain Z-scores. It is possible that measurement error could in some way be taken into account when a conditional weight gain Z-score is interpreted. However, even if there is way of quantifying the measurement error, there are still issues of biological variation and the possible haphazard nature of the growth process.

To conclude, we have explored a variety of approaches to using conditional weight gain Z-scores. It is felt that there is some value in using more than the conventional one conditional weight gain Z-score to assess a child's weight gain. For example, if we consider conditioning on the previous weight Z-score, then the use of criteria such as the first two or three conditional weight gain Z-scores 'negative' for detecting growth faltering in early infancy would be a valuable asset for monitoring a child's growth. Use of lowess to assess the trend in sequences of conditional weight gain Z-scores indicate that case infants do exhibit a markedly different growth pattern in infancy when contrasted with the remaining infants in the Newcastle cohort. It appears that case infants weight gain decelerates rapidly from birth with a slowing down of the deceleration in growth towards the end of infancy. An interesting, and important, result is that the use of the one variable Argyle model to model the correlation between weight Z-scores in infancy is that conditioning on the previous weight Z-score is the same as conditioning on all previous weight Z-scores. Thus to assess recent weight gain we only need to consider the previous weight Z-score. This result could have implications for other areas of research that could use the one variable Argyle model to approximate some process.

[1]The measurement error in a research study setting is addressed briefly in the measurement technique section of the paper on the UK 1990 reference, "the teams were trained and tested for within and between observer error which were $\leqslant$0.4cm for stature and $\leqslant$0.05Kg for weight in all cases" (Freeman et al. 1995, pp19)

Figure 7.1: Birth to 2 years: Contour plot of correlation between weight Z-scores generated from the Argyle model with two coefficients and constant ($c = 2$ weeks added to age). The line in dashes represents $t1 = t2$.

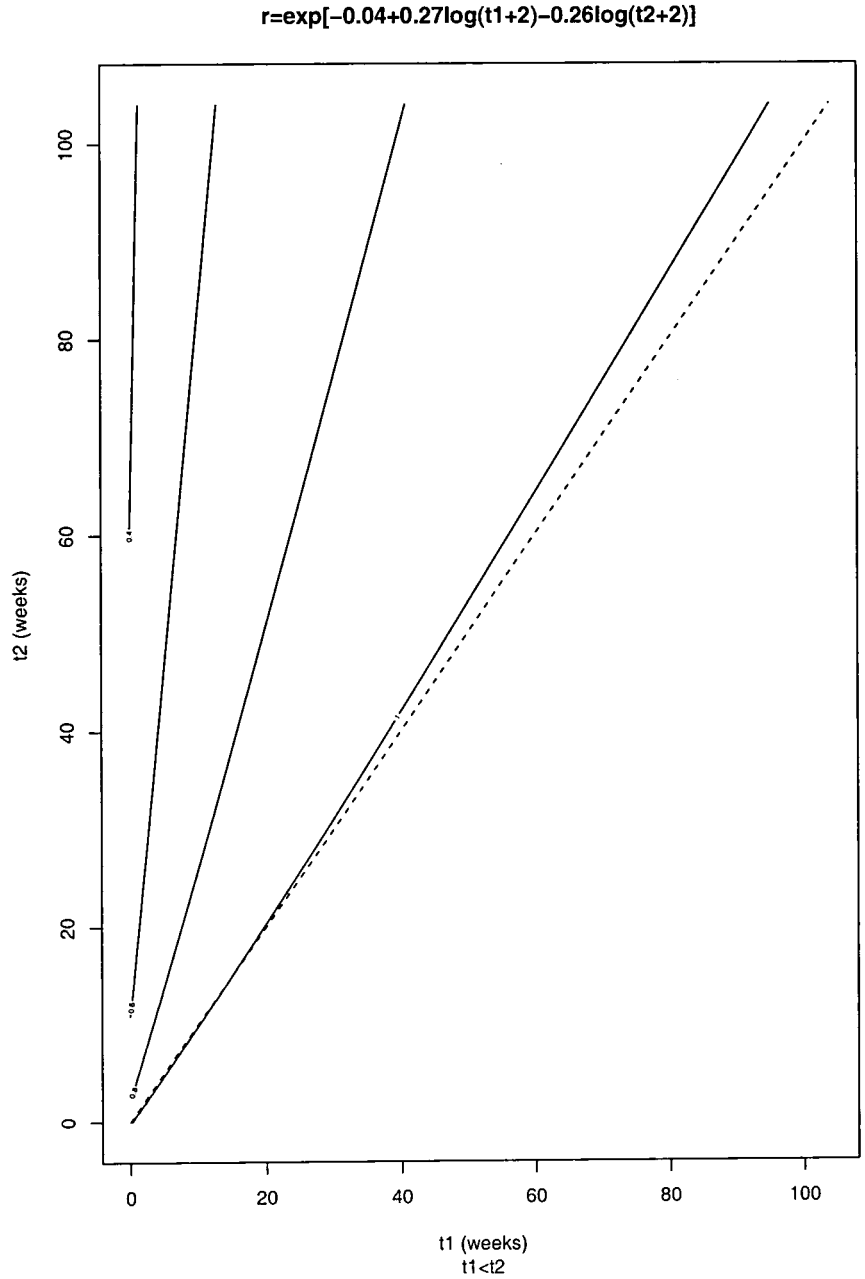


Figure 7.2: **Age to nearest fortnight:** Scatterplots of standardised residuals versus fitted values. Left Argyle model with two coefficients and constant ($c = 1.6$) Right Argyle model with one coefficient and constant ($c = 1.6$)

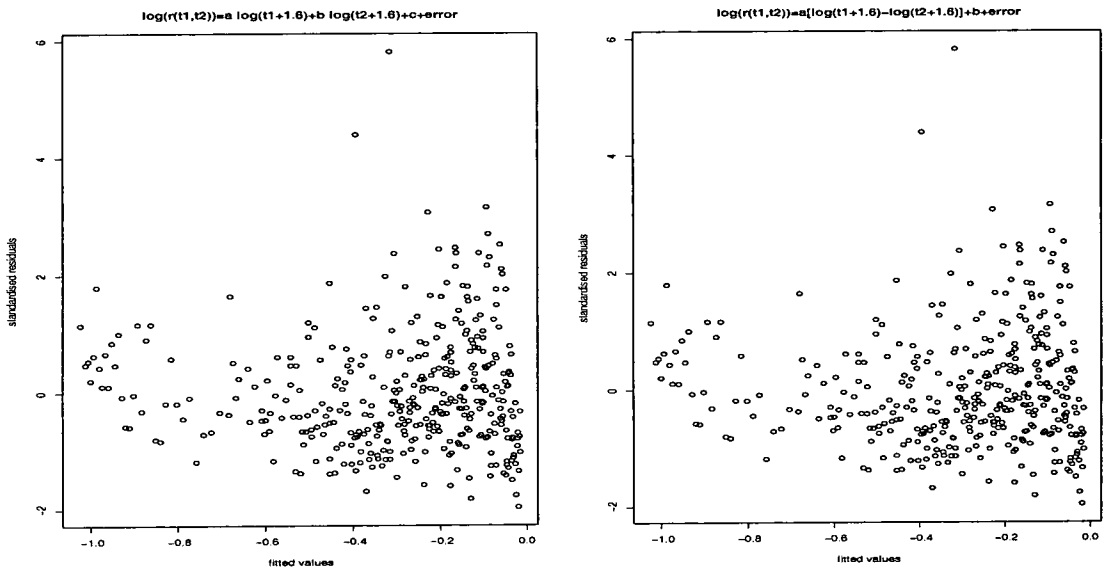



Figure 7.3: **Age to nearest fortnight**: Plot of Cook's distance versus index for Argyle model with one coefficient and constant ($c = 1.6$)

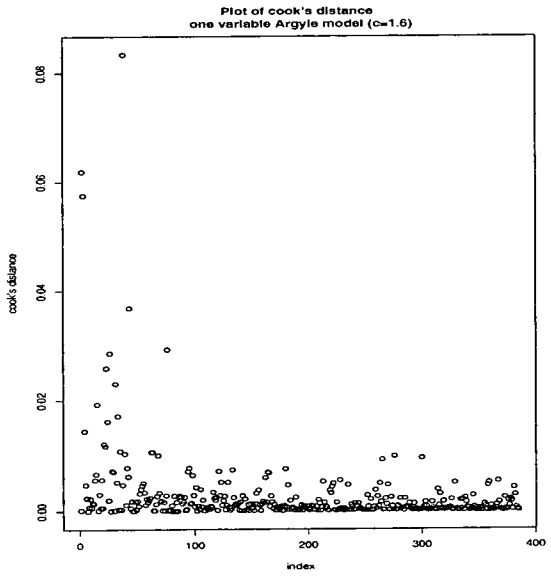


Figure 7.4: **Age to nearest fortnight:** Plot of standardised residuals versus fitted values for Argyle model with one coefficient ($c = 1.1$ weeks added to age) and with/without intercept term

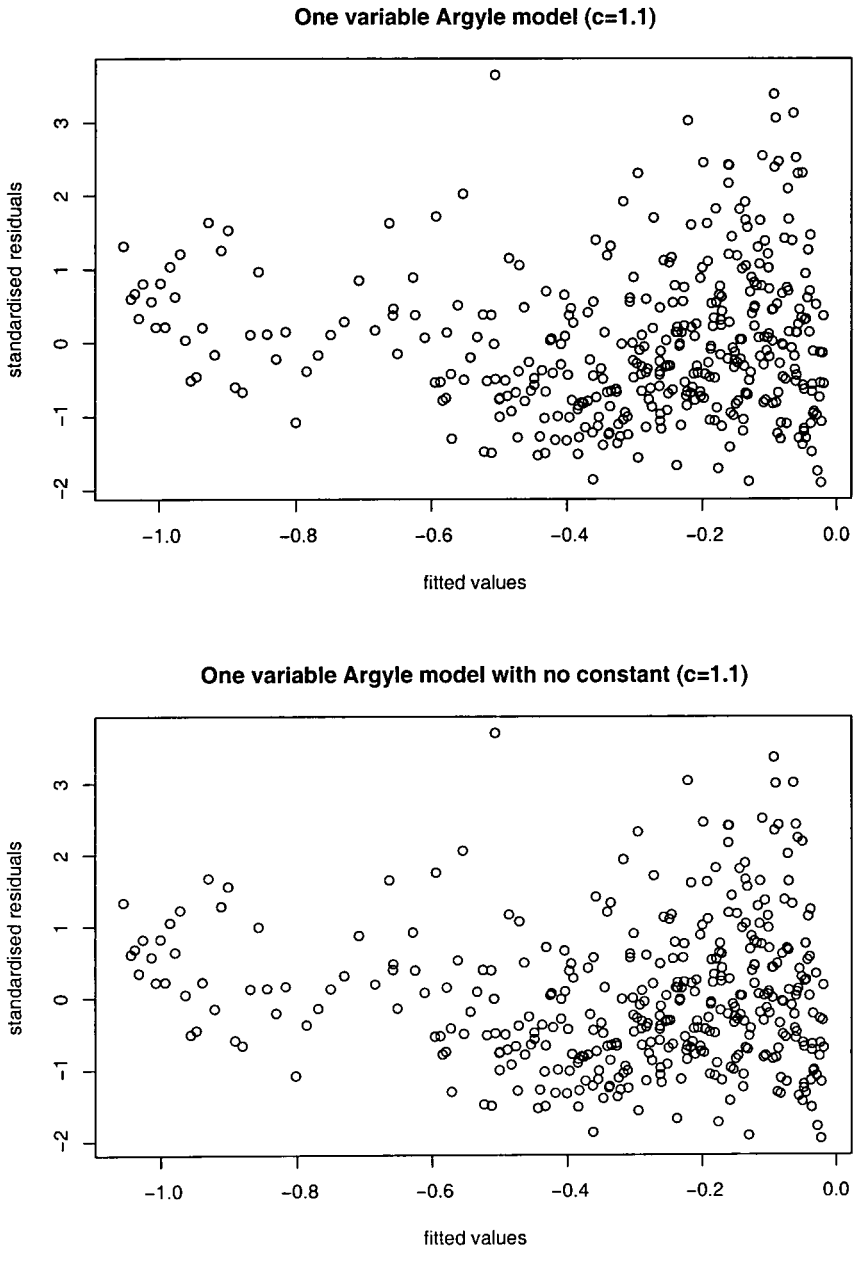


Figure 7.5: Birth to 2 years: Contour plot of correlation between weight Z-scores generated from the Argyle model with one coefficient and no constant ($c = 1.1$ weeks added to age)

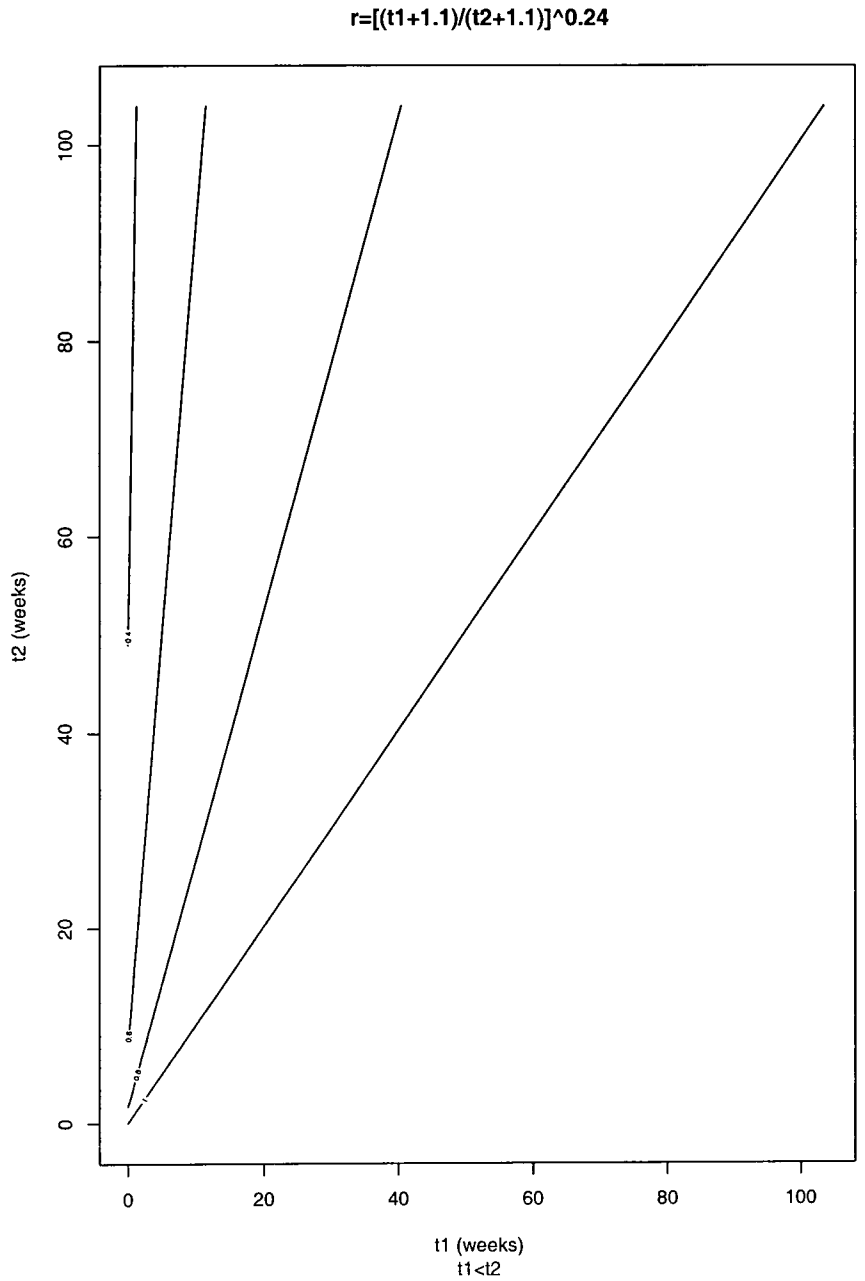


Figure 7.6: Plot of Newcastle fortnightly correlations with fitted curve (Argyle model, $c = 1.1$), confidence interval and prediction interval

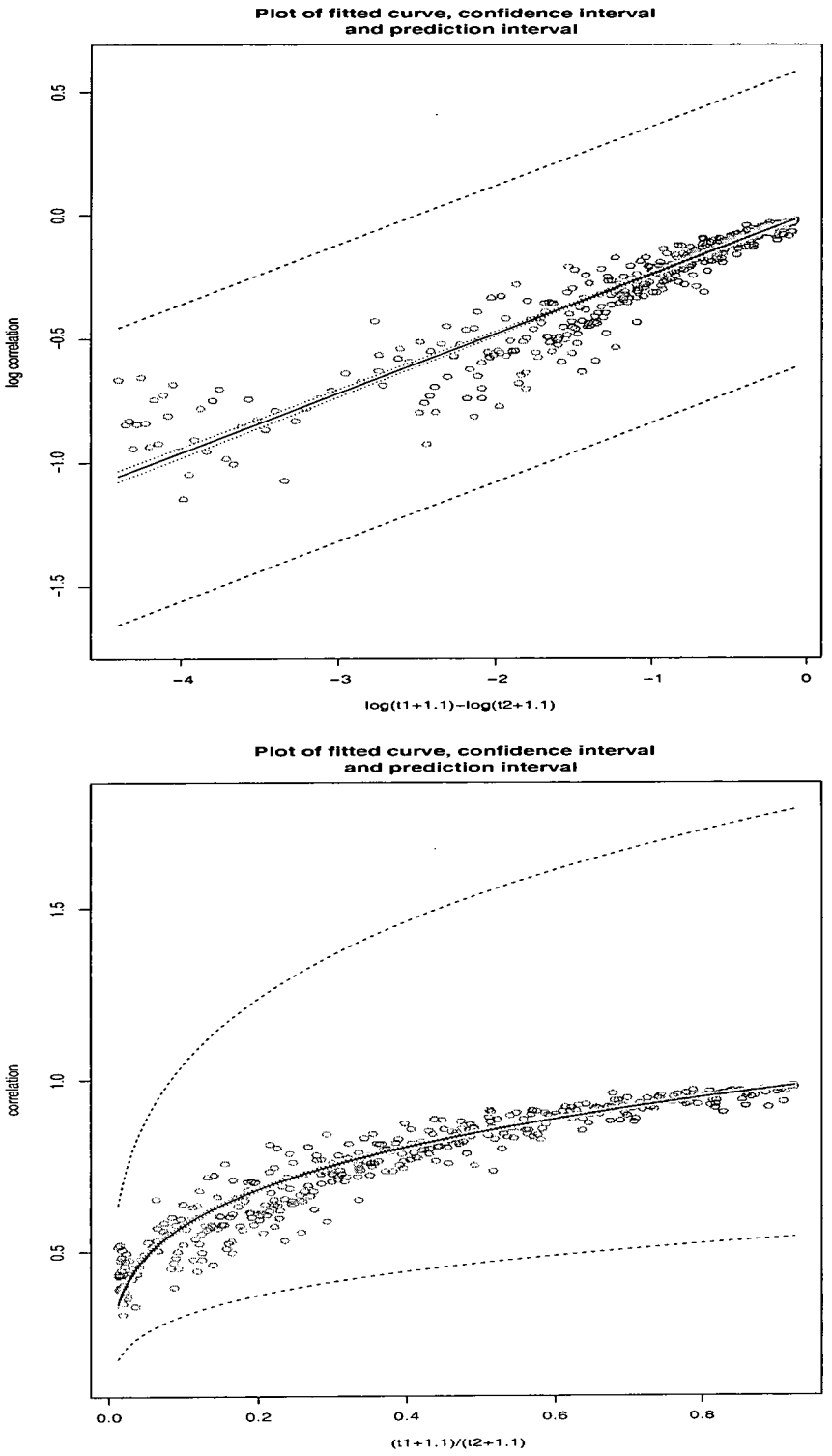


Figure 7.7: Plots of means of conditional weight gain Z-scores versus average age (on left) and time elapsed (on right). Age is in weeks

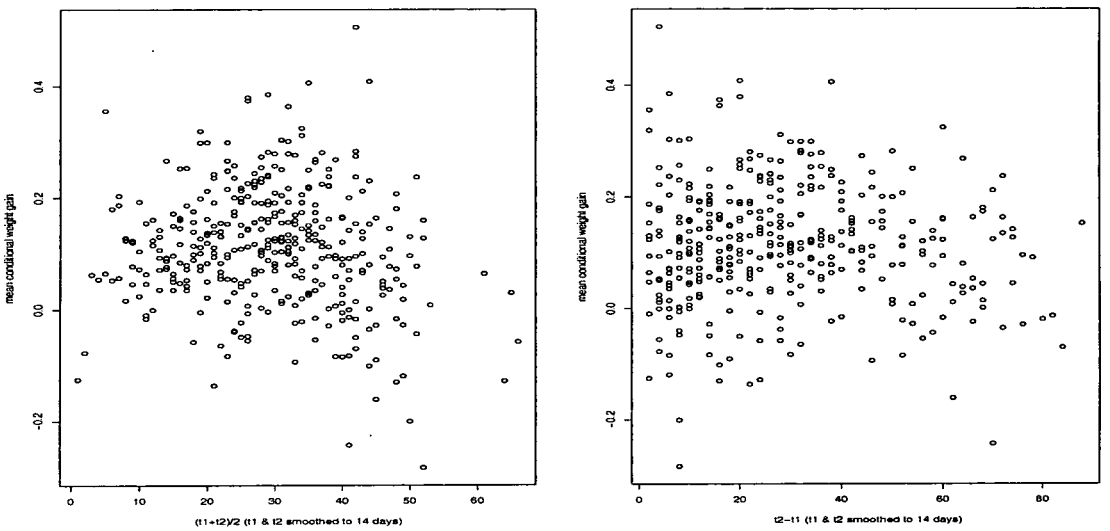



Figure 7.8: Plots of standard deviations of conditional weight gain Z-scores versus average age (on left) and time elapsed (on right). Age is in weeks

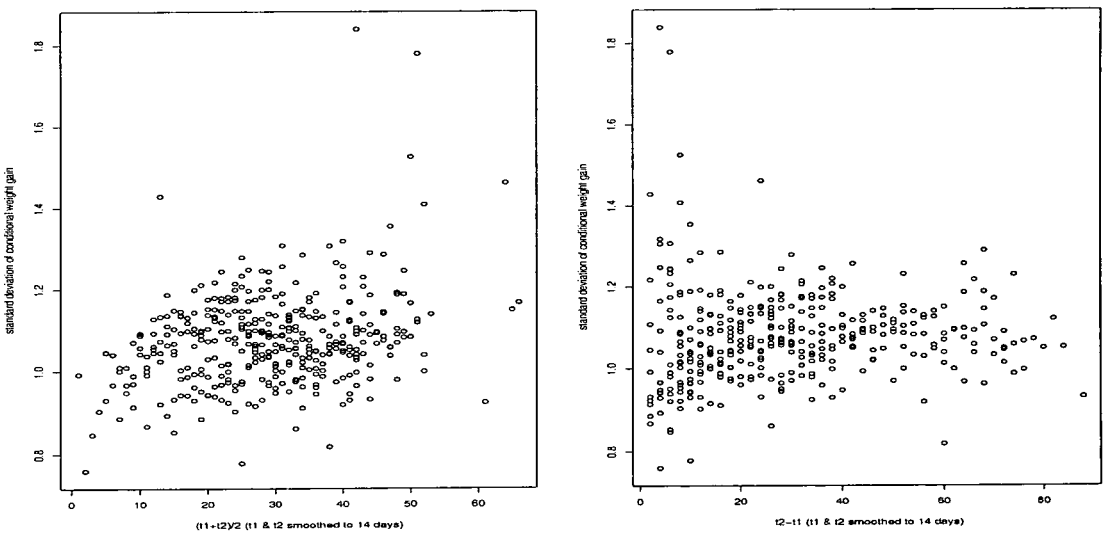


Figure 7.9: Plots of minimum values of conditional weight gain Z-scores versus average age (on left) and time elapsed (on right). Age is in weeks

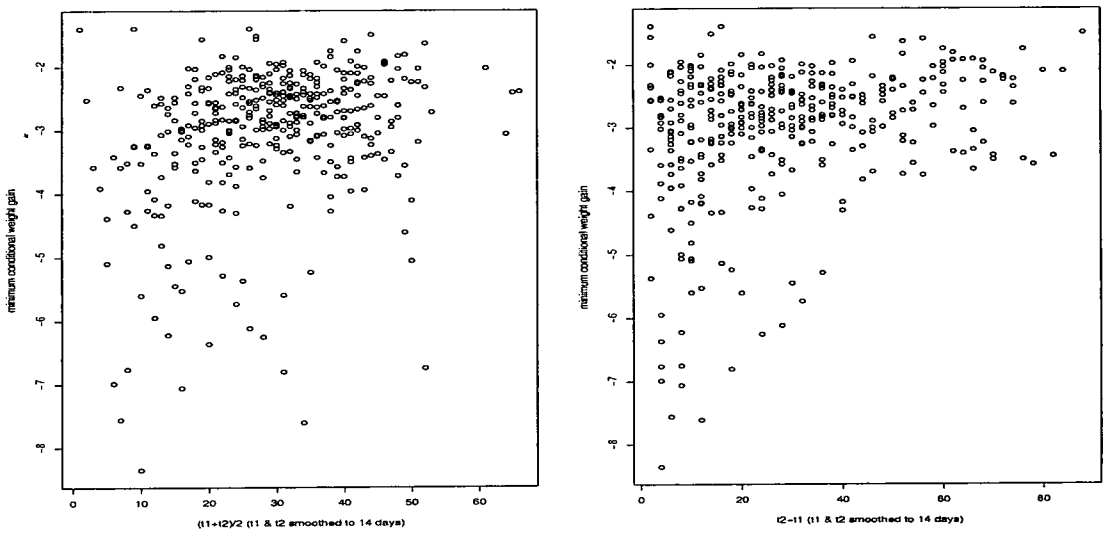



Figure 7.10: Plots of maximum values of conditional weight gain Z-scores versus average age and time elapsed. Age is in weeks

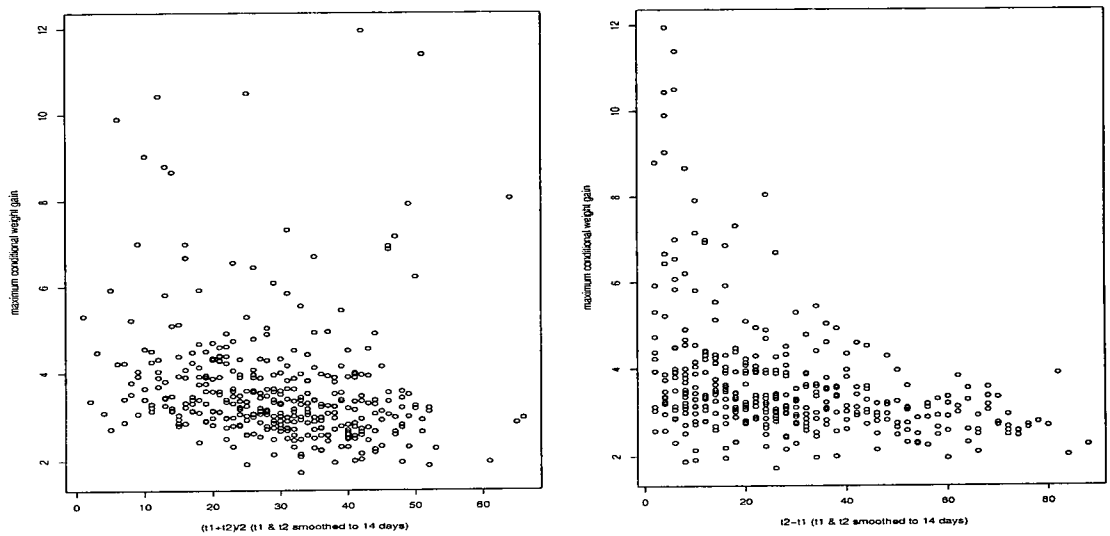


Figure 7.11: Graphical display of results from testing that the mean of the conditional weight gain Z-scores is zero: Plot of t-values versus time elapsed (on left) or average age (on right). Age is in weeks

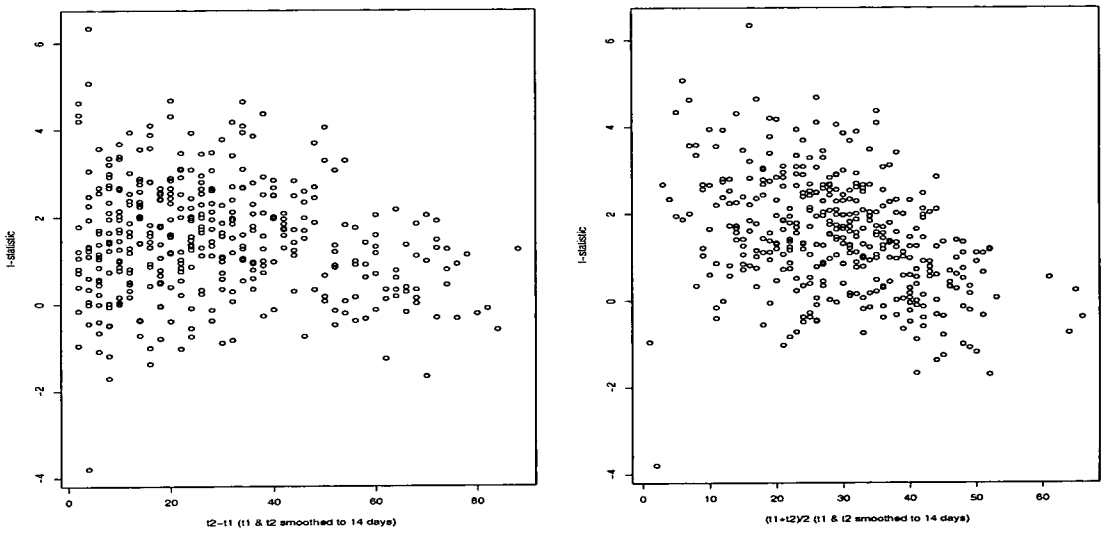



Figure 7.12: Graphical display of results from testing that the variance of the conditional weight gain Z-scores is one: Plot of approximate p-values versus time elapsed (on left) or average age (on right). Age is in weeks

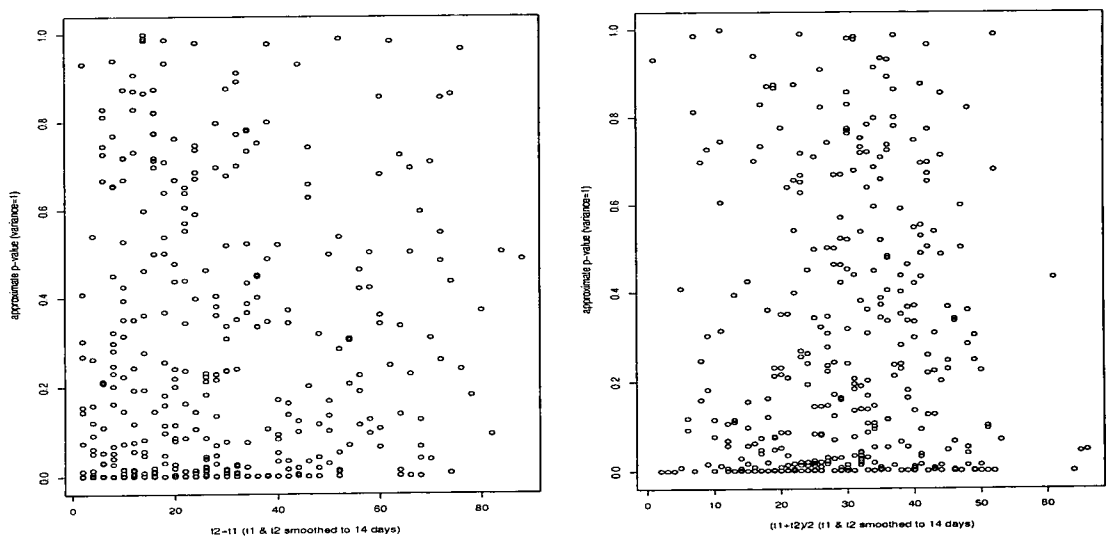


Figure 7.13: Graphical display of results from tests that variance of the conditional weight gain Z-score is one: Plots of approximate p-values from testing that variance is one versus sample size (on left) and t-value (on right). Age is in weeks

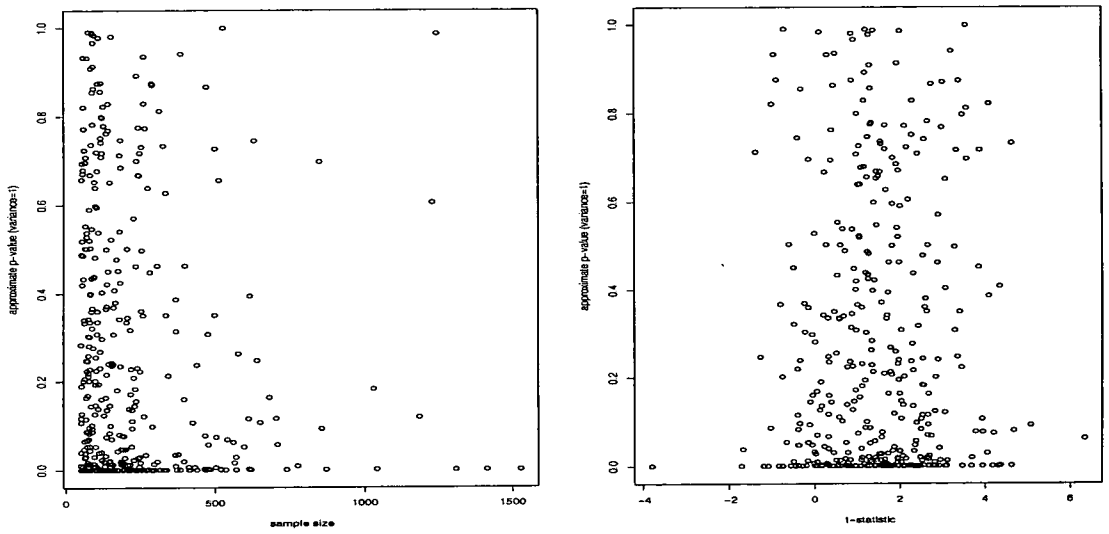



Figure 7.14: Graphical display of correlation between initial weight Z-score and conditional weight gain Z-score (age in weeks)

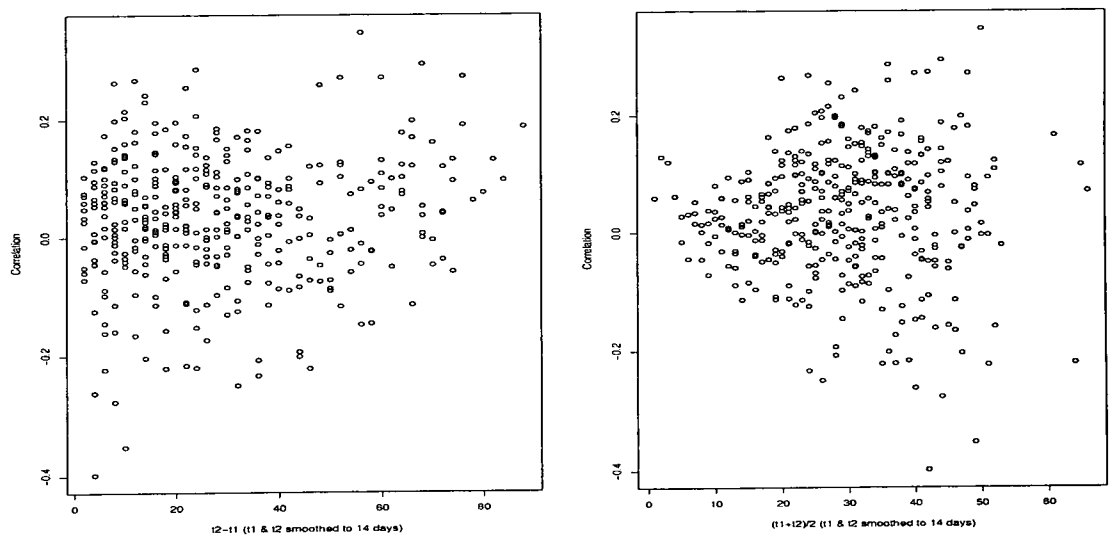


Figure 7.15: Graphical display of results from testing Normality of conditional weight gain Z-scores: Plot of p-values (Shapiro-Wilk Normality test) versus time elapsed (on left) or average age (on right). Age is in weeks

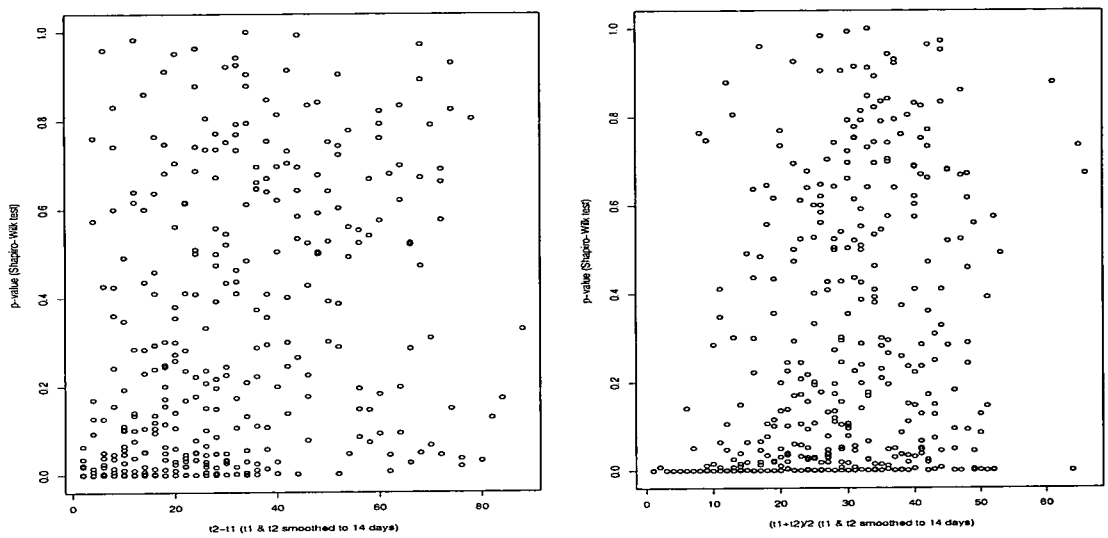


Figure 7.16: Conditioning on first weight Z-score: Plots of conditional weight gain Z-score versus age for full sample, cases, controls and others

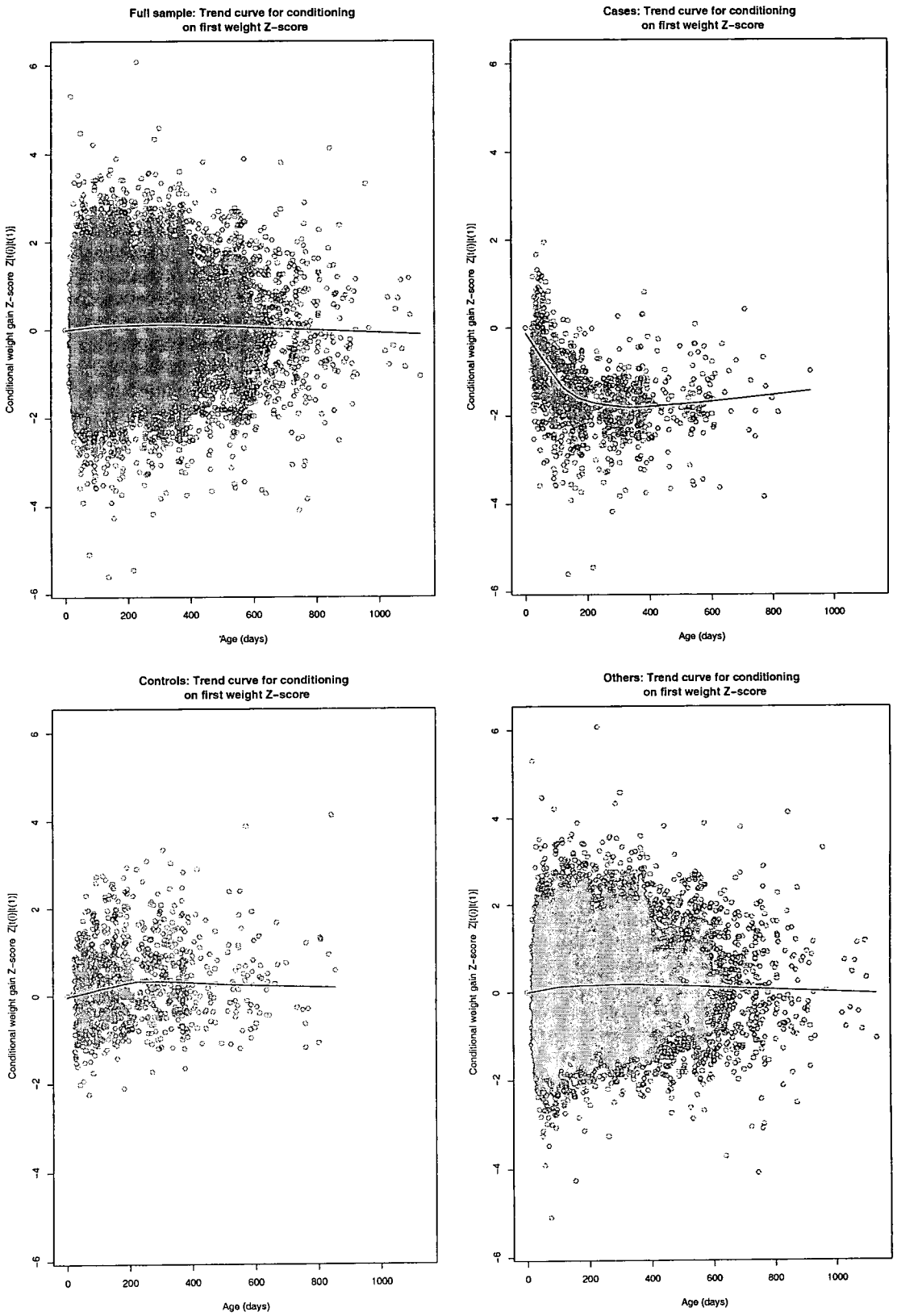


Figure 7.17: Conditioning on first weight Z-score after birth: Plots of conditional weight gain Z-score versus age for full sample, cases, controls and others

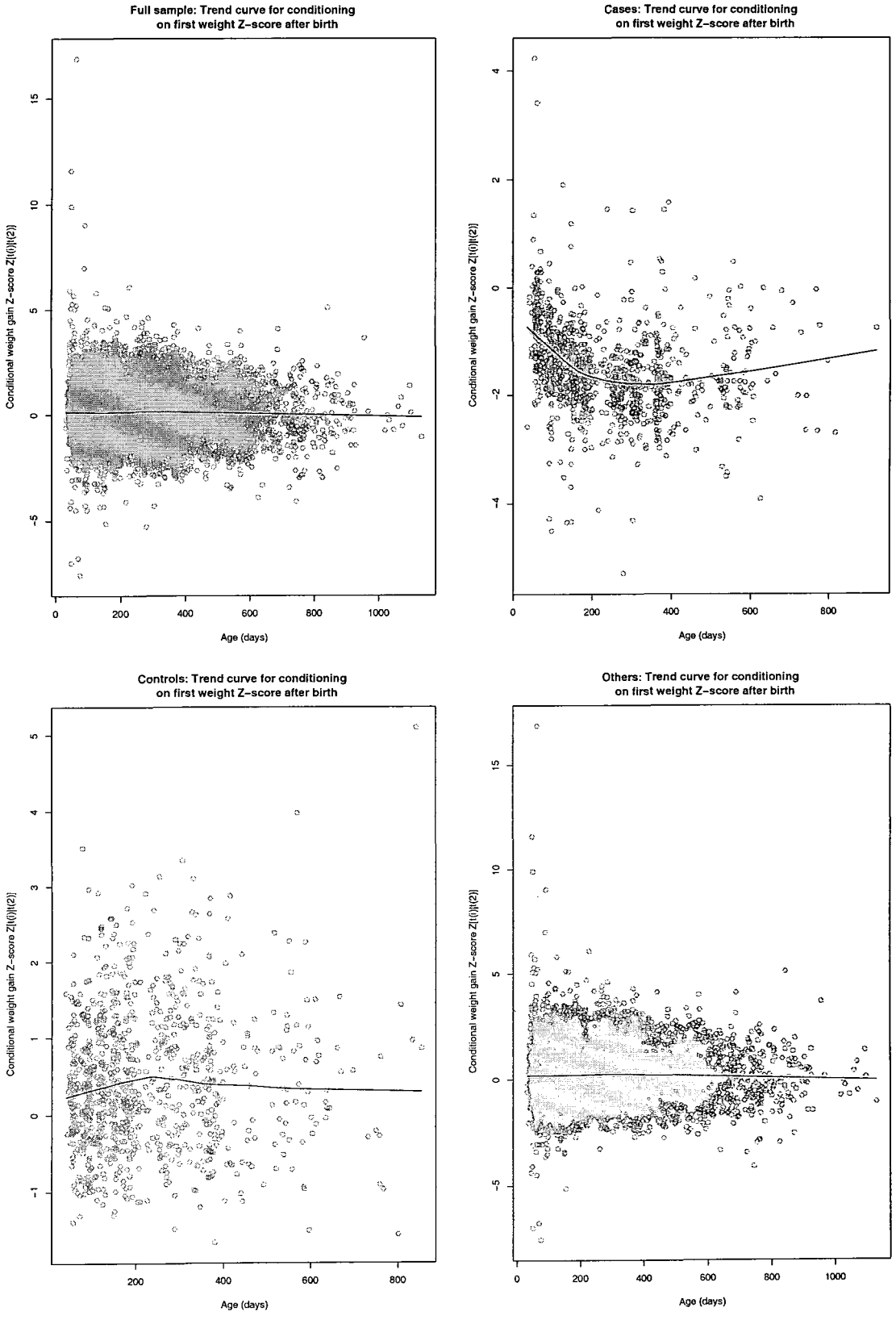


Figure 7.18: Conditioning on previous weight Z-score: Plots of conditional weight gain Z-score versus age for full sample, cases, controls and others

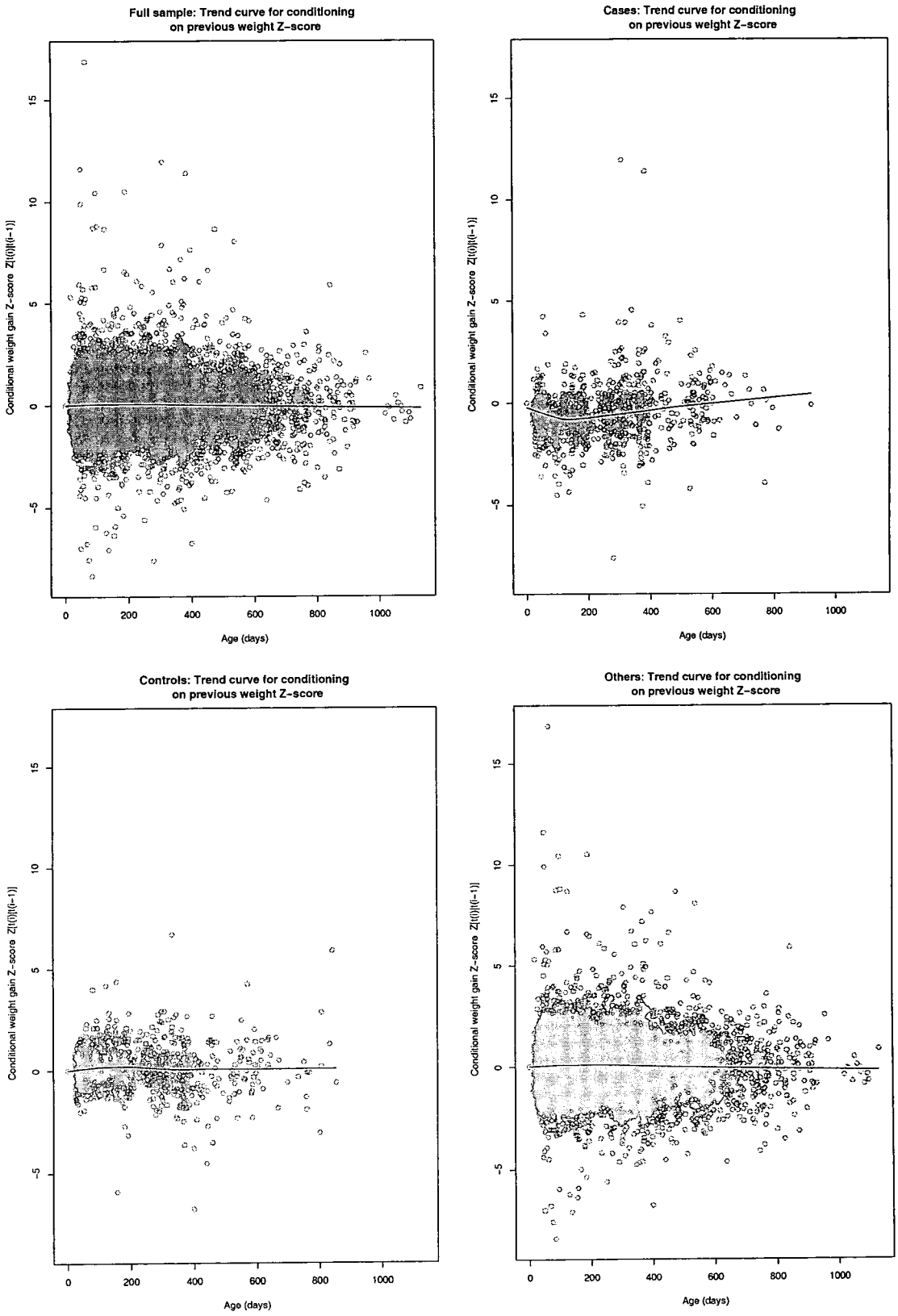


Figure 7.19: Connected plots of weight Z-score versus age for a random sample of 20 infants from the case and other groups

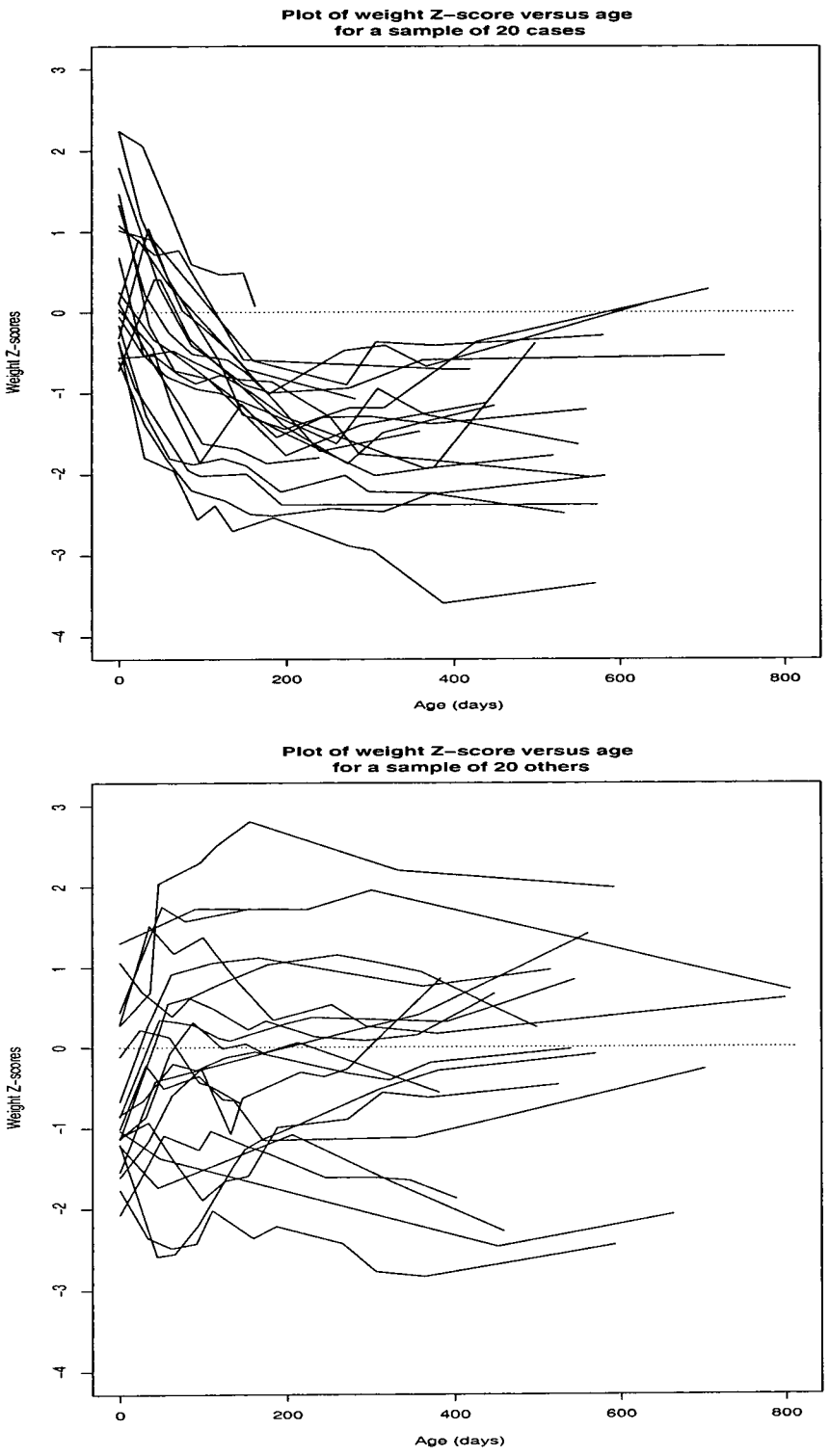


Figure 7.20: Connected plots of conditional weight gain Z-score (when conditioning on first weight Z-score) versus age for a random sample of 20 infants from the case and other groups

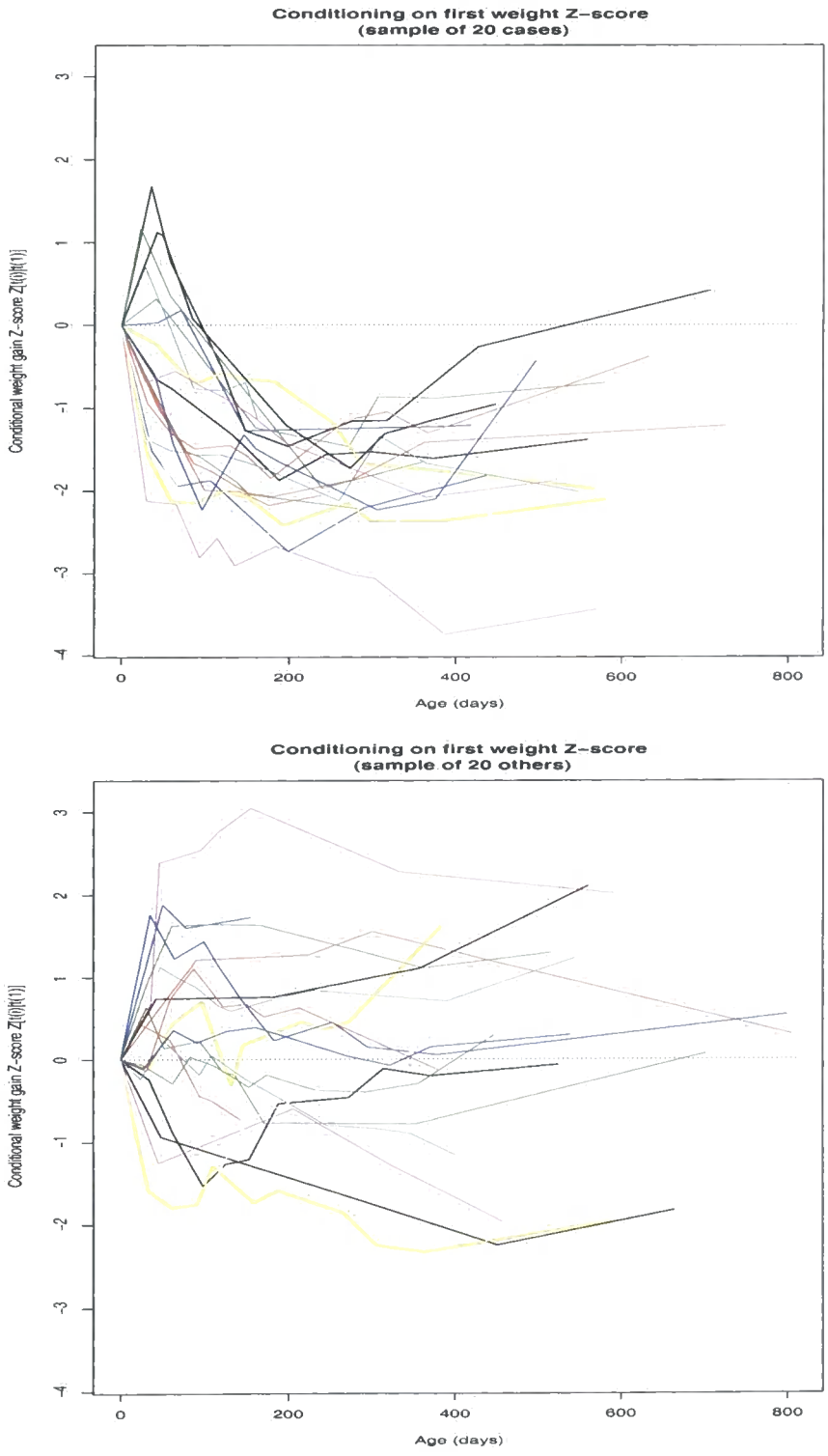


Figure 7.21: Connected plots of conditional weight gain Z-score (when conditioning on first weight Z-score after birth) versus age for a random sample of 20 infants from the case and other groups

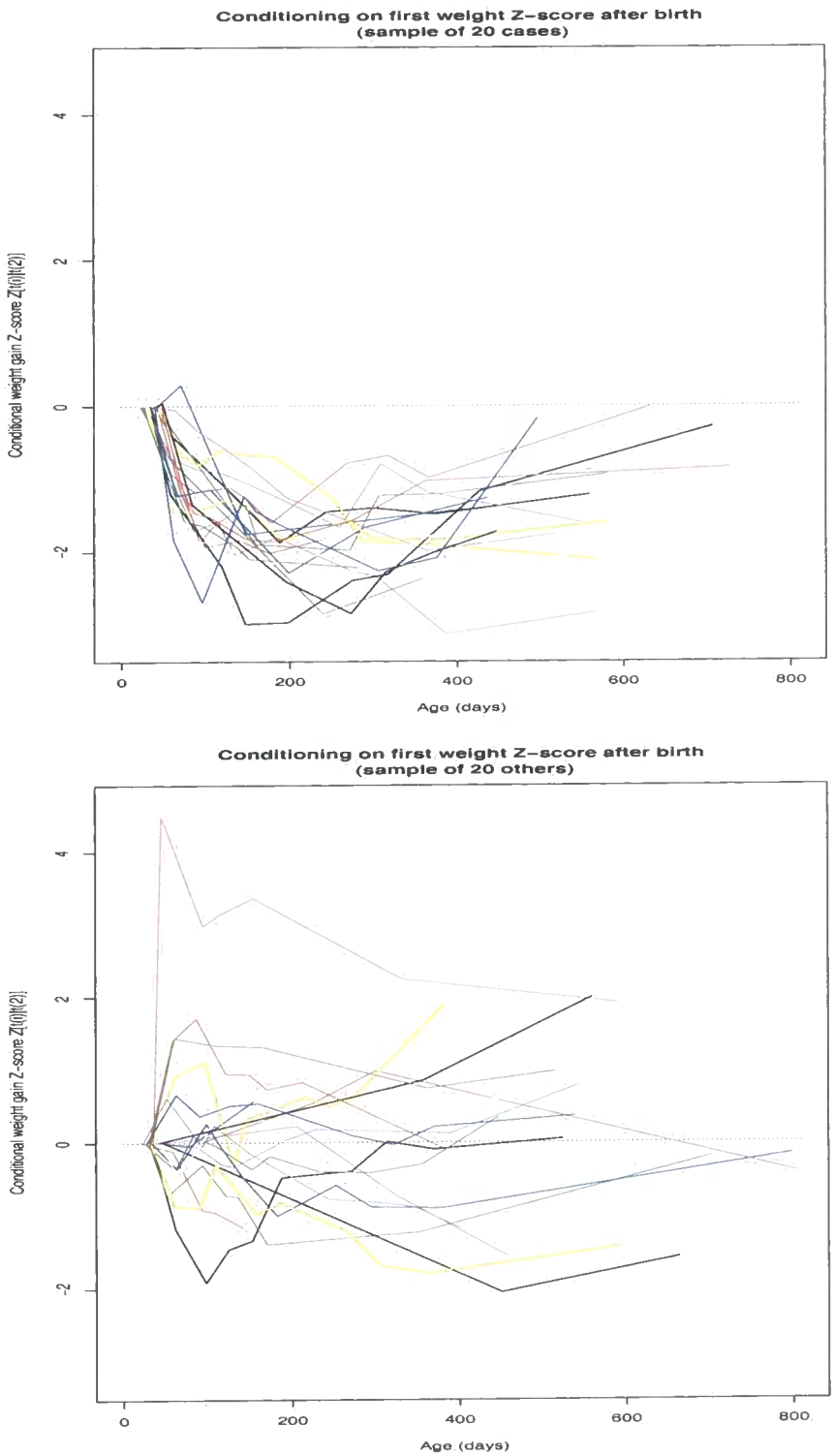


Figure 7.22: Connected plots of conditional weight gain Z-score (when conditioning on previous weight Z-score) versus age for a random sample of 20 infants from the case and other groups

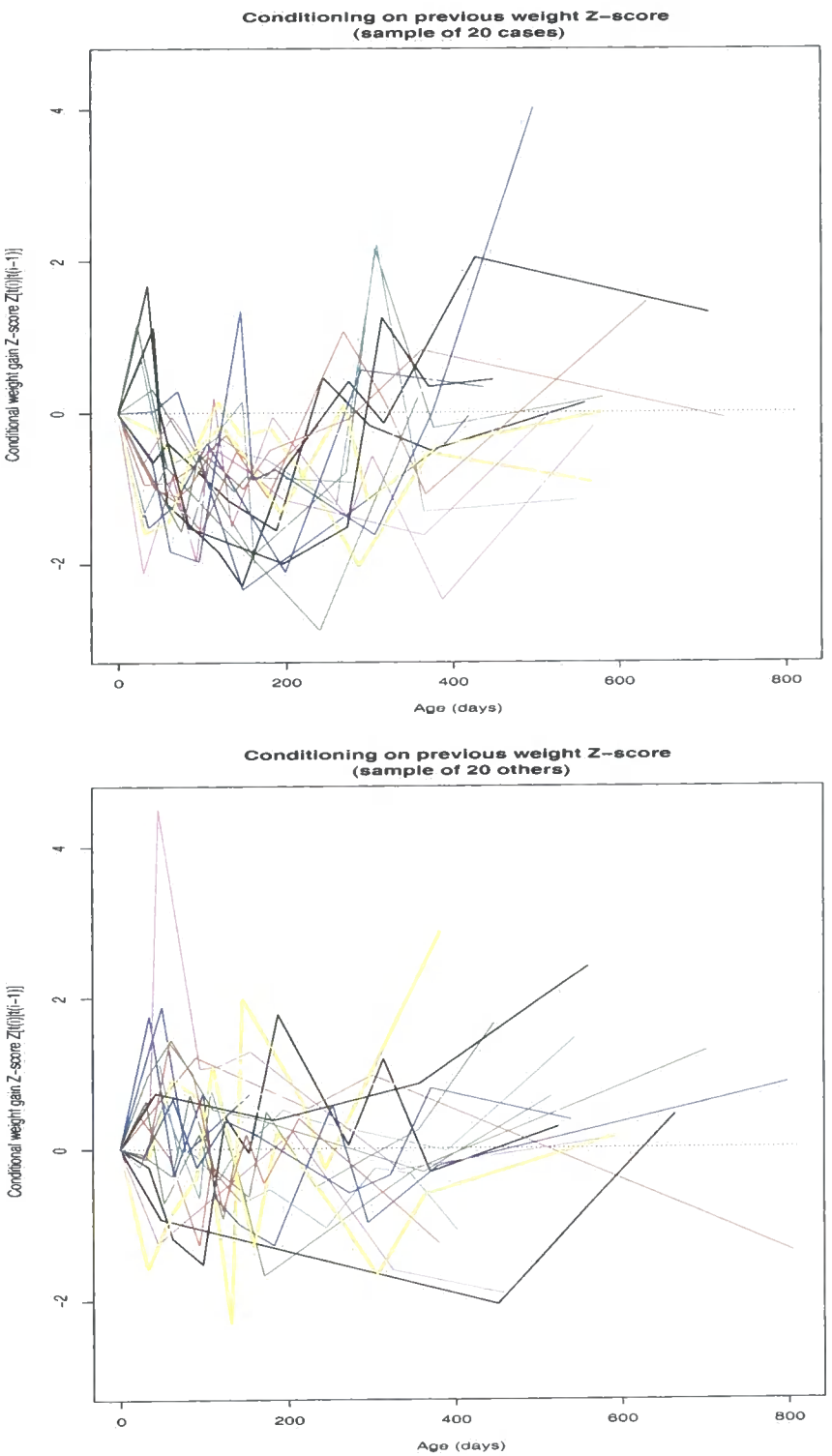


Figure 7.23: Four individuals with extreme conditional weight gain Z-scores

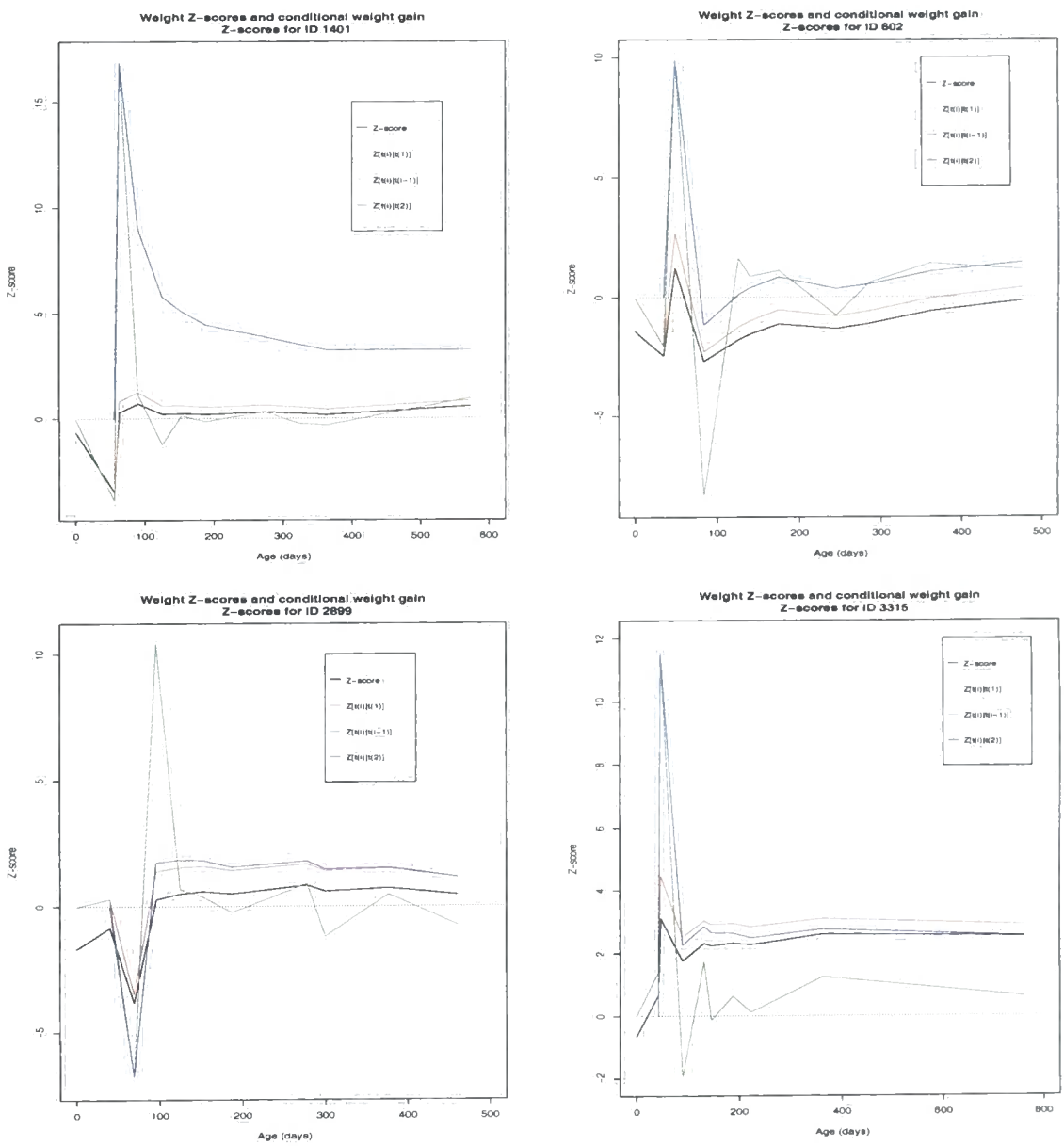


Figure 7.24: Conditioning on first weight Z-score: Lowess trend curves for conditional weight gain Z-score versus age by sex and level of deprivation

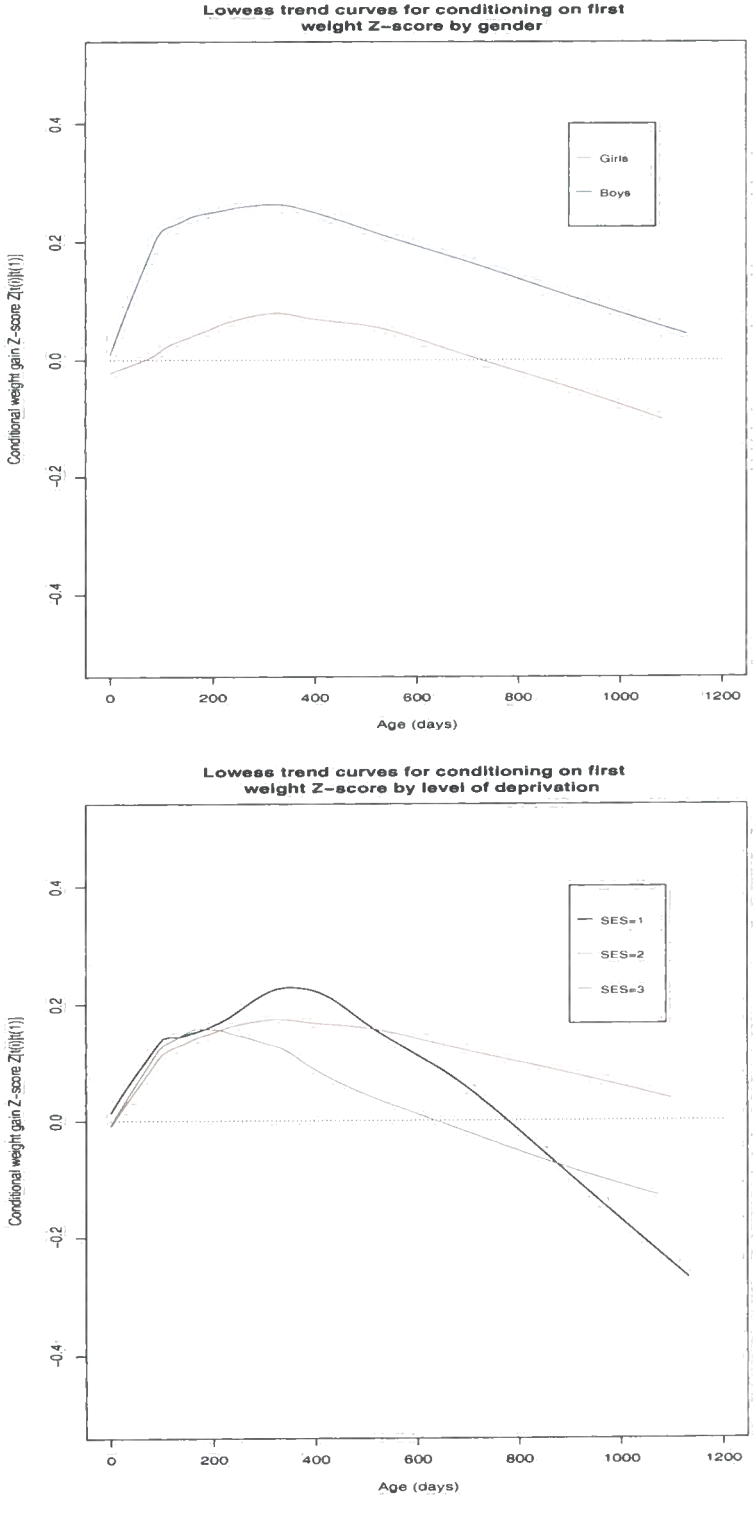


Figure 7.25: Conditioning on first weight Z-scores after birth: Lowess trend curves for conditional weight gain Z-score versus age by sex and level of deprivation

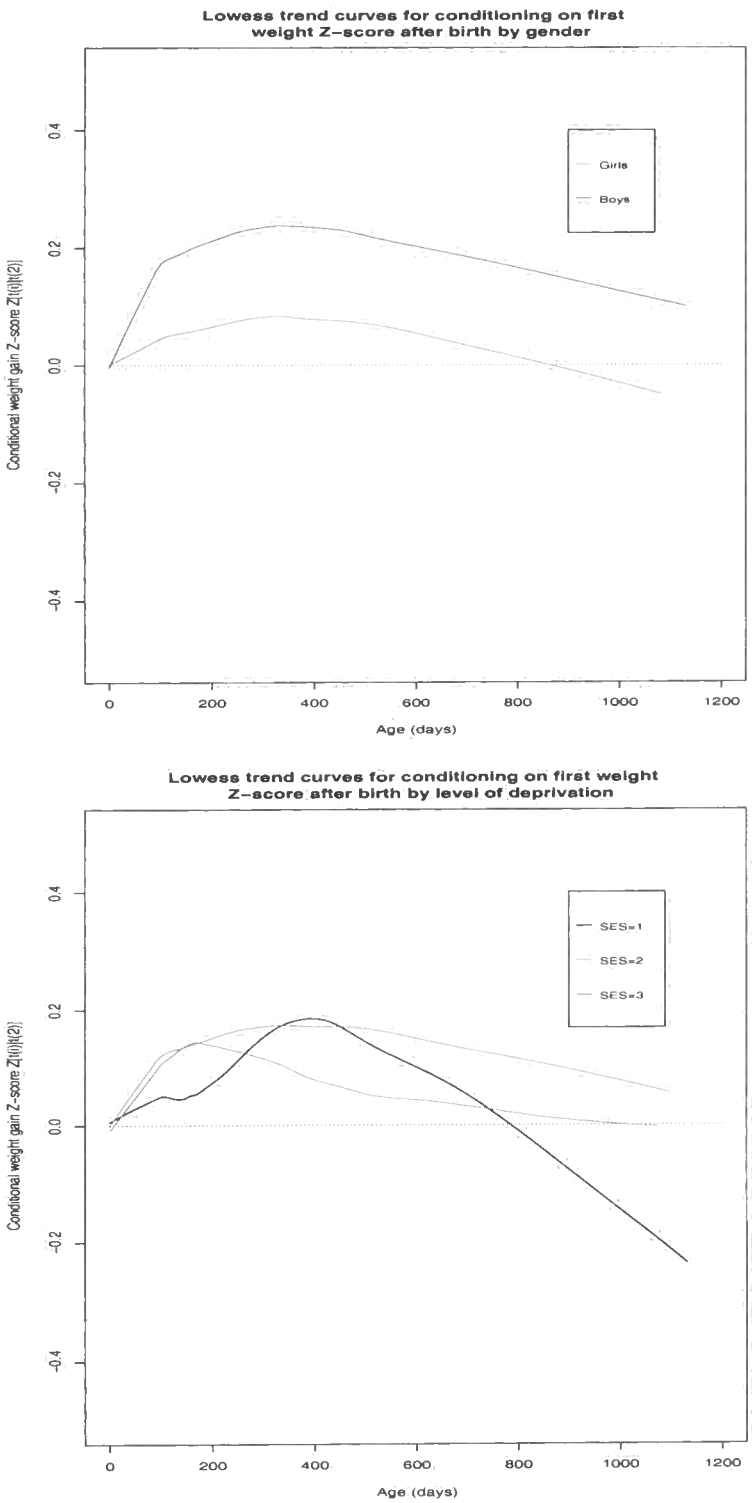


Figure 7.26: Conditioning on previous weight Z-score: Lowess trend curves for conditional weight gain Z-score versus age by sex and level of deprivation

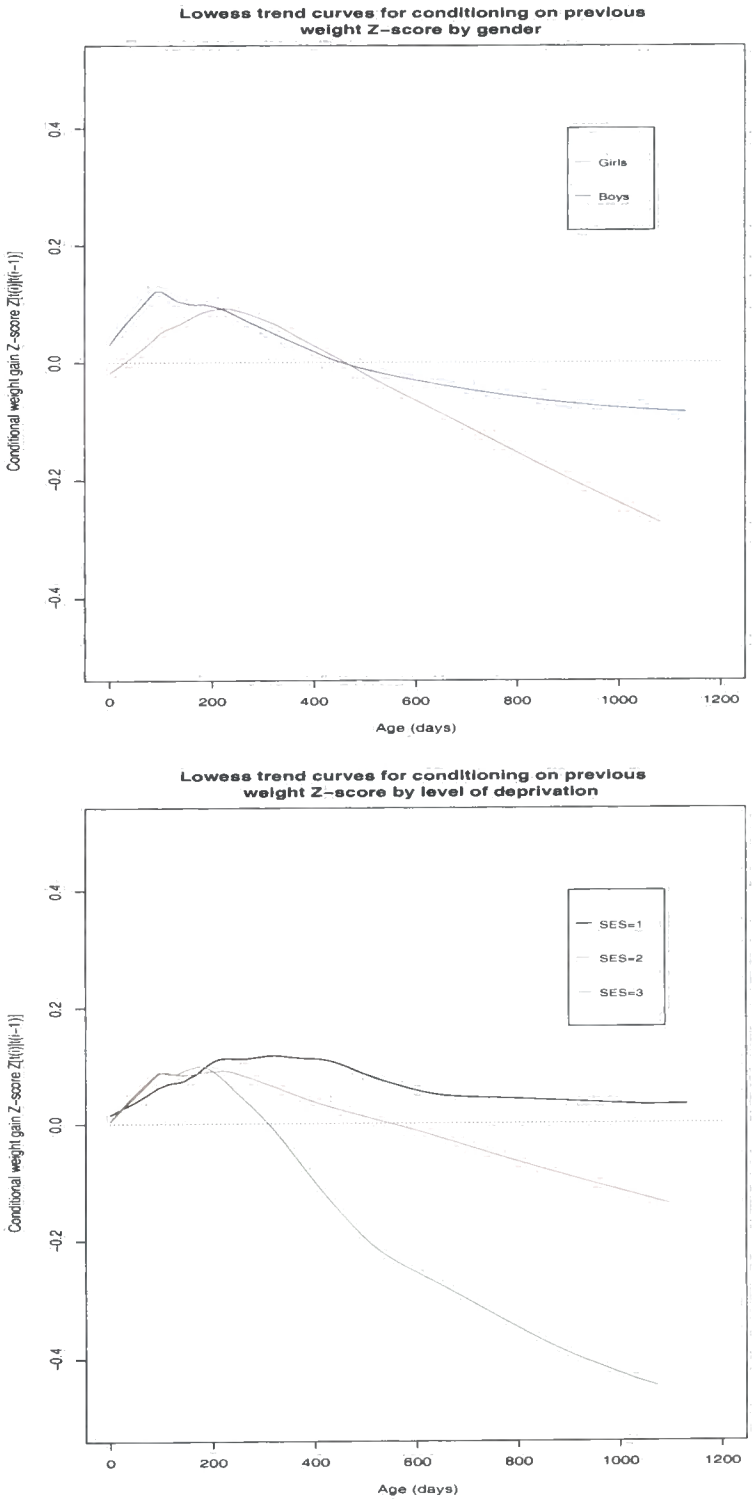


Figure 7.27: Conditioning on first weight Z-score: Histograms of proportion negative by case or other status (the same threshold is varied in steps of 0.2 until 1)

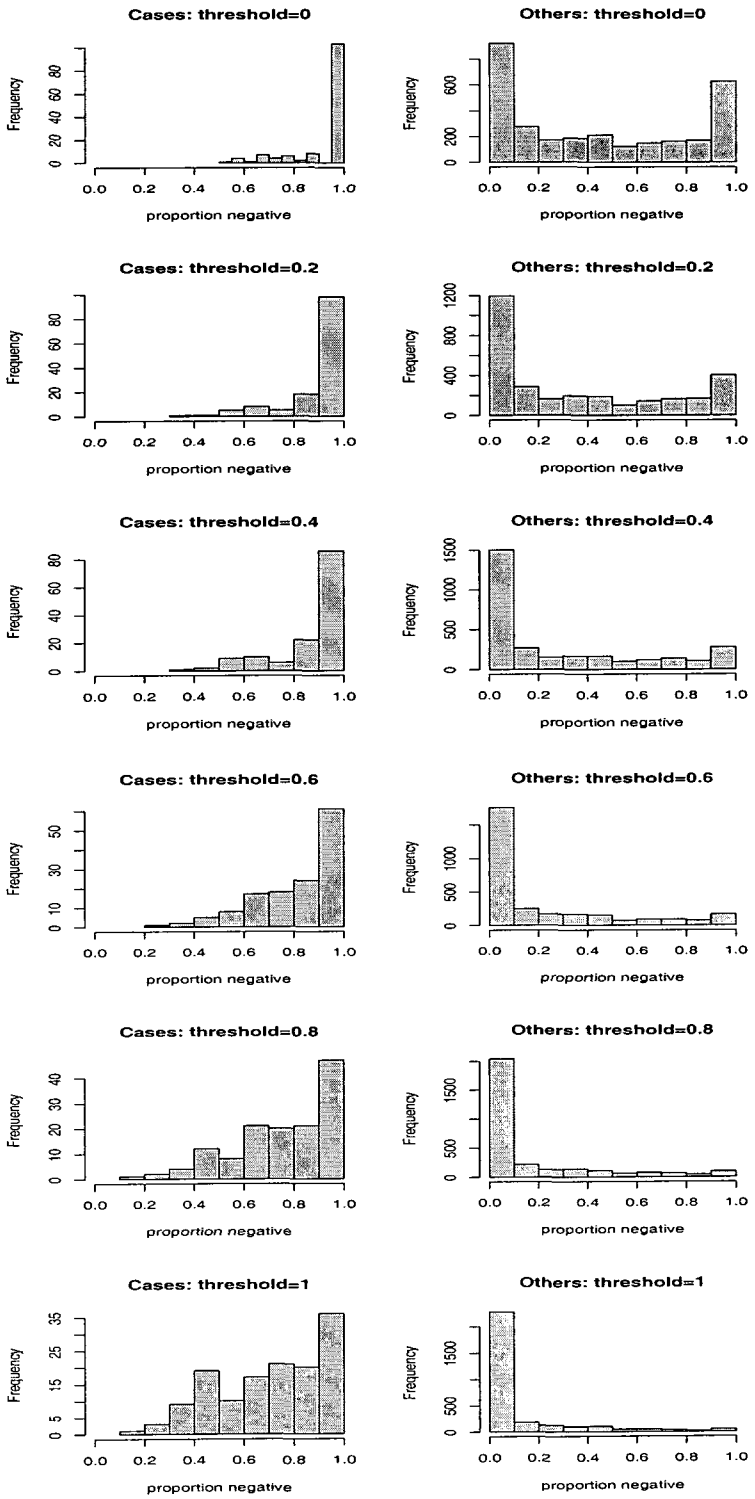


Figure 7.28: Conditioning on previous weight Z-score: Histograms of proportion negative by case or other status (the threshold is varied in steps of 0.2 until 1)

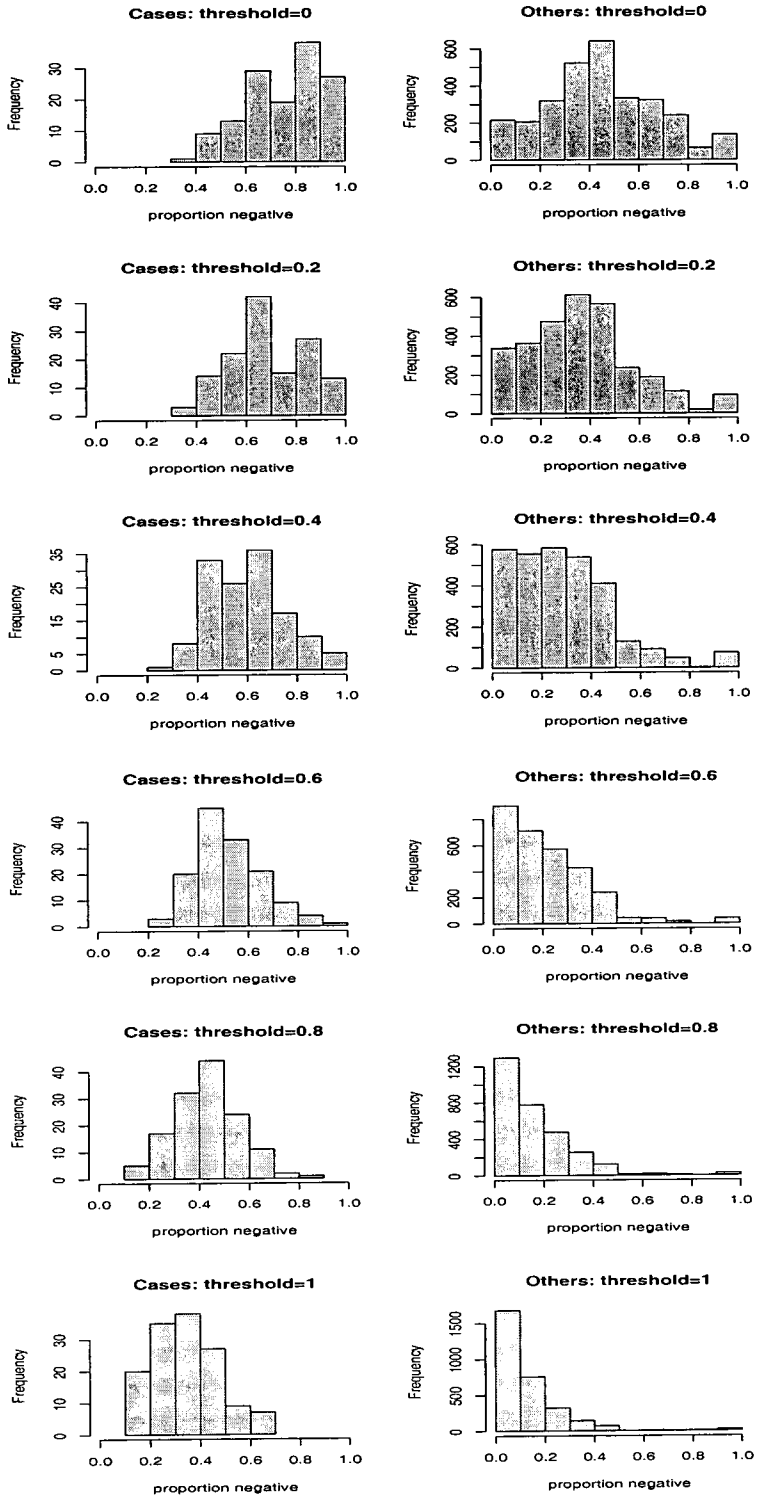


Figure 7.29: Conditioning on previous weight Z-score: Barplots of percentage of cases and others which have first two conditional weight gain Z-scores negative as the threshold varies between 0 and 1 in steps of 0.2 (the proportion of the bar shaded black represent the percentage of cases (others) where criterion 'first two conditional weight gain Z-scores negative' is true)

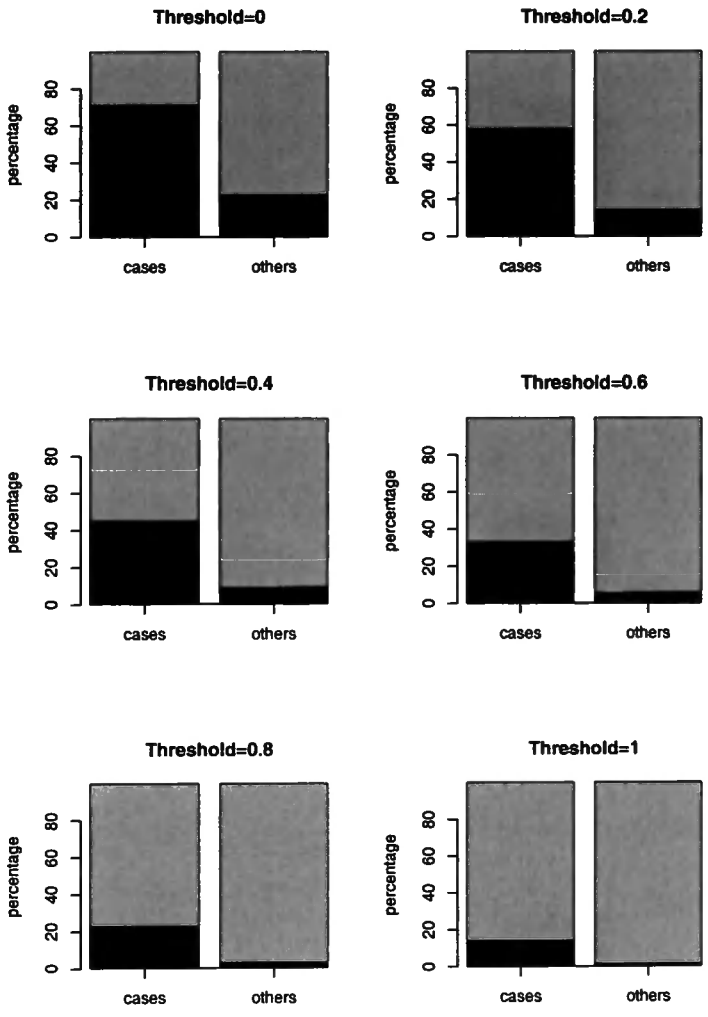


Figure 7.30: Conditioning on previous weight Z-score: Plots of conditional weight gain Z-score (conditioning on previous weight Z-score) versus age with lowess trend curves for individuals where 'two negatives' criterion is true and false (Threshold=0) Upper panel: Cases lower panel: Others

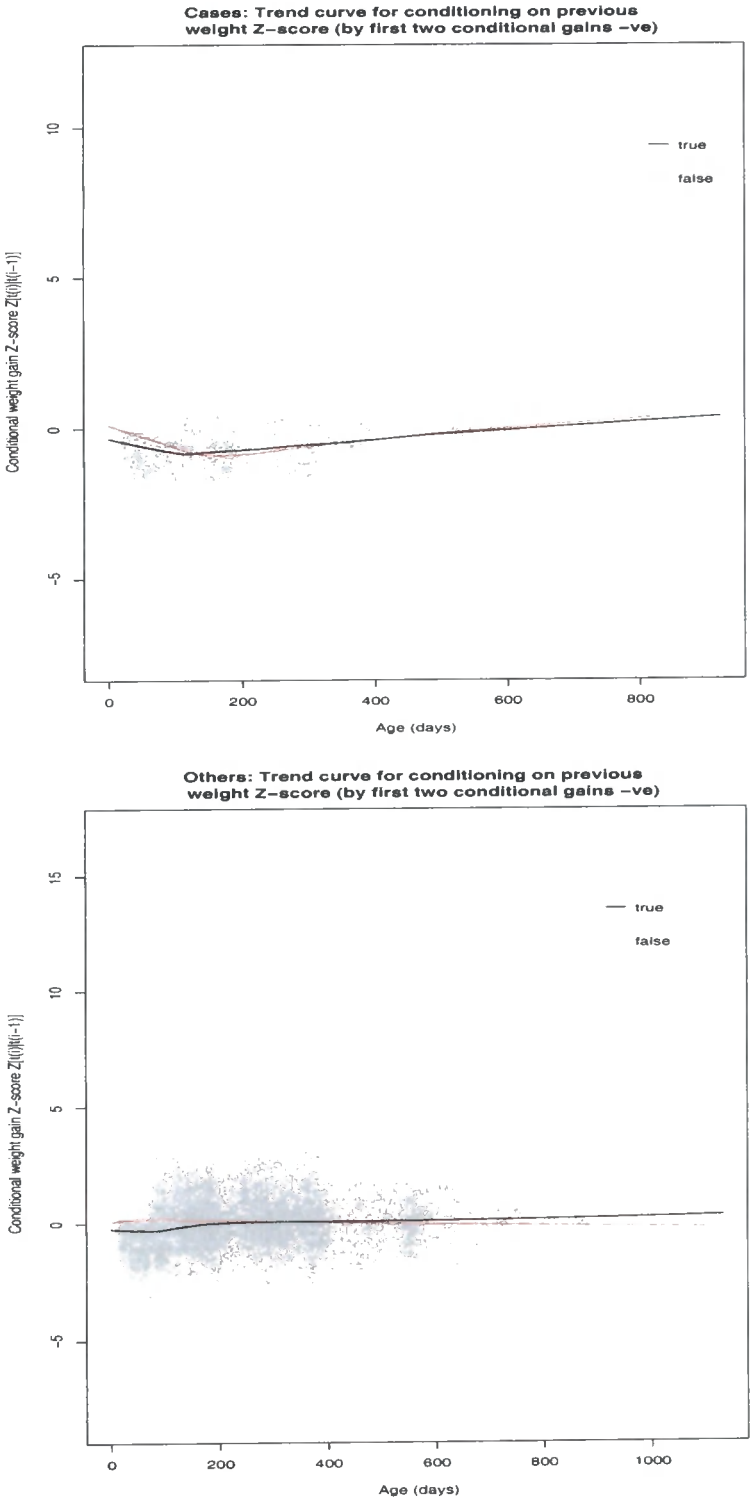


Figure 7.31: Conditioning on previous weight Z-score: Barplots of percentage of cases and others which have first three conditional weight gain Z-scores negative as the threshold varies between 0 and 1 in steps of 0.2 (the proportion of the bar shaded black represent the percentage of cases (others) where criterion 'first three conditional weight gain Z-scores negative' is true)

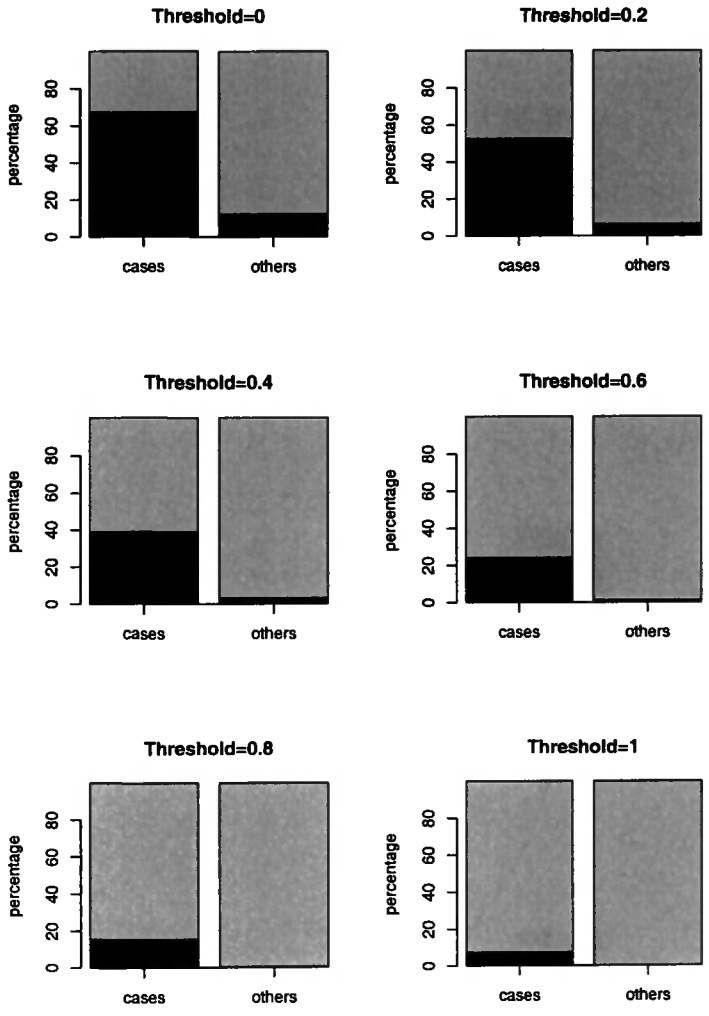


Figure 7.32: Conditioning on previous weight Z-score: Barplots of percentage of cases and others which have first four conditional weight gain Z-scores negative as the threshold varies between 0 and 1 in steps of 0.2 (the proportion of the bar shaded black represent the percentage of cases (others) where criterion 'first four conditional weight gain Z-scores negative' is true)

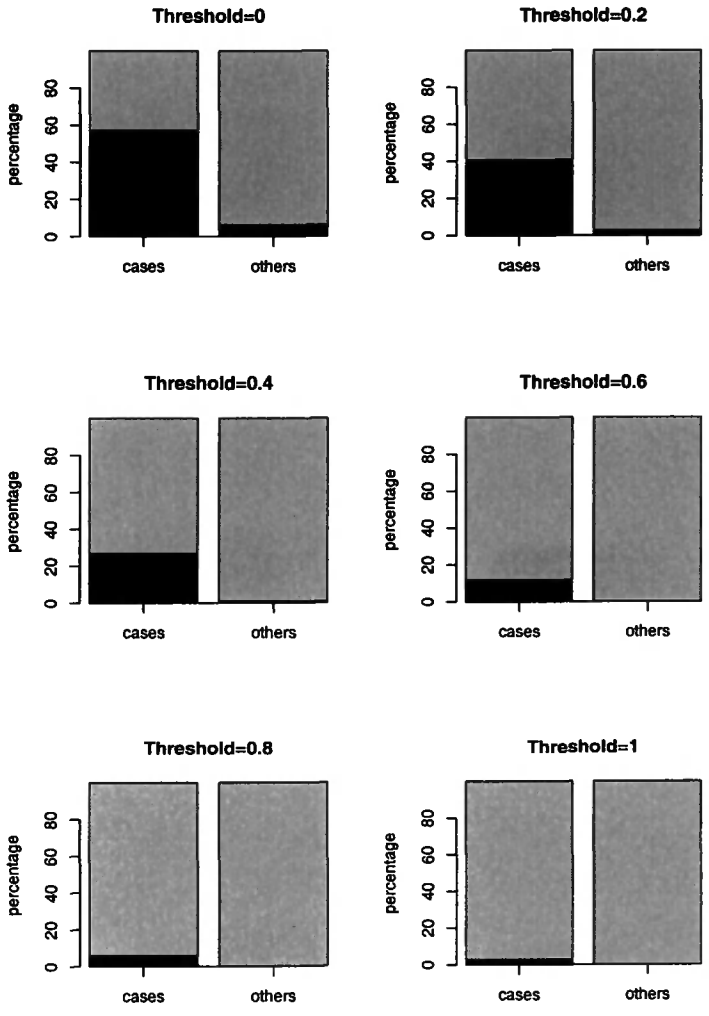


Figure 7.33: Conditioning on previous weight Z-score: Barplots of percentage of cases and others which have second and third conditional weight gain Z-scores negative as the threshold varies between 0 and 1 in steps of 0.2 (the proportion of the bar shaded black represent the percentage of cases (others) where criterion 'second and third conditional weight gain Z-scores negative' is true)

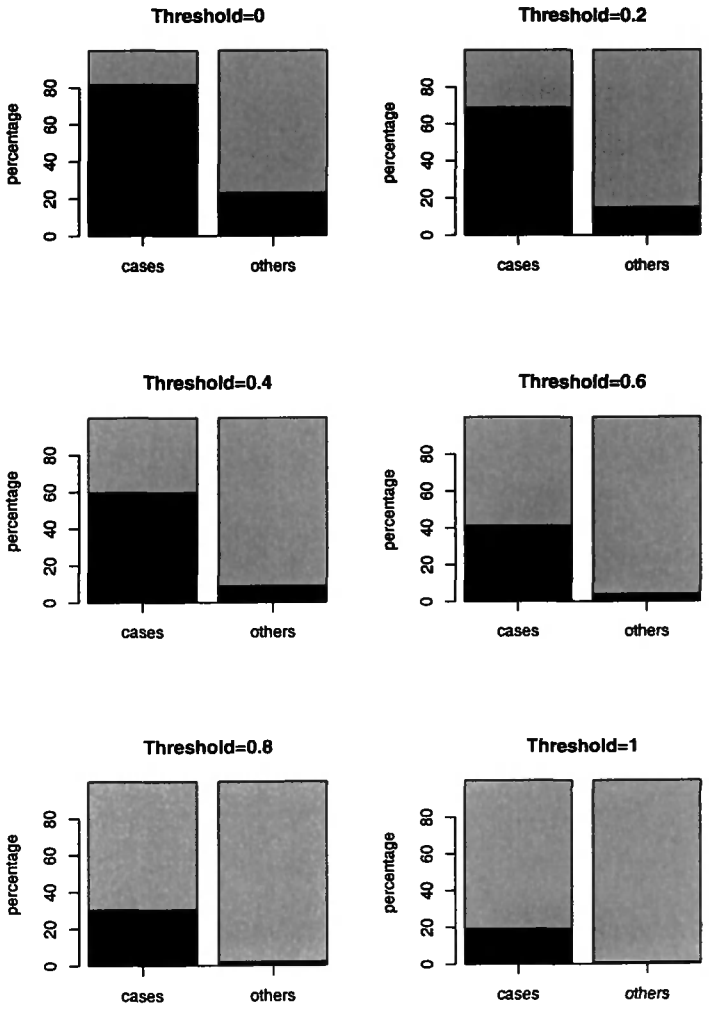


Figure 7.34: Plot of median pattern and fifth centile of weight gain generated by one variable Argyle model and Cole's (1998) model fitted to Cambridge correlations (including birthweight)

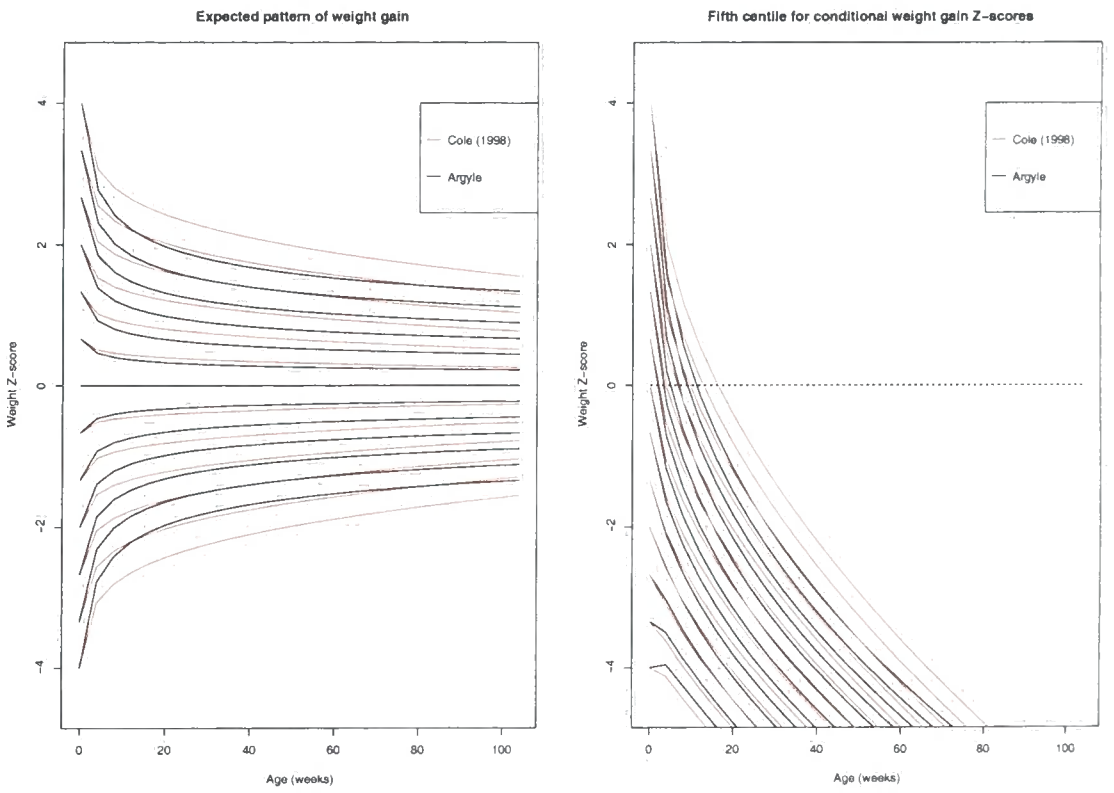


Figure 7.35: Plot of median pattern of weight gain generated by one variable Argyle model, along with conditional weight gain Z-scores for infant with ID 2149. Left Conditioning on previous weight Z-score. Right Conditioning on birth weight Z-score. The weight Z-scores are connected by a solid line. The conditional weight gain Z-scores are connected by dotted lines from age of first weight to age of second weight.

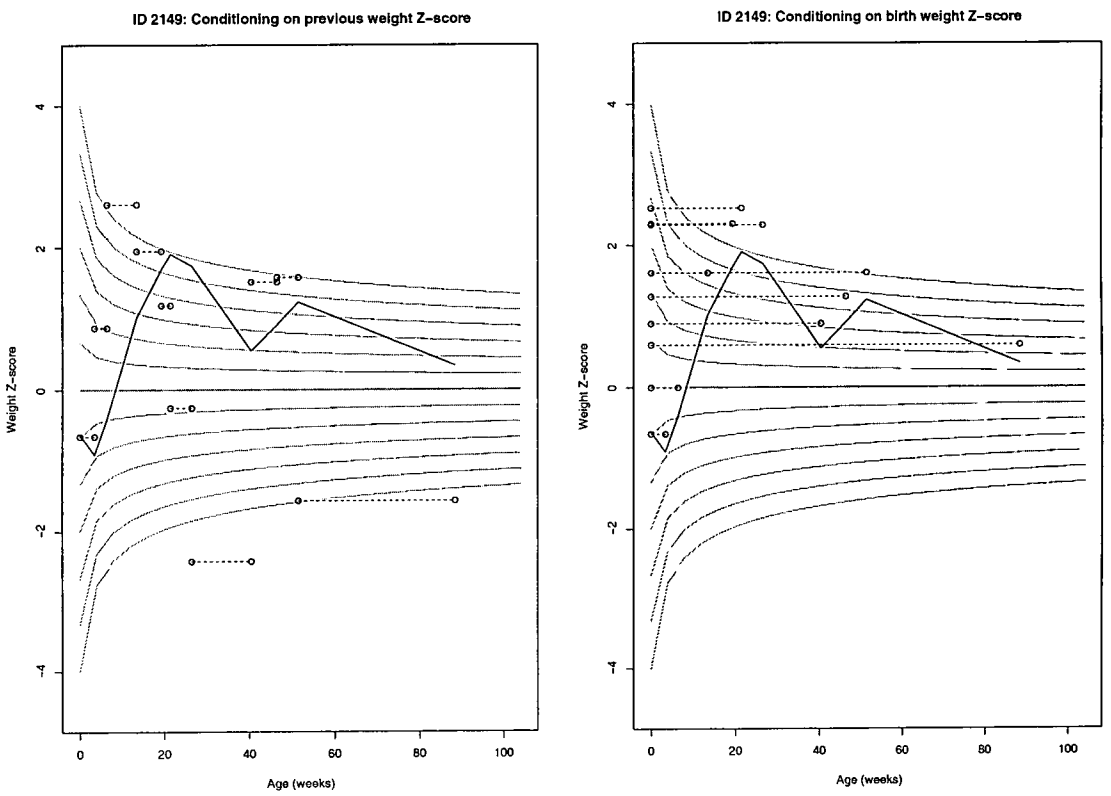
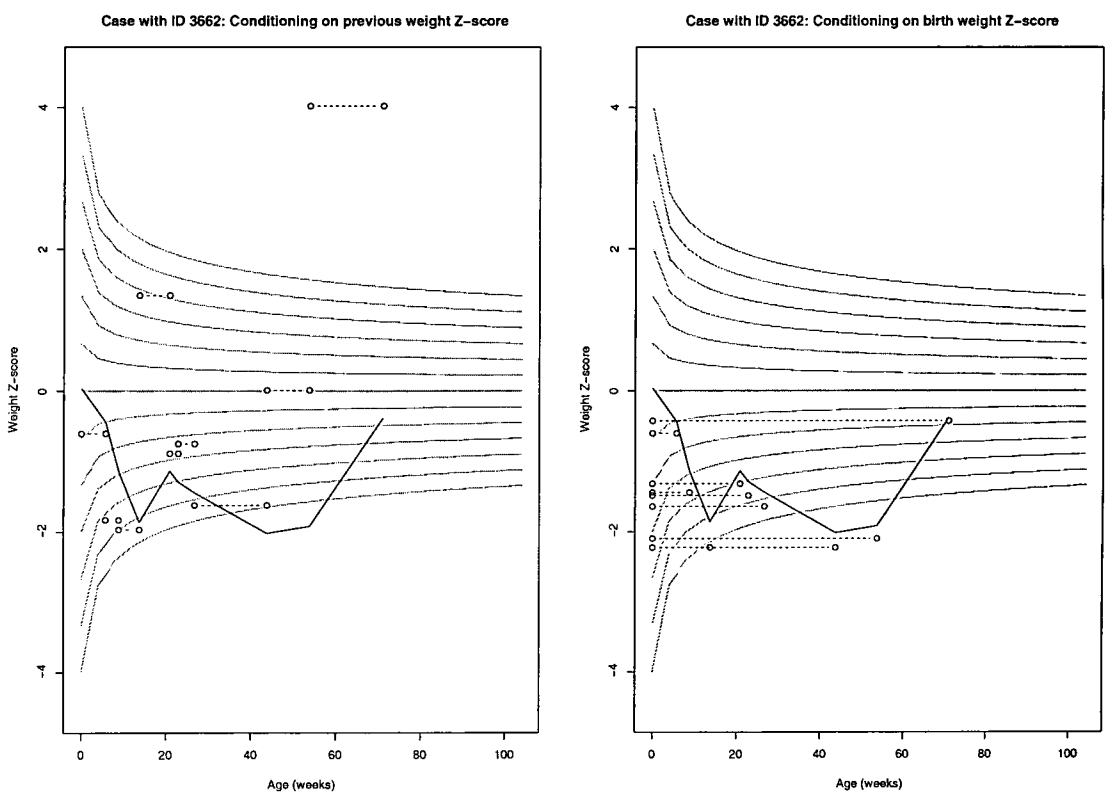


Figure 7.36: Plot of median pattern of weight gain generated by one variable Argyle model, along with conditional weight gain Z-scores for a case infant with ID 3662. Left Conditioning on previous weight Z-score. Right Conditioning on birth weight Z-score. The weight Z-scores are connected by a solid line. The conditional weight gain Z-scores are connected by dotted lines from age of first weight to age of second weight.



# Chapter 8

# School entry data

This chapter details the school entry height and weight data of children that participated in the follow-up study. I retrieved this data from school health records held in Newcastle clinics. This chapter also provides a preliminary analysis of this data.

In section 8.1 we discuss the role of school entry examination and discuss possible sources of error inherent in routine school entry data. Section 8.2 provides details of the data collection process and summarises the variables within the school entry data frame. Sections 8.3 to 8.5 provide a preliminary analysis of the school entry data. In section 8.4, we compare the attained heights and weights of the systematic sample to those children that contribute to the revised UK 1990 growth reference. In section 8.5 we compare attained heights and weights of children in the case-control study.

## 8.1 The role of the school entry examination

### 8.1.1 Routine height and weight monitoring after infancy

Children are routinely screened on entering school for their heights and weights. In addition the school entry examination provides the opportunity to screen for vision and hearing defects, and is useful for identifying children that may have missed out on pre-school health care (Hall and Stewart-Brown 1998). Over the years there has been much debate about the value of growth monitoring (GM) and whether height or weight should be monitored more frequently than at present. Recent guidelines suggest that school entry heights offer the best opportunity to identify conditions such as growth hormone deficiency and Turners syndrome (Hall 2000). The situation

with regard to weight monitoring after infancy is still less clear. However, recording both height and weight, in order to obtain the BMI of a child, would be of clinical value and public health interest (Hall 2000). Use of the BMI reference (Cole et al. 1995; Cole et al. 1998), for the school entry data appears to be appropriate (Voss and Mulligan 1999a). However guidelines for underweight children are not readily available. The main source of discussion and publications in the medical literature are with regard to the use of BMI for assessing obesity, because this affects a larger proportion of the population. There is no known literature on applying appropriate cut-offs for anorexic, very underweight and underweight children when using the UK 1990 BMI reference. This latter situation is likely to apply to a greater proportion of the school entry data, because children that failed to thrive in infancy may be wasted (low weight-for-height) and stunted at school entry.

High coverage is easily achieved at school entry because it is policy to assess all 5 year old children when starting school (Hall and Stewart-Brown 1998). Laing and Rossor (1996) examined the medical records of 327 children from eight schools in Lambeth, a deprived inner city area of London. Out of these, 262 children (80%) had their heights monitored at the age of 5 years, with coverage varying between 61% to 98% for different schools (Laing and Rossor 1996). A recent survey of current growth screening practice found:

> Pre-school and school age height screening took place in 75% and 81% of districts respectively, but most children were only measured once before school or at school (Hulse and Schilg (1995) in Schilg and Hulse (1997, pp3)).

The number of individuals with height and weight measures in the 20% systematic sample would provide some indication of the coverage achieved in the Newcastle area at that time, i.e. 1991-1993.

### 8.1.2 The accuracy of school entry measurements

> If a trained observer measures the height of a child to the nearest millimetre and then repeats the measurement, experience shows that the results will usually not be identical, indeed that differences of 3 to 5 mm are not uncommon. If the observer is inadequately trained larger differences will occur, and still larger differences are found when the two measurements are made by different observers. The height of a child does not change over a

> period of few minutes, and the discrepancies between measures must be due to failure of the recorded figures to correspond exactly to the true height; in brief to measurement error (Healy 1989b, pp893).

The equipment used to measure height and weight at school entry and its accuracy will vary from one school to the next. Laing and Rossor (1996) carried out a survey of measuring equipment in 59 schools in Lambeth; noting both the type and location of measuring equipment. A metre rule and 10 Kg were used to check the scales and height measures in each school (Laing and Rossor 1996). Laing and Rossor (1996) found that 27 scales weighed within $\pm 100$ g of the 10 Kg weight (range 7.5 Kg to 11.95 Kg) and 23 height measures were within $\pm 0.5$ cm of the metre rule (range 98 cm to 102 cm). Voss et al. (1990) used a metre rule to check the installation of 230 measuring instruments in health centres, hospitals, schools and GP surgeries in Wessex and found the readings obtained ranged from 90 to 108.5cm.

At school entry the measurements are taken by different nurses, so the measurement techniques used may be different and each nurse will have their own individual degree of error when measuring height. Cotterill et al. (1993) considered the reliability (reproducibility and accuracy) of 7 school nurses' measurements in comparison to a trained auxologist. Seven groups of 10 children, aged 5-6 years, were measured 4 times using the same technique (twice by both nurse and auxologist). Cotterill et al. (1993) found that the pooled standard deviation of the differences between repeat measures recorded by the school nurses were comparable to those of the auxologist (0.32cm versus 0.35cm, respectively). The difference between mean values of measurements made by the school nurse and the auxologist were taken to represent the accuracy of the school nurses measures; the range of the school nurses measurements were -0.53 cm to +0.64 cm in contrast to the auxologists (Cotterill et al. 1993).

A child's height is also affected by the time of day, children are taller in the morning than in the afternoon and this is termed diurnal variation. Voss and Bailey (1997) stated that most authors agreed that the loss in height over the day was between 2 and 3 cm. There is also the question of whether the child was stretched or not, community guidelines recommend that gentle upward pressure is applied to the mastoid process to encourage the child to stand tall (Schilg and Hulse 1997). Voss and Bailey (1997) found that by using this technique the recorded height was significantly affected by an average of 0.28 cm. Voss and Bailey (1997) saw no advantage in stretching as it didn't minimise diurnal variation and recommended that the child should be measured at the same time of day on each measurement

occasion, preferably in the afternoon.

Errors can be made when recording heights and weights. For example there may be the absence of a date, so the exact age can not be calculated or the nurse may only plot the height and weight data rather than writing the numerical value down. Laing and Rossor (1996) looked at the latter and found that for 85 out of 327 records, there were only plotted points with no numerical value recorded. When a measurement had both the numerical value noted and was also plotted on a child's growth chart, Laing and Rossor (1996) found that 10 measures were plotted incorrectly (error ranged from 2 to 10 cm).

Although the guidelines for weighing children over 2 indicate that children should wear no more than light clothing with shoes off (Schilg and Hulse 1997), children undergoing a school entry assessment are likely to be fully clothed. The date of the school entry assessment will provide some indication to whether the child is in summer or winter clothing. However thin children are more likely to wear thicker clothes and children from a lower socio-economic class are more likely to wear jeans. Children from the case-control study come predominantly from deprived areas of Newcastle. Sumner and Whitacre (1931) (in Alsop-Shields and Alexander (1997)) found that for half of their sample of 100 Texan children (aged 10-11 years), weight variations in a child's clothing could equal or exceed their monthly change in weight.

The data collected, see section 8.2, will be on children of a variety of ages and the accuracy of these may be influenced by the child's age, a young child is more likely to move when being weighed or having their height measured. The standard deviation of a single height measurement is around 0.2-0.3 cm in school age children, leading to 95% confidence limits of about $\pm$0.5cm (Hall 2000). However these confidence limits are wider for younger children; about $\pm$1cm for two year olds (Hall 2000). Ahmed et al. (1990) assessed the accuracy of height measurements made by two sets of four health visitors on two groups of ten children, aged 3 and 4.5 years, respectively. These measurements were then compared to a trained auxologists measurements on a different piece of equipment. Ahmed et al. (1990) found that a substantial part of the variance in the observed measurement bias was attributable to the three year olds; due to difficulty in obtaining accurate measurement at this age. A further consideration is that the time of school entry examination may be influenced by other factors, such as a delayed start to schooling.

As can be seen by the discussion in this section, much emphasis is placed on the accurate measurement of height by school nurses as this provides the best opportunity to detect growth problems of an organic aetiology. Little emphasis is placed on

the accurate measurement of weight as it is assumed that:

> ... weight is easy to determine accurately. Weight measurements do not, however, provide us with sufficient information for the assessment of linear growth. If a child's weight increases with time we only know that he has become heavier - but whether this is due to excess fat, increased bone or muscle, or retention of body fluid as a result of some pathological condition is unknown (Ahmed et al. 1990, pp1347).

However if it is viewed that BMI is an important quantity to monitor then surely the accuracy of a child's weight is vital too.

## 8.2 Practicalities of collecting school entry data

### 8.2.1 Motives for data retrieval

The motives for retrieving the school entry data were:

1. that it was policy to give all individuals entering school an assessment
2. this data was readily available in school medical records for the individuals involved in the follow-up study
3. this data would 'fill' the gap between the infancy and follow-up study data

The weights already obtained in infancy only indicate how well the child grew in infancy and tell us nothing about the length of the individual at this stage as these were not routinely taken. The BMI obtained from the follow-up data only tells us whether the child is underweight (overweight) at their present height and doesn't provide any indication of whether this child has always been underweight (overweight).

An additional motive, at the time, was to provide a crude indicator of the child's height velocity as this is often monitored by specialists when testing for growth hormone secretion (Brook and Hindmarsh 1991). Recent publications (Voss 1999; Hall 2000) concluded that monitoring height velocity was not useful because of the lack of precision (Voss et al. 1991). However, the inherent lack of precision would be

amplified in inexperienced hands (such as in a routine setting) because of elevated measurement error in individual height measurements.

In individuals that had failed to thrive in infancy, the school entry measure would provide some indication of whether the child had growth problems that had persisted beyond infancy and may provide some indication of presence or absence of stunting or wasting in this population at school entry.

The overall aim was to obtain school entry data (date measurement taken, height and weight) for individuals in the case-control study and the systematic 20% sample. In addition, if any other subsequent height and weight measurements were documented then these were also noted.

### 8.2.2 The location and collection of school entry data

The school entry data was retrieved over a period of 2-3 months; data collection was completed by the end of June 1998. In order to trace the children's records, the original database for the follow-up study was utilised, this had last been updated 18 months earlier. This database contained information on 808 individuals. From this, the child's school was used to identify the clinic where the child's school health record (M10) should be located. Some individuals had moved schools since the follow-up study, in particular pupils within the west area of Newcastle had moved from first school to middle school. Children that had moved schools were traced by using clinic out-books or the child health computer at information services, Newcastle General Hospital.

The most difficult records to locate were those children that had changed surname. Further some files were unavailable through being with clinic staff. Other children had moved just outside the Newcastle area and some records were 'centrally held' (this usually means the family have moved out of the district and the notes haven't been requested from their new health authority yet). Centrally held records were stored in date of birth order at Newcastle General Hospital.

In Newcastle the health services are divided into four regions, within these are a total of 11 clinics and 5 special schools. These clinics and special schools hold a child's M10 until the child moves to secondary school. The original follow-up study database was read into the software package ACCESS and a database file was created for school-clinic allocation. A query option was defined within ACCESS to create forms in alphabetic order for each clinic to aid data collection. All heights and weights documented since infancy were recorded. Within a child's M10 their

school entry height and weight were supposed to be recorded on the front page of their file along with any subsequent heights or weights. In practice, most of the records had a standard loose sheet attached with the height and weight recorded along with details of vision and hearing tests. Sometimes the heights and weights were in a school interview form. Some individuals had their heights and weights recorded and plotted on a growth chart (mainly east clinics and special schools) and others were within school nurse notes. In the east district area it was policy to give all children nine-pluses, these measurements were taken after the follow-up study and have also been documented. Children in special schools tended to have almost yearly measures, although the numerical value wasn't always recorded and often points were just plotted on a growth chart. In special schools, if the child had 'Downs' then their height and weight were plotted on a growth chart for children with Downs syndrome.

Out of the 807 records (1 child had died), 5 (0.6%) of these were not found, 68 (8.4%) had no 'school-entry' measure and 71 (8.8%) had moved outside the area. School entry assessment data for height and weight were obtained for 567 (249 boys, 318 girls) and 569 (250 boys, 319 girls) children, respectively. In general most individuals had a school entry assessment measure, with the exception of children attending private school.

The school entry data set has the following variables for each individual:

1. **ID**
2. **DOB**
3. **Sex**
4. **Date.ent** Date of school entry assessment; this was usually during the calendar years of 1991-93. This was used to derive **Age.ent**, the age at school assessment (which in general was between 4 and 6 years). Occasionally, only the month and year when height and weight measurements were monitored was recorded in the M10. In this situation, if a clinic visit was recorded within the notes in the same month then this date was used. In one instance the $15^{th}$ of that particular month was used. In using the middle of the month there would be less bias than taking either the beginning or the end of the month. In theory the age will be out by at most 15 days.
5. **Weight.ent** The child's weight was taken down as recorded, the majority of these were in metric. The few that were in imperial measures

were converted to Kilogrammes using 1 lb $\sim$ 0.4536 Kg.

6. **Height.ent** The child's height was taken down as recorded. If recorded in imperial units, the measurement was converted to centimetres.

7. **BMI.ent** This was derived from the heights and weights.

The heights, weights and BMI were converted to Z-scores using the revised UK 1990 growth reference (Cole et al. 1998), to give **ZscoreH.ent**, **ZscoreW.ent** and **ZscoreB.ent**, respectively.

## 8.3 Age distribution of school entry assessment

Originally the school entry data was taken to be the first weight and height measure after infancy. In this first data set the age range covered was 1.94 to 8.43 years. On further examination of the most extreme individuals:

1. ID 1870 has a weight and height measure at the end of infancy. This individual is identified as a case although the Z-score for height at 1.94 years is 0.91 (which is above the median of the reference) and the Z-score for BMI is -0.50 (which is below the median but would not classify the child as wasted). This is one of the cases that does show some kind of recovery by the end of infancy. No later height or weight measures were taken, so we have no information on this individual's growth between infancy and the follow up study.

2. ID 1197 has an assessment at 8. This individual is part of the systematic 20% sample but as there is no follow up data it is presumed that this individual's parents did not consent. These routine measures could potentially be included in the follow up study data set. This individual has extremely low height (Z-score = -4.05) and weight (Z-score = -3.40) but a BMI (Z-score = -0.76) in the normal range.

3. ID 3600 had 12 measurement occasions after infancy. On two occasions only the weight was monitored, indicating that there was some concern about this individual's weight (at 3.5 years this individual had a Z-score of 3.09 for BMI but at 5 years this was reduced to 2.26). This individual belongs to the systematic sample and had their first weight and height taken at 2.5 years. The measurements made closest to 5 years were selected.

The boxplot in the upper panel of figure 8.1 illustrates that most of the school entry assessments were made between 3.5 and 6 years, but there are still 12 outliers. These outliers were not excluded from the school entry data set because they provide a 'link' between the infancy and follow-up study data. Details of these outliers including their Z-scores can be found in table 8.1. These individuals tend to be members of the systematic sample or controls, with the exception of ID 3662, he has a height, weight and BMI below the second centile.

The final data frame for school entry, contains weights for 569 individuals (250 boys, 319 girls) and heights for 567 individuals (249 boys, 318 girls). A histogram of the distribution of the age at school entry assessment can be found in the lower panel of figure 8.1; age varies from 3 years and 2 months to 7 years and 4 months. This was of interest because:

> It is national policy to offer a child health surveillance review to all children at around 3.5 years of age and this is usually undertaken by the family's health visitor. The coverage varies widely, from as low as 40% to over 80% (Hall and Stewart-Brown 1998, pp941).

Therefore we may have expected to see a peak around 3.5 years as well as 5 years in the histogram of the age distribution. However this is certainly not the case, with only 3 individuals (only two shown in histogram plus ID 3600's earlier assessment)

Table 8.1: Details of 12 individuals with outlying age at school entry assessment (where age is in years, sex is 1 for male and 2 for female, CACO is case-control indicator defined in Chapter 6, HSDS is the height Z-score, WSDS is the weight Z-score and BSDS is the BMI Z-score)

| ID | age | sex | CACO | HSDS | WSDS | BSDS |
|---|---|---|---|---|---|---|
| 862 | 6.18 | 2 | NA | -0.9955 | -0.8817 | -0.415 |
| 940 | 6.16 | 1 | NA | 1.9977 | 1.7557 | 0.9054 |
| 1920 | 6.22 | 2 | NA | 1.6936 | 1.2324 | 0.4429 |
| 2131 | 6.38 | 1 | 2 | -0.2433 | -0.5302 | -0.6048 |
| 2416 | 6.46 | 1 | NA | 1.4587 | 1.5616 | 1.0579 |
| 2459 | 6.59 | 2 | NA | -0.4625 | -0.4203 | -0.2515 |
| 2825 | 6.20 | 2 | NA | 0.7936 | 0.8499 | 0.5368 |
| 3195 | 6.18 | 1 | 2 | 0.5027 | -0.0606 | -0.6518 |
| 3640 | 3.19 | 2 | NA | 0.1616 | 1.5934 | 1.9762 |
| **3662** | **7.34** | **2** | **1** | **-2.2459** | **-3.607** | **-2.8981** |
| 3702 | 6.22 | 2 | NA | 0.5765 | -0.001 | -0.5404 |
| 3741 | 6.37 | 1 | NA | 0.1753 | 0.7907 | 0.9621 |

having weight and height measures taken close to 3.5 years. The majority (96%) of school entry assessments are done between the ages of 4 and 5.75 years.

## 8.4 Preliminary analysis of school entry data for individuals within the systematic sample

### 8.4.1 Summary plots and statistics

There are 561 individuals in the systematic sample. There are 395 individuals (70%) that have a school entry assessment, which is ten percent lower than the coverage achieved by the research follow up study. The systematic sample contains some individuals that are cases and as these children have poor attained growth outcomes at 7-9 years (Drewett et al. 1999; Corbett 1998), we would expect a similar outcome at school entry. In order to assess the adequacy of the revised UK 1990 growth reference for Newcastle children at school entry, the data for the cases were excluded as the inclusion of these children may lower the mean Z-scores for the systematic sample. The school entry data is over an age range of 3.5 years, so in order to assess the adequacy of the growth reference, the children from the systematic sample were grouped according to the nearest full year. Grouping to nearest year was thought to be appropriate because of seasonal variation in growth. Children grow faster in height in spring and summer than in autumn and winter (Marshall 1971). Summary statistics at grouped ages of 4, 5 and 6 years can be found in tables 8.2 and 8.3 for boys and girls, respectively. For example, children that have school entry assessment in the age range (3.5,4.5] are considered to belong to the 4 year old group. However, for both boys and girls the 4 year old group represent ages 4 to 4.5 and anthropometric measures are taken mainly between September and December. Two girls had school entry data outside of these age groupings (ID 3640 and ID 2459 measured at 3 years and 2 months, and 6 years and 7 months respectively). Tables 8.2 and 8.3 indicate that the majority of school entry assessments lie within the grouping of 5 years and relatively few individuals have school entry assessments at the grouped age of 6 years. In each age-sex grouping the mean SES was calculated, this provides a rough indication of whether children within these groupings have a similar socio-economic status to the Newcastle population. Recall that from the birth cohort of 3415 infants, 379 individuals had SES indicator of 1, 2353 individuals had SES indicator of 2 and 683 individuals had SES indicator of 3. The mean SES within this birth cohort was 2.089, only the 5 year old boys had

a significantly different mean SES (t=2.59, p=0.011, 95% CI [2.120, 2.324]). This leads us to suspect that the 5 year old boys within the systematic sample may be more deprived than the general Newcastle population.

Variable width boxplots for the Z-scores of height, weight and BMI can be found in figure 8.2. The width of each box in these plots is made proportional to the square root of the number of children within the corresponding age-sex specific group (McGill et al. 1978). We should be aware that the 6 year olds only represent a small group and as discussed above the 4 year olds are a 6 month age grouping measured predominantly in the winter months. The boxplots for height Z-score of boys in all age groups are reasonably symmetric and there are no outliers. There

Table 8.2: Boys: Summary statistics for height, weight and BMI measures and corresponding Z-scores (systematic sample excluding cases)

**Age 4 to nearest year** (n=56, av. SES=2.071)

| | Min. | LQ | Median | Mean | UQ | Max. | SD |
|---|---|---|---|---|---|---|---|
| Age.ent | 4.070 | 4.250 | 4.360 | 4.328 | 4.420 | 4.500 | 0.1124 |
| Height.ent | 98.0 | 102.0 | 104.8 | 104.9 | 107.9 | 113.0 | 3.6197 |
| ZscoreH.ent | -1.8030 | -0.6329 | 0.0187 | 0.0231 | 0.5852 | 1.700 | 0.8331 |
| Weight.ent | 13.50 | 16.00 | 17.50 | 17.82 | 19.00 | 25.20 | 2.1635 |
| ZscoreW.ent | -2.1900 | -0.5504 | 0.1039 | 0.2180 | 0.7853 | 2.8570 | 0.9686 |
| BMI.ent | 13.93 | 14.99 | 15.97 | 16.17 | 16.78 | 19.84 | 1.4754 |
| ZscoreB.ent | -1.6240 | -0.5674 | 0.2630 | 0.2933 | 0.8252 | 2.6260 | 1.0708 |

**Age 5 to nearest year** (n=108, av. SES=2.222)

| | Min. | LQ | Median | Mean | UQ | Max. | SD |
|---|---|---|---|---|---|---|---|
| Age.ent | 4.510 | 4.765 | 4.965 | 4.964 | 5.125 | 5.480 | 0.2510 |
| Height.ent | 95.0 | 106.0 | 108.9 | 108.8 | 112.0 | 119.3 | 4.5319 |
| ZscoreH.ent | -2.6230 | -0.7261 | -0.0370 | -0.1001 | 0.5188 | 1.7860 | 0.9738 |
| Weight.ent | 13.00 | 17.00 | 19.00 | 18.88 | 20.13 | 27.60 | 2.4498 |
| ZscoreW.ent | -2.73900 | -0.7109 | 0.1567 | 0.0456 | 0.7823 | 3.1740 | 1.0348 |
| BMI.ent | 12.63 | 14.97 | 15.71 | 15.90 | 16.74 | 20.94 | 1.4121 |
| ZscoreB.ent | -3.0590 | -0.5242 | 0.1372 | 0.1558 | 0.8812 | 3.0960 | 1.0756 |

**Age 6 to nearest year** (n=16, av. SES=1.938)

| | Min. | LQ | Median | Mean | UQ | Max. | SD |
|---|---|---|---|---|---|---|---|
| Age.ent | 5.510 | 5.575 | 5.660 | 5.828 | 6.108 | 6.460 | 0.3280 |
| Height.ent | 103.0 | 111.4 | 114.6 | 115.3 | 119.1 | 126.7 | 6.3782 |
| ZscoreH.ent | -2.1100 | -0.4425 | 0.0726 | 0.0838 | 0.6074 | 1.9980 | 1.0398 |
| Weight.ent | 17.00 | 18.75 | 21.00 | 21.61 | 24.30 | 27.20 | 3.6446 |
| ZscoreW.ent | -1.2760 | -0.5912 | 0.1988 | 0.2873 | 1.3220 | 1.9370 | 1.1288 |
| BMI.ent | 13.97 | 14.71 | 16.45 | 16.15 | 16.99 | 18.86 | 1.5070 |
| ZscoreB.ent | -1.4030 | -0.6502 | 0.7002 | 0.3508 | 0.9993 | 2.0810 | 1.0817 |

Table 8.3: Girls: Summary statistics for height, weight and BMI measures and corresponding Z-scores (systematic sample excluding cases)

**Age 4 to nearest year** (n=65, av. SES=2.138)

| | Min. | LQ | Median | Mean | UQ | Max. | SD |
|---|---|---|---|---|---|---|---|
| Age.ent | 4.090 | 4.270 | 4.350 | 4.343 | 4.430 | 4.500 | 0.1061 |
| Height.ent | 90.0 | 100.0 | 102.4 | 102.9 | 106.5 | 114.0 | 4.6697 |
| ZscoreH.ent | -3.2010 | -0.8783 | -0.3612 | -0.2507 | 0.5239 | 2.4430 | 1.1131 |
| Weight.ent | 12.50 | 15.50 | 17.00 | 17.01 | 18.00 | 23.80 | 2.1731 |
| ZscoreW.ent | -2.7330 | -0.6754 | 0.0339 | -0.0448 | 0.5999 | 2.5800 | 1.0117 |
| BMI.ent | 13.39 | 15.00 | 15.96 | 16.02 | 17.00 | 20.40 | 1.4650 |
| ZscoreB.ent | -1.8410 | -0.4455 | 0.2728 | 0.2080 | 0.9258 | 2.6010 | 0.9921 |

**Age 5 to nearest year** (n=107, av. SES=2.15)

| | Min. | LQ | Median | Mean | UQ | Max. | SD |
|---|---|---|---|---|---|---|---|
| Age.ent | 4.510 | 4.745 | 4.920 | 4.948 | 5.120 | 5.460 | 0.2542 |
| Height.ent | 93.7 | 104.0 | 107.0 | 107.2 | 110.0 | 117.5 | 4.4442 |
| ZscoreH.ent | -2.9490 | -0.8988 | -0.2917 | -0.2723 | 0.3127 | 2.4500 | 0.9872 |
| Weight.ent | 13.00 | 16.45 | 18.00 | 18.13 | 19.60 | 25.40 | 2.4452 |
| ZscoreW.ent | -2.5750 | -0.8570 | -0.0085 | -0.1249 | 0.5228 | 2.6300 | 1.0271 |
| BMI.ent | 11.57 | 14.80 | 15.68 | 15.74 | 16.70 | 21.58 | 1.5451 |
| ZscoreB.ent | -3.6620 | -0.5324 | 0.1268 | 0.0583 | 0.7720 | 2.9990 | 1.0236 |

**Age 6 to nearest year** (n=17, av. SES=1.765)

| | Min. | LQ | Median | Mean | UQ | Max. | SD |
|---|---|---|---|---|---|---|---|
| Age.ent | 5.510 | 5.640 | 5.680 | 5.831 | 6.100 | 6.220 | 0.2583 |
| Height.ent | 104.5 | 110.0 | 115.5 | 114.2 | 118.5 | 125.0 | 5.8128 |
| ZscoreH.ent | -2.2550 | -0.8136 | 0.2223 | -0.0233 | 0.7936 | 1.6940 | 1.1330 |
| Weight.ent | 15.00 | 18.20 | 20.00 | 20.25 | 22.00 | 25.50 | 3.1213 |
| ZscoreW.ent | -2.4570 | -0.8533 | -0.0010 | -0.0804 | 0.7242 | 1.6780 | 1.0989 |
| BMI.ent | 13.19 | 14.31 | 15.17 | 15.47 | 16.51 | 19.12 | 1.5640 |
| ZscoreB.ent | -1.7890 | -0.8267 | -0.1930 | -0.1090 | 0.6405 | 1.8950 | 1.0102 |

is some evidence of skewness within the boxplots for height Z-scores of girls aged 4 and 6 years old. The boxplots for weight Z-scores are reasonably symmetric. The 6 years old boys appear to be more homogeneous group but these are from a winter 6-month period. There is evidence of skewness within the boxplot for BMI Z-scores of 6 year old boys.

### 8.4.2 Adequacy of UK 1990 growth reference for Newcastle children at school entry

To assess the adequacy of the revised UK 1990 reference we assume that the systematic sample excluding case children is representative of children in Newcastle

at school entry. In order to assess the adequacy of the revised UK 1990 growth reference for Newcastle children at school entry, we need to test the null hypotheses that the mean Z-score is zero and the variance of the Z-scores is one for each age-sex grouping. There are very few individuals with grouped age of 6 years, so it is unlikely that the boys and girls within this sample are representative of Newcastle as a whole. We produced quantile-quantile plots of Z-scores for height, weight and BMI respectively, these indicate that the Z-scores at school entry are reasonably normal. A t-test was used to assess whether the mean Z-score was zero in each age-sex grouping, the resulting t-values can be found in table 8.4. There may be some reason to doubt the null hypothesis that the mean height Z-score is zero for 5 year old girls. It appears that this sample of 5 year old girls from Newcastle are generally shorter than those children that contribute to the reference. There is some reason to doubt the null hypothesis that the mean BMI Z-score is zero for 4 year old boys. This indicates that 4 year old boys may have a higher BMI than those children that contribute to the reference. However, these 4 year old boys represent the age group 4 to 4.5 years measured in winter. We should be aware that these differences are not viewed to be clinically significant; less than half a centile space difference.

The standard deviations of the Z-scores for each age-sex grouping can be found in tables 8.2 and 8.3. According to WHO guidelines (WHO 1995), if the standard deviation of the Z-scores are above 1.3, this suggests inaccurate data either due to measurement error or incorrect age reporting. None of the standard deviations

Table 8.4: **Systematic sample**: Results from tests that mean Z-score is zero

| Z-score | Age | sex | t | p | 95% CI |
|---|---|---|---|---|---|
| ZscoreH.ent | 4 | Boys | 0.2073 | 0.8365 | [-0.2000, 0.2462] |
| | 4 | Girls | -1.816 | 0.0741 | [-0.5265, 0.0251] |
| | 5 | Boys | -1.0678 | 0.288 | [-0.2858, 0.0857] |
| | 5 | Girls | -2.853 | 0.0052 | [-0.4615, -0.0831] |
| ZscoreW.ent | 4 | Boys | 1.6842 | 0.09782 | [-0.0414, 0.4774] |
| | 4 | Girls | -0.3567 | 0.7225 | [-0.2954, 0.2059] |
| | 5 | Boys | 0.4574 | 0.6483 | [-0.1518, 0.2429] |
| | 5 | Girls | -1.2578 | 0.2112 | [-0.3217, 0.0720] |
| ZscoreB.ent | 4 | Boys | 2.0498 | 0.04516 | [0.0066, 0.5801] |
| | 4 | Girls | 1.69 | 0.0959 | [-0.0379, 0.4538] |
| | 5 | Boys | 1.5055 | 0.1351 | [-0.0494, 0.3610] |
| | 5 | Girls | 0.589 | 0.5571 | [-0.1379, 0.2545] |

Table 8.5: **Systematic sample**: Results from testing that the variance of Z-scores is one

| Z-score | Age | sex | $\frac{(n-1)s^2}{\sigma^2}$ | approx. p-value | 95% CI |
|---|---|---|---|---|---|
| ZscoreH.ent | 4 | Boys | 38.17 | 0.1086 | [0.4933, 1.0487] |
| | 4 | Girls | 79.29 | 0.1765 | [0.9010, 1.8113] |
| | 5 | Boys | 101.46 | 0.7048 | [0.7378, 1.2640] |
| | 5 | Girls | 103.31 | 0.8535 | [0.7575, 1.3011] |
| ZscoreW.ent | 4 | Boys | 51.60 | 0.7461 | [0.6669, 1.4177] |
| | 4 | Girls | 65.50 | 0.8943 | [0.7443, 1.4963] |
| | 5 | Boys | 114.58 | 0.6043 | [0.8332, 1.4275] |
| | 5 | Girls | 111.82 | 0.6896 | [0.8199, 1.4082] |
| ZscoreB.ent | 4 | Boys | 63.07 | 0.4417 | [0.8150, 1.7327] |
| | 4 | Girls | 62.99 | 0.9289 | [0.7158, 1.4389] |
| | 5 | Boys | 123.79 | 0.2511 | [0.9002, 1.5422] |
| | 5 | Girls | 111.06 | 0.7284 | [0.8143, 1.3987] |

for the Z-scores for height, weight, or BMI are above 1.3, indicating that any of the measuring error inherent in school entry data discussed above, doesn't appear to inflate the standard deviations of the Z-scores. In table 8.5 we summarise the results of testing that the variance of the Z-scores are one for each age-sex grouping. It would appear that at school entry there is no reason to doubt that the variance of the Z-scores for weight, height and BMI is one.

In table 8.6, we regress the Z-scores for height, weight and BMI on age; if the age interval is accounted for then the slope of these regressions would be zero. Plots of the Z-scores for height, weight and BMI versus age are produced in figure 8.3, these are included here as a curvilinear relationship will not be detected by linear regression. The results from table 8.6 indicate that there is only reason to doubt the null hypothesis of zero slope for BMI indicating that there may be a negative linear trend with age for girls. However, if we consider the bottom right hand plot of figure 8.3 there are two outliers which may be influential, so the linear regression was repeated without these two individuals (see row labelled Z(BMI.2) in table 8.6). To conclude, there is no apparent trend with age for the Z-scores of height, weight and BMI indicating that the school entry data are appropriately adjusted for age by the UK 1990 growth reference.

Table 8.6: Slope coefficients from regression of Z-scores on age of school entry assessment (Z(BMI.2) represents the Z-scores for BMI after excluding two outliers)

| **Boys** | Estimate | Std. Error | t value | Pr(> \|t\|) |
|---|---|---|---|---|
| Z(height) | -0.0148 | 0.1465 | -0.101 | 0.92 |
| Z(weight) | -0.0026 | 0.1600 | -0.016 | 0.987 |
| Z(BMI) | 0.0168 | 0.1678 | 0.100 | 0.92 |
| **Girls** | Estimate | Std. Error | t value | Pr(> \|t\|) |
| Z(height) | -0.0365 | 0.1495 | -0.244 | 0.807 |
| Z(weight) | -0.2065 | 0.1471 | -1.403 | 0.162 |
| Z(BMI) | -0.3122 | 0.1446 | -2.159 | 0.0321 |
| Z(BMI.2) | -0.2269 | 0.1437 | -1.579 | 0.1160 |

### 8.4.3 Child's height at school entry conditional on mid-parental height Z-score

As discussed in Chapter 6, in order to assess whether a child is an appropriate height for their parents heights then we need to condition the child's current height Z-score on their mid-parental height Z-score. Recall:

$$Z_{h|mph} = \frac{Z_h - rZ_{mph}}{\sqrt{1-r^2}} \tag{6.1}$$

where $r$ is the correlation between $Z_h$, the child's height Z-score and $Z_{mph}$, their mid-parental height Z-score. Correlations between a child's height at school entry and their mid-parental height, along with the confidence interval for this correlation can be found in table 8.7. These correlations show a similar pattern to the correlations obtained with child's height at follow-up, that girls heights tend to be more correlated with the reported parental heights than boys.

A scatterplot of the child's height Z-score at school entry versus their mid-parental height Z-score can be found in the upper panel of figure 8.4. The outlying high mid-parental height corresponds to the child whose mother reported a height of 6'5". The other outlier within the scatterplot is for a girl with ID 2510, she has a high mid-parental height Z-score but a negative school entry height Z-score (this child's weight Z-score dropped to -3.14 at 18.6 months).

A correlation of 0.48 was used in equation (6.1) to obtain the Z-score for height conditional on mid-parental height. Table 8.8 contains the summary statistics of

Table 8.7: **Correlation between Z-score for height at school entry and parental height Z-scores**: correlation coefficients and approximate 95% CI's

| Z-score | group | correlation | n | 95% CI |
|---|---|---|---|---|
| mothers height | all | 0.3228 | 328 | [0.2223, 0.4165] |
| | boys | 0.3077 | 162 | [0.1611, 0.4409] |
| | girls | 0.3562 | 166 | [0.2155, 0.4823] |
| fathers height | all | 0.4222 | 321 | [0.3279, 0.5082] |
| | boys | 0.3746 | 156 | [0.2310, 0.5022] |
| | girls | 0.4617 | 165 | [0.3323, 0.5740] |
| mid-parent height | all | 0.4766 | 321 | [0.3874, 0.5570] |
| | boys | 0.4476 | 156 | [0.3125, 0.5650] |
| | girls | 0.5135 | 165 | [0.3914, 0.6178] |

the Z-score for height conditional on mid-parental height. The Z-score for height conditional on mid-parental height should have zero mean, variance of one and be uncorrelated with the mid-parental height Z-score. Table 8.9 summarises the results from testing that the mean and variance of the Z-scores for height conditional on mid-parental height are zero and one, respectively. There is no reason to doubt that the mean of the Z-score for height conditional on mid-parental height is zero for girls. However, there is some indication that the mean of the Z-score for height conditional on mid-parental height is non-zero for boys. There is no reason to doubt that the variance of the Z-score for height conditional on mid-parental height is one for both sexes. The correlation of the Z-score for height conditional on mid-parental height with mid-parental height Z-score is small ($r$=-0.0845). Therefore, we should be cautious in using conventional cut-offs when conditioning a boy's height on their mid-parental height.

Boxplots of the Z-score for height conditional on mid-parental height at school entry can be found in the lower panel of figure 8.4. The outlier corresponds to child with ID 2510 and this is the only child below the 0.4th centile.

### 8.4.4 Does canalization occur between school entry and follow-up?

Tanner (1989) used the term 'canalization' to describe the process by which a child's growth tends to return to its original path or channel if environmental circumstances,

Table 8.8: Summary statistics of Z-score for height at school entry conditional on mid-parental height

| Z(h.ent\|mph) | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| All | -3.6490 | -0.6062 | -0.0125 | 0.0423 | 0.6587 | 2.9130 | 1.0063 | 321 |
| Boys | -2.3930 | -0.5302 | 0.1655 | 0.1839 | 0.8399 | 2.9130 | 0.9858 | 156 |
| Girls | -3.6490 | -0.7400 | -0.1209 | -0.0916 | 0.5485 | 2.4320 | 1.0101 | 165 |

Table 8.9: **Systematic sample (excluding cases)**: Results of testing that (a) the mean Z-score for height conditional on mid-parental height is zero (by sex) and (b) the variance of the Z-score for height conditional on mid-parental height is one (by sex)

| | t | p | 95% CI |
|---|---|---|---|
| All | 0.753 | 0.452 | [-0.0682, 0.1528] |
| Boys | 2.3298 | 0.0211 | [0.0280, 0.3398] |
| Girls | -1.1646 | 0.2459 | [-0.2468, 0.0637] |
| | $\frac{(n-1)s^2}{\sigma^2}$ | approx p-value | 95% CI |
| All | 324.05 | 0.8729 | [0.8724, 1.1899] |
| Boys | 150.63 | 0.8041 | [0.7872, 1.2304] |
| Girls | 167.33 | 0.8541 | [0.8310, 1.2829] |

such as starvation, have pushed it off course. In order to crudely investigate whether attained growth status 'tracks' from school entry to 7-9 years, indicator vectors were created for the Z-scores for height, weight and BMI at school entry and follow-up. The current UK 1990 reference chart has major centiles spaced approximately two-thirds of a Z-score apart and current guidelines recommend referral if height is below the 0.4th (Z=-2.67) centile. The indicator vectors created contained whole numbers from 1:10, these are the corresponding codings and range of Z-scores they represent:

**1** $-\infty < Z \leq -2.67$ **6** $0 < Z \leq 0.67$
**2** $-2.67 < Z \leq -2$ **7** $0.67 < Z \leq 1.33$
**3** $-2 < Z \leq -1.33$ **8** $1.33 < Z \leq 2$
**4** $-1.33 < Z \leq -0.67$ **9** $2 < Z \leq 2.67$
**5** $-0.67 < Z \leq 0$ **10** $2.67 < Z < \infty$

Table 8.10 summarises the number of children between each of the major centiles

for height, height conditional on mid-parental height, weight and BMI at school entry and follow-up. It would appear that for height most children maintain their height between the same centiles or move to a neighbouring centile range. A similar pattern is observed for weight and BMI although there might be a slight tendency for children to move to a neighbouring centile range above rather than below. The majority of children's height Z-scores conditional on mid-parental height Z-score lie on the diagonal or just off the diagonal within table 8.10; this is illustrated in figure 8.6. At follow-up there are no children within the systematic sample with height conditional on parental height below the 0.4th centile. At school entry there is one child below the height conditional on mid-parental height 0.4th centile; ID 2510 discussed above.

The number of individuals in the systematic sample that fall below the 0.4th centile for height would provide some indication of the number of individuals likely to be referred after school entry check-up in the Newcastle area. At school entry there are three individuals within the systematic sample below the 0.4th centile for height (ID's 2287, 2464, 3317); this represents 0.8% of children (with heights at school entry) within the systematic sample excluding cases.

Table 8.10: Correspondence between height, height conditional on mid-parental height, weight and BMI Z-scores at school entry and follow-up (coding 1 corresponds to a Z-score in range $(-\infty,-2.67]$, 2 is (-2.67,-2], 3 is (-2,-1.33], 4 is (-1.33,-0.67], 5 is (-0.67,0], 6 is (0,0.67], 7 is (0.67,1.33], 8 is (1.33,2], 9 is (2,2.67] and 10 is $(2.67,\infty]$)

| **Height** | follow up | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| school entry | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 2 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 4 | 13 | 7 | 2 | 0 | 1 | 0 | 0 | 0 |
| 4 | 0 | 0 | 7 | 33 | 21 | 2 | 1 | 0 | 0 | 0 |
| 5 | 0 | 0 | 1 | 11 | 56 | 22 | 1 | 1 | 0 | 0 |
| 6 | 0 | 0 | 0 | 3 | 15 | 42 | 17 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 8 | 22 | 6 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 1 | 7 | 12 | 4 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| **h\|mph** | follow up | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| school entry | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 3 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 1 | 7 | 8 | 1 | 0 | 1 | 0 | 0 | 0 |
| 4 | 0 | 0 | 4 | 25 | 12 | 3 | 0 | 0 | 0 | 0 |
| 5 | 0 | 1 | 2 | 17 | 43 | 24 | 3 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 1 | 13 | 39 | 24 | 1 | 1 | 0 |
| 7 | 0 | 0 | 0 | 0 | 1 | 12 | 19 | 11 | 1 | 0 |
| 8 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 14 | 5 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 5 | 2 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

| **Weight** | follow up | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| school entry | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 3 | 2 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 2 | 5 | 4 | 3 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 13 | 27 | 14 | 7 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 1 | 17 | 37 | 14 | 2 | 2 | 0 | 0 |
| 6 | 0 | 1 | 0 | 3 | 26 | 32 | 16 | 8 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 3 | 27 | 19 | 7 | 2 | 2 |
| 8 | 0 | 0 | 0 | 0 | 0 | 2 | 7 | 11 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 1 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |

| **BMI** | follow up | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| school entry | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 2 | 6 | 8 | 4 | 2 | 1 | 0 | 0 | 0 |
| 4 | 0 | 1 | 6 | 20 | 13 | 7 | 1 | 0 | 0 | 0 |
| 5 | 0 | 0 | 5 | 20 | 28 | 12 | 5 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 7 | 31 | 28 | 12 | 5 | 4 | 0 |
| 7 | 0 | 0 | 0 | 0 | 11 | 25 | 16 | 9 | 5 | 0 |
| 8 | 0 | 0 | 0 | 1 | 2 | 5 | 9 | 9 | 0 | 1 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 2 | 1 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |

## 8.5 Preliminary analysis of school entry data for individuals within the case-control study

In this section we consider the height, weight and BMI of children from the case-control study. Table 8.12 contains the summary statistics for the height, weight and BMI Z-scores by sex for case and control children. It would appear that the distribution of the case Z-scores is shifted to the left of the control Z-scores for all anthropometric measures. Quantile-quantile plots of the Z-scores for height, weight and BMI by sex and case-control status were produced and these plots indicated that the Z-scores were reasonably normal. There are 33 (63.5%) case boys and 35 (64.8%) control boys with school entry assessment, note that the control anthropometric measures are over a slightly wider age range. There are 64 (63.5%) case girls and 57 (64.8%) control girls with school entry assessment, note that the case anthropometric measures are over a slightly wider age range. Table 8.11 contains a summary of the number of individuals in the age groupings, 4, 5 and 6 by sex and case-control status for height.

At school entry it is recommended that a child whose height is below the 0.4th centile should be referred. None of the control boys have a height below the 0.4th centile. However, one control girl (ID 639) has a height below the 0.4th centile at both school entry and follow-up. There are two case children with heights below the 0.4th centile at both school entry and follow-up; a boy with ID 975 and a girl with ID 3288. As described above, a correlation of 0.48 was used in equation (6.1) to obtain the Z-score for height conditional on mid-parental height for cases and controls. The summary statistics of the Z-score for height conditional on mid-parental height for cases and controls can also be found in table 8.12. Quantile-quantile plots of the Z-score for height conditional on mid-parental height, see figure 8.5, indicate

Table 8.11: Age distribution of cases' and controls' school entry height assessment by sex

| | Group | 4 years | 5 years | 6 years |
|---|---|---|---|---|
| Boys | cases | 9 | 19 | 4 |
| | controls | 9 | 23 | 3 |
| Girls | cases | 22 | 39 | 1 |
| | controls | 21 | 32 | 4 |

that there is no reason to doubt normality for case boys and girls, and control girls. However there is some evidence of skewness within the quantile-quantile plot for control boys.

One of the questions of interest is whether failure to thrive in infancy affects attained growth status at school entry. A t-test for two independent samples was carried out on the weight, height and BMI Z-scores from the case and control samples; the results of which can be found in table 8.13, these indicate that the case boys and girls are significantly shorter, lighter and leaner than the control boys and girls.

There was evidence of skewness in the quantile-quantile plot of the Z-score for height conditional on mid-parental height for control boys. Thus, the Wilcoxon Mann-Whitney test was used to compare case and control boys, this gave a W value of 184 (p=0.0002). As there was no reason to doubt the normality of the Z-score for height conditional on mid-parental height for girls a t-test was used, the results of which can be found in the bottom row of table 8.13. Therefore, even

Table 8.12: **Case-control study**: summary statistics of Z-scores for height, weight and BMI at school entry by sex and case-control status

| **Boys** | Group | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|---|
| Age.ent | cases | 4.12 | 4.48 | 4.90 | 4.90 | 5.27 | 5.76 | 0.48 | 33 |
| | controls | 4.19 | 4.50 | 4.85 | 4.87 | 5.13 | 6.38 | 0.51 | 35 |
| ZscoreH.ent | cases | -2.91 | -1.65 | -0.90 | -0.99 | -0.24 | 1.59 | 0.99 | 32 |
| | controls | -2.14 | -0.64 | 0.23 | 0.13 | 0.58 | 2.52 | 0.96 | 35 |
| ZscoreW.ent | cases | -5.25 | -1.73 | -1.07 | -1.19 | -0.54 | 0.64 | 1.07 | 33 |
| | controls | -1.75 | -0.38 | 0.61 | 0.27 | 0.84 | 1.67 | 0.84 | 35 |
| ZscoreB.ent | cases | -3.24 | -1.11 | -0.60 | -0.56 | 0.02 | 1.50 | 1.02 | 32 |
| | controls | -1.93 | -0.27 | 0.19 | 0.26 | 0.82 | 2.55 | 0.97 | 35 |
| Z(h.ent\|mph) | cases | -2.97 | -1.70 | -1.01 | -0.99 | -0.31 | 0.85 | 1.02 | 28 |
| | controls | -1.30 | -0.52 | -0.11 | 0.24 | 0.53 | 2.97 | 1.10 | 30 |
| **Girls** | Group | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
| Age.ent | cases | 3.59 | 4.45 | 4.65 | 4.76 | 5.05 | 7.34 | 0.51 | 64 |
| | controls | 4.09 | 4.39 | 4.65 | 4.73 | 5.00 | 5.97 | 0.43 | 57 |
| ZscoreH.ent | cases | -3.10 | -1.66 | -0.62 | -0.82 | -0.24 | 2.02 | 1.01 | 63 |
| | controls | -2.92 | -0.66 | -0.16 | -0.13 | 0.50 | 2.05 | 1.00 | 57 |
| ZscoreW.ent | cases | -3.61 | -1.65 | -1.11 | -1.15 | -0.54 | 1.02 | 1.04 | 64 |
| | controls | -2.01 | -0.60 | -0.01 | 0.05 | 0.47 | 2.86 | 0.93 | 57 |
| ZscoreB.ent | cases | -3.21 | -1.36 | -0.61 | -0.82 | -0.16 | 1.71 | 0.95 | 63 |
| | controls | -1.62 | -0.30 | 0.09 | 0.20 | 0.64 | 2.92 | 0.89 | 57 |
| Z(h.ent\|mph) | cases | -3.02 | -1.25 | -0.79 | -0.68 | -0.08 | 2.24 | 1.04 | 53 |
| | controls | -2.09 | -0.49 | -0.07 | 0.07 | 0.49 | 2.38 | 0.92 | 50 |

after conditioning on mid-parental height case children are significantly shorter than control children.

Figure 8.7 contains variable width notch boxplots for the case-control Z-scores of height, weight, BMI and height conditional on mid-parental height by sex. These boxplots indicate that anthropometric differences between case and controls are more marked for boys than for girls, although there are fewer boys identified by the 'thrive index' methodology as FTT in infancy (52 boys and 84 girls). In table 8.14 we summarise the results of testing that the median Z-score of weight, height and BMI is the same in case and control children. Recall that the same pattern was observed in boxplots that were produced for the follow-up study. In order to establish whether this sex difference existed towards the end of infancy, a box-plot was created for the last weight Z-score in infancy, see figure 8.8. At the end of

Table 8.13: School entry: results from tests that mean Z-score of cases equals mean Z-score of controls

| Z-score | sex | t | p | 95% CI |
|---|---|---|---|---|
| ZscoreH.ent | Boys | -4.6557 | $1.672 \times 10^{-5}$ | [-1.5867, -0.6339] |
| | Girls | -3.7792 | 0.00025 | [-1.0548, -0.3294] |
| ZscoreW.ent | Boys | -6.2338 | $4.819 \times 10^{-8}$ | [-1.9236, -0.9892] |
| | Girls | -6.6919 | $7.621 \times 10^{-10}$ | [-1.5536, -0.8441] |
| ZscoreB.ent | Boys | -3.3619 | 0.00131 | [-1.3047, -0.3320] |
| | Girls | -6.0342 | $1.897 \times 10^{-8}$ | [-1.3498, -0.6828] |
| Z(h.ent\|mph) | Girls | -3.9125 | 0.000166 | [-1.1363, -0.3717] |

Table 8.14: **Case-Control study**: Results of testing that the median Z-score for height, weight and BMI is the same in cases and controls (by sex)

| | sex | Case notches | Control notches | W | p |
|---|---|---|---|---|---|
| ZscoreH.ent | Boys | (-1.216, -0.584) | (-0.035, 0.490) | 219 | $8.573 \times 10^{-6}$ |
| | Girls | (-0.868, -0.381) | (-0.370, 0.042) | 1121 | 0.0003971 |
| ZscoreW.ent | Boys | (-1.329, -0.808) | (0.345, 0.869) | 144 | $1.087 \times 10^{-8}$ |
| | Girls | (-1.304, -0.915) | (-0.199, 0.182) | 708.5 | $7.05 \times 10^{-9}$ |
| ZscoreB.ent | Boys | (-0.850, -0.340) | (-0.048, 0.421) | 305 | 0.001147 |
| | Girls | (-0.815, -0.403) | (-0.081, 0.252) | 766 | $6.387 \times 10^{-8}$ |

infancy the median weight Z-score of the case boys and girls are similar. Similarly the median weight Z-score of control boys and girls are similar at the end of infancy. However, there appears to be less variability in the last weight Z-score in infancy for control boys.

Some of the outliers within the boxplots attend special schools. Some of the children attending special schools have 'Downs', these children will follow a different growth pattern to normal children. Children with Downs syndrome end up with shorter than average stature; this is a result of slow growth during foetal life and in the early years after birth (Cronk 1978). The 'thrive-index' methodology (Wright et al. 1994) probably does pick up children with 'Downs' because their growth will falter and these children would be considered as individuals with organic causes of failure-to-thrive. Table 8.15 provides information on the 9 individuals attending special schools, 5 of the children attending special schools are cases. Only ID's 1766 and 1813 were identified as individuals that had organic conditions that could affect growth (personal communication, Dr S.C. Corbett, October 1999).

## 8.6 Discussion and Conclusions

There are 807 children (1 child had died) in the follow-up study data base and 569 (70.5%) had a school entry assessment height and weight. This coverage is

Table 8.15: Characteristics of 9 children that attend special schools (where W is the number of weights in infancy; CACO is the case-control indicator; ZH.ent and ZH.fol are the Z-scores for height at school entry and follow-up, respectively; and ZW.ent and ZW.fol are the Z-scores for weight at school entry and follow-up, respectively)

| ID | SES | Sex | W | CACO | ZH.ent | ZW.ent | ZH.fol | ZW.fol |
|---|---|---|---|---|---|---|---|---|
| 85 | 3 | 2 | 7 | NA | -2.25 | -2.58 | -2.70 | -2.98 |
| 639 | 3 | 2 | 8 | 2 | -2.92 | -2.02 | -2.92 | -1.72 |
| 1089 | 2 | 2 | 8 | 1 | -1.00 | -1.86 | -0.53 | -0.53 |
| 1432 | 3 | 2 | 8 | 1 | NA | -2.57 | NA | NA |
| 1609 | 3 | 1 | 7 | 1 | NA | -5.26 | NA | NA |
| 1766 | 3 | 2 | 6 | 1 | -1.48 | -0.92 | -0.56 | -0.03 |
| 1813 | 3 | 2 | 5 | 1 | -1.99 | -1.59 | -3.78 | -0.7 |
| 1920 | 2 | 2 | 5 | NA | 1.69 | 1.23 | 1.35 | 0.81 |
| 2287 | 2 | 2 | 5 | NA | -2.94 | -2.74 | -3.12 | -0.8 |

in reasonable agreement with other studies (Laing and Rossor (1996), Hulse and Schilg (1995) in Schilg and Hulse (1997)). Out of the school records located only 68 (8.4%) had no school entry assessment. In general most individuals had a school entry assessment measure, with the exception of children attending private school. The few children attending special schools had their growth monitored on almost a yearly basis. In the east district area a 9-plus height and weight were also obtained. There are 2 sets of twins within the data set. A girl and boy, ID 1621 (systematic) and 1622 (control) respectively and girl twins, ID 2164 (systematic) and 2615 (not followed up). In retrospect, at the time of data collection the nurses that recorded the height and weight measures could have been noted; because each nurse will have a different observer error. Additional height and weight data (other than school entry) was sparse: only 113 children had additional heights or weights and this data has not been used. The school entry measures seem in reasonable agreement with follow-up measures; namely, children appear to be 'canalised' with a tendency to remain between the same major centiles or move to neighbouring centiles. The standard deviation of the Z-scores for height and weight are not above 1.3, indicating that the effect of measuring error is negligible (de Onis and Blössner 1997).

The revised UK 1990 reference is adequate for use on Newcastle anthropometric data at school entry as there is no apparent linear trend with age in the Z-scores for height, weight and BMI, indicating that the school entry data are appropriately adjusted for age. However, 5 year old girls from Newcastle may be shorter than those children that contribute to the reference. There is some indication that 4 year old boys may have a higher BMI than those children that contribute to the reference but this conclusion is reached on the basis of a 6-month age group of children measured in winter. It should be noted that these differences are not viewed to be clinically significant. Using current guidelines of referring children below the 0.4th centile for height would lead to a referral of 6 children (ID's 2287, 2464 and 3317 from the systematic sample, ID 639 from the controls, ID's 975 and 3288 from the cases). Cotterill et al. (1996) considered school entry heights of children from Hackney (inner-city London) and found that the proportion of children below the 0.4th and 3rd centiles were as expected using the UK 1990 reference. However, Cotterill et al. (1996) felt that the use of the 0.4th centile would miss a significant proportion of children with abnormality and proposed that the 2nd centile be used instead.

The case-control data was split by sex as there was some indication that Newcastle girls were shorter than those children that contribute to the UK 1990 reference. At school entry, the case children were shorter, lighter and leaner than control children. Even after conditioning on reported mid-parental height, the significant

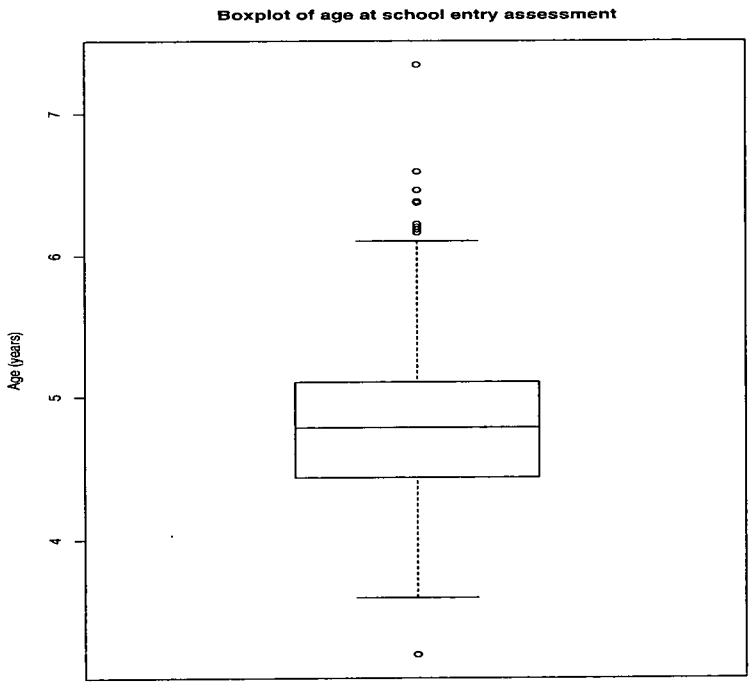
difference in heights persists. Variable width notch boxplots indicate that this attained growth difference may be more marked in boys. This observed difference may indicate that boys recover less well than girls after failing to thrive in infancy. Tanner (1989) observed that girls recovered more rapidly than boys after an insult to the growth process. Graham et al. (1982) (in Rudolf and Hochberg (1990)), in a study on infant malnutrition in Peru, found that the girls growth caught up with their sisters, whereas the boys lagged behind their brothers. Another explanation for this observed difference could be the excess of girls identified as failing to thrive, Skuse et al. (1994) (in Wright et al. (1996)) observed that failure-to-thrive has an even sex divide. However Rudolf and Hochberg (1990) found an excess of boys admitted with non-organic failure-to-thrive and suggested that boys may be more vulnerable to psycho-social deprivation. Boddy et al. (2000) considered the routine heights and weights at 6 years of 42 cases (matched with 42 controls) belonging to a 1-year birth cohort, from inner-city London, that had weights in infancy below the third centile for at least three months. Boddy et al. (2000) also noted that the physical stature, at 6 years, of the case children had been compromised but did not make gender comparisons.

Dowdney et al. (1987) used weight and height below the 10th centile according to the Tanner-Whitehouse standards (Tanner and Whitehouse 1976) at four years as part of their selection criteria for individuals that had failed to thrive in infancy, selecting individuals with persisting growth problems. The mean weight, height and BMI Z-scores of case children at school entry in Newcastle are not as low as those in the deprived inner city study (Dowdney et al. 1987; Dowdney et al. 1998). The Newcastle children are slightly older (mean age 4.81 years) than the London children (mean age 4.02 years) and cover a wider age range. The criteria employed by Dowdney et al. (1987) specified that both height and weight should be less than the tenth centile of the Tanner-Whitehouse reference at 4 years, but if we informally use the tenth centile of the revised UK 1990 reference at school entry, then only 27 case children (20 girls, 7 boys) have both heights and weights less than the tenth centile (Z-score $< -1.28$). In addition, Dowdney et al. (1987) required that the height at four years be less than the 10th centile when using "parent-allowed-for" charts (Tanner et al. 1970). If instead we use height conditional on mid-parental height Z-score, then the number of children below the tenth centile for height, weight and height conditional on mid-parental height is reduced to 16 children (4 boys, 12 girls). If children with possible organic causes are excluded, then this would leave 13 children (2 boys, 11 girls). Thus the Newcastle case children are not comparable with the cases selected in the inner city study, the cases in the inner-city study have

persistent growth faltering from infancy until school entry.

To conclude, the revised UK 1990 reference appears to be adequate for converting weights, heights and BMIs to Z-scores at school entry. The school entry measures seem to be in reasonable agreement with the follow-up measures, because children's anthropometric measures tend to remain between the same major centiles or move to neighbouring centiles. The use of the Z-score for height conditional on mid-parental height to assess a child's current height seems promising. However, in agreement with observation at 7-9 years, extreme values of this conditional Z-score should be interpreted cautiously for boys. Case children are significantly shorter, lighter and leaner than control children at school entry. Furthermore, it appears that the impact of growth faltering in infancy is more detrimental, in terms of attained growth status, for boys.

Figure 8.1: Upper panel Boxplot of age in years at school entry assessment
Lower panel Histogram of age in years at school entry assessment



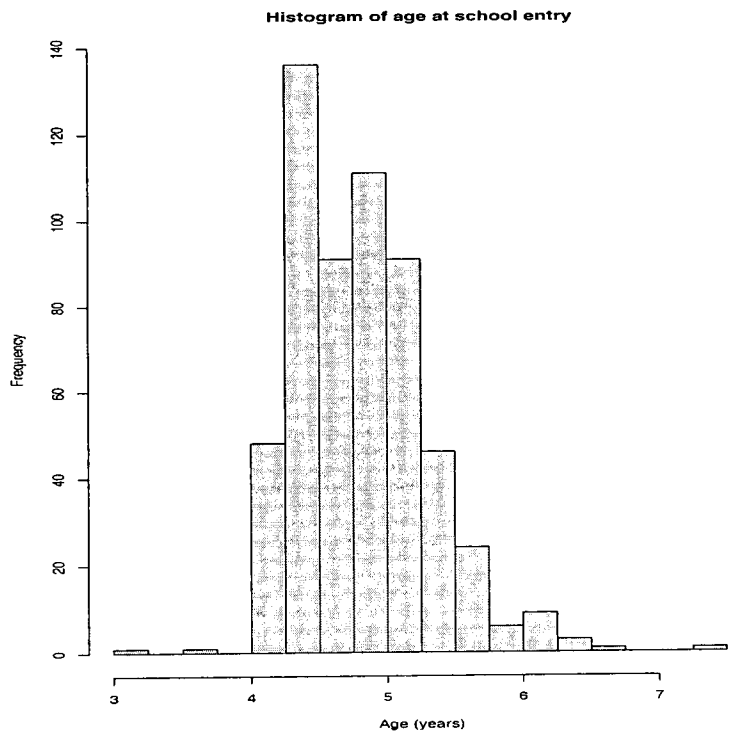



Figure 8.2: Variable width boxplots of Z-scores for height, weight and BMI at school entry assessment (grouped by age to nearest year and sex)

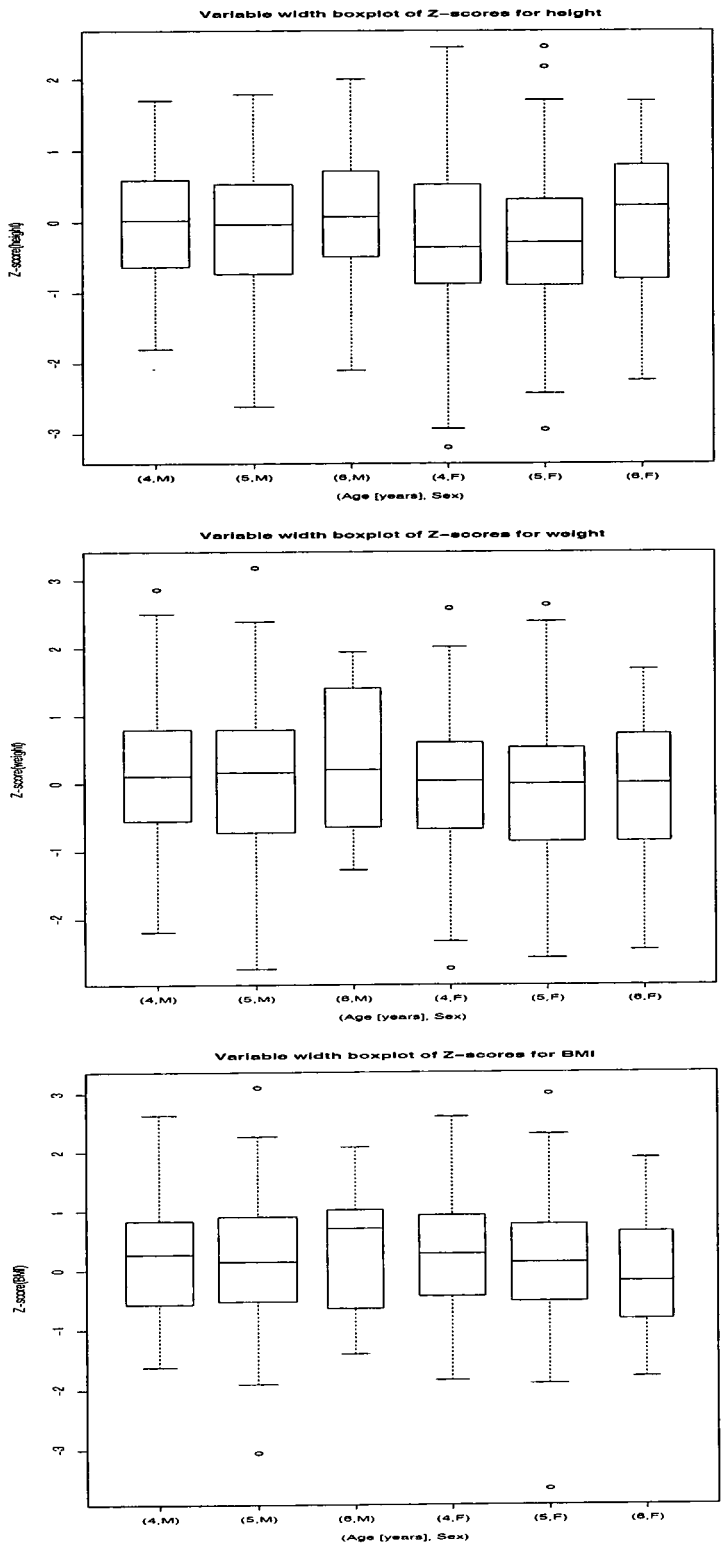


Figure 8.3: **Systematic sample (excluding cases)**: Plots of Z-scores for height, weight and BMI versus age at school entry assessment (Left Boys, Right Girls)

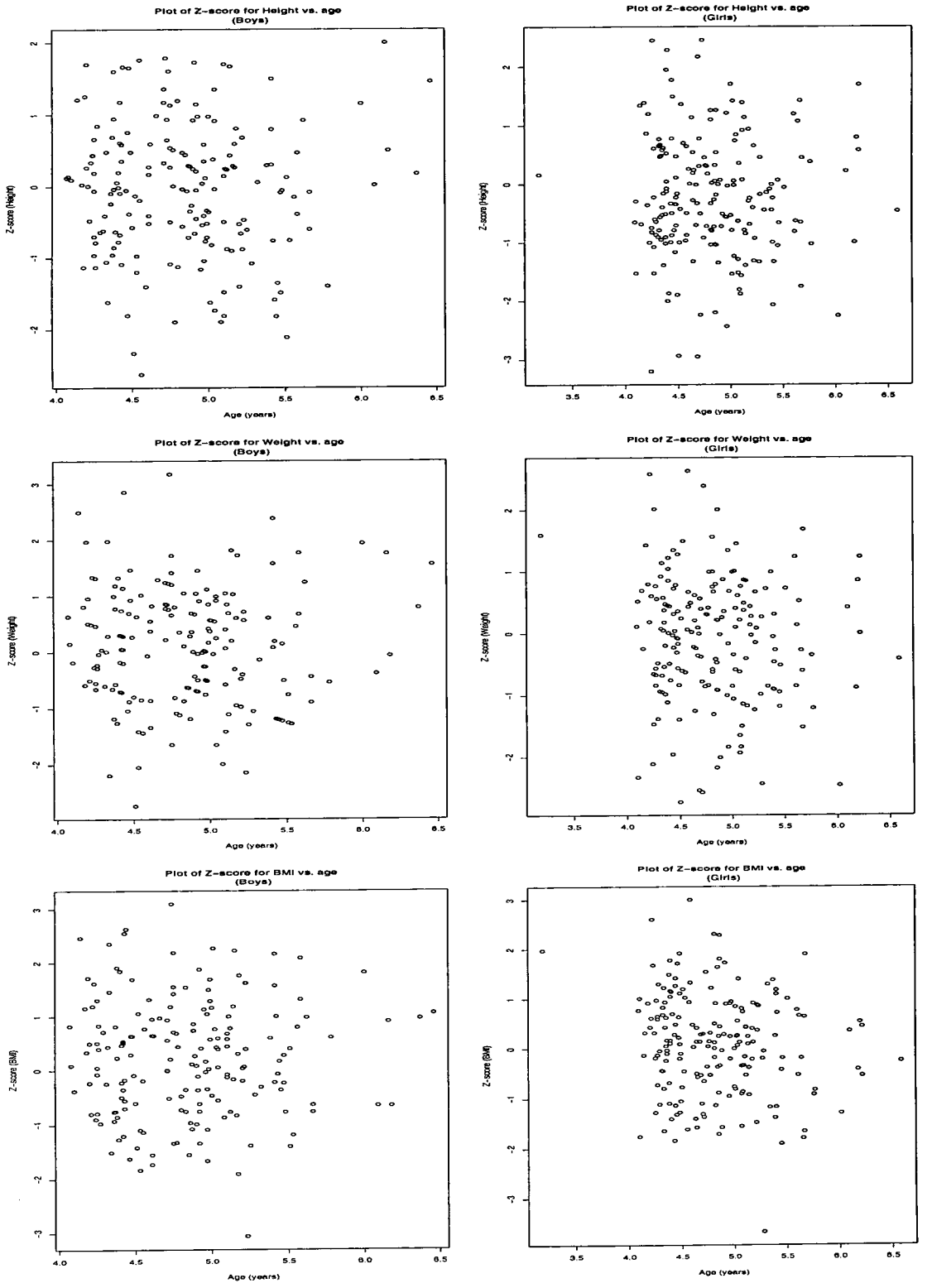


Figure 8.4: Upper panel: Plot of Z-score for height at school entry versus mid-parental height Z-score Lower panel: Boxplot of Z-score for height conditional on mid-parental height by sex

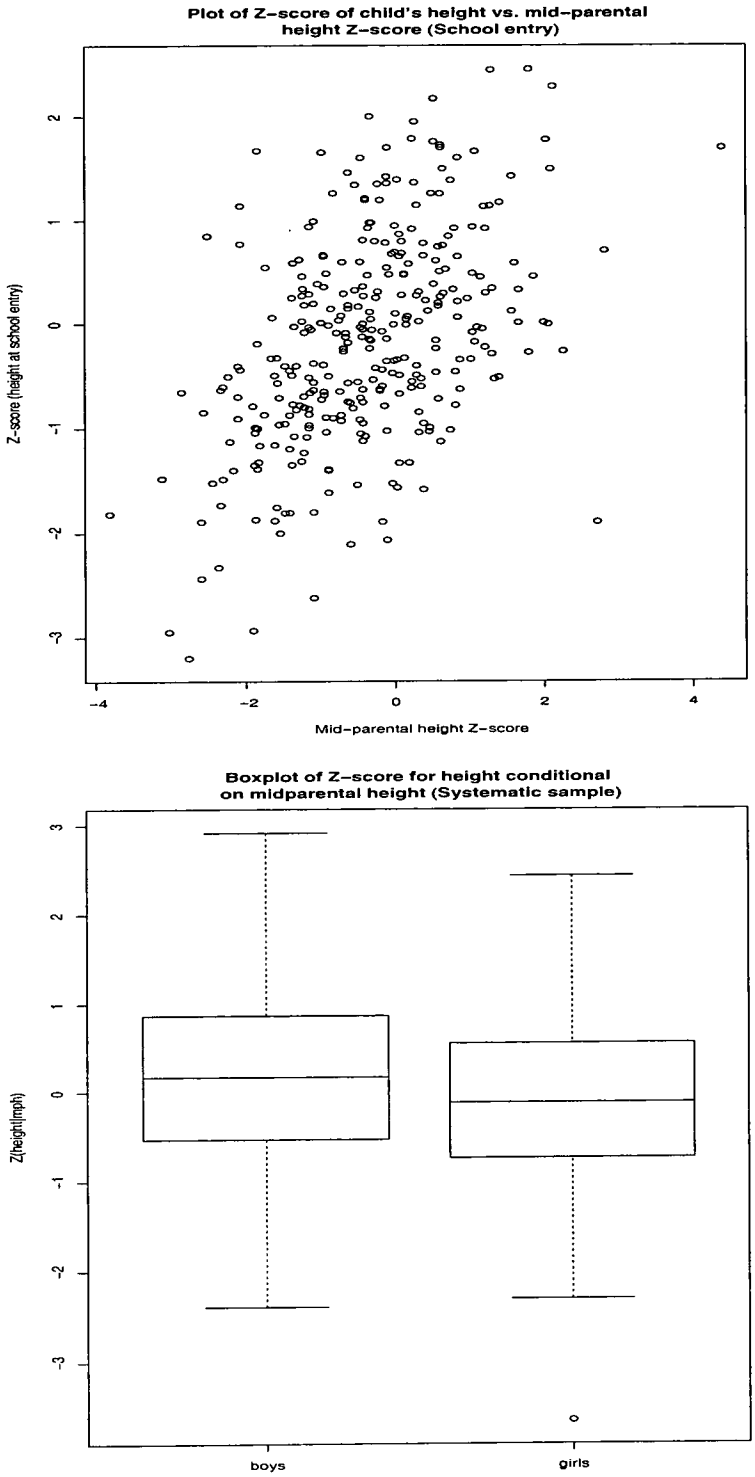


Figure 8.5: Case-control sample: Quantile-quantile plots of Z-score for height at school entry conditional on mid-parental height

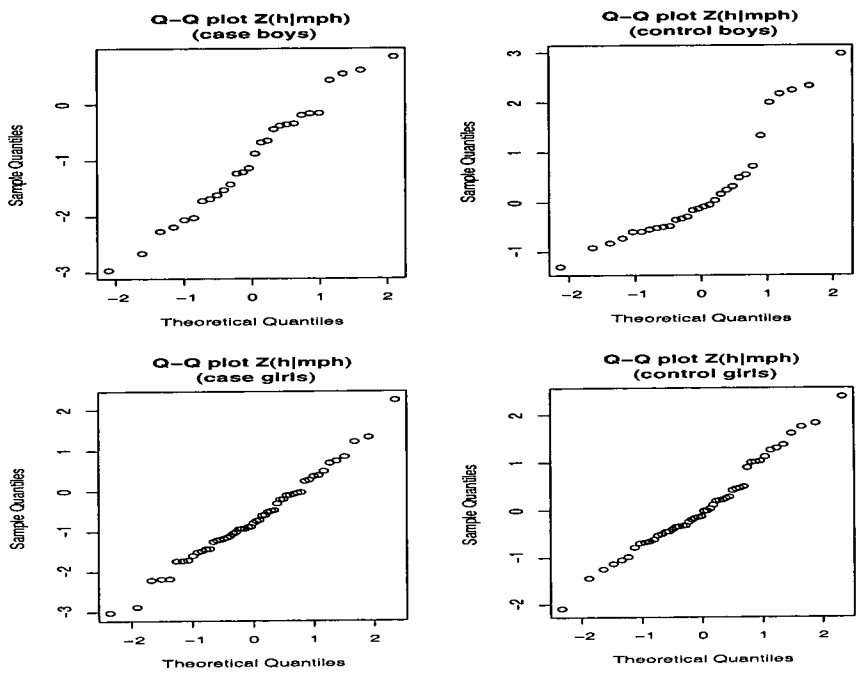


Figure 8.6: Plot of Z-score for child's height conditional on mid-parental height at follow-up versus school entry

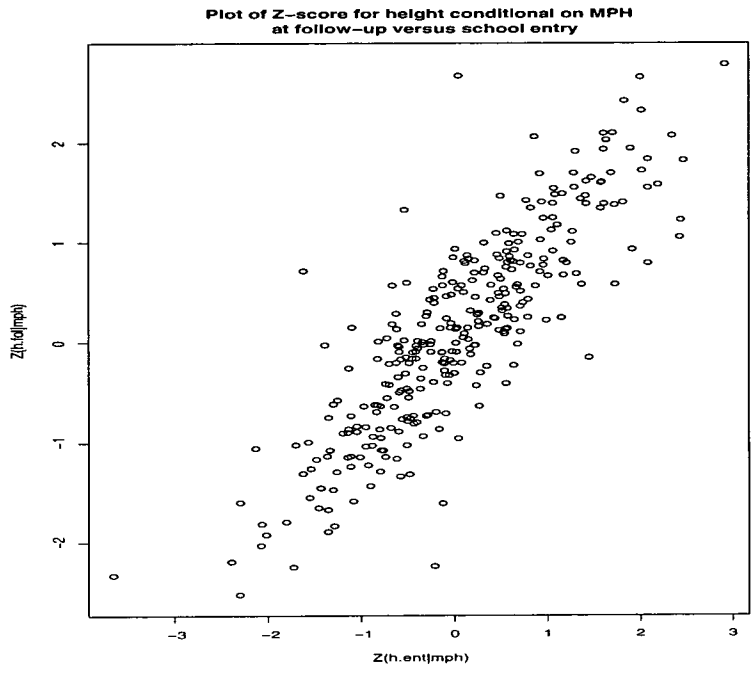


Figure 8.7: Case-control sample: Variable width notch boxplots of Z-scores for height, weight and BMI at school entry assessment (grouped by case-control status and sex)

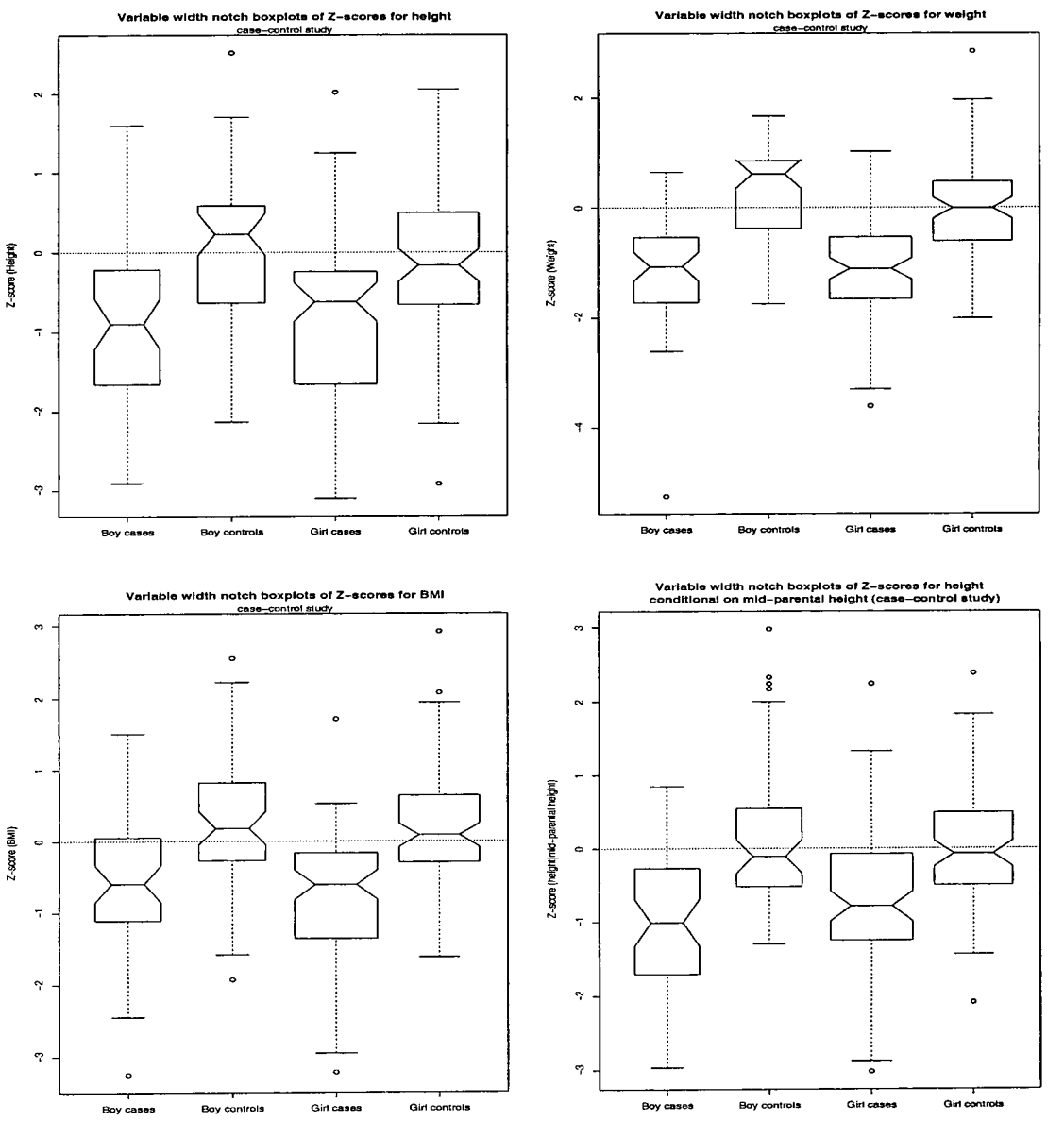
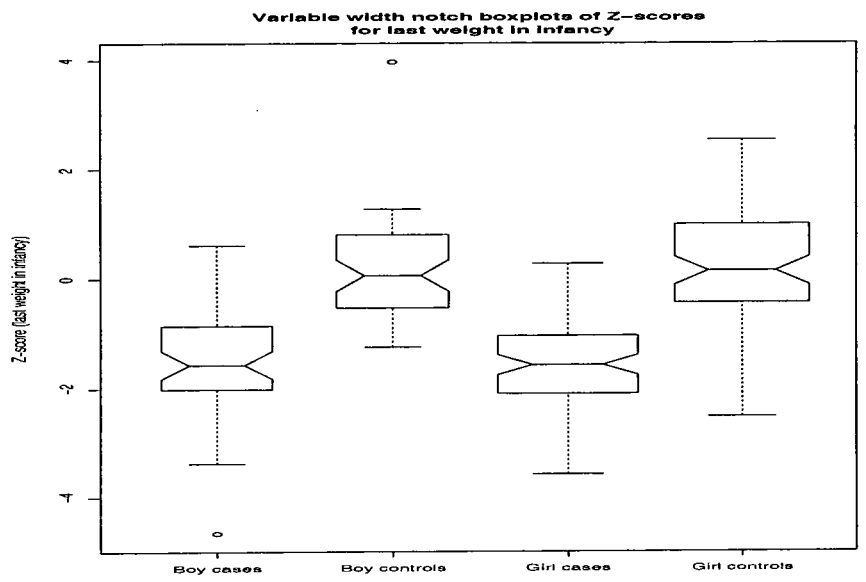


Figure 8.8: Variable width notch boxplots for last weight Z-score in infancy (grouped by case-control status and sex)



# Chapter 9

# Analysis of growth data beyond infancy

This chapter explores the relationships between data collected in the follow-up study (anthropometric data and systolic blood pressure) and earlier anthropometric data (school entry heights and weights and weights in infancy). In section 9.1 we briefly look at the variability of the three systolic blood pressures taken on each child, these measures were taken a matter of minutes apart. Concentrating then on the last systolic blood pressure, we investigate its relationship to current and earlier size.

In section 9.2 we explore the feasibility of obtaining conditional gain Z-scores. Height correlations from the French longitudinal growth study (Cole 1997a) are modelled as a function of the age at school entry and age at follow up. The model developed is then used to obtain conditional height gain Z-scores for the Newcastle data between school entry and follow-up. We then explore the feasibility of obtaining conditional weight gain Z-scores by conditioning on birth weight Z-score and weight Z-score at 1 year. Only conditioning on birth weight or weight at 1 year is considered because birth weight is noted routinely and weight at 1 year is the next most common. The conditional approach is also used to look at the relationship between weight and height at school entry and at follow-up. In section 9.3 lowess trend curves for weight Z-scores from birth to 7-9 years are produced for the cases, controls and systematic sample.

# 9.1 Analysis of blood pressure data

## 9.1.1 The fetal origins hypothesis

The 'fetal origins hypothesis' is a term coined by Barker et al. (1989);

> small body size or shape at birth (or subsequently) has been seen as a marker of poor fetal nutrition, which, it is suggested, results in fetal adaptations that programme future propensity to adult disease. (Lucas et al. 1999, pp245)

Barker (1992) (in Lucas et al. (1999)) have shown that small size at birth or in infancy is associated with an increased tendency to adverse outcomes in adulthood, which includes hypertension and death from ischaemic heart disease.

Barker et al. (1989) found that men who had the lowest weights at birth and at one year had the highest death rates from ischaemic heart disease. Postnatal factors add to the effects of low birth weight (Barker 1999). The highest prevalence of non-insulin dependent diabetes was found to be within individuals that had low birth weights but then became obese adults (Barker et al. (1991) in Barker (1999)). The risk of disease associated with size at birth is modified by growth during childhood; for example, men who were thin at birth but then went on to have accelerated weight gain during childhood were found to have the highest death rates from coronary heart disease (Eriksson et al. (1999) in Barker (1999)).

Many authors have explored the relationship between adult systolic or diastolic blood pressure with measures of current size and size at birth. A critique on the statistical methods used to explore the fetal origins of adult disease hypothesis was published by Lucas et al. (1999). The two main points were (Lucas et al. 1999):

- When size in early life (e.g. birth weight) was found to be related to later health outcomes, only after adjusting for current size (e.g. current weight or BMI), then it was probably the change in size between these two time-points rather than the fetal biology that was implicated.
- That even when early size is directly related to outcome, then we should explore whether this is partly or wholly explained by postnatal factors rather than prenatal factors. It was suggested that this could be achieved by introducing intermediate body sizes into the regression model.

Lucas et al. (1999) recommended fitting four models:

1. Early model: regression of later outcome, such as blood pressure, on early size.
2. Combined model: regression of later outcome on both early and later size.
3. Interaction model: Combined model with interaction of early and later size.
4. Late model: regression of later outcome on later size alone.

In the follow up study there are three systolic blood pressure (Systolic BP) measures for each individual. These measurements were taken in quick succession, with each measurement being a matter of minutes apart. At the time of data collection the child's blood pressure was taken three times, because children of this age are not used to having their blood pressure taken. Data analysis will concentrate on the last systolic blood pressure.

Regression will be used as a tool to explore the relationships between systolic blood pressure at 7-9 years and current weight, height, BMI and head circumference Z-scores. We hope to select the best regression equation. Stepwise regression is to be used to identify a potential model from the 4 variables. This is allowed to move in 'both directions', so if we start from the additive model with all the current Z-score variables in the model (ZscoreW.fol, ZscoreH.fol, ZscoreC.fol, ZscoreB.fol), we compare the magnitudes of the Akaike information criterion (AIC) statistic generated if we drop each of the individual variables one at a time or add an interaction of two of the variables one at a time. If the current model has the lowest AIC statistic then we stop. If not then the model with the variable dropped or added that has the lowest AIC is selected. This process is then repeated until we reach the 'best' regression model. Allowing the stepwise procedure to move in 'both directions' has advantages over the forward selection procedure, as a variable can be dropped again even after addition to the model (Draper and Smith 1998). This process was undertaken on the additive model with weight Z-scores at birth and 1 year, school entry height and weight Z-scores and follow-up height and weight Z-scores. In starting with this model we are hoping to determine which Z-scores are most important in determining systolic blood pressure at 7-9 years. Is the current weight Z-score and birth weight Z-score the most important or some other intermediate weight or height Z-score? A potential problem is that all the Z-scores in this initial model are highly correlated.

### 9.1.2 Variability of systolic blood pressure measures

At the time of collecting the data, Ms. J. Callum (research nurse that measured children) suspected that the initial observation may be slightly elevated in contrast to later recorded systolic blood pressures. To explore the issue of variability within the systolic blood pressure measures we concentrated initially on the recordings from the systematic sample excluding cases. Case children were excluded because their poor weight gain in infancy may exert some unknown influence on systolic BP.

A matrix of scatterplots for the systolic blood pressures suggested that the variance of the initial systolic blood pressure measure was greater than the variance of the final systolic blood pressure measure (F = 0.6882, p = 0.0036 for boys; F = 0.6669, p = 0.0019 for girls). In the systematic sample (excluding cases), 407 children have a systolic blood pressure on all three occasions. If for simplicity we just contrast the first and last systolic blood pressure: 256 children have a lower final systolic BP, 111 children have a higher final systolic BP and 49 have the same initial and final systolic BP. A box-plot of the difference in systolic blood pressure between initial and final measurement occasion (SYS3-SYS1) can be found in figure 9.1. The noticeable feature is that there are more outlying differences for boys than girls.

At the moment no allowance is being made for current size and we are just tentatively comparing the distributions of the systolic blood pressure by gender and case-control status. The upper panel in figure 9.2 contains a notch boxplot of the final systolic blood pressure (SYS3) by gender for the systematic sample. The notches on the boxplots overlap, indicating at the 5% level that there is no reason to doubt that the median systolic blood pressure for boys and girls is the same. The lower panel in figure 9.2 contains a box-plot of the systolic blood pressure (SYS3) by gender for individuals within the case-control study. The notches on the boxplots overlap for case and control girls; indicating that the median systolic BP's for these two groups are not significantly different at the 5% level. However the notches on the boxplots for the boys don't overlap ([88.5598, 95.4402] for control boys, [83.5898,88.4102] for case boys) indicating that the median systolic BP's are significantly different in the case and control boys (Mann Whitney test W = 584, p-value = 0.001340).

### 9.1.3 Relationship of systolic blood pressure to current size

If we consider the 420 children that have a final systolic blood pressure within the systematic sample (excluding cases), then the correlation between the systolic BP and the Z-scores of anthropometric measures at 7-9 years are 0.197, 0.244, 0.274 and 0.177, for height, BMI, weight and head circumference, respectively. Thus systolic BP is most correlated with current weight Z-score and least correlated with Z-score for head circumference. Figure 9.3 contains scatterplots of the systolic BP versus the Z-scores for height, weight, BMI and head circumference at follow-up. The discrete nature within these scatterplots is due to systolic blood pressure being recorded to the nearest mm Hg.

Systolic BP was regressed on all the anthropometric Z-scores at follow-up. Although it seems unlikely that head circumference would be an important predictor it was included here because Berkey et al. (1998) found that the systolic BP of females at age 30 was related to head circumference at birth. Using this regression model as a starting point, a stepwise procedure was then used. A 'both ways' procedure was used to allow the addition of an interaction or the dropping of a single variable. The final model arrived at using this procedure (see table 9.1) can be found in the upper table of table 9.2, this is essentially the same model derived by Wright et al. (1999). However the Z-scores for weight and BMI, and weight and height are highly correlated. Thus the initial starting model would be ill conditioned due to collinearity, leading in the second step of the stepwise procedure to the exclusion of the Z-score for weight at follow-up. The lower table in table 9.2 summarises the fit when we regress systolic BP on just the weight Z-score at follow up. The adjusted multiple correlation coefficient for both models in table 9.2 are of the same magnitude. Neither of the models are that predictive, only explaining about 7.3% of the variation in the systolic blood pressure data. Both these models essentially serve the purpose of providing a description of the relationship between systolic blood pressure and current weight Z-score or height and BMI Z-scores. Residual plots for both models given in table 9.2 were produced, both plots looked fairly similar and there was no evidence of heteroscedasticity.

### 9.1.4 Relationship of systolic blood pressure to current size and earlier size

The standard approach described in section 9.1.1 is to regress adult blood pressure on current size and birth weight. As discussed in this same section it could be that blood

Table 9.1: **Systematic sample (excluding cases)**: Sequence of steps taken in the reduction of the Akaike information criterion (AIC) in the 'both-ways' stepwise procedure. Starting model regressed systolic BP on Z-scores for weight, height, BMI and head circumference at follow-up

| | Step | Df | Deviance | Resid. Df | Resid. Dev | AIC |
|---|---|---|---|---|---|---|
| 1 | | | | 415 | 35355.5 | 1871.8 |
| 2 | drop ZscoreC.fol | 1 | 34.8 | 416 | 35390.2 | 1870.3 |
| 3 | drop ZscoreW.fol | 1 | 45.3 | 417 | 35435.5 | 1868.8 |

Table 9.2: **Systematic sample (excluding cases)**: Regression of systolic BP on: upper table Z-scores for BMI and height at follow-up lower table Z-score for weight at follow-up

Regression of SYS3 on ZscoreB.fol and ZscoreH.fol

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 87.8784 | 0.4540 | 193.562 | $< 2 \times 10^{-16}$ |
| ZscoreB.fol | 1.8400 | 0.4397 | 4.185 | $3.48 \times 10^{-5}$ |
| ZscoreH.fol | 1.4045 | 0.4937 | 2.845 | 0.00466 |

$R^2$=0.07754, $R^2$(adj)=0.07311, residual SE=9.218 on 417 df

Regression of SYS3 on ZscoreW.fol

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 87.8748 | 0.4504 | 195.100 | $< 2 \times 10^{-16}$ |
| ZscoreW.fol | 2.5452 | 0.4363 | 5.834 | $1.09 \times 10^{-8}$ |

$R^2$=0.07529, $R^2$(adj)=0.07308, residual SE=9.218 on 418 df

pressure is partly or wholly explained by postnatal factors, such as intermediate size variables or weight gain, rather than prenatal factors. In order to investigate whether birth weight Z-score, weight Z-score at 1 year, weight or height Z-scores at school entry were in anyway related to systolic blood pressure at 7-9 years a stepwise regression procedure was used. This provided an exploratory tool, starting with the regression of systolic blood pressure on weight Z-scores at birth, 1 year, school entry and at follow-up, and height Z-scores at school entry and at follow-up.

It wasn't feasible to consider all of the weight Z-scores at grouped age allocated by Dr C.M. Wright in infancy because only 122 children had weight Z-scores at all the grouped ages in infancy along with school entry and follow-up heights and weights.

The correlations between infancy weight Z-scores and systolic BP at 7-9 years can be found in table 9.3. The weight Z-score at 1 year is the most correlated with systolic BP at 7-9 years and birth weight Z-score is the least correlated. Therefore weight Z-score at 1 year was considered as a candidate because it was the most correlated with systolic BP, birth weight Z-score was included as an indicator of prenatal growth, intermediate heights and weights at school entry were also considered as it might be the weight gain in childhood that is influential on systolic blood pressure at 7-9 years. However there is likely to be a high degree of collinearity in this starting model as all the weight and height Z-scores are highly correlated. An alternative would be to regress on the first few principal components of the correlation matrix for all variables in the starting model, if we believed that all of these variables would be related to the systolic blood pressure at 7-9 years. This approach is considered in section 9.1.5.

A summary of the sequence of steps employed by the stepwise procedure can be found in table 9.4. A summary of the fit can be found in the upper table of table 9.5, neither of the coefficients for the weight Z-score at 1 year (Zscore6) or birth weight Z-score (Zscore1) are significantly different from zero at the conventional 5% level. The addition of the birth weight Z-score and Z-score at 1 year leads to a small improvement in the adjusted multiple correlation coefficient. It would seem that current weight Z-score is more important than birth weight Z-score in terms of predicting systolic BP at 7-9 years. Within this model the birth weight Z-score is contrasted with weight Z-score at 1 year. Therefore, the weight gain in infancy may also have some influence on systolic blood pressure at 7-9 years. In order to explore this issue further, the conditional weight gain Z-score between birth and 1 year was calculated for each individual in the systematic sample using equation (7.2) to obtain the correlation between weight Z-scores. The systolic blood pressure, was regressed on weight Z-score at follow-up and this conditional weight gain Z-score (see

Table 9.3: **Systematic sample excluding cases**: Correlation of weight Z-scores in infancy with systolic BP at 7-9 years (where N denotes the number of individuals contributing to each correlation coefficient)

| Z(weight): | birth | 6wks | 3mths | 6mths | 9mths | 12mths | 18mths |
|---|---|---|---|---|---|---|---|
| systolic BP | 0.0297 | 0.1116 | 0.1450 | 0.1677 | 0.1397 | 0.1766 | 0.0926 |
| N | 410 | 399 | 395 | 387 | 320 | 332 | 238 |

middle table in table 9.5), this led to a small reduction in the multiple correlation coefficient in contrast to the model selected by the step-wise procedure. Comparable adjustable multiple correlation coefficients are obtained if systolic blood pressure is regressed on birth weight Z-score, conditional weight gain Z-score and Z-score for weight at follow-up (see lower table in table 9.5).

Consider now, for comparative purposes, the models suggested by Lucas et al. (1999) discussed in section 9.1.1. The fit of the 'late model' can be found in table 9.2, whereas the fit of the 'early', 'combined' and 'interaction' models can be found in table 9.6. Regressing systolic BP on birth weight Z-score alone provides a very poor fit. The addition of weight Z-score at 7-9 years and the interaction Zscore1:ZscoreW.fol leads to an increase in the multiple correlation coefficient. Residual plots were produced for all models in tables 9.5 and 9.6 and there were no indication of heteroscedasticity.

### 9.1.5 Regression of systolic blood pressure on principal components

One approach to dealing with intercorrelated predictor variables is to regress on the principal components of these variables instead. The obvious advantage to regressing on the principal components is that they are uncorrelated, because they are orthogonal to each other by construction. Above we found that weight Z-scores in infancy and at follow-up were related to systolic blood pressure. Furthermore, the weight Z-score at school entry is highly correlated with the weight Z-score at

Table 9.4: **Systematic sample (excluding cases)**: Sequence of steps taken in the reduction of the Akaike information criterion (AIC) in the 'both-ways' stepwise procedure. Starting model: regressed systolic BP on Z-scores for weight and height at school entry and follow-up, weight Z-scores at birth and 1 year

| | Step | Df | Deviance | Resid. Df | Resid. Dev | AIC |
|---|---|---|---|---|---|---|
| 1 | | | | 242 | 21587.2 | 1125.1 |
| 2 | drop ZscoreW.ent | 1 | 2.0 | 243 | 21589.2 | 1123.2 |
| 3 | drop ZscoreH.fol | 1 | 6.5 | 244 | 21595.7 | 1121.2 |
| 4 | drop ZscoreH.ent | 1 | 3.6 | 245 | 21599.3 | 1119.3 |

Table 9.5: **Systematic sample (excluding cases)**: Upper table Regression of systolic BP on birth weight Z-score, weight Z-scores at 1 year and follow-up Lower tables Regression of systolic BP on weight Z-score at follow-up, birth weight Z-score and conditional weight gain Z-score between birth and 1 year (Z(1yr|birth))

Regression of SYS3 on ZscoreW.fol, Zscore1 and Zscore6

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 87.8447 | 0.5293 | 165.951 | $< 2 \times 10^{-16}$ |
| Zscore1 | -0.9495 | 0.5767 | -1.647 | 0.101 |
| Zscore6 | 0.7878 | 0.6479 | 1.216 | 0.225 |
| ZscoreW.fol | 2.6576 | 0.5954 | 4.464 | $1.12 \times 10^{-5}$ |

$R^2$=0.09526, $R^2$(adj)=0.08681, residual SE=9.002 on 321 df

Regression of SYS3 on ZscoreW.fol and Z(1yr|birth)

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 88.0396 | 0.5050 | 174.332 | $< 2 \times 10^{-16}$ |
| ZscoreW.fol | 2.4243 | 0.5652 | 4.289 | $2.37 \times 10^{-5}$ |
| Z(1yr\|birth) | 0.7444 | 0.5961 | 1.249 | 0.213 |

$R^2$=0.0911, $R^2$(adj)=0.08546, residual SE=9.009 on 322 df

Regression of SYS3 on ZscoreW.fol, Zscore1 and Z(1yr|birth)

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 87.8442 | 0.5293 | 165.956 | $< 2 \times 10^{-16}$ |
| ZscoreW.fol | 2.6551 | 0.5954 | 4.459 | $1.14 \times 10^{-5}$ |
| Zscore1 | -0.6389 | 0.5223 | -1.223 | 0.222 |
| Z(1yr\|birth) | 0.7293 | 0.5957 | 1.224 | 0.222 |

$R^2$=0.09532, $R^2$(adj)=0.08686, residual SE=9.002 on 321 df

follow-up. Therefore, it seem reasonable to take weight Z-scores at birth, 1 year, school entry and follow-up to provide some indication of size. Principal component analysis has already been discussed in detail in Chapter 5. The principal components for the correlation matrix in table 9.7 were obtained, these can be found in table 9.8. The first principal component is an average of all weight Z-scores, providing a measure of overall size and the second principal component contrasts weight Z-scores in childhood with weight Z-scores in infancy, reflecting some measure of weight gain. The third principal component contrasts weight Z-score at 1 year with weight Z-scores at birth and follow-up, and the fourth principal component contrasts school entry weight Z-score with weight Z-scores at 1 year and follow-up.

Draper and Smith (1981) suggested that some selection rule, such as the principal components that explain more than 75% of the variation in the data, could be used

Table 9.6: **Systematic sample (excluding cases)**: Upper table Regression of systolic BP on birth weight Z-score ('early model') middle table Regression of systolic BP on weight Z-score at follow-up and birth weight Z-score ('combined model') lower table Regression of systolic BP on weight Z-score at follow-up, birth weight Z-score and their interaction ('interaction model')

Regression of SYS3 on Zscore1

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 88.0010 | 0.4987 | 176.455 | $< 2 \times 10^{-16}$ |
| Zscore1 | 0.2844 | 0.4745 | 0.599 | 0.549 |

$R^2$=0.0008797, $R^2$(adj)=-0.001569, residual SE=9.589 on 408 df

Regression of SYS3 on Zscore1 and ZscoreW.fol

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 87.5771 | 0.4848 | 180.651 | $< 2 \times 10^{-16}$ |
| Zscore1 | -0.6971 | 0.4857 | -1.435 | 0.152 |
| ZscoreW.fol | 2.7866 | 0.4740 | 5.879 | $8.63 \times 10^{-9}$ |

$R^2$=0.07907, $R^2$(adj)=0.07455, residual SE=9.218 on 407 df

Regression of SYS3 on Zscore1*ZscoreW.fol

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 87.3934 | 0.5085 | 171.859 | $< 2 \times 10^{-16}$ |
| ZscoreW.fol | 3.0180 | 0.5121 | 5.893 | $7.98 \times 10^{-9}$ |
| Zscore1 | -0.7021 | 0.4855 | -1.446 | 0.149 |
| Zscore1:ZscoreW.fol | 0.5123 | 0.4305 | 1.190 | 0.235 |

$R^2$=0.08227, $R^2$(adj)=0.07549, residual SE=9.213 on 406 df

to arrive at principal components that could then be utilised in the least squares procedure. Therefore, it seems sensible to regress systolic BP on the first and second principal components. A summary of the fit for this model can be found in table 9.9, this actually explains less of the variation than regressing on current weight Z-score alone. However, both the coefficients of the first two principal components are significantly different from zero. When regressing systolic BP on the first two principal components we arrive at equation (9.1). Therefore, the following weight Z-scores (in order of importance) are related to systolic BP at 7-9 years: the current weight Z-score, weight Z-score at school entry, weight Z-score at birth and small contribution from weight Z-score at 1 year.

$$\begin{aligned} \text{SYS3} = 87.6008 - 1.0336\text{Zscore1} + 0.2197\text{Zscore6} + 1.2368\text{ZscoreW.ent} \\ + 1.5306\text{ZscoreW.fol} \end{aligned} \quad (9.1)$$

It would appear that regressing systolic blood pressure on first two principal components is not worthwhile as this leads to a poorer fitting model. To conclude, it is better to regress systolic blood pressure on current weight Z-score alone. If birth weight Z-score is also included in the regression there is a very slight improvement in fit and the coefficient of birth weight Z-score is negative. These observations are in reasonable agreement with other authors (Berkey et al. 1998; Barker et al. 1989; Law et al. 1993).

Table 9.7: **Systematic sample (excluding cases)**: Correlation matrix for weight Z-scores at birth, 1 year, school entry and follow-up (252 children - 119 boys & 133 girls)

| | Zscore1 | Zscore6 | ZscoreW.ent | ZscoreW.fol |
|---|---|---|---|---|
| Zscore1 | 1.0000 | 0.5472 | 0.4371 | 0.3516 |
| Zscore6 | | 1.0000 | 0.6884 | 0.5343 |
| ZscoreW.ent | | | 1.0000 | 0.7986 |
| ZscoreW.fol | | | | 1.0000 |

Table 9.8: Results from principal component analysis of correlation matrix for weight Z-scores at birth, 1 year, school entry and follow-up (prop. var. and cum. prop. denote 'Proportion of Variance' and 'Cumulative Proportion' respectively)

| | Comp.1 | Comp.2 | Comp.3 | Comp.4 |
|---|---|---|---|---|
| Zscore1 | 0.4130 | -0.7863 | -0.4587 | -0.0262 |
| Zscore6 | 0.5160 | -0.1942 | 0.7807 | 0.2942 |
| ZscoreW.ent | 0.5531 | 0.3143 | 0.0033 | -0.7715 |
| ZscoreW.fol | 0.5072 | 0.4951 | -0.4244 | 0.5635 |
| SD's | 1.6435 | 0.8596 | 0.6253 | 0.4113 |
| prop. var. | 0.6752 | 0.1847 | 0.0977 | 0.0423 |
| cum. prop. | 0.6752 | 0.8600 | 0.9577 | 1.0000 |

Table 9.9: **Systematic sample (excluding cases):** Regression of systolic BP on first and second principal components (where PC1=0.4130Zscore1+0.5160Zscore6+0.5531ZscoreW.ent+0.5072ZscoreW.fol and PC2=-0.7863Zscore1-0.1942Zscore6+0.3143ZscoreW.ent+0.4951ZscoreW.fol

| | Value | Std. Error | t-value | $Pr(>\|t\|)$ |
|---|---|---|---|---|
| (Intercept) | 87.6008 | 0.6241 | 140.352 | $< 2 \times 10^{-16}$ |
| PC1 | 1.1469 | 0.3602 | 3.184 | 0.00164 |
| PC2 | 1.9167 | 0.6749 | 2.840 | 0.00489 |
| $R^2$=0.06691, $R^2$(adj)=0.05932, residual SE=9.386 on 246 df | | | | |

## 9.2 Correlation of infancy weight Z-scores with childhood height and weight Z-scores

### 9.2.1 Growth monitoring

The Newcastle data set has weights in infancy, weights and heights at school entry and follow-up. In chapters 4, 6 and 8 we concentrated on analysing the data up to 18 months, at school entry and at follow-up, respectively. Here we attempt to say something about the longitudinal element of the Newcastle data.

Recall from chapter 3, that the raw change in height (weight) Z-score can not be used to assess growth from one occasion to the next because this does not take into account the correlation between the two height (weight) Z-scores. Equations (3.11) and (3.5) represent the unconditional and conditional weight gain Z-scores, respectively. Equation (3.11) takes into account regression to the mean. Equations (3.11) and (3.5) give similar results if $r$ is close to 1, however if $r$ is smaller then the conditional gain Z-score should be used in preference (Cole 1996). As the anthropometric measurements taken after infancy are several years apart then we will use the conditional gain Z-score approach.

$$Z_{(Z_2-Z_1)} = \frac{Z_2 - Z_1}{\sqrt{2(1-r)}} \tag{3.5}$$

$$Z_{2|1} = \frac{Z_2 - rZ_1}{\sqrt{1-r^2}} \tag{3.11}$$

where $r$ is the correlation between $Z_1$ and $Z_2$, this depends both on the child's age and the interval between measurements.

The only information required to use the conditional approach is a growth reference to convert the measurements to Z-scores and the correlation $r$ between the two measurements. The correlation will depend on the age and sex of the child. Although Cole (1993) states this may well be available in the literature, this is not the case for weight Z-scores beyond infancy. However, yearly correlations are available for height (Cole 1997a; Cole 1994a; Cameron 1980; Bailey 1994; Berkey et al. 1983a).

### 9.2.2 Correlation observed between infancy weights and later anthropometric measures

Tanner (1989) reported that the correlation coefficient between length at birth and adult height is only about 0.3. However, during the first year of life this coefficient rises sharply and by the age of 2 this has reached 0.80 (Tanner 1989). Weight is correlated with height, so we may expect weights in infancy to be reasonably correlated with later anthropometric measures. The correlations between weight Z-scores at birth and 1 year, and all follow-up anthropometric Z-scores are presented in table 9.10 for the systematic sample excluding cases. The weight Z-scores in infancy are reasonably correlated with weight and height Z-scores at 8 years, the correlation between weight Z-score at 1 year and later weight Z-score or height Z-score is about 0.54. The correlation between weight Z-scores in infancy and BMI Z-score is the lowest, possibly reflecting that high weights observed in infancy are not a strong predictor of later obesity.

Table 9.10: Correlation of weight Z-scores in infancy with Z scores for weight, height, BMI and head circumference in follow-up study (based on 328 individuals)

| | Z(1 year) | ZscoreW.fol | ZscoreH.fol | ZscoreB.fol | ZscoreC.fol |
|---|---|---|---|---|---|
| Z(birth) | 0.5257 | 0.3460 | 0.2759 | 0.2733 | 0.3543 |
| Z(1 year) | | 0.5449 | 0.5406 | 0.3727 | 0.5151 |
| ZscoreW.fol | | | 0.7071 | 0.8548 | 0.5301 |
| ZscoreH.fol | | | | 0.2468 | 0.4587 |
| ZscoreB.fol | | | | | 0.3988 |

Scatterplots of all pair-wise combinations of Z-scores for height, weight, BMI and head circumference at follow-up, for the systematic sample excluding cases, can be found in figure 9.4. There is some evidence of heteroscedasticity within the plot of weight Z-score versus BMI Z-score, children with negative BMI Z-scores have a more variable weight Z-score. The correlations for every pair-wise combination of Z-scores at follow-up can be found above the diagonal in table 9.11. The body mass index (BMI) is commonly used in adults for assessing adiposity. Weight and height are strongly correlated during childhood, so a child's weight centile tends to be strongly influenced by their height centile. In children the BMI changes substantially with age, rising steeply in infancy, then falling during the pre-school years and again rising until adulthood (Cole et al. 1995). The correlation between weight Z-score at 8 and BMI Z-score is high whereas correlation between height Z-score at 8 and BMI Z-score is low. We may have expected this as the BMI is supposed to reflect adiposity and it is desirable that an index of weight-for-height should be independent of height (Freeman et al. 1995). The correlations between the Z-scores of the anthropometric measures tend to be higher than the correlations observed between the raw data, with the exception of the correlation between weight and BMI. In agreement with Cole (1997a), the correlation between the Z-scores for height and weight at 7-9 years is about 0.7.

A similar matrix of scatterplots were produced for individuals within the case-control study, labelled by organic case, case and control; see figure 9.5. All these scatterplots, show two overlapping clusters, with cases tending to occupy the lower left hand corner and controls tending to occupy the upper right hand corner. Some of the children that have case status possibly due to organic causes can be found on the edges of the case cluster within these plots.

Table 9.11: Correlation between weight, height, BMI and headcircumference at 7-9 years (based on 426 individuals) Correlations between Z-scores are above diagonal and correlations between raw anthropometric data are below diagonal

| | weight | height | BMI | HC |
|---|---|---|---|---|
| weight | 1 | 0.7053 | 0.8696 | 0.5254 |
| height | 0.6610 | 1 | 0.2735 | 0.4754 |
| BMI | 0.8852 | 0.2424 | 1 | 0.3890 |
| HC | 0.5051 | 0.4876 | 0.3498 | 1 |

Table 9.12 contains the correlation matrix for weight Z-scores in infancy and weight and height Z-scores at school entry and follow-up. The weight (or height) Z-scores at school entry are over a 3 year age range. Similarly weight (or height) Z-scores at follow-up are over a 1.5 year age range. The correlation of later weight Z-scores with birth weight Z-score exhibits the expected feature, as the time between weight measures increases the correlation decreases. The correlations between weight Z-score at 1 year and weight and height Z-score at follow-up are very similar. A scatterplot matrix for all the Z-scores in table 9.12 can be found in figure 9.6. There is a strong linear relationship between the Z-scores for height at school entry and follow-up.

### 9.2.3 Conditional height gain Z-score

Deriving a model for the correlation structure of heights in Newcastle is not a viable option, owing to the small sample size. Cole (1997a) gave annual height correlations from the French longitudinal growth study, see table 9.13. These correlations were based on between 204 and 318 children, aged 2-9 years, with the sexes averaged. Height is normally distributed, so the correlation between heights is the same as the correlation between height Z-scores. It would be preferable to have correlations specific to each gender, however these are only available for consecutive 6-monthly or yearly correlations (Bailey 1994; Cole 1994a). A plot of the French height correlations versus later time and time difference can be found in figure 9.7. If height measures are made close in time then correlation is high. In the age group of interest, for the Newcastle height data, the correlation of height at 4 years with subsequent

Table 9.12: Correlation matrix for weight Z-scores at birth, 1 year, school entry and follow-up, and height Z-scores at school entry and follow-up (based on 252 individuals in systematic sample excluding cases)

| | Z(1 year) | ZscoreW.ent | ZscoreH.ent | ZscoreW.fol | ZscoreH.fol |
|---|---|---|---|---|---|
| Z(birth) | 0.547 | 0.437 | 0.326 | 0.352 | 0.291 |
| Z(1 year) | | 0.688 | 0.560 | 0.534 | 0.532 |
| ZscoreW.ent | | | 0.730 | 0.799 | 0.697 |
| ZscoreH.ent | | | | 0.597 | 0.880 |
| ZscoreW.fol | | | | | 0.719 |

heights is above 0.9.

> The timing of measurements was sufficiently close to the nominal ages for adjustment to be unnecessary. (Cole 1994a, pp2479)

In order to obtain a conditional height gain Z-score between school entry and the follow-up study for a child, we would need to interpolate between the correlations in table 9.13, to obtain the 'exact' correlation for a child. However we don't know the sample size for each of the correlations within table 9.13, but we do know that the sample size was greater for height measurements made a year apart than for those height measures made several years apart (personal communication, Prof. T.J. Cole). Using a similar approach to chapter 5, we will consider plausible models for the correlation structure in table 9.13. All the correlations in table 9.13 are near one, so any model would predict a correlation close to one. Therefore expansions for Fishers transformation of correlation may converge slowly as the correlation is above 0.9 (Ghosh 1966). The height correlation is highly correlated with initial time ($r = 0.772$) and time elapsed ($r = -0.851$) but has negligible correlation with later time point ($r = -0.079$). The response variable could be correlation, log-correlation or Fishers transformed correlation. In table 9.14 we present the correlations between transformations of the French height correlations and functions of the initial and later ages in years ($t1$ and $t2$, respectively). As in infancy all functions of $t1$ are highly correlated with transformations of the height correlation. The highest observed correlation is between Fishers transformed correlation and $\sqrt{t1}$. The highest correlation, involving $t2$ is also the square root function, namely $\sqrt{t2 - t1}$. Exploratory analysis revealed that the height correlations could be modelled well by

Table 9.13: Correlation matrix of height for 204 to 318 children belonging to the French longitudinal study (Age in years)

| Age | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| 2 | 1 | | | | | | | |
| 3 | 0.941 | 1 | | | | | | |
| 4 | 0.888 | 0.958 | 1 | | | | | |
| 5 | 0.860 | 0.934 | 0.975 | 1 | | | | |
| 6 | 0.835 | 0.912 | 0.958 | 0.981 | 1 | | | |
| 7 | 0.824 | 0.893 | 0.944 | 0.971 | 0.985 | 1 | | |
| 8 | 0.800 | 0.875 | 0.926 | 0.957 | 0.970 | 0.986 | 1 | |
| 9 | 0.772 | 0.857 | 0.907 | 0.939 | 0.954 | 0.975 | 0.988 | 1 |

several two variable models; all giving a multiple R-squared of around 95-97%.

As there was little difference in the predictive capability of several models, the model for correlation developed in chapter 5, was applied to the French height correlations. The fit of this model using unweighted least squares can be found in the upper table of table 9.15. The sample size for each of the French height correlations is unknown, so the assumption was made that the sample size for each of the correlations was 'similar' and only the magnitude of the correlation was taken account of in the weighting. The fit of the Argyle model using weighted least squares can be found in the lower table of table 9.15, there is a slight reduction in the multiple correlation coefficient. It is interesting that the same simple model, the Argyle model, is suitable for modelling the correlation between weight Z-scores (0 to 2 years) and the correlation between heights (2 to 9 years).

Table 9.14: Correlation between transformations of French height correlations and potential predictor variables in $t1$ and $t2$

| variable | $r$ | $\log(r)$ | Fishers$(r)$ |
|---|---|---|---|
| $t1$ | 0.772 | 0.761 | 0.859 |
| $\log(t1)$ | 0.818 | 0.810 | 0.859 |
| $\sqrt{t1}$ | 0.799 | 0.789 | 0.864 |
| $1/t1$ | -0.831 | -0.825 | -0.829 |
| $\exp(-t1)$ | -0.814 | -0.813 | -0.768 |
| $\exp(t1)$ | 0.422 | 0.410 | 0.597 |
| $t2$ | -0.079 | -0.088 | 0.027 |
| $\log(t2)$ | -0.079 | -0.088 | 0.031 |
| $\sqrt{t2}$ | -0.079 | -0.088 | 0.030 |
| $1/t2$ | 0.079 | 0.088 | -0.032 |
| $\exp(-t2)$ | 0.073 | 0.080 | -0.026 |
| $\exp(t2)$ | -0.076 | -0.084 | -0.001 |
| $t2 - t1$ | -0.851 | -0.849 | -0.832 |
| $\log(t2 - t1)$ | -0.791 | -0.784 | -0.839 |
| $\sqrt{t2 - t1}$ | -0.829 | -0.824 | -0.843 |
| $1/(t2 - t1)$ | 0.696 | 0.685 | 0.793 |
| $\exp(t1 - t2)$ | 0.666 | -0.653 | 0.776 |
| $\exp(t2 - t1)$ | -0.715 | -0.732 | -0.553 |
| $(t1 + t2)/2$ | 0.400 | 0.388 | 0.512 |
| $\log[(t1 + t2)/2]$ | 0.341 | 0.329 | 0.451 |
| $\sqrt{(t1 + t2)/2}$ | 0.373 | 0.361 | 0.484 |
| $2/(t1 + t2)$ | -0.267 | -0.255 | -0.375 |
| $\exp[-(t1 + t2)/2]$ | -0.113 | -0.103 | -0.213 |
| $\exp[(t1 + t2)/2]$ | 0.427 | 0.416 | 0.563 |

Table 9.15: **French height correlations [N=28]** (204 to 318 children belonging to the French longitudinal study): Regression of log correlation on $\log(t1)$ and $\log(t2)$

| unweighted | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.018621 | 0.015311 | -1.216 | 0.235 |
| $\log(t1)$ | 0.171941 | 0.005927 | 29.012 | $< 2\times10^{-16}$ |
| $\log(t2)$ | -0.151281 | 0.009107 | -16.612 | $5.11 \times10^{-15}$ |
| $R^2$=0.9714, $R^2$(adj)=0.9691, residual SE=0.01198 on 25 df | | | | |

| weighted | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.012098 | 0.013678 | -0.884 | 0.385 |
| $\log(t1)$ | 0.151311 | 0.006468 | 23.393 | $< 2\times10^{-16}$ |
| $\log(t2)$ | -0.139759 | 0.009697 | -14.412 | $1.29 \times10^{-13}$ |
| $R^2$=0.9571, $R^2$(adj)=0.9537, residual SE=0.03213 on 25 df | | | | |

In order to see if we could improve the fit of the Argyle model the impact of adding a constant to age was explored. Initially two weeks were added to the age in years to see if the model developed for weight in infancy was of the same format for height correlations in childhood. The fit of this model can be found in the upper table of table 9.16, this model provides a similar fit to the model with no constant. Using a similar approach to chapter 5, we then explored the effect of varying the constant added to the time point on the fit of the model. Adding a constant to the time points led to an increase in the deviance, whereas subtracting a constant led to a decrease in the deviance. Exploratory work suggested that the minimum deviance was achieved when 70 weeks was subtracted from the time points (see figure 9.8), the fit of this model can be found in the lower table of table 9.16. A plot of standardised residuals versus fitted values can be found on the left in figure 9.9, this suggests that there may be slightly more variability for correlations close to 1. A plot of Cook's distance versus index can be found on the right in figure 9.9, this suggests that there are three influential points: the most extreme corresponds to the correlation between heights at 2 and 3 years which is underestimated; the correlation between heights at 3 and 4 years which is overestimated and the correlation between heights at 2 and 9 years which is also overestimated. The impact of excluding these influential observations is considered in table 9.17. So the final model for the correlation between height Z-scores is given by equation (9.2).

$$\log(r) = 0.0158 + 0.0929\log(t1 - 70/52) - 0.0997\log(t2 - 70/52) \qquad (9.2)$$

where $t1$ and $t2$ are the initial age and final age in years, respectively and $t1 < t2$.

Table 9.16: **French height correlations [N=28]** (204 to 318 children belonging to the French longitudinal study): Regression of log correlation on $\log(t1 + c)$ and $\log(t2 + c)$

$\log(r(t1, t2)) = a\log(t1 + 2/52) + b\log(t2 + 2/52) + c + \text{error}$

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.020158 | 0.015620 | -1.29 | 0.209 |
| $\log(t1 + 2/52)$ | 0.173818 | 0.006057 | 28.70 | $< 2\times10^{-16}$ |
| $\log(t2 + 2/52)$ | -0.152322 | 0.009267 | -16.44 | $6.44 \times10^{-15}$ |

$R^2$=0.9562, $R^2$(adj)=0.9527, residual SE=0.03246 on 25 df

$\log(r(t1, t2)) = a\log(t1 - 70/52) + b\log(t2 - 70/52) + c + \text{error}$

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.020524 | 0.005621 | 3.651 | 0.00121 |
| $\log(t1 - 70/52)$ | 0.094375 | 0.002282 | 41.358 | $< 2\times10^{-16}$ |
| $\log(t2 - 70/52)$ | -0.103543 | 0.004145 | -24.978 | $< 2\times10^{-16}$ |

$R^2$=0.9859, $R^2$(adj)=0.9847, residual SE=0.01845 on 25 df

Table 9.17: Fit of Argyle model to height correlations after excluding influential observations

Excluding correlation between heights at (2,3) years

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.013159 | 0.006181 | 2.129 | 0.0437 |
| $\log(t1 - 70/52)$ | 0.094567 | 0.002122 | 44.566 | $< 2\times10^{-16}$ |
| $\log(t2 - 70/52)$ | -0.099628 | 0.004234 | -23.532 | $< 2\times10^{-16}$ |

$R^2$=0.9882, $R^2$(adj)=0.9872, residual SE=0.01714 on 24 df

Excluding correlation between heights at (2,3) and (3,4) years

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.019201 | 0.007121 | 2.696 | 0.0129 |
| $\log(t1 - 70/52)$ | 0.094896 | 0.002070 | 45.844 | $< 2\times10^{-16}$ |
| $\log(t2 - 70/52)$ | -0.103131 | 0.004672 | -22.073 | $< 2\times10^{-16}$ |

$R^2$=0.9894, $R^2$(adj)=0.9885, residual SE=0.01663 on 23 df

Excluding correlation between heights at (2,3), (3,4) and (2,9) years

| | Value | Std. Error | t-value | Pr(> \|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.015821 | 0.006340 | 2.496 | 0.0206 |
| $\log(t1 - 70/52)$ | 0.092932 | 0.001938 | 47.961 | $< 2\times10^{-16}$ |
| $\log(t2 - 70/52)$ | -0.099692 | 0.004261 | -23.395 | $< 2\times10^{-16}$ |

$R^2$=0.9908, $R^2$(adj)=0.99, residual SE=0.01455 on 22 df

The model developed for the French height correlations, given by equation (9.2), was used to calculate a conditional height gain Z-score between school entry and follow-up for each child within the systematic sample (excluding cases). Boxplots of the conditional height gain Z-scores by gender can be found in figure 9.10; both these boxplots are reasonably symmetric with four and three outlying points for the boys and girls, respectively. There are two individuals with conditional height gain Z-scores above 6: ID 2210 has a conditional height gain Z-score of 6.90 (he has a height Z-score of -0.52 and 1.84 at school entry and follow up, respectively) and ID 3489 has a conditional height gain Z-score of 6.39 (she has a height Z-score of -1.33 and 0.72 at school entry and follow up, respectively). It is possible that the height measurements for these two individuals at school entry are subject to measurement or recording errors. The presence of these two individuals will elevate both the mean and standard deviation of the conditional height gain Z-scores.

In figure 9.11 we plot the conditional height gain Z-score versus: age at school entry, age at follow up and time elapsed. There is no strong trend in any of these scatterplots suggesting that the conditional height gain Z-score appears to be uncorrelated with initial age, final age and time elapsed. The conditional height gain Z-score should be uncorrelated with the initial height Z-score. A plot of the conditional height gain Z-score versus initial height Z-score can also be found in figure 9.11. There is no apparent linear trend in this scatterplot and the correlation between height Z-scores at school entry and conditional height gain Z-scores was found to be small ($r$=-0.10).

The summary statistics of the conditional height gain Z-scores by gender, initial age group and time elapsed can be found in table 9.18. The age at school entry and time elapsed was rounded to the nearest year to make groups of a viable size for obtaining summary statistics. It would appear that the mean of the conditional height gain Z-scores is significantly different from zero for the girls. However there is no reason to doubt that the mean of the conditional height gain Z-scores is zero for the boys (see table 9.19). The variance of the conditional height gain Z-scores are significantly different from 1 (see table 9.19). However, we should be aware that this test is sensitive to non-normality and quantile-quantile plots (see figure 9.13) indicate that the distribution of conditional height gain Z-scores has heavy tails. To conclude, we should be wary when interpreting extreme conditional height gain Z-scores because we will have more extreme gains than expected.

Ideally we would like to look at the distributional properties of the conditional height gain Z-scores grouped by age and time elapsed. Figure 9.12 summarises the distribution of the measurement ages at school entry and follow-up. The time

elapsed between height measures can vary between 1.73 and 5.98 years. However, from figure 9.12, there appears to be two clusters of data points: age at school entry of 4 years and time elapsed of 4-5 years, and age at school entry of 5 years and time elapsed of 3-4 years. The summary statistics for these two groups of children are considered in table 9.20. There is no reason to doubt that the mean of the conditional height gain Z-scores is zero for these two clusters of children ($t = 0.5987$, $p = 0.5506$ when age at school entry is 4 years and time elapsed is 4 to 5 years; $t = 0.5987$, $p = 0.5506$ when age at school entry is 5 years and time elapsed is 3 to 4 years). Again there is reason to suggest that the variance of the conditional height gain Z-scores is significantly different from 1 ($S = 155.53$, $p \approx 0.0012$ when age at school entry is 4 years and time elapsed is 4 to 5 years; $S = 323.30$, $p \approx 0$ when age at school entry is 5 years and time elapsed is 3 to 4 years).

Table 9.21 contains the summary statistics of the conditional height gain Z-scores for children in the case and control samples by gender. The noticeable feature is that the variance of the conditional height gain Z-scores of case girls is higher than the control girls ($F = 1.8077$, $p = 0.02799$). However there is no reason to doubt that variance of the conditional height gain Z-scores for case and control boys are the same ($F = 1.0617$, $p = 0.864$). There is no reason to doubt that the mean conditional height gain Z-scores for case and control children are the same ($t = -1.3604$, $p = 0.1791$ for boys; $t = 1.0554$, $p = 0.2936$ for girls). Thus indicating that there is no difference in the rate of height gain between school entry and follow-up

Table 9.18: Systematic sample: Summary statistics of conditional height gain Z-scores by gender, age-group at school entry and time elapsed between height measures

| | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| All | -4.604 | -0.641 | 0.136 | 0.182 | 0.982 | 6.899 | 1.347 | 334 |
| Boys | -4.604 | -0.728 | 0.133 | 0.097 | 0.932 | 6.899 | 1.358 | 161 |
| Girls | -3.314 | -0.590 | 0.159 | 0.262 | 1.086 | 6.393 | 1.336 | 173 |
| Initial age | | | | | | | | |
| 4 years | -4.604 | -0.538 | 0.038 | 0.074 | 1.030 | 2.406 | 1.196 | 110 |
| 5 years | -3.790 | -0.696 | 0.126 | 0.194 | 0.868 | 6.899 | 1.410 | 194 |
| 6 years | -1.855 | -0.396 | 0.259 | 0.467 | 1.199 | 4.846 | 1.482 | 28 |
| Time elapsed | | | | | | | | |
| 2 years | -1.855 | -0.128 | 0.673 | 0.770 | 1.310 | 4.846 | 1.582 | 17 |
| 3 years | -3.790 | -0.407 | 0.370 | 0.410 | 1.143 | 6.393 | 1.364 | 112 |
| 4 years | -3.314 | -0.821 | -0.081 | 0.021 | 0.831 | 6.899 | 1.306 | 150 |
| 5 years | -4.604 | -0.732 | 0.083 | -0.039 | 0.701 | 2.406 | 1.264 | 54 |

Table 9.19: Systematic sample: Results of testing that the mean conditional height gain Z-score is zero and variance is one (by gender, age-group at school entry and time elapsed between height measures)

| | t | p | 95% CI |
|---|---|---|---|
| All | 2.4752 | 0.0138 | [0.0374, 0.3274] |
| Boys | 0.9055 | 0.3665 | [-0.1144, 0.3082] |
| Girls | 2.5798 | 0.01072 | [0.0615, 0.4625] |
| Initial age | | | |
| 4 years | 0.6476 | 0.5186 | [-0.1521, 0.2998] |
| 5 years | 1.9173 | 0.0567 | [-0.0056, 0.3939] |
| 6 years | 1.6682 | 0.1068 | [-0.1075, 1.0421] |
| Time elapsed | | | |
| 2 years | 2.0075 | 0.0619 | [-0.0431, 1.5833] |
| 3 years | 3.1831 | 0.0019 | [0.1549, 0.6657] |
| 4 years | 0.1961 | 0.8448 | [-0.1898, 0.2317] |
| 5 years | -0.2256 | 0.8224 | [-0.3839, 0.3062] |
| | $\frac{(n-1)s^2}{\sigma^2}$ | approx. p-value | 95% CI |
| All | 604.21 | 0 | [1.5676,2.1250] |
| Boys | 294.95 | $4.5741 \times 10^{-14}$ | [1.4978,2.3248] |
| Girls | 306.99 | $3.3884 \times 10^{-13}$ | [1.4604,2.2313] |
| Initial age | | | |
| 4 years | 155.83 | 0.0015 | [1.1148,1.9004] |
| 5 years | 383.95 | 0 | [1.6453,2.4546] |
| 6 years | 59.33 | $1.0877 \times 10^{-5}$ | [1.3735, 4.0708] |
| Time elapsed | | | |
| 2 years | 40.03 | $2.1592 \times 10^{-5}$ | [1.3877,5.7949] |
| 3 years | 206.54 | $1.4328 \times 10^{-10}$ | [1.4540,2.4666] |
| 4 years | 254.26 | $1.0790 \times 10^{-9}$ | [1.3767,2.1713] |
| 5 years | 84.70 | 0.0021 | [1.1293,2.4355] |

Table 9.20: Systematic sample: Summary statistics of conditional height gain Z-scores by age-group at school entry and time elapsed between height measures

| | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| 4, 4-5 years | -4.604 | -0.542 | 0.038 | 0.069 | 1.031 | 2.406 | 1.200 | 109 |
| 5, 3-4 years | -3.790 | -0.702 | 0.107 | 0.221 | 0.850 | 6.899 | 1.483 | 148 |

for case and control children.

A notch boxplot of the conditional height gain Z-scores by gender and case-control status can be found in figure 9.14. The notches on the boxplots for the boys overlap, indicating that there is no reason to doubt that the median conditional height gain Z-scores are the same for the case and control boys. The notches on the boxplots for the girls don't overlap ([-0.50794, 0.03439] for control girls, [0.05736, 0.6401] for case girls), suggesting that the median conditional height gain Z-scores are significantly different in the case and control girls. However, if the Wilcoxon Mann-Whitney test is used then W = 2016 and p = 0.0638, so difference just fails to reach statistical significance. There are four case girls with outlying conditional height gain Z-scores: ID 559 (her height Z-score drops from 0.39 at school entry to -1.68 at follow-up); ID 1813 (her height Z-score drops from -1.98 at school entry to -3.78 at follow-up, this child was a definite organic FTT); ID 2289 (her height Z-score rises from -0.33 at school entry to 1.11 at follow-up) and ID 3121 (her height Z-score rises from -1.68 at school entry to -0.03 at follow-up). There are two control children with extreme high conditional height gain Z-scores: ID 1185 (her height Z-score rises from 0.36 at school entry to 1.45 at follow-up) and ID 2016 (his height Z-score rises from -0.81 at school entry to 0.32 at follow-up).

We will use the suggested value of -2.67 as a cut-off, which in theory should represent the 0.4th centile. Within the systematic sample there are five children (1.5%) with conditional height gain Z-scores less than -2.67 (ID's 235, 313, 989, 1621 and 3689). Three case children (the two discussed above plus ID 3624) and one control child (ID 1380) have a conditional height gain Z-score less than -2.67. If we now consider 2.67 as a cut off, which in theory should represent the 99.6th centile. Within the systematic sample there are nine children (2.7%) with conditional height gain Z-scores greater than 2.67 (ID's 838, 1060, 1158, 1402, 1543, 2210, 2280, 2825,

Table 9.21: Case-control study: Summary statistics of conditional height gain Z-scores

| | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| cases | -5.4410 | -0.6841 | 0.2747 | 0.0856 | 1.0210 | 4.8060 | 1.5426 | 88 |
| controls | -3.0170 | -0.7029 | 0.0101 | 0.0343 | 0.7611 | 3.2340 | 1.2182 | 89 |
| Case boys | -2.6050 | -0.9039 | -0.0721 | -0.1244 | 0.7933 | 2.0420 | 1.1576 | 28 |
| Control boys | -2.0060 | -0.4649 | 0.2960 | 0.2747 | 0.8752 | 2.9850 | 1.1234 | 33 |
| Case girls | -5.4410 | -0.5947 | 0.3487 | 0.1836 | 1.0680 | 4.8060 | 1.6926 | 60 |
| Control girls | -3.0170 | -0.9272 | -0.2368 | -0.1074 | 0.5685 | 3.2340 | 1.2589 | 56 |

3489). Two case children (ID's 2289 and 3121) and two control children (ID's 1185 and 2016) have conditional height gain Z-scores greater than 2.67. Assuming that the systematic sample is representative of children in Newcastle; it would seem that using cut-offs of -2.67 and 2.67 are not appropriate. As mentioned above, the conditional height gain Z-scores have a variance greater than 1; thus leading to more than expected individuals with extreme gains. However, if sample size allows, the use of boxplots appears to be a viable means of detecting individuals with unusual height gains.

### 9.2.4 Conditional weight gain Z-scores

In infancy weight is routinely monitored and it would be nice to link these weights up with weights at school entry and at follow-up. One approach would be to use a conditional weight gain Z-score. However in order to do this would need to know the correlation structure for weights from birth to 9 years. This is not available in the literature, Berkey et al. (1983a) published yearly correlations for log weight from 1 to 6 years (see table 9.22).[1] However over this time period the power term in the LMS method decreases from 0.04 to -0.71 for boys and from -0.081 to -0.68 for girls. An additional consideration is that these correlations may not be calculated using the actual weights. In infancy the Harvard weights and heights were adjusted to exact age using the Jenss model (Jenss and Bayley 1937) to generate correlations and it is suspected that the same is true for the weights and heights in childhood.

Earlier work indicated that weight Z-scores needed to be grouped to the nearest 6 months at follow-up in order to achieve sample sizes of more than 50, that had

[1]The log transformation was used to achieve approximate normality for the weights

Table 9.22: Correlation matrix of log(weight) for 229 Boston children (Age in years). Boys above diagonal and girls below

| Age | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 1 | 0.925 | 0.844 | 0.766 | 0.697 | 0.638 |
| 2 | 0.929 | 1 | 0.970 | 0.908 | 0.839 | 0.776 |
| 3 | 0.854 | 0.976 | 1 | 0.980 | 0.940 | 0.896 |
| 4 | 0.791 | 0.928 | 0.984 | 1 | 0.988 | 0.965 |
| 5 | 0.736 | 0.871 | 0.950 | 0.989 | 1 | 0.993 |
| 6 | 0.686 | 0.812 | 0.905 | 0.963 | 0.992 | 1 |

both a birth weight and weight at follow up. In order to get an handle on the correlation structure beyond infancy, pair-wise correlations between weight Z-scores at birth (one year) and weight Z-scores at school entry and follow-up were obtained. The resulting correlations are tabulated in tables 9.23 and 9.24 (for the moment we are ignoring the small sample sizes). Plots of $t2$ versus correlation for the tabulated values in tables 9.23 and 9.24 can be found on the left and right of figure 9.15, respectively. There appears to be a linear relationship in the correlations with weight Z-score at one year. One possibility would be to fit the line $r = 0.929 - 0.046t2$ (obtained using weighted least squares) to these correlations and then use this line to predict the correlation between weight Z-score at one year and weight Z-scores at school entry.

Table 9.23: Correlation between weight Z-scores in childhood and birth weight Z-score (where $r$ and $n$ denote correlation and sample size, respectively)

| Age | 6wks | 3mths | 6mths | 9mths | 12mths | 18mths | 4yr |
|---|---|---|---|---|---|---|---|
| $r$ | 0.689 | 0.599 | 0.511 | 0.469 | 0.523 | 0.513 | 0.455 |
| $n$ | 484 | 480 | 467 | 388 | 394 | 296 | 24 |
| Age | 4.5yr | 5yr | 5.5yr | 6yr | 8yr | 8.5yr | 9yr |
| $r$ | 0.445 | 0.400 | 0.437 | 0.164 | 0.252 | 0.370 | 0.399 |
| $n$ | 144 | 134 | 45 | 12 | 62 | 225 | 129 |

Table 9.24: Correlation between weight Z-scores in childhood and weight Z-score at 1 year (where $r$ and $n$ denote correlation and sample size, respectively)

| Age | 18mths | 4yr | 4.5yr | 5yr | 5.5yr |
|---|---|---|---|---|---|
| $r$ | 0.8676 | 0.7446 | 0.7174 | 0.6576 | 0.6535 |
| $n$ | 247 | 21 | 112 | 97 | 38 |
| Age | 6yr | 8yr | 8.5yr | 9yr | |
| $r$ | 0.6651 | 0.6093 | 0.5281 | 0.5438 | |
| $n$ | 11 | 51 | 178 | 106 | |

### 9.2.5 Conditioning weight on height

A conditional index, equation (3.11), can be used to compare the current height and weight of a child. It is suggested that the correlation between weight and height is 0.7. The correlation between weight and height Z-scores is 0.7184 (95% CI is [0.6652,0.7643]) at school entry and 0.7145 (95% CI is [0.6647,0.7579]). So letting $r = 0.7$, equation (9.3) can be used to compare current height and weight.

$$Z_{w|h} = \frac{Z(\text{weight}) - 0.7Z(\text{height})}{\sqrt{(1 - 0.7^2)}} \tag{9.3}$$

The summary statistics of the Z-score for weight conditional on height at school entry and follow-up can be found in table 9.25. The Z-score for weight conditional on height appears to cover a wider range at follow-up than at school entry. In table 9.26 we summarise the result of testing that the Z-score for weight conditional on height has a mean of zero and variance of one. At school entry there is reason to doubt that the Z-score for weight conditional on height has a mean of zero for both genders. Similarly, at follow-up there is reason to doubt that the Z-score for weight conditional on height has a mean of zero for girls. However, for boys, there is no reason to doubt that the mean of the Z-score for weight conditional on height is zero at follow-up. Furthermore, there is no reason to doubt that the variance is one for the Z-score for weight conditional on height at school entry and follow-up. Therefore, caution is needed when using this conditional index at school entry.

A plot of this index, given by equation (9.3), against the Z-score for BMI at follow-up and school entry can be found in figure 9.16. There is generally a linear relationship between the conditional weight on height Z-score and the Z-score for

Table 9.25: Systematic sample (excluding cases): Summary statistics of Z-score for weight conditional on height by gender

| school entry | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
|---|---|---|---|---|---|---|---|---|
| All | -2.8270 | -0.8873 | -0.2400 | -0.2188 | 0.4146 | 2.7740 | 0.9669 | 371 |
| Boys | -2.8270 | -0.8073 | -0.1822 | -0.1819 | 0.4568 | 1.6730 | 0.9082 | 180 |
| Girls | -2.4420 | -0.9911 | -0.2715 | -0.2537 | 0.3759 | 2.7740 | 1.0203 | 191 |
| follow-up | Min. | LQ | Median | Mean | UQ | Max. | SD | no. |
| All | -3.5850 | -0.7103 | -0.1400 | -0.1343 | 0.4670 | 3.2790 | 0.9554 | 428 |
| Boys | -3.5390 | -0.6298 | -0.0735 | -0.0899 | 0.4649 | 2.1160 | 0.9239 | 214 |
| Girls | -3.5850 | -0.7929 | -0.2233 | -0.1787 | 0.4624 | 3.2790 | 0.9860 | 214 |

Table 9.26: Systematic sample (excluding cases): Results of testing that the mean Z-score for weight conditional on height is zero and variance is one (by gender)

| | | t | p | 95% CI |
|---|---|---|---|---|
| school entry | All | -4.3594 | $1.693 \times 10^{-5}$ | [-0.3176, -0.1201] |
| | Boys | -2.6866 | 0.0079 | [-0.3155,-0.0483] |
| | Girls | -3.4362 | 0.0007 | [-0.3993,-0.1081] |
| follow-up | All | -2.9083 | 0.0038 | [-0.2251,-0.0435] |
| | Boys | -1.4228 | 0.1563 | [-0.2144,0.0346] |
| | Girls | -2.652 | 0.0086 | [-0.3116,-0.0459] |
| | | $\frac{(n-1)s^2}{\sigma^2}$ | approx p-value | 95% CI |
| school entry | All | 345.94 | 0.3765 | [0.8136,1.0858] |
| | Boys | 147.66 | 0.0976 | [0.6775,1.0265] |
| | Girls | 197.81 | 0.6889 | [0.8598,1.2868] |
| follow-up | All | 389.73 | 0.2022 | [0.8017,1.0487] |
| | Boys | 181.83 | 0.1309 | [0.7122,1.0421] |
| | Girls | 207.06 | 0.7736 | [0.8110,1.1868] |

BMI. At follow-up there is slightly more variability in the conditional index as the BMI Z-score increases. The unconditional index, equation (3.5), could be used to compare current height and weight centiles. It has been suggested that if the height and weight centiles differ by more than three major centiles then this would be a cause for concern. However as the correlation between weight and height is smaller, Cole (1997a) states the conditional index should be used in preference.

## 9.3 Lowess trend curves for weight growth from infancy to 7-9 years

The lowess procedure was used to draw trend curves for the full curve of growth. A potential problem in using the lowess procedure is that the data is in three distinct clusters. Altering the span has little impact on smoothing the unusual appearance of the lowess curve as it enters or leaves a cluster, see figure 9.17. The case children have lowess curves that drop away from zero after birth, reaching a minimum around the first year, there is then a gradual improvement in childhood; see figures 9.18 and 9.19. However, the weight Z-scores for case children at 7 to 9 years are still substantially below zero. The control boys have lowess curves that increase above

zero in early infancy and remains just above the zero line into childhood; see figure 9.18. The control girls have similar lowess curves in infancy to control boys, but in childhood their lowess curve runs along the zero line; see figure 9.19. After 1 year girls and boys within the systematic sample have lowess curves that are just below and above the zero line, respectively (see figure 9.20). Figure 9.21 summarises all the default lowess trend curves for the cases, controls and systematic samples.

## 9.4 Discussion and Conclusions

Relationships between the systolic blood pressure at 7-9 years and current Z-scores of anthropometric measures have been explored previously (Wright et al. 1999). For individuals within the systematic sample, Wright et al. (1999) regressed the systolic blood pressure on the Z-score for BMI and height. The adjusted blood pressure was obtained using the results from this regression to arrive at the following equation[2]:

$$\text{Adjusted BP} = \text{systolic BP} - 1.56\text{Height SDS} + 1.75\text{BMI SDS}$$

Even after adjusting for height and BMI, Wright et al. (1999) found that the blood pressure was still slightly lower in the FTT group. This lead to the conclusion that children that had failed to thrive in infancy had no tendency to increased blood pressure (Wright et al. 1999). The study of the systematic sample was not designed to compare the relative influence of birth weight and subsequent weight gain for children growing within normal limits, however weak associations would suggest that accelerated weight gain in infancy may predict better later hypertension than absolute size at birth (Wright et al. 1999).

Final systolic blood pressures were similar in boys and girls. This is in agreement with Barker et al. (1989), they found that men had higher blood pressure than women but at 10 years the blood pressure of boys and girls were little different. If a stepwise procedure is used to identify a model for the systolic blood pressure then the same model choice as derived by Wright et al. (1999) is arrived at. However, the Z-score for Weight is highly correlated with the Z-scores for BMI and height. Regressing systolic blood pressure on the current weight Z-score leads to simpler model with an identical adjusted multiple correlation coefficient to the model derived when regressing systolic BP on the Z-scores for height and BMI. Barker et al. (1989) found that systolic blood pressure in 10 year old boys and girls

[2]coefficients slightly different to model obtained on systematic sample (excluding cases) as some children's data were not included in Wright et al. (1999), as they were not coded correctly as belonging to this sample

was associated with current weight. If weight Z-scores from infancy are considered then the weight Z-score at birth and 1 year make a much smaller contribution to explaining the variation in the systolic BP than current weight Z-score alone. The sign of the coefficient in this model for the birth weight Z-score is negative, which is an agreement with other authors, that after allowing for current size, a low weight at birth would lead to a higher systolic BP at 7-9 years.

Singhal et al. (2001) suggested that the nutritional programming of blood pressure may not manifest itself until puberty or adulthood. Barker et al. (1989) reported a stronger inverse relationship between systolic blood pressure and birth weight at 36 years than at 10 years. Berkey et al. (1998) arrived at a model for systolic blood pressure at 17 and 30 years using a similar stepwise approach, considering: height, weight and BMI measures at 17 and 30; length, weight, ponderal index[3], head and chest circumference at birth. Berkey et al. (1998) regressed systolic BP at 17 for women on BMI at 17, whereas at 30 systolic BP was regressed on head circumference at birth, systolic BP at 17 and change in BMI between 17 and 30 years. For males, Berkey et al. (1998) regressed systolic BP at 17 on BMI at 17 and ponderal index at birth, whereas at 30 systolic BP was regressed on BMI at 17 and birth weight. The sample sizes in this exploratory work were much smaller than considered here; between 49 and 67 for females and 60 to 62 for males. An additional point, is the issue of 'spurious correlation', the correlation between ratios that exists even if all the component variables are uncorrelated (Kronmal 1993). Kronmal (1993) cautions against regressing on ratios such as the BMI. If BMI is included in any regression model, then its lower order terms, $(\text{height})^{-2}$ and weight, such be incorporated in the model too.

Although the use of conditional index for monitoring height or weight gain in childhood is advocated (Cole 1994a; Cole 1997a) it hasn't been put into practice in the growth literature. Published work tends to look at the raw height change or the change in height Z-score with no allowance made for the correlation between the height measures. Mulligan et al. (1998) used community height data collected on 486 children (247 boys, 239 girls) at a mean age of 4.91 years and 7.87 years, to see how many children had an abnormal growth rate. As pointed out in this paper there is no empirical standard for defining slow growth. Mulligan et al. (1998) took a change in height Z-score greater than 0.67 to be abnormal. The community data identified 11 'slow growing' and 9 'fast growing' children. Rudolf et al. (2000) looked at the change in height, weight and BMI Z-score of 694 children from Leeds participating in the APPLES project. These children were between 7 and 10 years

[3]ponderal index is weight divided by cubed length at birth

old and they were followed longitudinally over a period of 3 years. Rudolf et al. (2000) considered the number of children that had a change in Z-score greater than 0.67 (one channel width). Rudolf et al. (2000) found that 1% of sample experienced a change (increase or decrease) in height Z-score greater than 0.67. Whereas for weight and BMI no children experienced a decrease in Z-score less than -0.67, but 2% and 5% of sample experienced an increase in Z-score greater than 0.67 for weight and BMI, respectively.

Conditional height gain Z-scores were calculated (using the Argyle model to estimate correlation) between school entry and follow-up. It was found that there was no reason to doubt that the mean of the conditional height gain Z-scores obtained was zero for boys but this was not the case for the girls. There was also strong evidence to suggest that the variance of the conditional height gain Z-scores was not one. Therefore the use of conventional cut-offs (0.4th percentile) with conditional height gain Z-scores was not advocated. It appears that there is no significant difference in the mean of the conditional height gain Z-scores for cases and controls. Thus suggesting that the rate of height gain is similar in case and control children. However there was some indication that the variance of the conditional height gain Z-scores for case girls was more variable than for control girls. This difference in variance was partly attributable to the four extreme conditional height gain Z-scores obtained within the case girls group.

Conditioning weight on height was also considered. This was found to produce comparable results to the Z-score for BMI, but with slightly more variability for the conditional index at follow up. The plausibility of conditioning follow-up weight on previous weight Z-scores was considered. However the Newcastle sample is too small in childhood to provide viable sample sizes in order to provide a reliable correlation. Furthermore, there are no adequate correlation matrices for weight Z-scores covering this age range within the research literature. An alternative to using a conditional index would be to utilise the conditional growth Norms in LGROW (see Chapter 3). These are derived using multilevel modelling on several data sources from European growth studies.

To conclude, we have considered the use of conditional height gain Z-scores and conditioning weight on height for childhood data. There was several potential models for the correlation between height Z-scores, but with similar predictive capabilities. It was found that the correlation between height Z-scores in childhood was also adequately modelled by the Argyle model. It would be interesting to see if this model was adequate for describing the correlation between weight Z-scores in childhood. However, as noted above, the Newcastle childhood weight data is too sparse to obtain

the correlations between weight Z-scores. The use of conditional height gain Z-scores seems promising, we only need to be cautious in using conventional cut-offs because there is some indication that the variance is greater than one. It was interesting to find that the conditional height gain Z-scores in case and control children were similar. The conditional Z-score for weight on height was found to be comparable to the Z-score for BMI. Systolic blood pressure at 7 to 9 years appears to be associated most with current weight Z-score.

Figure 9.1: **Systematic sample (excluding cases)**: Boxplot of difference between systolic blood pressure on initial and final measurement occasion (SYS3-SYS1) by gender

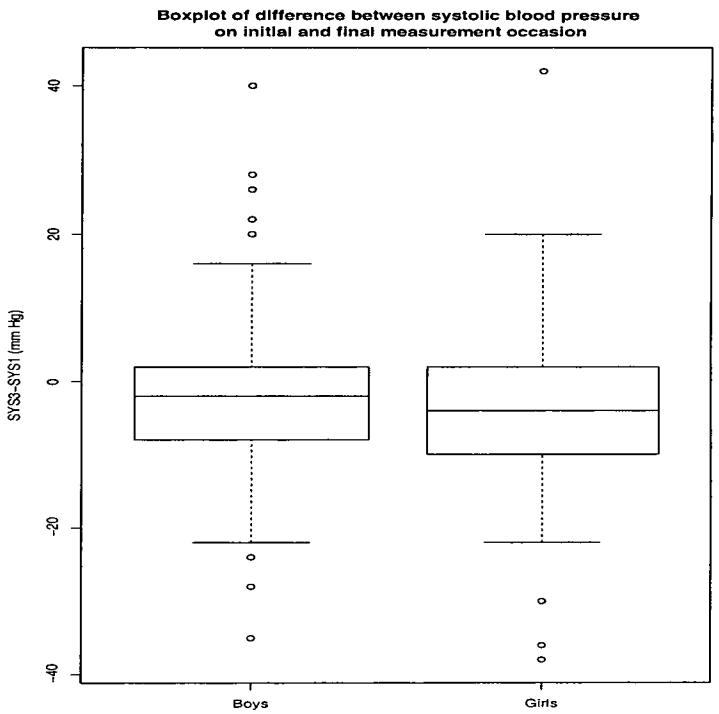


Figure 9.2: Notch boxplots of systolic blood pressure by gender for: upper panel systematic sample excluding cases lower panel case-control study

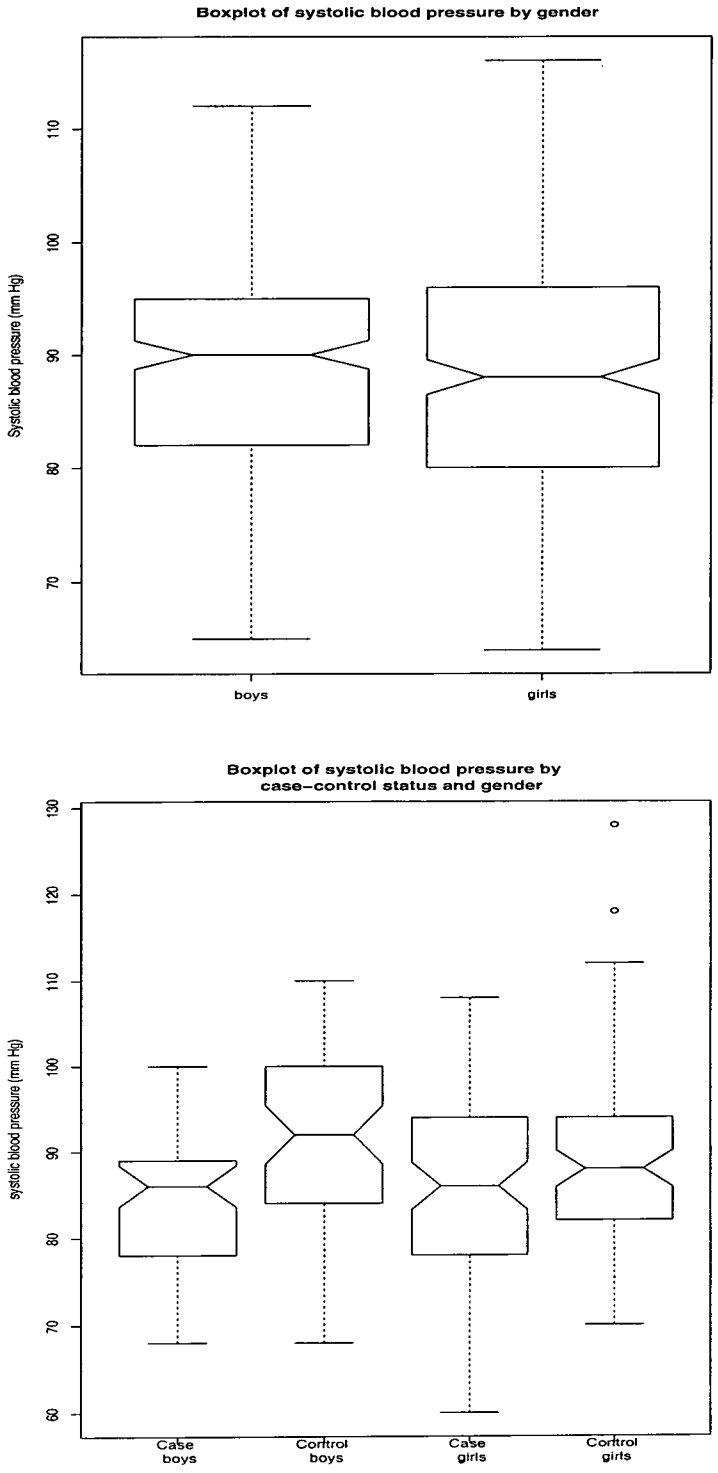


Figure 9.3: **Systematic sample (excluding cases)**: Scatterplots of systolic blood pressure versus Z-scores for height, weight, BMI and head circumference at follow-up

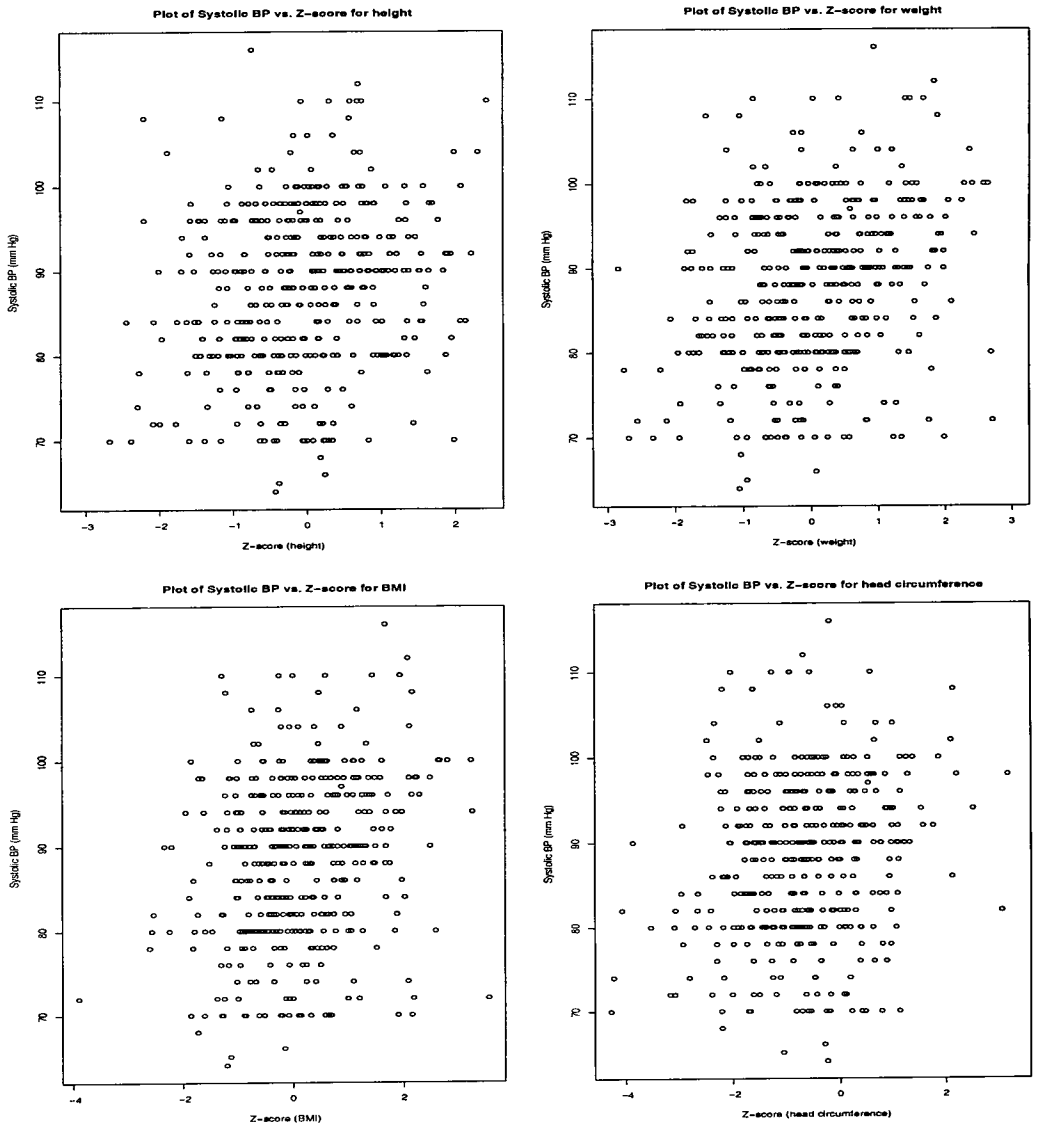


Figure 9.4: Scatterplots of all pair-wise combinations of Z-scores for height, weight, BMI and head circumference for individuals within systematic sample (excluding cases)

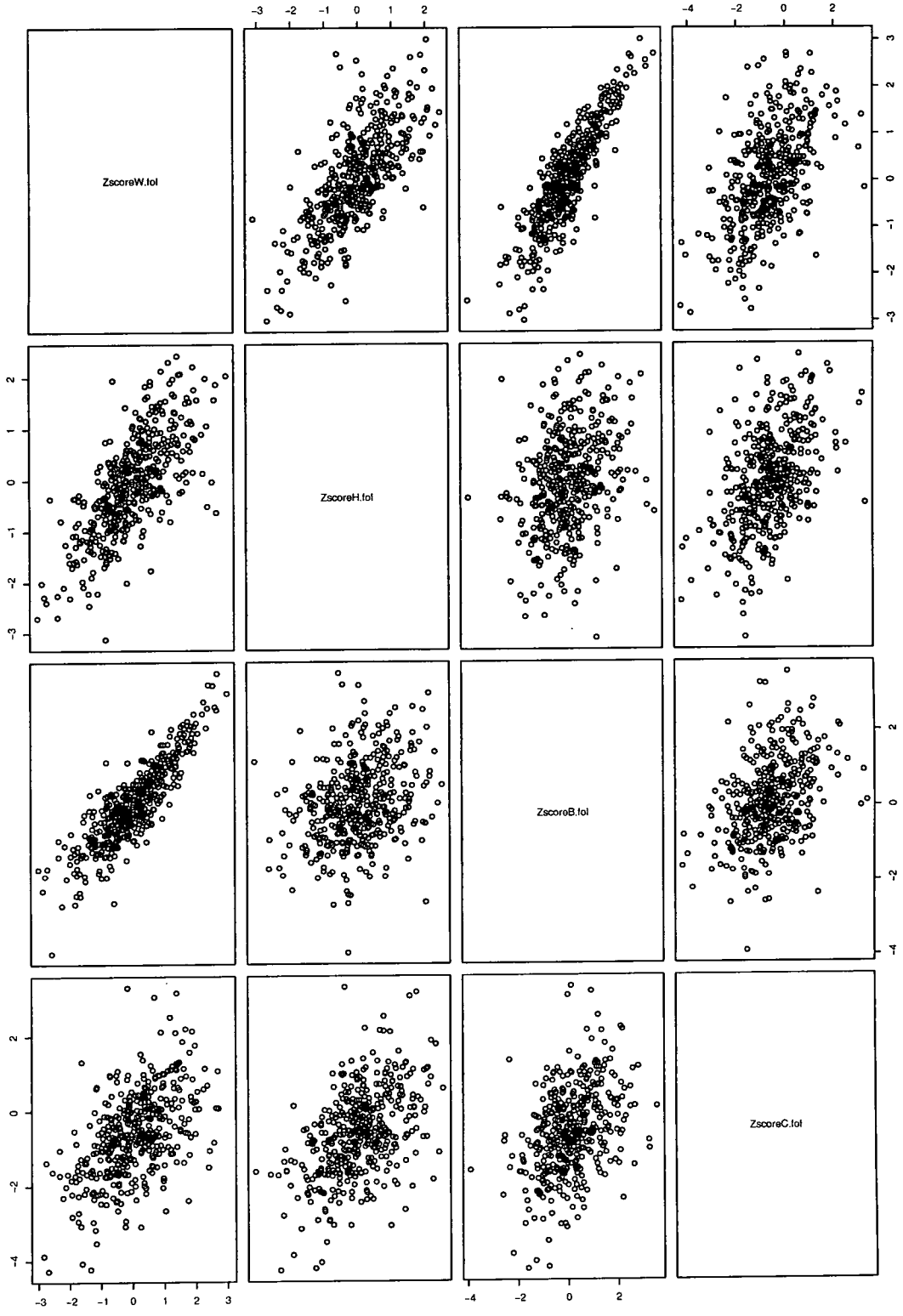


Figure 9.5: Scatterplots of all pair-wise combinations of Z-scores for height, weight, BMI and head circumference for individuals within case-control study

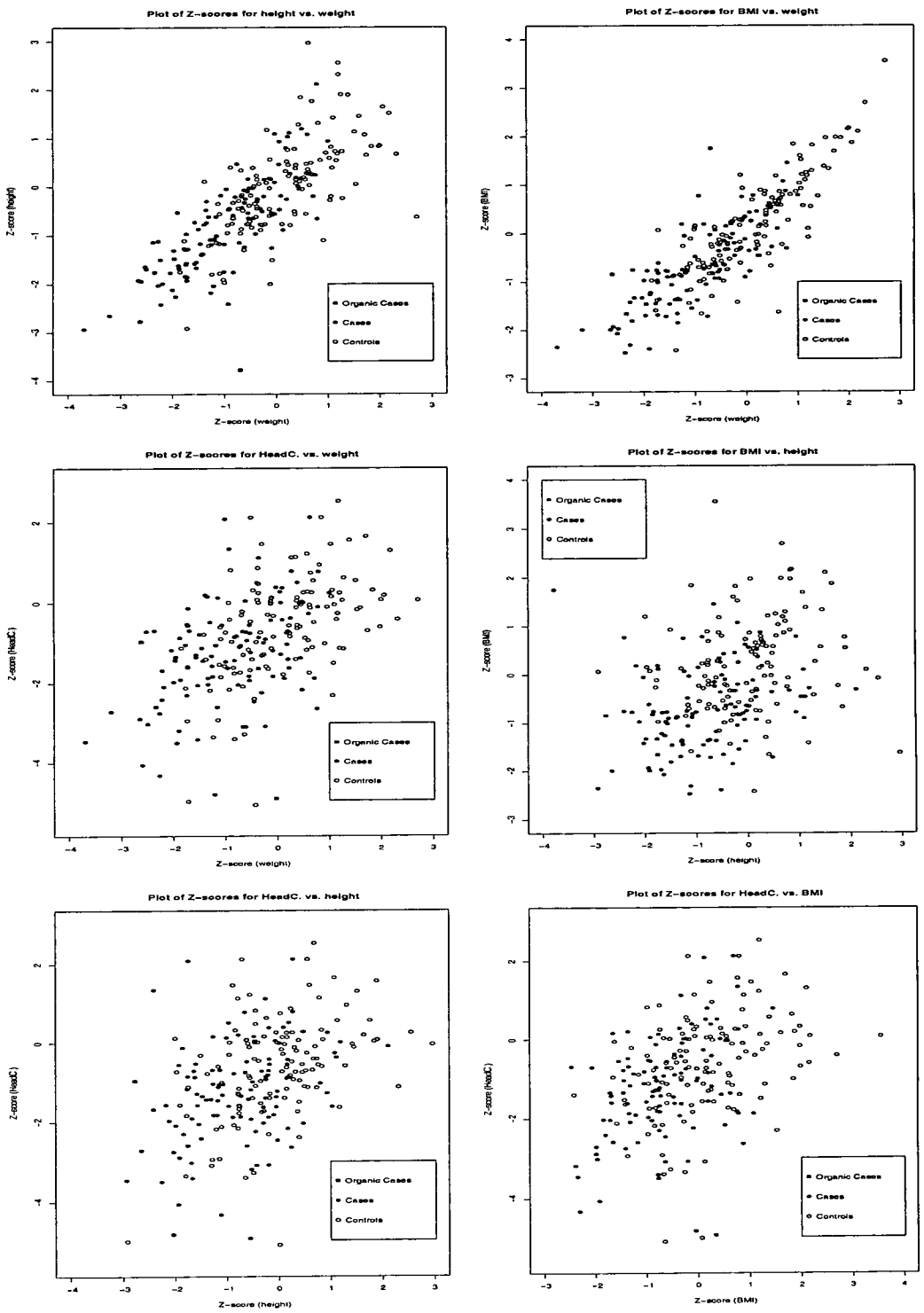


Figure 9.6: **Systematic Sample (excluding cases)** Scatterplots of all pairwise combinations of Z-scores for: birth weight, weight at 1 year, heights and weights at school entry and follow-up

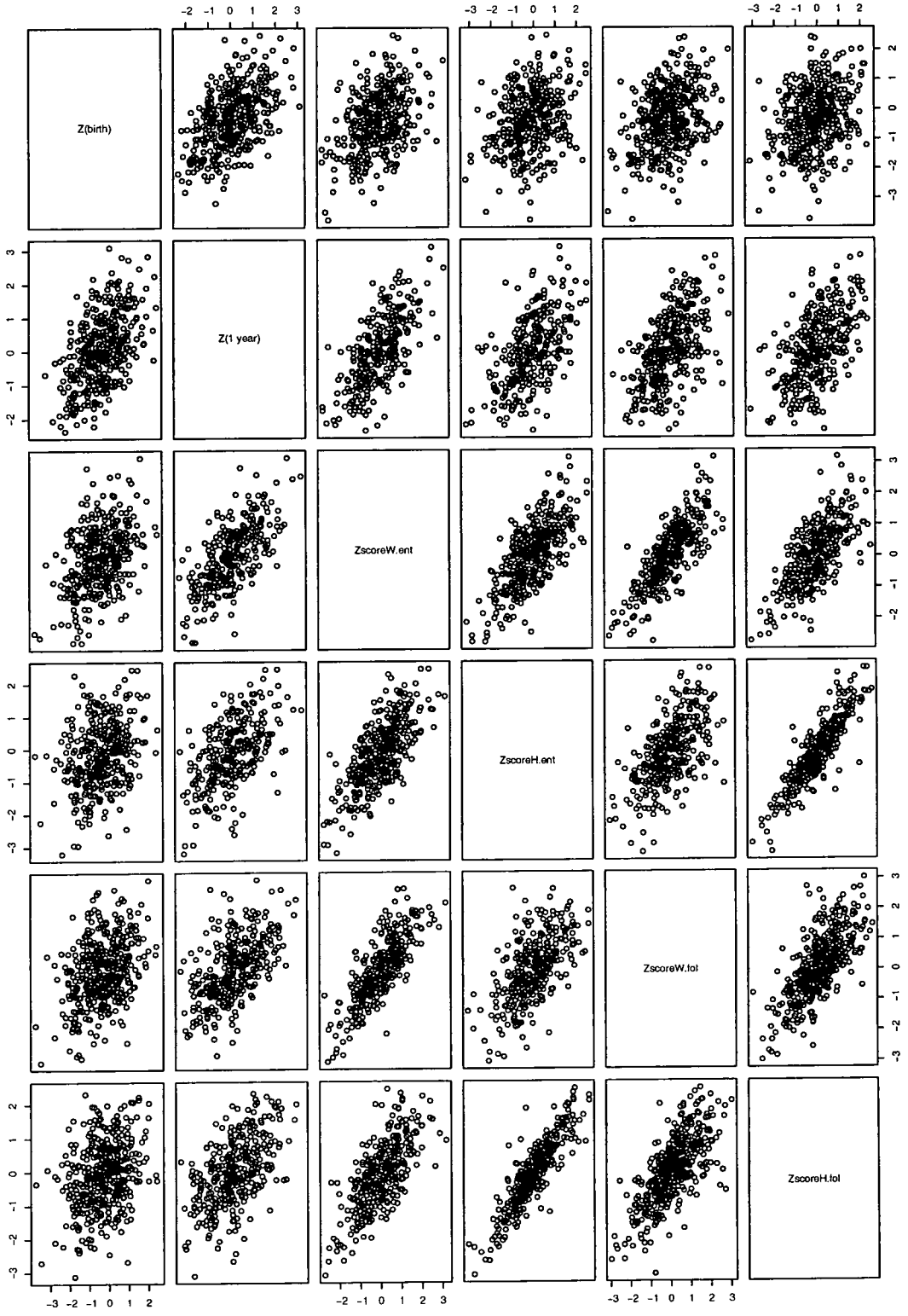


Figure 9.7: **French height correlations:** Scatterplots of correlations versus $t2$ and $t2 - t1$ (points that take on the same value of $t1$ are connected)

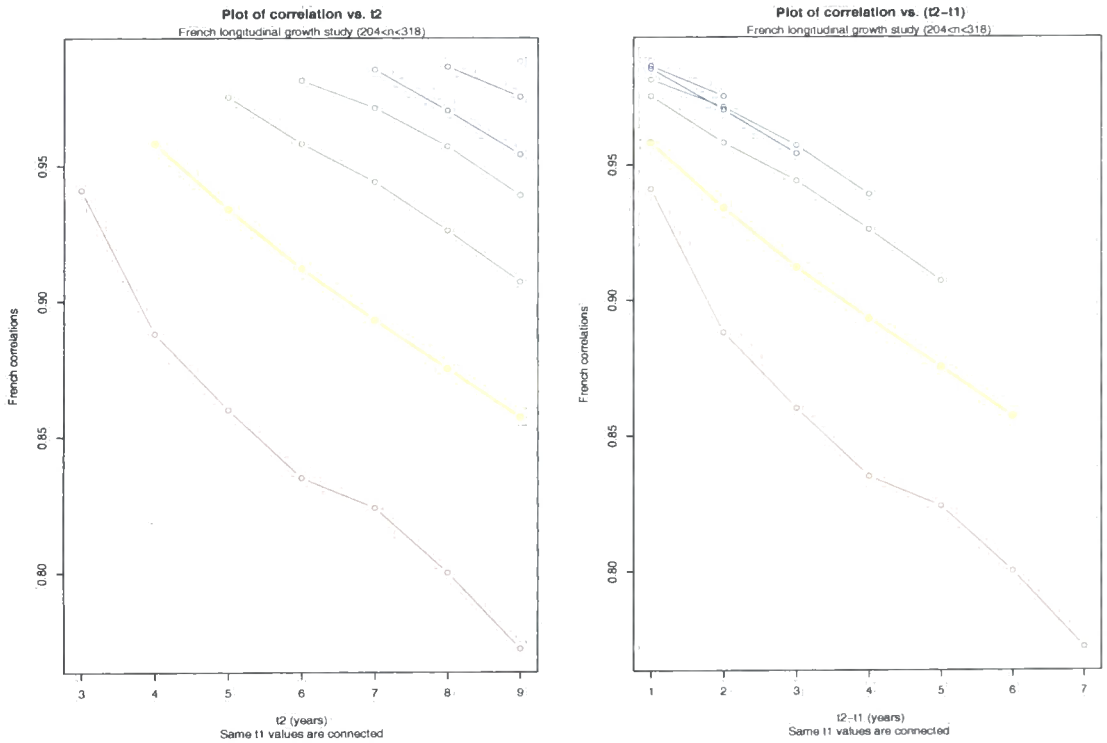


Figure 9.8: **French height correlations**: Exploratory plots to see how constant (c) added in model $\log(r(t1,t2)) = A\log(t1+c)+B\log(t2+c)+C+\epsilon$ effects model fit, term coefficients and intercept

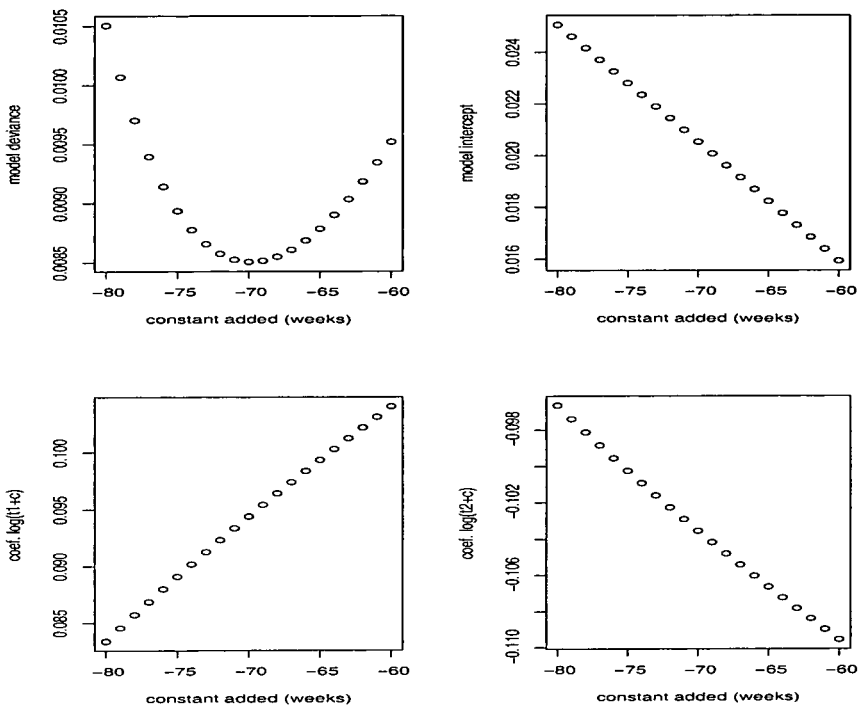


Figure 9.9: Regressing log correlation on $\log(t1 - 70/52)$ and $\log(t2 - 70/52)$: Left Plots of standardised residuals versus fitted values Right Plot of Cook's distance versus index

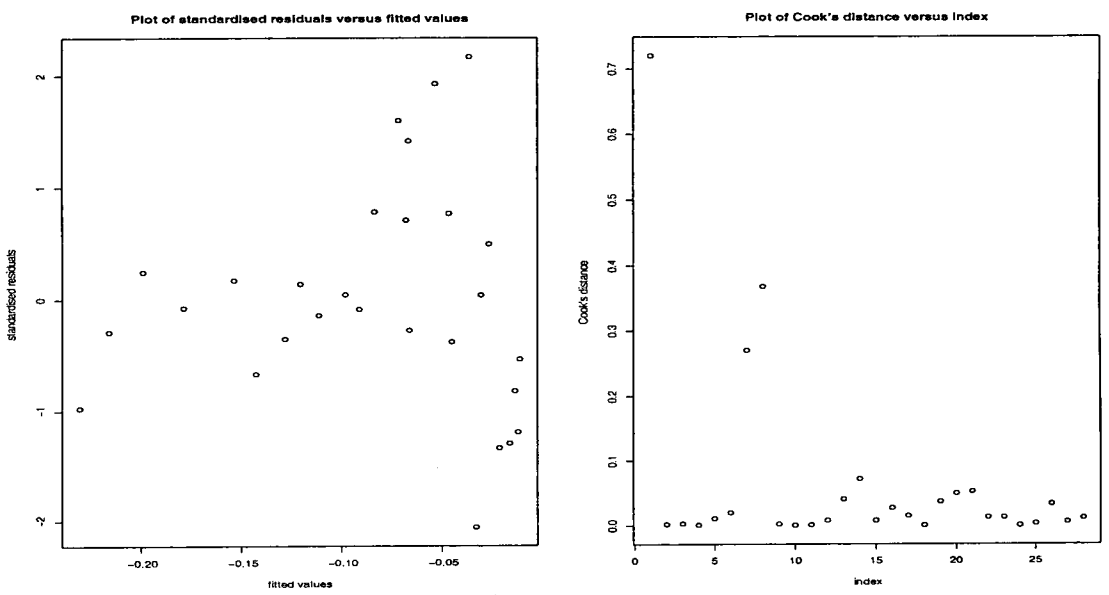


Figure 9.10: **Systematic sample (excluding cases)**: Boxplots of conditional height gain Z-score by gender

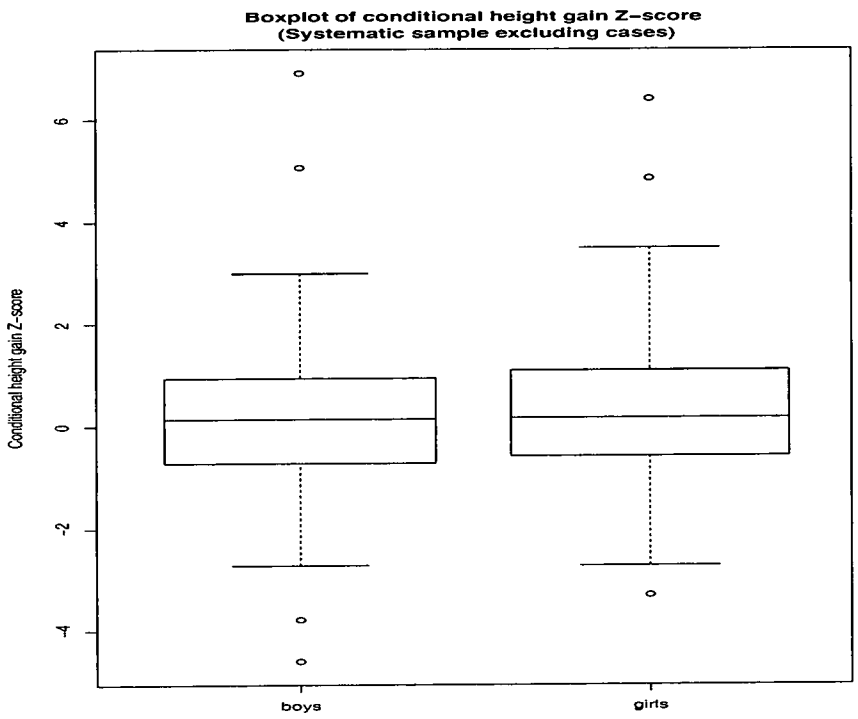


Figure 9.11: **Systematic sample (excluding cases)**: Plot of conditional height gain Z-score versus: initial time (upper left), final time (upper right), time elapsed (lower left) and initial height Z-score (lower right)

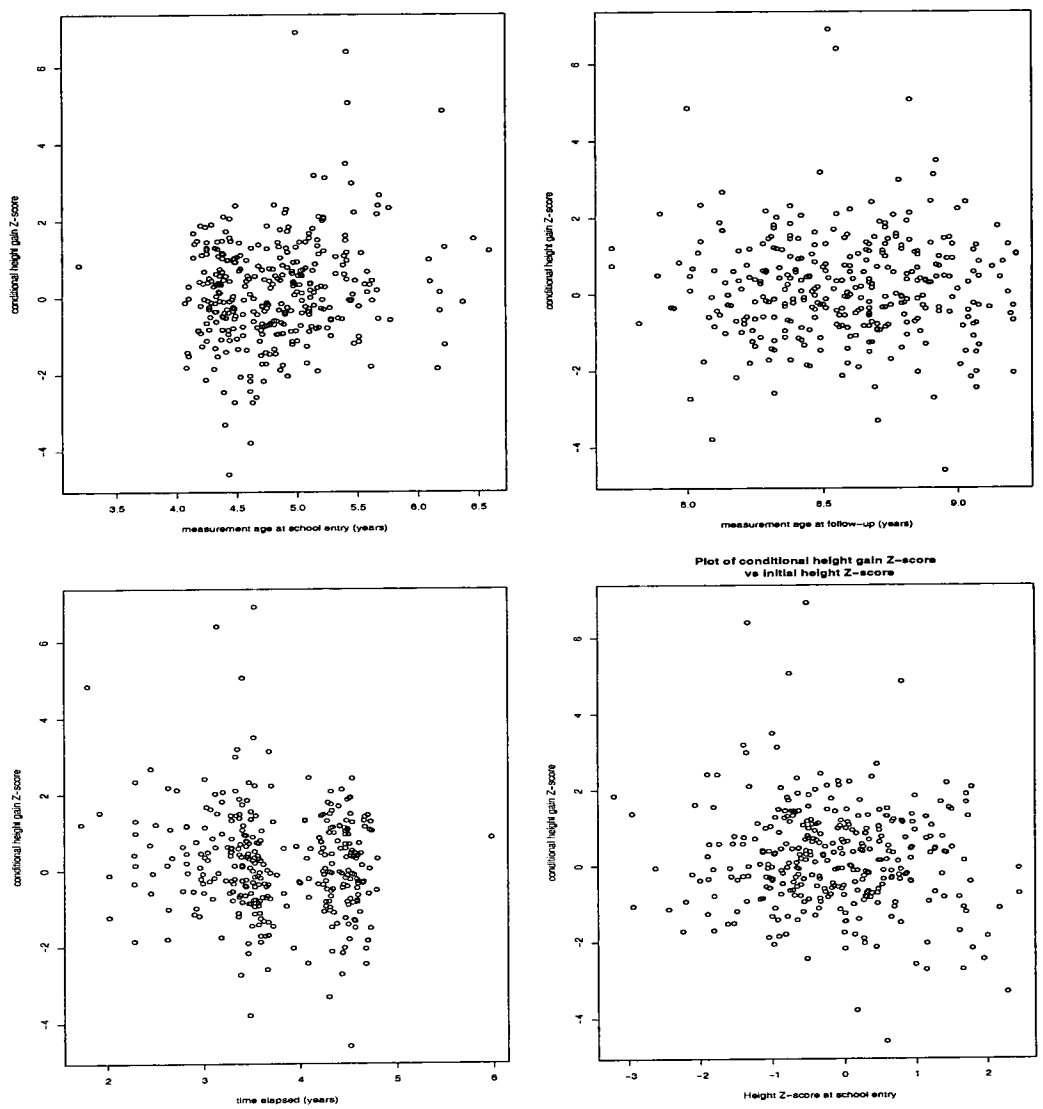


Figure 9.12: **Systematic sample (excluding cases)**: Plot of age at school entry versus time elapsed between school entry and follow up measure

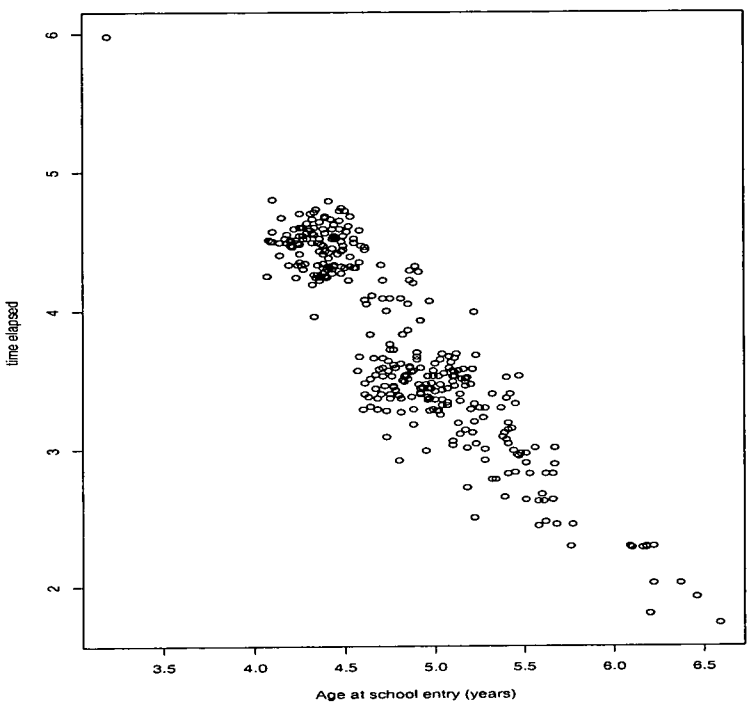


Figure 9.13: **Systematic sample (excluding cases)**: Quantile-Quantile plot of conditional height gain Z-scores by gender

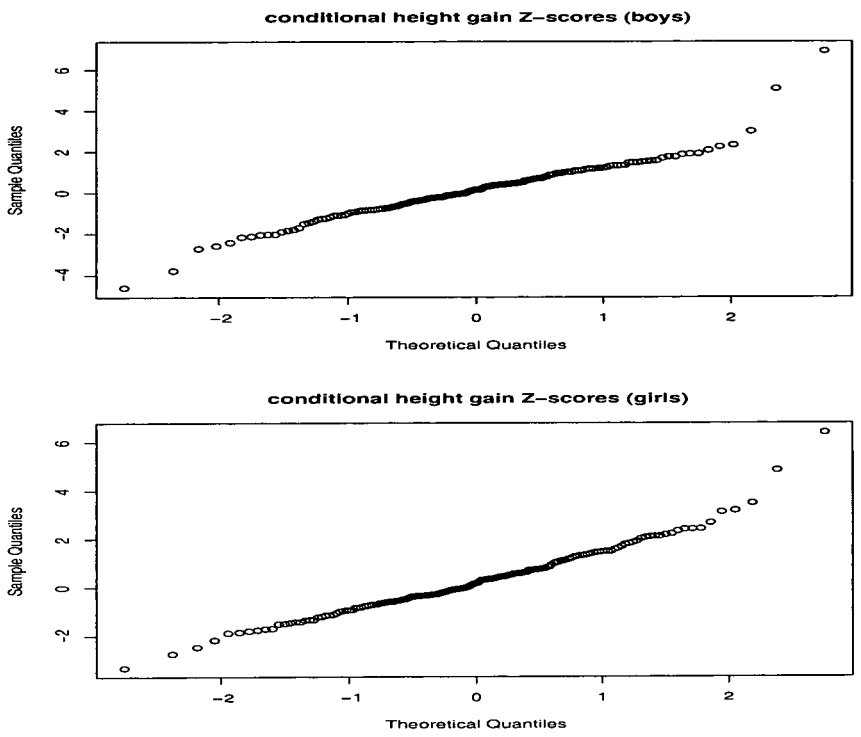


Figure 9.14: Case-control sample: Boxplots of conditional height gain Z-scores

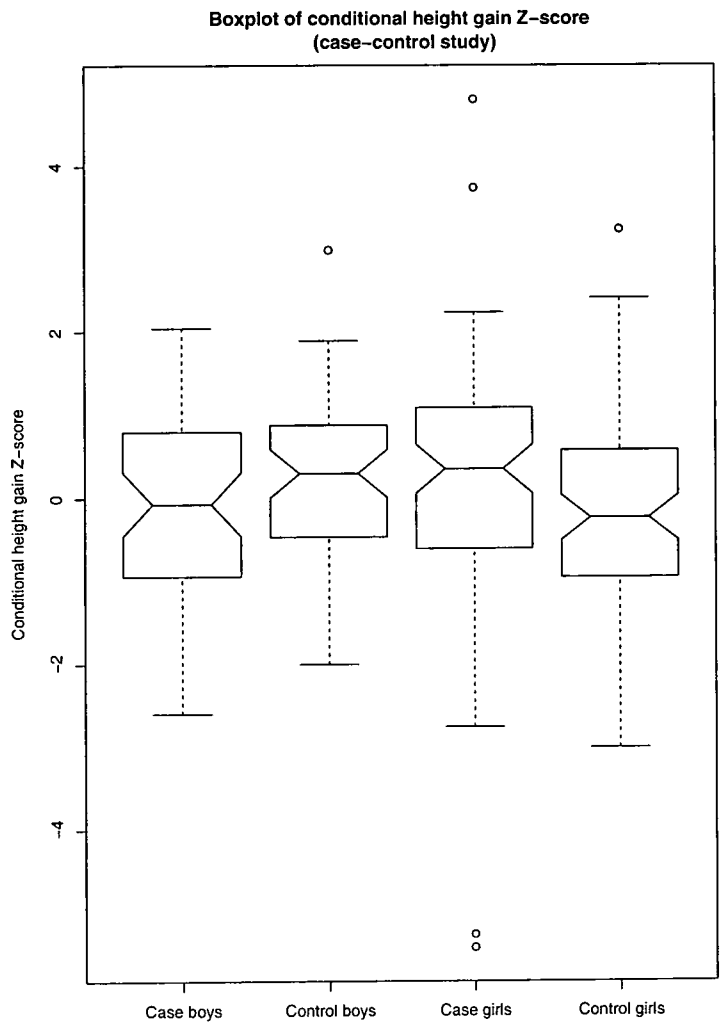


Figure 9.15: **Systematic sample (excluding cases)**: Plot of correlation between later weight Z-scores and weight Z-scores at birth and 1 year versus later time

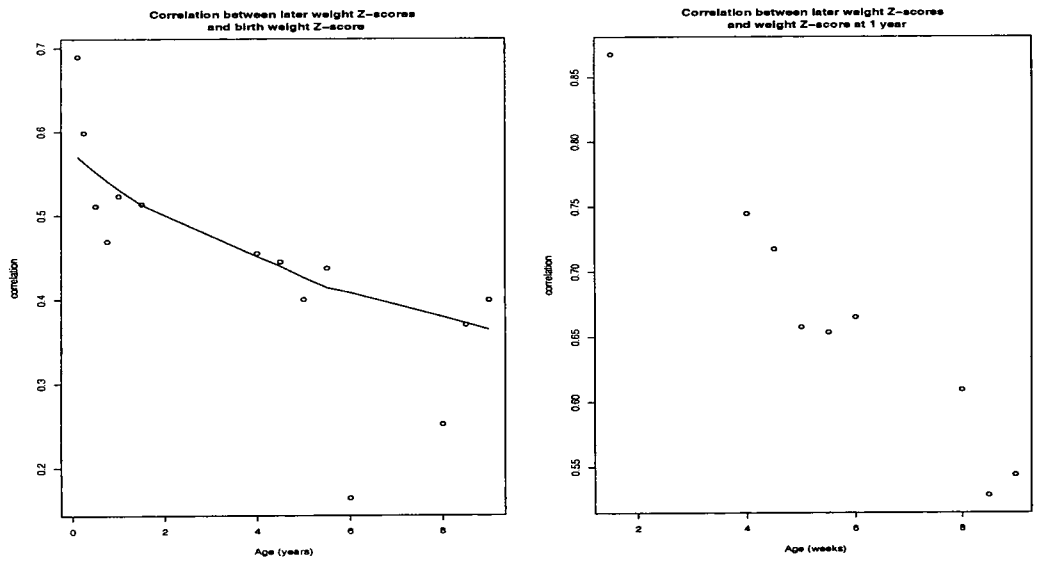



Figure 9.16: Left: Plot of $Z(w|h)$ versus Z-score for BMI for all follow-up study data Right: Plot of $Z(w|h)$ versus Z-score for BMI for all school entry data

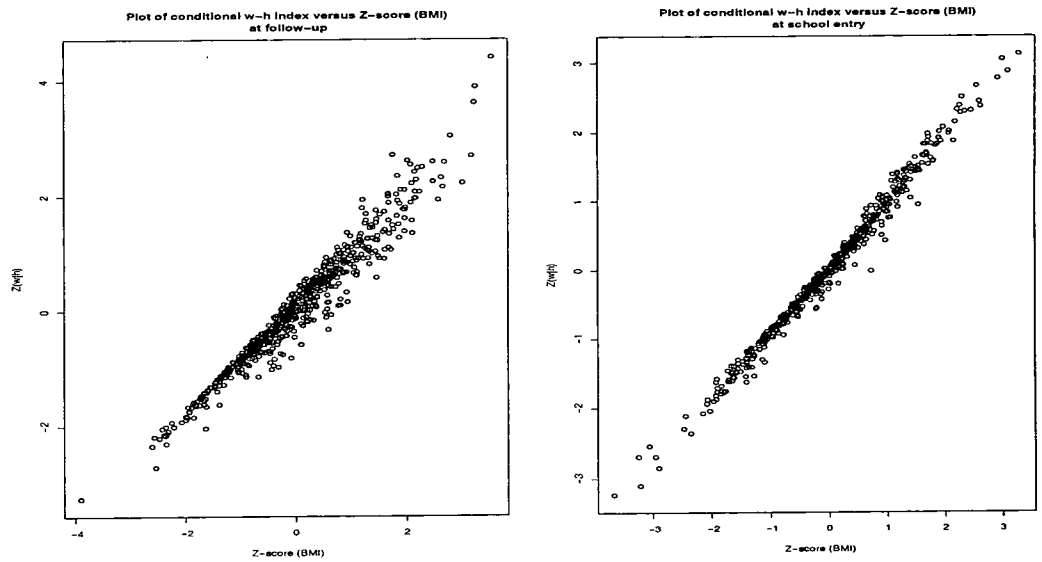


Figure 9.17: **Systematic sample** Effect of varying span on trend curves for weight Z-scores from birth to 9 years

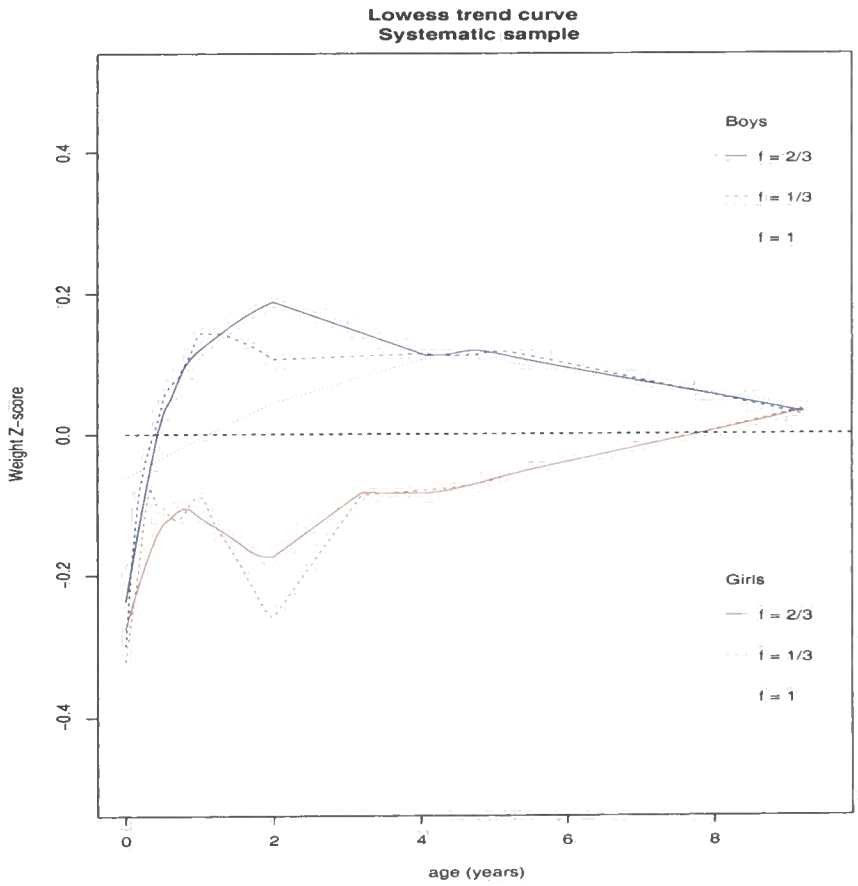


Figure 9.18: **Boys** Upper panel Plot of weight Z-score versus age for cases with default lowess curve. Lower panel Plot of weight Z-score versus age for controls with default lowess curve.

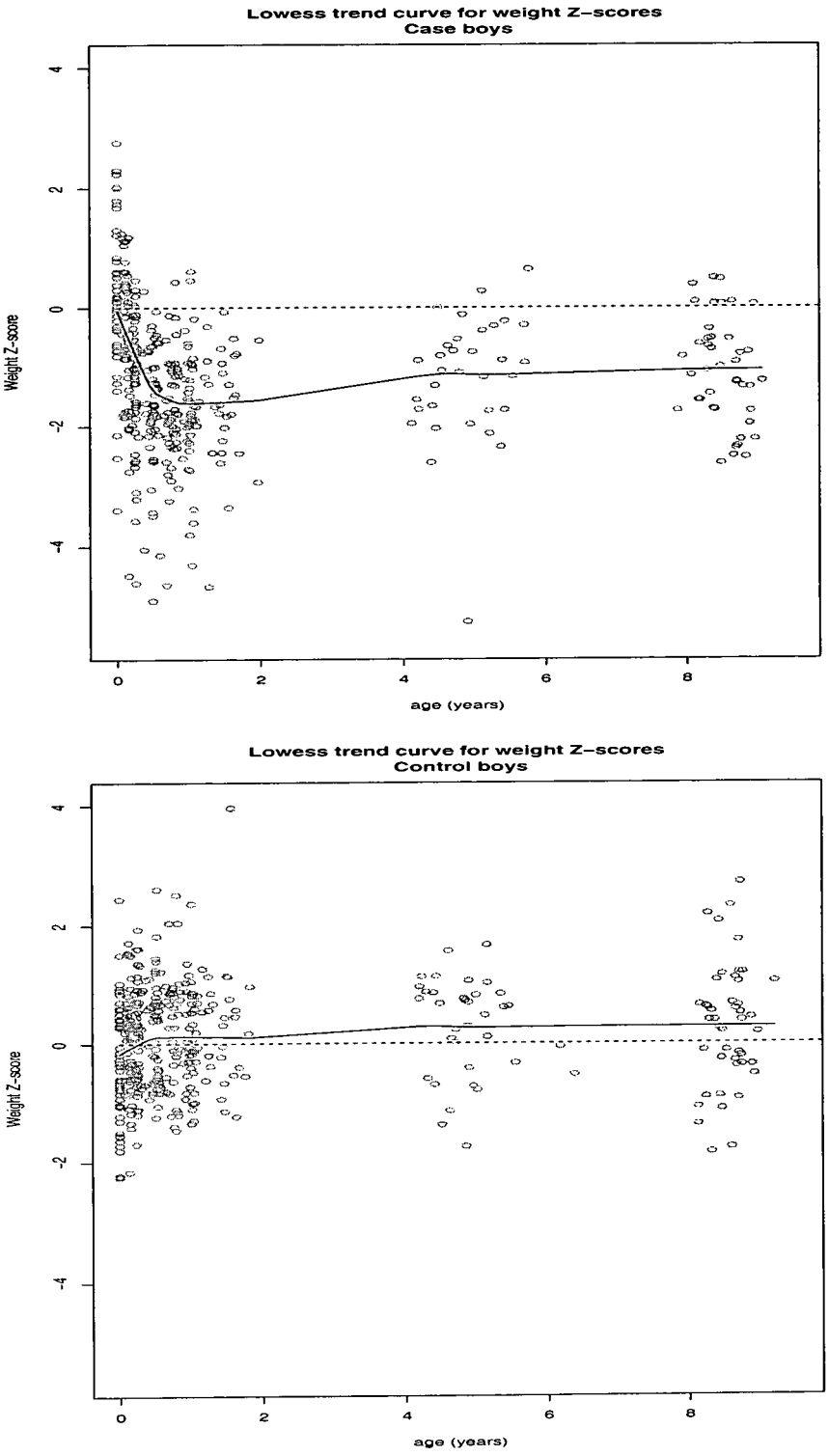


Figure 9.19: **Girls** Upper panel Plot of weight Z-score versus age for cases with default lowess curve. Lower panel Plot of weight Z-score versus age for controls with default lowess curve.

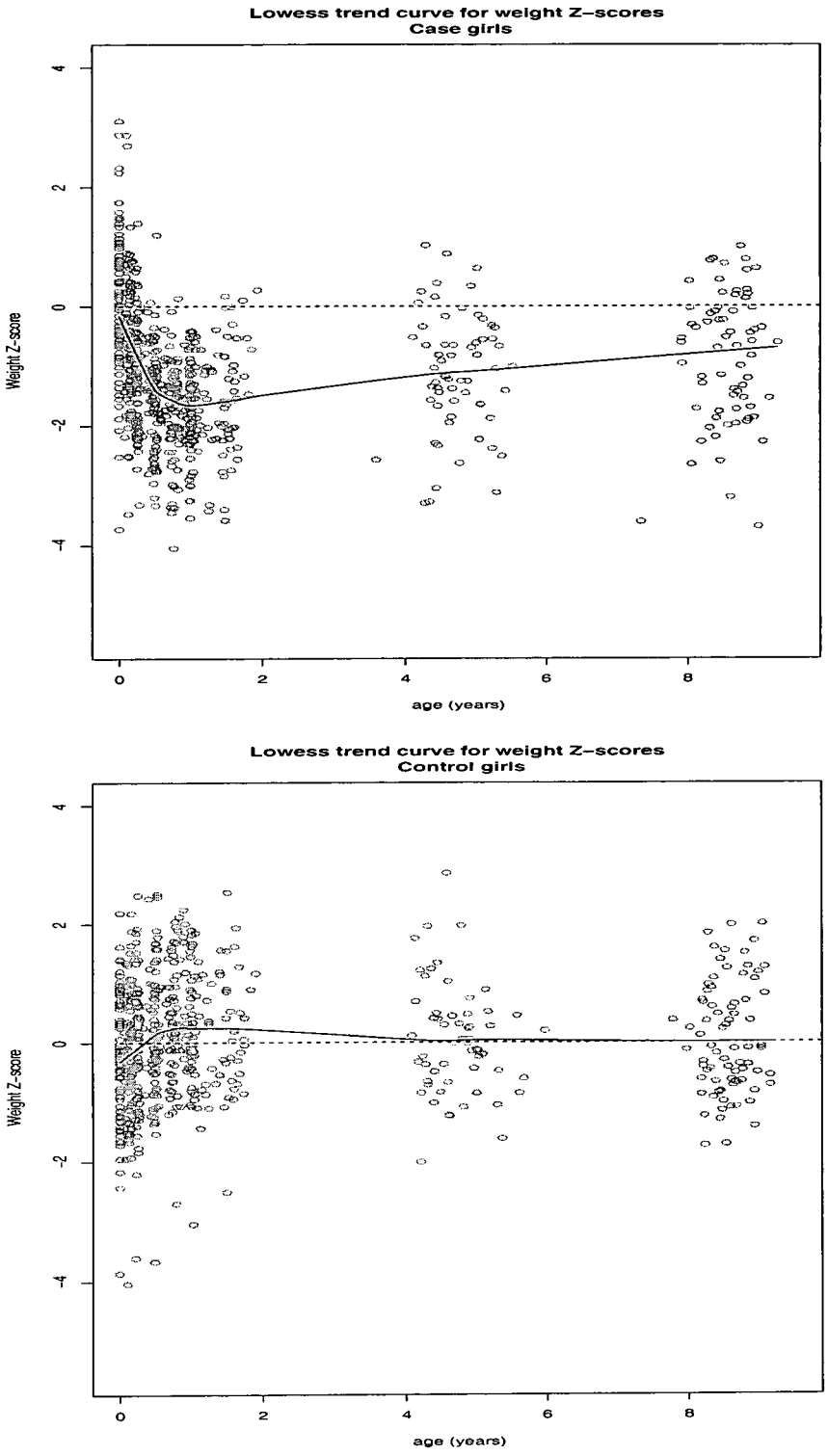


Figure 9.20: **Systematic sample (excluding cases)** Plot of weight Z-score versus age for boys (upper panel) and girls (lower panel) with default lowess curve

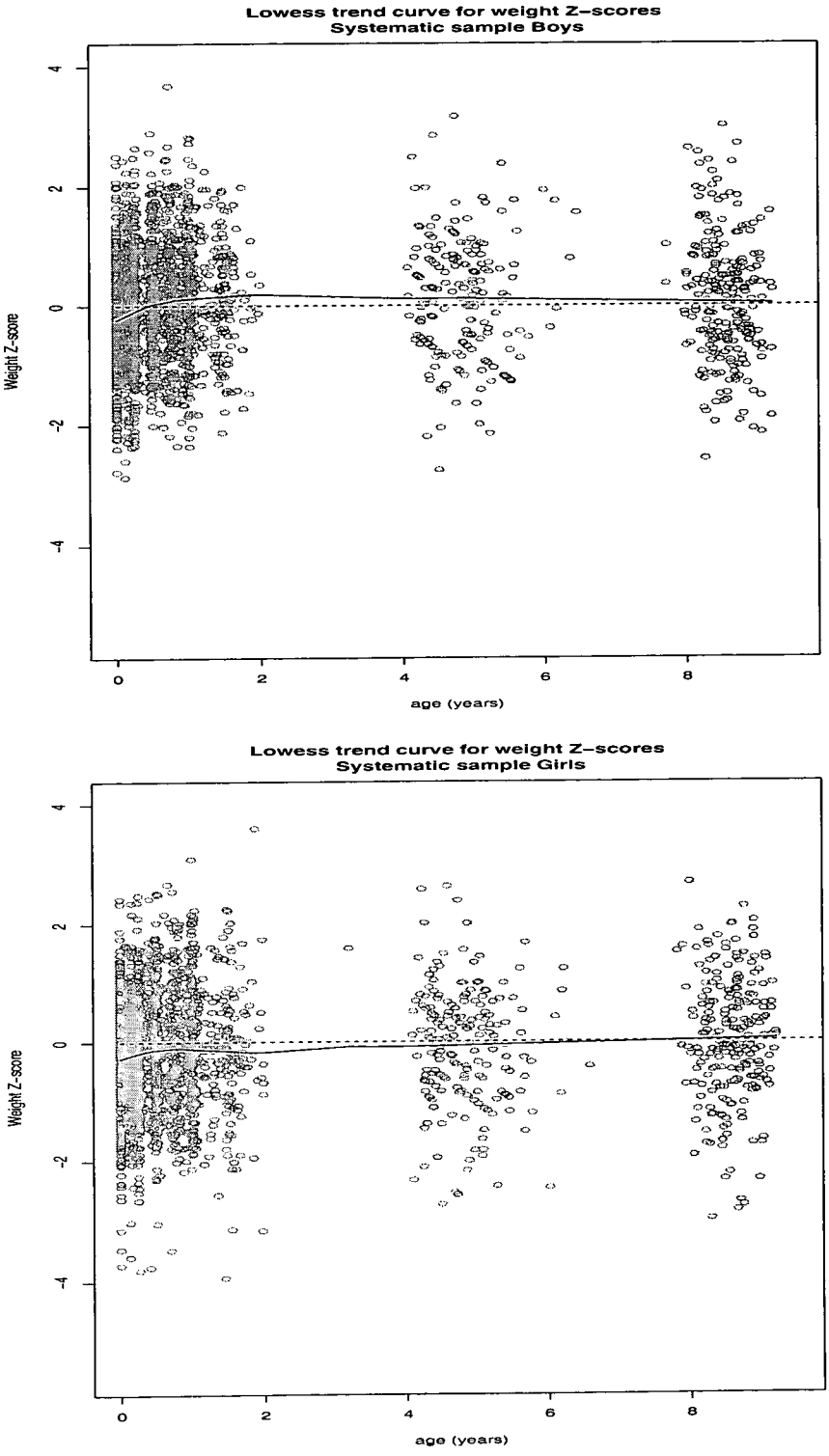
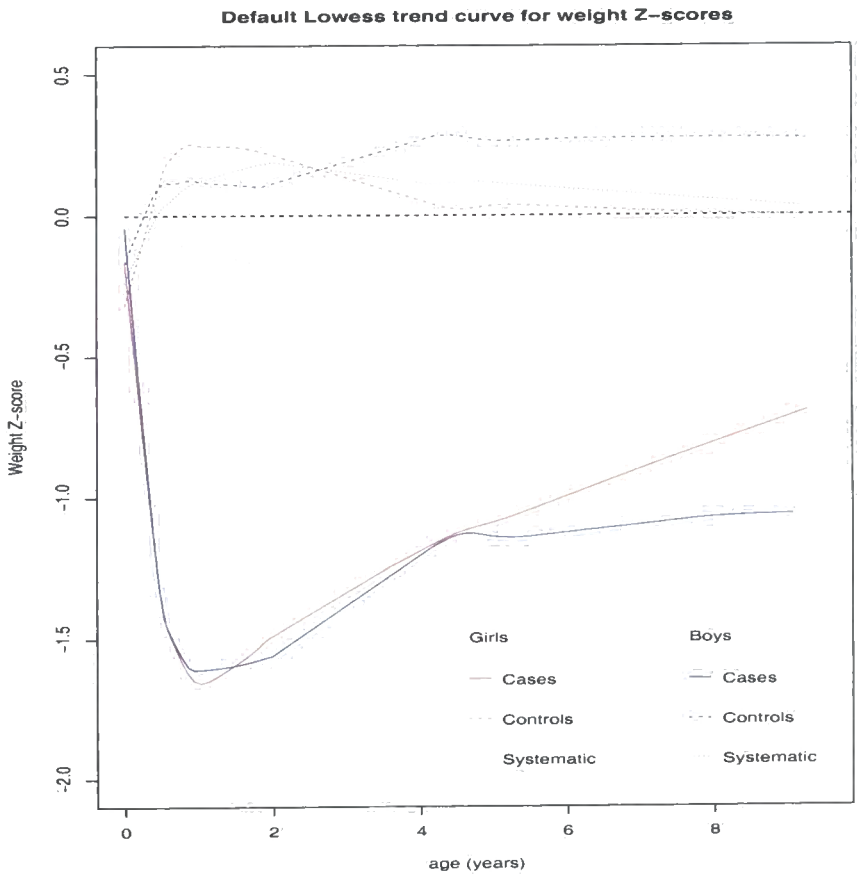


Figure 9.21: Birth to 9 years: Summary lowess trend curves of weight Z-scores for cases, controls and systematic sample by gender



# Chapter 10

# Conclusions

The statistical analysis of child growth data has been of interest to many academics, from a wide range of disciplines, over the last century. In this thesis we have focused mainly on one aspect of child growth, namely assessing a child's weight or height gain. An approach based on the conditional weight gain Z-score was used in infancy to assess an infant's current weight Z-score given their previous weight Z-scores. We also explored how conditional weight gain Z-scores in infancy could be utilised as they evolved. The conditional gain Z-score was also used in childhood to assess a child's current height Z-score given their previous height Z-score or mid-parental height Z-score.

## 10.1 Modelling correlation

In order to calculate a conditional weight gain Z-score all that is needed is a growth reference to convert weights to Z-scores and the correlation structure of weight Z-scores in infancy. In this thesis we have modelled the correlation structure of weight Z-scores in infancy to allow us to calculate the correlation for any pair of ages within infancy. We have arrived at a very simple model, termed the Argyle model for convenience, for the correlation between weight Z-scores: $r = \left(\frac{t1+1.1}{t2+1.1}\right)^{0.24}$ where $t1$ and $t2$ are age in weeks and $t1 < t2$. An interesting, and important, result is that the use of this one variable Argyle model to model the correlation between weight Z-scores in infancy is that conditioning on the previous weight Z-score is the same as conditioning on all previous weight Z-scores (Markov property). Thus to assess recent weight gain we only need to consider the previous weight Z-score.

Cole (1995) and Cole (1998a) had modelled the correlations generated from the

weight Z-scores in the Cambridge infant study. Cole (1995) applied the Fisher's transformation to these correlations and developed a model that had five terms and an intercept. The Argyle model developed here has fewer terms than the model proposed by Cole (1995). The Argyle model also provides a reasonable fit, in terms of the multiple correlation coefficient, to the correlation between weight Z-scores from the Cambridge infant study. However, the Argyle model is a monotonic function and the Cambridge correlations exhibit an upward trend towards the end of infancy indicating the need for a quadratic term. The upward trend in the Cambridge correlations at the end of infancy is counterintuitive and is not a feature observed in the Newcastle correlations. Furthermore, Heimendinger and Laird (1983) highlight the importance of using correlations pertinent to the target population.

The Newcastle infancy data generated a large number of correlations and there is scope for further research into models for these correlations. In this thesis we have focused on deriving the simplest model, a more complex model may result in a closer fit to the Newcastle correlations. A further improvement of the fit of the Argyle model could be achieved by additional work on the weightings of the correlations.

In childhood the Newcastle data is too sparse to generate correlation matrices for weight or height Z-scores. There are no published correlation matrices for weight Z-scores in childhood; so calculation of conditional weight gain Z-scores in childhood was not possible. However, Cole (1997a) published a correlation matrix for heights from a French longitudinal study. It was found that the Argyle model also provided a reasonable fit to the height correlations: $\log(r) = 0.0158 + 0.0929\log(t1 - 70/52) - 0.0997\log(t2 - 70/52)$ where $t1$ and $t2$ are age in years and $t1 < t2$. This model for correlation was used to calculate conditional height gain Z-scores in childhood for the Newcastle data. It was interesting that the form of the model developed for the correlation between weight Z-scores in infancy was equally valid for the correlation between height Z-scores in childhood. Seventy weeks was subtracted from age in the model for correlation, this has no biological significance and was deduced from exploratory data analysis. It would be interesting to see if this same functional form would be suitable for the correlation between weight Z-scores in childhood.

## 10.2 Functional Data analysis

As discussed in chapter 2, the aim of curve registration is to display growth data in a way that highlights the characteristics of the growth process, i.e. spurts, troughs

and levelling off. Once this is achieved, it is then possible to explore important sources of pattern and variation. At present functional data analysis (structural averaging) has only been used to characterise growth in childhood (Gasser et al. 1989; Gasser et al. 1990; Gasser et al. 1991; Gasser et al. 1991; Gasser et al. 1993; Gasser et al. 1994; Gasser et al. 1994). Therefore, there use of functional data analysis to characterise growth in infancy is a potential area of further research. Once characterisation of the growth process in infancy has been achieved it would be of potential value in comparing groups of infants e.g. boys and girls, impact of deprivation. Functional data analysis also has potential scope for discriminating between infants that are failing to thrive and infants that are growing 'normally'.

In curve registration there is initially a need to determine a set of characteristic times in infancy. In infancy it is expected that characteristic points may be conception, birth and period of weight loss immediately after birth. Further research is needed to determine suitable characteristic points in mid to late infancy. The non-parametric approach used in functional data analysis is heavily reliant on having large quantities of data for each individual. Usually this approach uses splines to model growth curves and the quality of these splines are highly dependent on the number of measurements. The Newcastle infancy data frame has too few weights for each individual to directly employ the techniques outlined in Ramsay and Silverman (1997) using their software. However, recently James et al. (2000) presented techniques for addressing the situation where data is at irregular time points and sparse. Another possibility may be apply any of the infancy models or multi-level modelling to the weight data of each individual and then apply functional data analysis techniques to the individual curves to align and analyse them.

## 10.3 Characteristics of Newcastle data

The Newcastle infancy data frame is formed from recorded routine weights for all infants resident in Newcastle in November 1989. It is suspected that these weights in infancy are likely to be typical of routine weights recorded in community clinics in other northern cities in the UK. Research suggests that the tempo of growth is influenced by many factors: regional variations, urbanisation, ethnic, socio-economic, disease and nutrition (Tanner 1989). The mode of feeding of infants from the Newcastle study is unknown, there will be a mix of bottle fed and breast fed infants within the sample.

Infants from the Cambridge study are breast fed and likely to be from a middle

class or affluent background. Weights from the Cambridge study are measured in a research setting and thus will have a lower measurement error than community weights. Therefore, the correlation between weight Z-scores is likely to be higher than those from a community setting, but Cole (1998a) argues that this increase will be likely to be offset to some extent by the homogeneous nature of the selection of Cambridge infants. Thus it could be argued that the research on correlations between weight Z-scores in Newcastle may be more relevant to weight measures on infants in the community.

## 10.4 Implications of research for clinicians

Overall it would appear that the revised UK 1990 growth reference is suitable for converting heights, weights and BMIs of Newcastle children to Z-scores. In infancy females from Newcastle tend to be lighter than those children that contribute to the reference. Whereas male infants from Newcastle tend to be heavier than those children that contribute to the reference. However, these differences are viewed to be not clinically significant. The variance of weight Z-scores in late infancy appears to be slightly greater than one. Therefore, we should be cautious in interpretation of weight Z-scores in the tails of the distribution in late infancy. At 4 to 5 years the revised UK 1990 reference appears to be adequate for converting weights, heights and BMIs to Z-scores. There may be some indication that girls at school entry are shorter than those children that contribute to the reference. At 7 to 9 years the use of the revised UK 1990 reference to convert weight, height and BMI to Z-scores is not entirely inappropriate. However, some caution may be needed when interpreting the BMI Z-scores of boys. However the UK 1990 growth reference is unsuitable for converting head circumference measures to Z-scores. Children in Newcastle have significantly lower head circumferences than those children that contribute to the reference (predominantly from Edinburgh growth study).

In contrast to the Wessex growth study (Voss and Mulligan 1999b), the body mass indices of Newcastle children from 4 to 9 years are in reasonable agreement with the children that contribute to the revised UK 1990 reference. However, some caution may be needed in interpreting the Z-score for BMI at the extremes at 7 to 9 years for boys, because there is reason to doubt that the variance of the Z-score for BMI is one. The Z-score for weight conditional on height correlates reasonably well with the Z-score for BMI. However, the use of the Z-score for BMI is advocated over the Z-score for weight conditional on height because there is evidence to suggest

that this conditional index has a non-zero mean.

In this thesis we have highlighted the benefits of using conditional weight gain Z-scores to assess weight gain in infancy. The Argyle model is a simple model that is suitable for use in calculating the correlation between weight Z-scores up to the age of about 18 months. The Argyle model developed for correlation does have implications for utilising conditional weight gain Z-scores: conditioning on the previous weight Z-score is the same as conditioning on all previous weight Z-scores.

A cautionary note is that the conditional approach assumes that the weight conditioned on itself is not abnormal (Healy 1986). It only takes one 'unusual' weight Z-score (in relation to rest of weight Z-scores of that individual) to produce very extreme conditional weight gain Z-scores. Therefore it would be unwise to take action after one extreme conditional weight gain Z-score. Furthermore, it is suggested that three weights are the bare minimum before any decision can be made about a child's weight gain. We have not being able to address here suitable spacing between weight measures and this is an area which could merit further research. Where growth is rapid, such as in early infancy, weight measures could be closer in time than say in late infancy. A further concern is that conditional gain Z-scores may have an elevated variance (greater than one) because of an elevated variance in either or both of the original Z-scores. Therefore the use of conventional cut-offs (0.4th centile or 2nd centile) is not advocated.

The reported parental heights were available for children that participated in the follow-up study. The girls' heights were found to be more correlated with the reported parental heights than the boys' heights. Preliminary research suggests that the use of the Z-score for height conditional on mid-parental height to assess a child's current height looks promising. There is no reason to doubt that the variance of the Z-score for child's height conditional on mid-parental height is one. However, this conditional index should be interpreted cautiously for boys as there is evidence to suggest that the mean is non-zero.

Research suggests that conditioning on previous weight Z-score is of value for detecting failure to thrive or growth faltering. An alternative criterion was proposed, the first two (or three) consecutive conditional weight gain Z-scores are 'negative' would identify infants that are 'at risk'. It may also be worth ignoring birth weight and employing the criterion that second and third consecutive conditional weight gain Z-scores are 'negative'.

The lowess procedure was used to characterise the growth patterns of infants that were identified as failing to thrive by the 'thrive index' approach (Wright et al.

1994) and infants growing 'normally' (full Newcastle data frame excluding cases). Infants that fail to thrive in infancy, in general, experience a rapid deceleration in weight gain from birth and the deceleration slows towards the end of infancy. This means that children continue to gain less weight than predicted for the individual, but with smaller systematic differences as the child ages.

It would appear that individuals that fail to thrive in infancy have significantly lower attained heights (even after adjusting for mid-parental height) and weights at school entry and 7-9 years. Furthermore, it would appear that failing to thrive in infancy has a greater impact on boys attained heights and weights in childhood. This observation may suggest that boys may be more vulnerable to growth faltering in infancy. It would appear that the conditional height gain Z-scores, between school entry and follow-up, are similar for case and control children. If we consider the measured mothers heights, then there is no significant difference between the heights of case and control mothers. However, the reported heights of fathers, indicated that fathers of case children were significantly shorter than fathers of control children. Thus suggesting that there may be some genetic component to the growth faltering in early infancy. This observation has not been noted elsewhere.

# Bibliography


Abidi, H., J. Borms, W. Duquet, and J. Pontier (1996). Bayesian estimation of the parameters of a non-linear model. an application to human height. *Growth Development and Aging 60*, 113–129.

Ahmed, M. L., P. L. Yudkin, J. A. Macfarlane, K. McPherson, and D. B. Dunger (1990). Are measurements of height made by health visitors sufficiently accurate for routine screening of growth? *Archives of Disease in Childhood 65*, 1345–1348.

Alsop-Shields, L. and H. Alexander (1997). A study of the errors that can occur when weighing infants. *Journal of Advanced Nursing 25*, 587–594.

Armitage, P. and G. Berry (1987). *Statistical methods in medical research* (2nd ed.). Blackwell, Oxford.

Atkinson, A. C. (1985). *Plots, Transformations and Regression : an introduction to graphical methods of diagnostic regression analysis.* Clarendon Press, Oxford.

Ayoub, C., D. Pfeifer, and L. Leichtman (1979). Treatment of infants with non-organic failure to thrive. *Child Abuse and Neglect 3*, 937–941.

Bailey, B. J. R. (1994). Monitoring the heights of prepubertal children. *Annals of Human Biology 21*(1), 1–11.

Barker, D. J. P. (Ed.) (1992). *Fetal and infant origins of adult disease.* British Medical Journal Publishing.

Barker, D. J. P. (1999). Early growth and cardiovascular disease. *Archives of Disease in Childhood 80*, 305–310.

Barker, D. J. P., C. Fall, and C. Osmond (1991). Fetal and infant growth and impaired glucose intolerance. *British Medical Journal 303*, 1474–75.

Barker, D. J. P., C. Osmond, J. Golding, D. Kuh, and M. E. J. Wadsworth (1989). Growth in utero, blood pressure in childhood and adult life, and mortality from cardiovascular disease. *British Medical Journal 298*, 564–567.

Barker, D. J. P., P. D. Winter, C. Osmond, B. Margetts, and S. J. Simmonds (1989). Weight in infancy and death from ischaemic heart disease. *Lancet i*, 577–580.

Batchelor, J. and A. Kerslake (1990). *Failure to find failure to thrive.* London: Whiting and Bush.

Baumgartner, R. N., A. F. Roche, and J. H. Himes (1986). Incremental growth tables. *American Journal of Clinical Nutrition 43*, 711–722.

Bayley, N. and S. R. Pinneau (1952). Tables for predicting adult height from skeletal age. revised for use with the greulich-pyle hand standards. *Journal of Pediatrics 40*, 423–441.

Bee, H. (1995). *The growing child.* Harper Collins.

Berglund, G. and E. Rabo (1973). A long term follow up investigation of patients with hypotrophic pyloric stenosis - with special reference to the physical and mental development. *Acta Paediatrica Scand 62*, 125–129.

Berkey, C. S. (1982a). Bayesian approach for a non-linear growth model. *Biometrics 38*, 953–961.

Berkey, C. S. (1982b). Comparison of two longitudinal growth models for preschool children. *Biometrics 13*(2), 111–128.

Berkey, C. S., D. W. Dockery, X. Wang, D. Wypij, and B. Ferris (1993). Longitudinal height velocity standards for U.S. adolescents. *Statistics in Medicine 12*, 403–414.

Berkey, C. S., J. Gardner, and G. A. Colditz (1998). Blood pressure in adolescence and early adulthood related to obesity and birth size. *Obesity Research 6*, 187–195.

Berkey, C. S. and R. L. Kent (1983). Longitudinal principal components and non-linear regression models of early childhood growth. *Annals of Human Biology 10*(6), 523–536.

Berkey, C. S. and N. M. Laird (1986). Nonlinear growth curve analysis: estimating the population parameters. *Annals of Human Biology 13*(2), 111–128.

Berkey, C. S., N. M. Laird, I. Valadian, and J. Gardner (1989). The analysis of longitudinal growth data with covariates. In J. Tanner (Ed.), *Auxology 88: Perspectives in the science of growth and development*, pp. 31–39. London: Smith-Gordon.

Berkey, C. S. and R. B. Reed (1987). A model for describing normal and abnormal growth in early childhood. *Human Biology 59*, 973–987.

Berkey, C. S., R. B. Reed, and I. Valadian (1983a). Longitudinal growth standards for preschool children. *Annals of Human Biology 10*(1), 57–67.

Berkey, C. S., R. B. Reed, and I. Valadian (1983b). Midgrowth spurt in height of boston children. *Annals of Human Biology 10*, 25–30.

Berkowitz, C. D. (1985). Comprehensive management of failure to thrive: an interdisciplinary approach. In D. Drotar (Ed.), *New developments in failure to thrive*, pp. 193–210. Plenum Press, New York.

Berwick, D. M., J. C. Levy, and R. Kleinerman (1982). Failure to thrive: diagnostic yield of hospitalisation. *Archives of Disease in Childhood 57*, 347–351.

Beyer, R. and H. Doerr (1998). Observations of reported and measured heights of mothers of short statured children. *Annals of Human Biology 25*(4), 387–390.

Bialik, O., E. Peritz, and A. Arnon (1973). Weight growth in infancy. *Human Biology 45*(1), 81–93.

Bithoney, W. G., J. McJunkin, J. Mchalek, H. Egan, J. Snyder, and A. Munier (1989). Prospective evaluation of weight gain in both non-organic and organic failure to thrive children: an outpatient trial of a multidisciplinary team intervention strategy. *Developmental Behaviour in Pediatrics 10*, 27–31.

Blair, P. S., P. Nadin, T. J. Cole, P. J. Fleming, I. J. Smith, M. W. Platt, P. J. Berry, J. Golding, and the CESDI SUDI research group (2000). Weight gain and sudden infant death syndrome: changes in weight Z-scores may identify infants at increased risk. *Archives of Disease in Childhood 82*, 462–469.

Bland, J. M. and D. G. Altman (1994a). Statistics notes: Regression towards the mean. *British Medical Journal 308*, 1499.

Bland, J. M. and D. G. Altman (1994b). Statistics notes: Some examples of regression towards the mean. *British Medical Journal 309*, 780.

Bock, R. D. (1995). Growth models. In R. Hauspie, G. Lindgren, and F. Falkner (Eds.), *Essays on Auxology Presented to James Mourilyan Tanner by Former colleagues and fellows*, pp. 28–38. Welywyn Garden City: Castlemead Publications.

Bock, R. D. and D. Thissen (1976). Fitting multi-component models for growth in stature. In *Proceedings of the 9th International Biometric Conference*, pp. 431–442.

Bock, R. D. and D. Thissen (1980). Statistical problems of fitting individual growth curves. In F. E. Johnston, A. F. Roche, and C. Susanne (Eds.), *Human physical growth and maturation*, pp. 265–290. New York: Plenum.

Bock, R. D., S. H. C. Toit, and D. Thissen (1994). *AUXAL: Auxological analysis of longitudinal measurements of human stature.* Chicage: Scientific Software International.

Bock, R. D., H. Wainer, A. Peterson, D. Thissen, J. Murray, and A. Roche (1973). A parameterization for individual human growth curves. *Human Biology 45*, 63–80.

Boddy, J., D. Skuse, and B. Andrews (2000). The development sequelae of nonorganic failure to thrive. *Journal of Child Psychology and Psychiatry 41*(8), 1003–1014.

Boddy, J. M. and D. H. Skuse (1994). Annotation: The process of parenting in failure to thrive. *Journal of Child Psychology and Psychiatry 35*(3), 401–424.

Boryslawski, K. (1988). Structure of monthly increments of length, weight and head circumference in the first year: a pure longitudinal study of 200 Wroclaw infants. *Annals of Human Biology 15*(3), 205–212.

Box, G. E. P. and D. R. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society (series B) 26*, 211–252.

British Standards Institution (1990). Body measurements of boys and girls from birth up to 16.9 years.

Brook, C. G. D. (1983). Earlier recognition of abnormal stature. *Archives of Disease in Childhood 58*, 840.

Brook, C. G. D. (1998). Identification of abnormal growth. In *The Cambridge Encyclopaedia of Human Growth and Development*, pp. 264–265. Cambridge University Press.

Brook, G. C. D. and P. C. Hindmarsh (1991). Tests for growth hormone secretion. *Archives of Disease in Childhood 66*, 85–87.

Buckler, J. and J. Tanner (1997). Revision and update of the Tanner-Whitehouse clinical longitudinal charts for height and weight. *European Journal of Pediatrics 156*, 248–249. (letter).

Butler, G. E., M. McKie, and S. G. Ratcliffe (1989). An analysis of the phases of mid-childhood growth by synchronization of growth spurts. In J. M. Tanner (Ed.), *Auxology 88 : Perspectives in the science of growth and development*, pp. 77–84. London: Smith-Gordon.

Butler, G. E., M. McKie, and S. G. Ratcliffe (1990). The cyclical nature of prepubertal growth. *Annals of Human Biology 17*(3), 177–198.

Byrad, P., R. Siervogel, and A. Roche (1983). Familial correlations for serial measurements of recumbent length and stature. *Annals of Human Biology 10*(3), 281–293.

Byrad, P. J., S. Guo, and A. F. Roche (1991). Model fitting to early childhood length and weight data from the Fels longitudinal study of growth. *American Journal of Human Biology 3*, 33–40.

Cameron, N. (1980). Conditional standards for growth in height of British children from 5.0 to 15.99 years of age. *Annals of Human Biology 7*(4), 331–337.

Cameron, N. (2002). British growth charts for height and weight with recommendations concerning their use in auxological assessment. *Annals of Human Biology 29*(1), 1–10.

Carpenter, R. G. (2000). Weight gain and sudden infant death syndrome: changes in weight Z-scores may identify infants at increased risk (commentary). *Archives of Disease in Childhood 82*, 468–469.

Child Growth Foundation (1996a). Boys four-in-one growth charts (birth to 20 years).

Child Growth Foundation (1996b). The British 1990 growth reference. Revised September 1996.

Child Growth Foundation (1996c). Girls four-in-one growth charts (birth to 20 years).

Child Growth Foundation (1996d). Three-in-one infant weight monitoring chart (birth - 60 weeks).

Chinn, S., T. J. Cole, M. A. Preece, and R. J. Rona (1996). Growth charts for ethnic populations in the UK. *Lancet 347*, 839–840.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association 79*, 829–836.

Cole, T. (1997a). Growth monitoring with the British 1990 growth reference. *Archives of Disease in Childhood 76*, 47–49.

Cole, T. (1998a). Presenting information on growth distance and velocity in one chart: practical issues of chart design. *Statistics in Medicine 17*, 2697–2707.

Cole, T., J. Freeman, and M. Preece (1998). British 1990 growth reference centiles for weight, height, BMI and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine 17*, 407–429.

Cole, T. J. (1985). A critique of the NCHS weight for height standard. *Human Biology 57*(2), 183–196.

Cole, T. J. (1988). Fitting smooth centile curves to reference data. *Journal of the Royal Statistical Society (series A) 151*(3), 385–418.

Cole, T. J. (1989a). Relating growth rate to environmental factors: methodological problems in the study of growth-infection interaction. *Acta Paediatrica Supplement 350*, 14–20.

Cole, T. J. (1989b). Using the LMS method to measure skewness in the NCHS and dutch national height standards. *Annals of Human Biology 16*, 407–419.

Cole, T. J. (1990). The LMS method for constructing normalized growth standards. *European Journal of Clinical Nutrition 44*, 45–60.

Cole, T. J. (1993). The use and construction of anthropometric growth reference standards. *Nutrition Research Reviews 6*, 19–50.

Cole, T. J. (1994a). Centile charts for both cross-sectional and longitudinal data. *Statistics in Medicine 13*, 2477–2492.

Cole, T. J. (1994b). Do growth chart centiles need a face lift? *British Medical Journal 308*, 641–2.

Cole, T. J. (1995). Conditional reference charts to assess weight gain in British infants. *Archives of Disease in Childhood 73*, 8–16.

Cole, T. J. (1996). Some questions about how growth standards are used. *Hormone Research 45*(suppl. 2), 18–23.

Cole, T. J. (1997b). 3-in-1 weight-monitoring chart. *Lancet 349*(9045), 102–103.

Cole, T. J. (1998b). The three-in-one infant weight monitoring chart. In *The Cambridge Encyclopaedia of Human Growth and Development*, pp. 66–67. Cambridge University Press.

Cole, T. J. (2000a). Galton's midparent height revisited. *Annals of Human Biology 27*(4), 401–405.

Cole, T. J. (2000b). A simple chart to identify non-familial short stature. *Archives of Disease in Childhood 82*, 173–176.

Cole, T. J., C. Bellizzi, K. M. Flegal, and W. H. Dietz (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. *British Medical Journal 320*, 1240–1243.

Cole, T. J., J. V. Freeman, and M. A. Preece (1995). Body mass index reference curves for UK, 1990. *Archives of Disease in childhood 75*, 23–29.

Cole, T. J. and P. J. Green (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine 11*, 1305–1319.

Cole, T. J., A. A. Paul, M. Eccles, and R. G. Whitehead (1989). The use of a multiple growth standard to highlight the effects of diet and infection on growth. In J. M. Tanner (Ed.), *Auxology 88 : Perspectives in the science of growth and development*, pp. 91–100. London: Smith-Gordon.

Cole, T. J., C. Power, and M. A. Preece (1999). Child obesity and body mass index. *Lancet 353*(9159), 1188. (letter).

Cole, T. J. and M. J. Roede (1999). Centiles of body mass index for dutch children aged 0-20 years in 1980 - a baseline to assess recent trends in obesity. *Annals of Human Biology 26*(4), 303–308.

Colley, N. V., J. M. Tremble, G. L. Henson, and T. J. Cole (1991). Head circumference/abdominal circumference ratio, ponderal index and fetal malnutrition. Should head circumference/abdominal circumference ratio be abandoned. *British Journal of Obstetrics and Gynaecology 98*, 524–527.

Corbett, S., R. F. Drewett, and C. M. Wright (1996). Does a fall down a centile chart matter? the growth and developmental sequelae of mild failure to thrive. *Acta Paediatrica 85*, 1278–1283.

Corbett, S. S. (1994). Cognitive and behavioural outcomes of non-organic failure to thrive. Master's thesis, Psychology, University of Durham.

Corbett, S. S. (1998). *Intelligence and reading abilities in eight year old children who failed to thrive in infancy.* Ph. D. thesis, Psychology, University of Durham.

Cotterill, A. M., W. H. Majrowski, S. Hearn, S. Jenkins, M. A. Preece, and M. O. Savage (1996). The potential effect of the uk 1990 height centile chart on community growth surveillance. *Archives of Disease in Childhood 74*, 452–454.

Cotterill, A. M., W. H. Majrowski, S. J. Hearn, and S. Jenkins (1993). Assessment of the reliability of school nurse height measurements in an inner city population. *Child: Care, Health and Development 19*(3), 159–165.

Count, E. (1943). Growth patterns of the human physique: an approach to kinetic anthropometry. *Human Biology 15*, 1–32.

Cronk, C. E. (1978). Growth of children with Down's syndrome: birth to age 3 years. *Pediatrics 61*, 564–568.

Darby, S. C. and T. Fearn (1979). The Chatham blood pressure study. an application of Bayesian growth curve models to a longitudinal study of blood pressure in children. *International Journal of Epidemiology 8*(1), 15–21.

Davies, D. P. and T. Williams (1983). Is weighing babies in clinics worthwhile? *British Medical Journal 286*, 860–863.

de Onis, M. and M. Blössner (1997). WHO global database on child growth and malnutrition. Technical report, World Health Organisation, Geneva, Switzerland.

de Onis, M. and C. Garza (1997). Time for a new growth reference. *Pediatrics 100*(5), E81–E82.

Deming, J. (1957). Application of the Gompertz curve to the observed pattern of growth in length of 48 individual boys and girls during the adolescent cycle of growth. *Human Biology 29*, 83–122.

Dewey, K. G., M. J. Heining, and L. A. Nommsen (1992). Growth of breast-fed infants from 0 to 18 months: The DARLING study. *Pediatrics 89*, 1035.

Dietz, W. H. and T. N. Robinson (1998). Use of body mass index (BMI) as a measure of overweight in children and adolescents. *Journal of Pediatrics 132*, 191–3.

Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics 44*, 959–971.

Diggle, P. J. and A. P. Verbyla (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics 54*, 401–415.

Donnelly, C., N. Laird, and J. Ware (1995). Prediction and creation of smooth curves. Temporally correlated longitudinal data. *Journal of American Statistical Association 90*(431), 984–989.

Doull, I. J. M., E. S. McCaughey, B. J. R. Bailey, and P. R. Betts (1995). Reliability of infant length measurements. *Archives of Disease in Childhood 72*, 520–521.

Dowdney, L., D. Skuse, E. Heptinstall, C. Puckering, and S. Zur-Szpiro (1987). Growth retardation and developmental delay among inner-city children. *Journal of Child Psychology and Psychiatry 28*, 529–541.

Dowdney, L., D. Skuse, K. Morris, and A. Pickles (1998). Short normal children and environmental disadvantage: A longitudinal study of growth and cognitive development from 4 to 11 years. *Journal of Child Psychology and Psychiatry 39*(7), 1017–1029.

Downie, A. B., J. Mulligan, R. J. Stratford, P. R. Betts, and L. D. Voss (1997). Are short normal children at a disadvantage? the Wessex growth study. *British Medical Journal 314*, 97–100.

Draper, N. R. and H. Smith (1981). *Applied Regression Analysis* (2 ed.). Wiley.

Draper, N. R. and H. Smith (1998). *Applied Regression Analysis* (3 ed.). Wiley.

Drewett, R. F., S. S. Corbett, and C. M. Wright (1999). Cognitive and educational attainments at school age of children who failed to thrive in infancy: a population based study. *Journal of Child Psychology and Psychiatry 40*(9), 551–561.

Drotar, D. (1990). Sampling issues in research with non-organic failure to thrive children. *Journal of Pediatric Psychology 15*(2), 255–272.

Editorial (1990). Failure to thrive revisited. *Lancet 336*, 662–663.

Editorial (1992). What happened to growth monitoring. *Lancet 340*, 149–150.

Edwards, A. G. K., P. C. Halse, J. M. Parkin, and A. J. R. Waterston (1990). Recognising failure to thrive in early childhood. *Archives of Disease in Childhood 65*, 1263–1265.

Ellerstein, N. S. and B. E. Ostrov (1985). Growth patterns in children hospitalised because of caloric-deprivation failure to thrive. *American Journal of Disease in Childhood 139*, 164–166.

Emery, J. L., A. Waite, R. Carpenter, S. R. Limerick, and D. Blake (1985). Apnoea monitors compared to weighing scales for siblings after cot death. *Archives of Disease in Childhood 60*, 1055–1060.

Eriksson, J. G., T. Forsen, J. Tuomileheto, P. D. Winter, C. Osmond, and D. J. P. Barker (1999). Catch up growth in childhood and death from coronary heart disease: a longitudinal study. *British Medical Journal 318*, 7–11.

Eveleth, P. B. and J. M. Tanner (1990). *Worldwide variation in human growth* (2 ed.). Cambridge University Press.

Falkner, F. (Ed.) (1960). *Child Development. An International Method of Study*. Basal: Karger.

Falkner, F. (1961). Croissance et developpement de l'énfant normal, une méthod internationale d'étude. Travaux et documents X11, Masson CIE, Paris.

Fatti, L., E. Senaona, and M. Thomson (1998). Bayesian updating in reference centile charts. *Journal of the Royal Statistical Society (series A) 161*(1), 103–115.

Fearn, T. (1975). A Bayesian approach to growth curves. *Biometrika 62*, 89–100.

Fomon, S. J. (1974). Normal growth, failure to thrive and obesity. In *Infant Nutrition*, pp. 34–94. W. B. Saunders, Philadelphia.

Fomon, S. J. (1991). Reference data for assessing growth of infants. *Journal of Pediatrics 119*, 415–416.

Foulkes, M. and C. Davis (1981). A index of tracking for longitudinal data. *Biometrics 37*, 439–446.

Foundation for the study of infant deaths (1985). Sheffield weight chart. London.

Freeman, J. V., T. J. Cole, S. Chinn, P. R. M. Jones, E. M. White, and M. A. Preece (1995). Cross sectional stature and weight reference curves for UK, 1990. *Archives of Disease in Childhood 73*, 17–24.

Freeman, J. V., C. Power, and B. Rodgers (1995). Weight-for-height indices of adiposity: Relationships with height in childhood and early adult life. *International Journal of Epidemiology 24*(5), 970–976.

Frongillo, E., G. Rothe, and J. Lambert (1990). Determining growth faltering with a tracking score. *American Journal of Human Biology 2*, 491–501.

Fryer, J. G. and J. Karlberg (1985). An approach to the estimation of growth standards. *Annals of Human Biology 12*, 83. (suppl1).

Gairdner, D. and J. Pearson (1971). A growth chart for premature and other infants. *Archives of Disease in Childhood 46*, 783–787.

Gairdner, D. and J. Pearson (1985). Revised Gairdner-Pearson growth charts. *Archives of Disease in Childhood 60*, 1202.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute 15*, 246–263.

Garn, S. M. and M. LaVelle (1984). Interaction between maternal size and birth size and subsequent weight gain. *American Journal of Clinical Nutrition 40*, 1120–1121. (letter).

Garn, S. M., W. R. Leonard, and V. M. Hawthorne (1986). Three limitations of the body mass index. *American Journal of Clinical Nutrition 44*, 996–997.

Garner, P., R. Panpanich, and S. Logan (2000). Is routine monitoring effective? A systematic review of trials. *Archives of Disease in Childhood 82*, 197–201.

Gasser, T. (1985). An analysis of the mid-growth and adolescent spurts of height based on acceleration. *Annals of Human Biology 12*(2), 129–148.

Gasser, T., A. Kneip, A. Binding, R. Largo, A. Prader, and L. Molinari (1989). Flexible methods for nonparametric fitting of individual and

sample curves. In J. M. Tanner (Ed.), *Auxology 88 : Perspectives in the science of growth and development*, pp. 23–30. London: Smith-Gordon.

Gasser, T., A. Kneip, A. Binding, A. Prader, and L. Molinari (1991). The dynamics of linear growth in distance, velocity and acceleration. *Annals of Human Biology 18*(3), 187–205.

Gasser, T., A. Kneip, P. Ziegler, R. Largo, L. Molinari, and A. Prader (1991). The dynamics of growth of width in distance, velocity and acceleration. *Annals of Human Biology 18*(5), 449–461.

Gasser, T., A. Kneip, P. Ziegler, R. Largo, and A. Prader (1990). A method for determining the dynamics and intensity of average growth. *Annals of Human Biology 17*(6), 459–474.

Gasser, T., A. Kneip, P. Ziegler, L. Molinari, A. Prader, and R. Largo (1994). Development and outcome of indices of obesity in normal children. *Annals of Human Biology 21*(3), 275–286.

Gasser, T., H.-G. Müller, W. Kohler, L. Molinari, and A. Prader (1984). Nonparametric regression analysis of growth curves. *The Annals of statistics 12*, 210–229.

Gasser, T., H.-G. Müller, and V. Mammitzch (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society (series B) 47*, 238–252.

Gasser, T., P. Ziegler, A. Kneip, A. Prader, L. Molinari, and R. Largo (1993). The dynamics of growth of weight, circumferences and skinfolds in distances, velocity and acceleration. *Annals of Human Biology 20*(3), 239–259.

Gasser, T., P. Ziegler, R. Largo, L. Molinari, and A. Prader (1994). A longitudinal study of lean and fat areas of the arm. *Annals of Human Biology 21*(4), 303–314.

Gasser, T., P. Ziegler, B. Seifert, A. Prader, L. Molinari, and R. Largo (1994). Measures of body mass and obesity from infancy to adulthood and their appropriate transformations. *Annals of Human Biology 21*, 111–126.

Geisser, S. (1970). Bayesian analysis of growth curves. *Sankhya*, 53–64.

Geisser, S. (1980). Growth curve analysis. In P. R. Kris (Ed.), *Handbook of Statistics*, Volume 1, pp. 89–115. Amsterdam : North-Holland.

Ghosh, B. K. (1966). Asymptotic expansions for the moments of the distribution of correlation coefficient. *Biometrika 53*, 258–262.

Giani, U., A. Filosa, and P. Causa (1996). A nonlinear model of growth in the first year of life. *Acta Paediatrica 85*, 7–13.

Gladstone, M., Y. Lasne, L. Sann, and G. Putet (1998). Non-linear patterns of growth in very low birth weight infants. *American Journal of Human Biology 10*, 637–646.

Goldstein, H. (1981a). Effects of age grouping on the estimate of a correlation coefficient. *Annals of Human Biology 8*(2), 181–183.

Goldstein, H. (1981b). Measuring the stability of individual growth patterns. *Annals of Human Biology 8*(6), 549–557.

Goldstein, H. (1986a). Efficient statistical modelling of longitudinal data. *Annals of Human Biology 13*, 129–141.

Goldstein, H. (1986b). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika 7*, 43–56.

Goldstein, H. (1989a). Efficient prediction of adult height. In J. M. Tanner (Ed.), *Auxology 88 : Perspectives in the science of growth and development*, pp. 41–48. London: Smith-Gordon.

Goldstein, H. (1989b). Models for multilevel response variables with an application to growth curves. In D. Bock (Ed.), *Multilevel analysis of educational data*, Chapter 6, pp. 107–125. Academic Press, New York.

Goldstein, H. (1995). *Multilevel Statistical Models* (2 ed.). Kendall's Library of Statistics 3. Arnold.

Goldstein, H., M. J. R. Healy, and J. Rabash (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine 13*, 1643–1655.

Goldstein, H. and H. Q. Pan (1992). Percentile smoothing using piecewise polynomials with covariates. *Biometrics 48*, 1057–1068.

Goldstein, H. and H. Q. Pan (1998, June). A general system for longitudinal growth norms. Seminar at London school of Hygiene & Tropical Medicine.

Gompertz, B. (1825). On the nature of the function of expressive law of human mortality and on a new method of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London 115*, 513–515.

Grady, J. J. and R. W. Helms (1995). Model selection techniques for the covariance matrix for incomplete longitudinal data. *Statistics in Medicine 14*, 1397–1416.

Graham, G. G., B. Adrianzen, J. Rabold, and E. D. Mellits (1982). Later growth in malnourished infants and children. *American Journal of Diseases of Children 136*, 348–352.

Greco, L., A. Capasso, C. De Fusco, and R. Paludetto (1990). Pulsatile weight increases in very low birth weight babies appropriate for gestational age. *Archives of Disease in Childhood 65*, 373–376.

Greco, L., C. Power, and C. Peckham (1995). Adult outcome of normal children who are short or underweight at age 7 years. *British Medical Journal 310*, 696–700.

Gregory, J. R., D. L. Collins, P. S. W. Davies, J. M. Hughes, and P. C. Clarke (1995). *National Diet and Nutrition Survey: children aged 1.5 to 4.5 years*, Volume 1. HMSO, London.

Grizzle, J. and D. Allen (1969). Analysis of growth and dose response curves. *Biometrics*, 357–381.

Guo, S., A. F. Roche, and W. M. Moore (1988). Reference data for head circumference and 1-month increments from 1 to 12 months of age. *Journal of Pediatrics 113*(3), 490–494.

Guo, S., R. M. Siervogel, A. F. Roche, and W. C. Chumlea (1992). Mathematical modelling of human growth: a comparative study. *American Journal of Human Biology 4*, 93–104.

Guo, S. M., A. F. Roche, S. J. Fomon, S. E. Nelson, W. C. Chumlea, R. R. Rogers, R. N. Baumgartner, E. E. Ziegler, and R. M. Siervogel (1991). Reference data on gains in weight and length during the first two years of life. *Journal of Pediatrics 119*(3), 355–362.

Guo, S. S., A. F. Roche, W. M. C. Chumlea, C. Johnson, R. J. Kuczmarski, and R. Curtin (2000). Statistical effects of varying sample size on the precision of percentile estimates. *American Journal of Human Biology 12*, 64–74.

Hall, D. M. B. (1996). *Health for all children.* Oxford University Press.

Hall, D. M. B. (2000). Growth monitoring. *Archives of Disease in Childhood 82*(1), 10–15.

Hall, D. M. B. and S. Stewart-Brown (1998). Screening in child health. *British Medical Bulletin 54*(4), 929–943.

Hamill, P. V. V., T. A. Drizd, C. L. Johnson, R. B. Reed, and A. F. Roche (1977). NCHS growth curves for children, birth - 18 years, united states. Vital and Health Statistics series 11. Data from the National Health Survey; number 165. DHEW publication number (PHS) 78-1650. Hyattsville: National Center for Health Statistics.

Hamill, P. V. V., T. A. Drizd, C. L. Johnson, R. B. Reed, A. F. Roche, and W. M. Moore (1979). Physical growth: National Center for Health Statistics percentiles. *American Journal of Clinical Nutrition 32*, 607–629.

Hannaway, P. J. (1970). Failure to thrive: a study of 100 infants and children. *Clinical Pediatrics 9*, 96–99.

Hauspie, R. C., A. Wachholder, G. Baron, F. Cantraine, C. Susanne, and M. Graffar (1980). A comparative study of the fit of four different functions to longitudinal data of growth in height of Belgian girls. *Annals of Human Biology 7*(4), 347–358.

Heagerty, P. J. and M. S. Pepe (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height for age in us children. *Journal of Royal Statistical Society (series C) 48*(4), 533–551.

Healy, M. J. R. (1962). The effect of age-grouping on the distribution of a measurement subject to growth. *American Journal of Anthropology 20*, 49–50.

Healy, M. J. R. (1974). Notes on the statistics of growth standards. *Annals of Human Biology 1*(1), 41–46.

Healy, M. J. R. (1986). Statistics of growth standards. In F. Falkner and J. M. Tanner (Eds.), *Human Growth* (2 ed.), Volume 3, pp. 47–58. New York: Plenum.

Healy, M. J. R. (1989a). Growth curves and growth standards - the state of the art. In J. M. Tanner (Ed.), *Auxology 88 : Perspectives in the science of growth and development*, pp. 13–21. London: Smith-Gordon.

Healy, M. J. R. (1989b). Measuring measuring errors. *Statistics in Medicine 8*, 893–906.

Healy, M. J. R. (1992). Normalizing transformations for growth standards. *Annals of Human Biology 19*(5), 521–526.

Healy, M. J. R. and H. Goldstein (1978). Regression to mean. *Annals of Human Biology 5*(32), 277–280.

Healy, M. J. R., J. Rasbash, and M. Yang (1988). Distribution free estimation of age related centiles. *Annals of Human Biology 15*, 17–22.

Healy, M. J. R., M. Yang, J. M. Tanner, and F. Y. Zumrawi (1988). The use of short term increments in length to monitor growth in infancy. In J. C. Waterlow (Ed.), *Linear growth retardation in less developed countries*, pp. 41–55. Raven, New York.

Heimendinger, J. and N. Laird (1983). Growth changes: measuring the effect of an intervention. *Evaluation Review 7*(1), 80–95.

Heptinstall, E., C. Puckering, D. Skuse, K. Start, S. Zur-Szpiro, and L. Dowdney (1987). Nutrition and mealtime behaviour in families of growth-retarded children. *Human Nutrition: Applied Nutrition 41A*, 390–402.

Hermanussen, M. (1995). No evidence for saltation in human growth. *Annals of Human Biology 22*, 341–345.

Hermanussen, M., C. Thiel, E. Von Buren, M. De Los Angeles Rol De Lama, A. Perez Romero, J. Ariznaverreta Ruiz, C.and Burmeister, and J. A. F. Tresguerres (1998). Micro and macro perspectives in auxology - findings and considerations upon the variability of short term and individual growth and the stability of population derived parameters. *Annals of Human Biology 25*(4), 359–385.

Himes, J. H. and A. F. Roche (1981). Parent specific adjustments for assessment of recumbent length and stature. *Monographs for pediatrics 13*, 1–88.

Himes, J. H. and A. F. Roche (1982). Reported versus measured adult statures. *American Journal of Physical Anthropology 53*, 335–341.

Hooper, P. M., D. C. Mayes, and N. N. Demianczuk (2002). A model for foetal growth and diagnosis of intrauterine growth restriction. *Statistics in Medicine 21*(1), 95–112.

Hufton, I. W. and R. K. Oates (1977). Non-organic failure to thrive: a long term follow up. *Pediatrics 59*, 73–77.

Hulse, J. A. and S. Schilg (1995). United Kingdom community growth screening 1994: a survey of current practice. *Journal of Medical Screening 2*, 154–156.

Ihaka, R. and R. Gentleman (1996). R: A language for data analysis and graphics. *Journal of Computional and Graphical Statistics 5*(3), 299–314.

James, G. M., T. J. Hastie, and C. A. Sugar (2000). Principal component models for sparse functional data. *Biometrika 87*(3), 587–602.

Jellinek, D. and D. M. B. Hall (1994). How are childrens growth problems diagnosed. *Child: Care, Health and Development 20*, 371–377.

Jenss, R. and N. Bayley (1937). A mathematical method for studying the growth of a child. *Human Biology 9*, 556–563.

Jolicoeur, P., J. Pontier, and H. Abidi (1992). Asymptotic models for longitudinal growth of human stature. *American Journal of Human Biology 4*, 461–468.

Jolicoeur, P., J. Pontier, M. O. Pernin, and M. Sempe (1988). A lifetime asymptotic growth curve for human height. *Biometrics 44*, 995–1003.

Jones, P. R. M., N. G. Norgan, M. G. Hunt, and R. H. Hooper (1993). British size surveys. Loughborough: Loughborough Consultants.

Joossens, J. V. and E. Brems-Heyns (1975). High power polynomial regression for the study of distance, velocity and acceleration of growth. *Growth 39*, 535–551.

Kanefuji, K. and T. Shohoji (1990). On a growth model of human height. *Growth Development and Ageing 54*, 155–165.

Karlberg, J. (1987). On the modelling of human growth. *Statistics in Medicine 6*, 185–192.

Karlberg, J., I. Engstrom, P. Karlberg, and J. G. Fryer (1987). Analysis of linear growth using a mathematical model. 1. from birth to 3 years. *Acta Paediatrica Scand 76*, 478–488.

Kelnar, C. J. H. (2000). Cost-benefit analysis is the key. *Archives of Disease in Childhood 83*, 176–177.

Kendall, M. and A. Stuart (1979). *The advanced theory of statistics : Inference and Relationship* (4 ed.), Volume 2. Charles Griffin & co. ltd.

Kent, R. L. (1975). *Principal component analysis of longitudinal growth data.* Ph. D. thesis, Harvard University.

Khamis, H. J. and S. Guo (1993). Improvement in the Roche-Wainer-Thissen stature prediction model: A comparative study. *American Journal of Human Biology 5*, 669–679.

Kolsteren, P. W., J. A. Kusin, and S. Kardjati (1997). Pattern of linear growth velocities of infants from birth to 12 months in Madura, Indonesia. *Tropical medicine and International Health 2*(3), 291–301.

Kouchi, M., A. F. Roche, and D. Mukherjee (1985a). Curve fitting for growth in weight during infancy with relationships to adult status and familial associations of the estimated parameters. *Human Biology 57*(2), 245–265.

Kouchi, M., A. F. Roche, and D. Mukherjee (1985b). Growth in recumbent length during infancy with relationships to adult status and familial associations of the estimated parameters of fitted curves. *Human Biology 57*, 449–472.

Koziel, S. M. (1997). Combined effects of the tempo of maturation and mid-parent height on the shape of individual growth curves. *American Journal of Human Biology 9*, 555–563.

Kristiansson, B. and S. P. Failstrom (1987). Growth at the age of 4 years subsequent to early failure to thrive. *Child Abuse and Neglect 11*, 35–40.

Kristiansson, B., J. Karlberg, and S. P. Failstrom (1981). Infants with low rate of weight gain. i. a study of organic factors and growth patterns. *Acta Paediatrica Scand 70*, 655–662.

Kronmal, R. A. (1993). Spurious correlation and the fallacy of the ratio standard revisited. *Journal of the Royal Statistical Society (series A) 156*(3), 379–392.

Kshirsagar, A. and W. Smith (1995). *Growth Curves*. Marcel Dekker, New York.

Kuczmarski, R. J., C. L. Ogden, L. M. Grummer-Strawn, K. M. Flegal, S. S. Guo, R. Wei, Z. Mei, L. R. Curtin, A. F. Roche, and C. L. Johnson (2000, December). CDC growthcharts: United states. Advance Data from Vital and Health Statistics; No. 314. Hyattsville, Maryland: National Center for Health Statistics.

Lacey, K. A. and J. M. Parkin (1974). The normal short child: Community study of children in Newcastle upon Tyne. *Archives of Disease in Childhood 49*, 417–424.

Laing, G. J. and E. B. Rossor (1996). Growth screening in schools: an evaluation of the programme in one district. *Child: Care, Health and Development 22*(1), 11–17.

Laird, A. K. (1967). Evolution of the human growth curve. *Growth 31*, 345–355.

Laird, N. M. and J. H. Ware (1982). Random effects models for longitudinal data. *Biometrics 38*, 963–974.

Lampl, M. and M. L. Johnson (1993). A case study of daily growth during adolescence: a single spurt or changes in dynamics of saltatory growth? *Annals of Human Biology 20*(6), 595–603.

Lampl, M. and M. L. Johnson (1997). Identifying saltatory growth patterns in infancy: A comparison of results based on measurement protocol. *American Journal of Human Biology 9*, 343–355.

Lampl, M., J. Velhuis, and M. Johnson (1992). Saltation and stasis: a model of human growth. *Science 258*, 801–803.

Largo, R. H., T. Gasser, A. Parader, W. Stützle, and P. Huber (1978). Analysis of adolescent growth spurt using smoothing spline functions. *Annals of Human Biology 5*, 421–434.

Lasker, G. W. and C. G. N. Mascie-Taylor (1989). Effects of social class differences and social mobility on growth in height, weight and body mass index in a British cohort. *Annals of Human Biology 16*(1), 1–8.

Law, C. M., M. de Swiet, C. Osmond, P. M. Fayers, D. J. P. Barker, A. M. Cruddas, and C. H. D. Fall (1993). Initiation of hypertension in utero and its amplification throughout life. *British Medical Journal 303*(6869), 24–27.

Ledford, A. and T. Cole (1998). Mathematical models of growth in stature throughout childhood. *Annals of Human Biology 25*(2), 101–115.

Li, J., W. J. Park, and A. F. Roche (1998). Decanalization of weight and stature during childhood and adolescence. *American Journal of Human Biology 10*, 351–359.

Lindley, D. V. and A. F. M. Smith (1972). Bayes estimates for linear model. *Journal of the Royal Statistical Society (series B) 34*, 1–41.

Lucas, A., S. M. Gore, T. J. Cole, M. F. Bamford, J. F. B. Dossetor, I. Barr, L. Dicarlo, S. Cork, and P. J. Lucas (1984). Multicentre trial

on feeding of low birth weight infants - effects of diet on early growth. *Archives of Disease in Childhood 59*, 722–730.

Lucas, E., M. S. Fewtrell, and T. J. Cole (1999). Fetal origins of adult disease - the hypothesis revisited. *British Medical Journal 319*(7204), 245–249.

Luo, Z. C., D. Albertsson-Wikland, and J. Karlberg (1998). Target height as predicted by parental heights in a population based study. *Pediatric Research 44*(4), 563–571.

Lyon, A. J., M. A. Preece, and D. B. Grant (1985). Growth curve for girls with Turner syndrome. *Archives of Disease in Childhood 60*, 932–5.

Maggioni, A. and F. Lifshitz (1995). Nutritional management of failure to thrive. *Pediatric Clinics of North America 42*(4), 791–810.

Mardia, K. V., J. T. Kent, and J. M. Bibby (1979). *Multivariate Analysis.* London: Academic Press.

Marshall, W. A. (1971). Evaluation of growth rate in height over periods of less than one year. *Archives of Disease in Childhood 46*, 414–420.

Marubini, E., L. F. Resele, and G. Barghini (1971). A comparative fitting of the Gompertz and logistic functions to longitudinal height data during adolescence in girls. *Human Biology 43*, 237–252.

McGill, R., J. W. Tukey, and W. A. Larsen (1978). Variations of boxplots. *American Statistician 32*, 12–16.

McMahan, C. A. (1981). An index of tracking. *Biometrics 37*, 447–455.

Merrell, M. (1931). The relationship of individual growth to average growth. *Human Biology 3*, 36–70.

Milani, S. (2000). Kinetic models for normal and impaired growth. *Annals of Human Biology 27*(1), 1–17.

Moore, T., C. B. Hindley, and F. Falkner (1954). A longitudinal research in child development and some of its problems. *British Medical Journal 2*, 1132.

Morrison, D. F. (1976). *Multivariate Statistical methods* (2 ed.). McGraw-Hill.

Mulligan, J. (2000, May). Definition of childhood overweight/obesity. (e–letter).

Mulligan, J., L. D. Voss, E. S. McCaughey, B. J. R. Bailey, and P. R. Betts (1998). Growth monitoring: testing the new guidelines. *Archives of Disease in Childhood 79*, 318–321.

Oates, R., A. Peacock, and D. Forest (1984). Development in children following abuse and non-organic failure to thrive: a long term follow-up. *American Journal of Disease in Childhood 138*, 765–767.

Ogden, C. L., R. J. Kuczmarski, K. M. Flegal, S. Guo, R. Wei, L. M. Grummer-Strawn, L. R. Curtin, A. F. Roche, and C. L. Johnson (2002). Centers for disease control and prevention 2000 growth charts. *Pediatrics 109*(1), 45–60.

Onat, T. (1995). Validation of methods for predicting adult stature in Turkish girls. *American Journal of Human Biology 7*, 757–767.

Palmer, S., S. Byford, and J. Raftery (1999). Types of economic evaluation. *British Medical Journal 318*(7194), 1349.

Pan, H. and H. Goldstein (1998). Multilevel repeated measures growth modelling using extended spline functions. *Statistics in Medicine 17*, 2755–2770.

Pan, H. Q. (1995, March). *Multilevel Models in Human Growth and Development research.* Ph. D. thesis, University College London, Institute of Education, School of mathematics and computing.

Pan, H. Q. and H. Goldstein (1997). Multilevel models for longitudinal growth norms. *Statistics in Medicine 16*, 2665–2678.

Pan, H. Q., H. Goldstein, and G. Di (1992). A two level cross-sectional model using grafted polynomials. *Annals of Human Biology 19*(4), 337–346.

Pan, H. Q. and S. G. Ratcliffe (1992). A new method of deriving velocity and acceleration curves for height from kernel estimation of distance. *Annals of Human Biology 19*(3), 303–316.

Park, W. J., J. Li, and A. F. Roche (1997). The decanalization of weight, recumbent length and head circumference during infancy. *American Journal of Human Biology 9*, 689–698.

Peerson, J. M., M. J. Heinig, L. A. Nommsen, B. O. Lonnerdal, and K. G. Dewey (1993). Use of growth models to describe patterns of length, weight and head circumference among breast fed and formula fed infants. *Human Biology 65*, 611–626.

Pere, A. (2000). Comparison of two methods for transforming height and weight to normality. *Annals of Human Biology 27*(1), 35–45.

Piwoz, E. G., J. M. Peerson, and K. H. Brown (1992). Potential for misclassification of infants growth increment by using existing reference data. *American Journal of Clinical Nutrition 56*, 58–64.

Pollitt, E. and R. Leibel (1980). Biological and social correlates of failure to thrive. In *Social and Biological predictors of nutritional status Physical growth and neurological development.* Academic press, New York.

Potthoff, R. F. and S. N. Roy (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika 51*, 237–252.

Power, C., J. K. Lake, and T. J. Cole (1997). Measurement and long-term health risks of child and adolescent fatness. *International Journal of Obesity 21*, 507–526.

Prader, A., J. M. Tanner, and G. A. Von Harnack (1963). Catch-up growth following illness or starvation: An example of developmental canalization in man. *Journal of Pediatrics 62*(3), 646–659.

Preece, M. A. (Ed.) (1998). *Cross-sectional versus longitudinal growth references*, Number 31. Proceedings of the Annual Spring meeting. York: Royal College of Paediatrics and Child Health.

Preece, M. A. and M. J. Baines (1978). A new family of mathematical models describing the human growth curve. *Annals of Human Biology 5*, 1–24.

Preece, M. A., J. V. Freeman, and T. J. Cole (1996). Sex differences in weight in infancy: published centile charts have been updated. *British Medical Journal 313*, 1486.

Preece, M. A. and I. Heinrich (1981). Mathematical modelling of individual growth curves. *British Medical Bulletin 37*(3), 247–252.

Quo, S. K. (1953). Mathematical analysis of the growth of man, with special reference to formosans. *Human Biology 25*, 333–358.

Ramsay, J., R. Bock, and T. Gasser (1995). Comparison of height acceleration curves in the Fels, Zurich and Berkley growth data. *Annals of Human Biology 22*(5), 413–426.

Ramsay, J. and X. Li (1998). Curve registration. *Journal of the Royal Statistical Society (series B) 60*(2), 351–363.

Ramsay, J. and B. W. Silverman (1997). *Functional Data Analysis.* Springer.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics 14*, 1–17.

Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis. In P. Krishnaiah (Ed.), *Multivariate Analysis*, pp. 87–103. Academic Press, New York.

Ratcliffe, S. G., N. Masera, H. Pan, and M. McKie (1994). Head circumference and IQ of children with sex chromosome abnormalities. *Developmental Medicine and Child Neurology 36*, 533–544.

Ratkowsky, D. A. (1983). *Nonlinear regression modelling - A unified practical approach.* Marcel Dekker, New York.

Raynor, P. and M. C. J. Rudolf (2000). Anthropometric indices of failure-to-thrive. *Archives of Disease in Childhood 82*, 364–365.

Reed, R. B. and C. S. Berkey (1989). Linear statistical model for growth in stature from birth to maturity. *American Journal of Human Biology 1*, 257–262.

Reilly, J. J. and A. R. Dorosty (1999). Epidemic of obesity in UK children. *Lancet 354*, 1874–1875.

Reilly, J. J., A. R. Dorosty, and P. M. Emmett (1999). Prevalence of overweight and obesity in British children: cohort study. *British Medical Journal 319*, 1039.

Reilly, J. J., S. A. H. Savage, C. H. S. Ruxton, and T. R. Kirk (1999). Suitability of revised UK reference data for the assessment of nutritional status in 7 year old children. *Journal of Human Nutrition and Dietetics 12*, 529–531.

Rice, J. A. (1995). *Mathematical Statistics and Data Analysis* (2 ed.). Duxbury Press.

Rice, N. and A. Leyland (1996). Multilevel models: applications to health data. *Journal of Health Services Research and policy 1*(3), 154–164.

Roche, A. F. (1992). *Growth, maturation and body composition: Fels longitudinal study 1929-1991.* Cambridge University Press.

Roche, A. F., S. Guo, and W. M. Moore (1989). Weight and recumbent length from 1 to 12 months of age: reference data for 1 month increments. *American Journal of Clinical Nutrition 49*, 599–607.

Roche, A. F. and J. H. Himes (1980). Incremental growth charts. *American Journal of clinical nutrition 33*, 2041–2052.

Roche, A. F., D. Mukherjee, and A. S. Guo (1986). Head circumference growth patterns: birth to 18 years. *Human Biology 58*(6), 893–906.

Roche, A. F., H. Wainer, and D. Thissen (1975). The RWT method for the prediction of adult stature. *Pediatrics 56*, 1026–1033.

Roddam, A. (1998). Childhood growth: A graphical modelling approach. Research students conference, Lancaster.

Rona, R. J. and S. Chinn (1986). National Study of Health and Growth: social and biological factors associated with height of children from ethnic groups living in England. *Annals of Human Biology 13*, 453–471.

Rona, R. J. and S. Chinn (1987). National Study of Health and Growth: social and biological factors associated with weight-for-height and triceps skinfold of children from ethnic groups in England. *Annals of Human Biology 14*, 231–248.

Rosenbaum, S., R. K. Skinner, I. B. Knight, and J. W. Garrow (1985). A survey of heights and weights of adults in great britain, 1980. *Annals of Human Biology 12*, 115–127.

Rosenn, D. W., L. S. Loeb, and M. B. Jura (1980). Differentiation of organic from non-organic failure to thrive syndrome in infancy. *Pediatrics 66*, 698–704.

Rossiter, J. E. (1991). Calculating centile curves using kernel density estimation with application to infant kidney lengths. *Statistics in Medicine 10*, 1693–1701.

Royston, P. (1991). Constructing time-specific reference ranges. *Statistics in Medicine 10*, 675–690.

Royston, P. (1995). Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal measurements. *Statistics in Medicine 14*, 1417–1436.

Royston, P. and D. G. Altman (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics 43*, 429–467.

Royston, P. and E. M. Wright (1998). A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials

and exponential transformation. *Journal of the Royal Statistical Society, (series A) 161*(1), 79–101.

Rudolf, M. C. J., T. J. Cole, A. J. Krom, P. Sahota, and J. Walker (2000). Growth of primary school children: a validation of the 1990 references and their use in growth monitoring. *Archives of Disease in Childhood 83*(4), 298–301.

Rudolf, M. C. J. and Z. Hochberg (1990). Are boys more vulnerable to psychosocial growth retardation. *Developmental Medicine and Child Neurology 32*, 1022–1025.

Savage, S. A. H., J. J. Reilly, C. A. Edwards, and J. V. G. A. Durnin (1999). Adequacy of standards for assessment of growth and nutritional status in infancy and early childhood. *Archives of Disease in Childhood 80*, 121–124.

Schilg, S. and T. Hulse (1997, June). Growth monitoring & assessment in the community: a guide to good practice. Child Growth Foundation.

Scott, J. A. (1961). *Report on the heights and weights (and other measurements) of school pupils in the County of London in 1959.* London County Council, London.

Sempé, M., G. Pédron, and M. Roy-Pernot (1979). Auxologie: Méthode et séquences. Theraplix, Paris.

Shaheen, E., D. Alexander, and M. Truskowsky (1968). Failure to thrive - a retrospective profile. *Clinical Pediatrics 7*, 255–261.

Sheskin, D. J. (1997). *Handbook of parametric and nonparametric statistical procedures.* CRC Press.

Shi, M., R. E. Weiss, and J. M. G. Taylor (1996). An analysis of paediatric CD4 counts for Acquired Immune Deficiency Syndrome using flexible random curves. *Applied Statistics 45*, 151–163.

Shohoji, T. and H. Sasaki (1985). An aspect of growth analysis of weight in savannah baboon. *Growth 49*, 500–509.

Shohoji, T. and H. Sasaki (1987a). An aspect of growth analysis of weight in savannah baboon. ii. Gender comparison by adjusting age. *Growth 51*, 425–431.

Shohoji, T. and H. Sasaki (1987b). Individual growth of stature of Japanese. *Growth 51*, 432–450.

Shrimpton, R., C. G. Victora, M. de Onis, R. Costa Lima, M. Blössner, and G. Clugston (2001). Worldwide timing of growth faltering: Implications for nutritional interventions. *Pediatrics 107*(5), U61–67.

Siegel, P. T., R. Clopper, and B. Stabler (1991). Psychological impact of significantly short stature. *Acta Paediatrica Scandinavia (Supplement) 377*, 14–18.

Simondon, K. B., F. Simondon, F. Delpeuch, and A. Cornu (1992). Comparitive study of five growth models applied to weight data from Congolese infants between birth and thirteen months of age. *American Journal of Human Biology 4*, 327–335.

Sinclair, D. (1985). *Human growth after birth* (4 ed.). Oxford University Press, Oxford.

Sinclair-Smith, C., F. Dinsdale, and J. Emery (1976). Evidence of duration and type of illness in children found unexpectedly dead. *Archives of Disease in Childhood 51*, 424–429.

Singhal, A., T. J. Cole, and A. Lucas (2001). Early nutrition in preterm infants and later blood pressure: two cohorts after randomised trials. *Lancet 357*, 413–419.

Skuse, D. (1987). The psychological consequences of being small. *Journal of Child Psychology and Psychiatry 28*(5), 641–650.

Skuse, D., A. Pickles, D. Wolke, and S. Reilly (1994). Postnatal growth and mental development: evidence for a 'sensitive period'. *Journal of Child Psychology and Psychiatry 35*(3), 521–545.

Skuse, D., S. Reilly, and D. Wolke (1994). Psychosocial adversity and growth during infancy. *European Journal of Clinical Nutrition 47*(i), 113–130.

Skuse, D. H. (1985). Nonorganic failure to thrive: a reappraisal. *Archives of Disease in Childhood 60*, 173–178.

Skuse, D. H. (1998). Growth and psychosocial stress. In *The Cambridge Encyclopaedia of Human Growth and Development*, pp. 341–342. Cambridge University Press.

Smith, C. A. and W. Berenberg (1970). The concept of failure to thrive. *Pediatrics 46*, 661–663.

Smith, D. W., W. Truog, J. E. Rogers, L. J. Greitzer, A. L. Skinner, J. J. McCann, and M. A. S. Harvey (1976). Shifting linear growth during

infancy: illustration of genetic factors in growth from fetal life through infancy. *Journal of Pediatrics 89*(2), 225–230.

Smith, E. O., R. J. Schanler, C. Garza, and B. Nichols (1983). Modeling the growth patterns of premature infants. *Growth 47*, 340–347.

Sorribas, A., J. March, and E. O. Voit (2000). Estimating age-related trends in cross-sectional studies using s-distributions. *Statistics in Medicine 19*(5), 697–713.

Steward, J. A. (1994, March). *A statistical analysis of the longitudinal growth of a cohort of pre-school children.* Ph. D. thesis, University College of Wales, Cardiff, School of Mathematics.

Stratford, R., J. Mulligan, B. Downie, and L. Voss (1999). Threats to validity in the longitudinal study of psychological effects: the case of short stature. *Child: Care, Health and Development 25*(6), 401–419.

Strenio, J. F., H. I. Weisberg, and A. S. Bryk (1983). Empirical Bayes estimation of individual growth-curve parameters and their relationships to covariates. *Biometrics 39*, 71–86.

Stützle, W., T. Gasser, L. Molinari, R. H. Largo, A. Prader, and P. J. Huber (1980). Shape-invariant modelling of human growth. *Annals of Human Biology 7*, 507–528.

Sumner, E. E. and J. Whitacre (1931). Some factors affecting the accuracy on the collection of data on the growth in weight of school children. *Journal of Nutrition 4*, 15–23.

Taitz, L. S. and J. M. King (1988). Growth patterns in child abuse. *Acta Paediatrica Scandinavia supplement 343*, 62–72.

Tango, T. (1998). Estimation of age-specific reference ranges via smoother AVAS. *Statistics in Medicine 17*(11), 1231–1243.

Tanner, J. (1981). *A history of the study of human growth.* Cambridge University Press, Cambridge.

Tanner, J. (1986a). Physical development. *British Medical Bulletin 42*(2), 131–138.

Tanner, J. M. (1958). The evaluation of physical growth and development. In A. Holzel and J. P. M. Tizard (Eds.), *Modern trends in pediatrics*, 2, pp. 325. Butterworth, London.

Tanner, J. M. (1986b). Use and abuse of growth standards. In F. Falkner and J. M. Tanner (Eds.), *Human Growth* (2 ed.), Volume 3, pp. 95–109. New York: Plenum.

Tanner, J. M. (1989). *Foetus into Man: Physical Growth from Conception to Maturity* (2 ed.). Castlemead publications.

Tanner, J. M. (1999). The growth and development of the 'Annals of Human Biology': a 25-year retrospective. *Annals of Human Biology 26*(1), 3–18.

Tanner, J. M. and P. Davies (1985). Clinical longitudinal standards for height and height velocity for North American children. *Journal of Pediatrics 107*, 317–328.

Tanner, J. M., H. Goldstein, and R. H. Whitehouse (1970). Standards for childrens height at ages two to nine allowing for height of parents. *Archives of Disease in childhood 45*, 755–762.

Tanner, J. M., K. W. Land, N. Cameron, B. S. Carter, and J. Patel (1983). Prediction of adult height from height and bone age in childhood: A new system of equations (TW Mark ii) based on a sample including very tall and very short children. *Archives of Disease in Childhood 58*, 767–776.

Tanner, J. M., H. Lejarraga, M. J. R. Healy, and G. Turner (1972). Within-family standards for birth-weight. *Lancet 2*, 193–197, 1314–1315.

Tanner, J. M., R. A. Whitehouse, W. A. Marshall, and B. S. Carter (1975). Prediction of adult height from height, bone age and occurrence of menarche, at ages 4 to 16 with allowance for mid-parent height. *Archives of Disease in Childhood 50*, 14–26.

Tanner, J. M. and R. H. Whitehouse (1976). Clinical longitudinal standards for height, weight, height velocity, weight velocity and stages of puberty. *Archives of Disease in Childhood 51*, 170–179.

Tanner, J. M., R. H. Whitehouse, and M. Takaishi (1966). Standards from birth to maturity for height, weight, height velocity and weight velocity: British children 1965. *Archives of Disease in Childhood 41*, 454–471, 613–635.

Thalange, N. K. S., P. J. Foster, M. S. Gill, D. A. Price, and P. E. Clayton (1996). Model of normal prepubertal growth. *Archives of Disease in Childhood 75*, 427–431.

Thisted, R. A. (1988). *Elements of statistical computing: Numerical computation.* Chapman and Hall: New York ; London.

Tillmann, V., N. K. S. Thalange, P. J. Foster, M. S. Gill, D. A. Price, and P. E. Clayton (1998). The relationship between stature, growth, and short term changes in height and weight in normal prepubertal children. *Pediatric Research 44*(6), 882–886.

Van't Hof, M. A., J. M. Wit, and M. J. Roede (1985). A method to construct age references for skewed skinfold data using Box-Cox transformations to normality. *Human Biology 57*, 131–139.

Venables, W. N. and B. D. Ripley (1997). *Modern applied statistics with S-PLUS* (2 ed.). Springer.

Von Rosen, D. (1991). The growth curve model: a review. *Communications in Statistics - Theory and Methods 20*(9), 2791–2822.

Voss, L. D. and B. J. R. Bailey (1997). Diurnal variation in stature: is stretching the answer? *Archives of Disease in Childhood 77*, 319–322.

Voss, L. D., B. J. R. Bailey, K. Cumming, T. J. Wilkin, and P. R. Betts (1990). The reliability of height measurement (the Wessex growth study). *Archives of Disease in Childhood 65*, 1340–1344.

Voss, L. D. and J. Mulligan (1994). The short normal child in school: self esteem, behaviour, and attainment before puberty (the Wessex growth study). In B. Stabler and L. E. Underwood (Eds.), *Growth, Stature and Adaption*, pp. 47–64. The University of North Carolina, Chapel Hill, NC.

Voss, L. D. and J. Mulligan (1999a). Child obesity and Body Mass Index. *Lancet 353*(9169), 2070. (letter).

Voss, L. D. and J. Mulligan (1999b). Too short or too fat: should we be monitoring weight? *Lancet 353*(915), 413–414.

Voss, L. D., J. Mulligan, P. R. Betts, and T. J. Wilkin (1992). Poor growth in school entrants as an index of organic disease. *British Medical Journal 305*, 1400–2.

Voss, L. D., J. M. Walker, H. Lunt, T. J. Wilkin, and P. R. Betts (1989). The Wessex growth study - first report. *Acta Paediatricaica Scandinavia Supplement 349*, 65–72.

Voss, L. D., T. J. Wilkin, B. J. R. Bailey, and P. R. Betts (1991). The reliability of height and height velocity in the assessment of growth.

*Archives of Disease in Childhood 66*, 833–837.

Voss, L. D., T. J. Wilkin, and P. R. Betts (1987). Do we need new growth charts? *Lancet (ii)*, 447–448. (letter).

Voss, L. V. (1999). Changing practice in growth monitoring: No evidence exists that monitoring height velocity is useful. *British Medical Journal 318*, 344–345.

Wade, A. M. and A. E. Ades (1994). Age-related reference ranges: significance tests for models and confidence intervals for centiles. *Statistics in Medicine*, 2359–2367.

Wade, A. M. and A. E. Ades (1998). Incorporating correlations between measurements into the estimation of age related reference ranges. *Statistics in Medicine 17*, 1989–2002.

Wainer, H., A. F. Roche, and S. Bell (1978). Predicting adult stature without skeletal age and without paternal data. *Pediatrics 61*(4), 569–572.

Wales, J. K. H. (1996). Sex differences in weight in infancy: Data from Sheffield support authors' findings. *British Medical Journal 313*, 1486.

Wales, J. K. H. (1998). A brief history of the study of human growth dynamics. *Annals of Human Biology 25*(2), 175–184.

Ware, J. H. and M. C. Wu (1981). Tracking: Prediction of future values from serial measurements. *Biometrics 37*, 427–437.

Waterlow, J. C. (1988). Observations on the natural history of stunting. In J. C. Waterlow (Ed.), *Linear Growth Retardation in Less Developed Countries*, Volume 14 of *Nestle Nutrition Workshop Series*, pp. 1–12. Raven Press, New York.

Waterlow, J. C., R. Buzina, W. Keller, J. M. Lane, M. Z. Nichaman, and J. M. Tanner (1977). The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years. *Bulletin of the World Health Organisation 55*(4), 489–498.

Wellek, S. and E. Merz (1995). Age-related reference ranges for growth parameters. *Methods of Information in Medicine 34*(5), 523–528.

Wenlock, D. R., M. M. Diselduff, R. K. Skinner, and I. Knight (1986). The diets of British school children. Technical report, HMSO, London.

Whitehead, R. G., A. A. Paul, and E. A. Ahmed (1989a). DHSS 'present-day infant feeding practice' and its influence on infant growth. In J. M. Tanner and M. A. Preece (Eds.), *The physiology of human growth*, pp. 69–79. Cambridge University Press: Cambridge.

Whitehead, R. G., A. A. Paul, and E. A. Ahmed (1989b). Diet and growth of healthy infants. *Journal of Human Nutrition and Dietetics 2*, 73–84.

Whitehouse, R. H., J. M. Tanner, and M. J. R. Healy (1974). Diurnal variation in stature and sitting height in 12-14 year old boys. *Annals Human Biology 1*, 103–106.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics.* John Wiley & Sons; Chichester.

WHO (1995). *Physical Status: the Use and Interpretation of Anthropometry*, Geneva. WHO.

Wilcox, W. D., P. Nieburg, and D. S. Miller (1989). Failure to thrive: A continuing problem of definition. *Clinical Pediatrics 28*, 391–394.

Willett, W. (1990). *Nutritional epidemiology*, Volume 15 of *Monographs in Epidemiology and Biostatistics.* Oxford University Press, New York.

Woolston, J. L. (1991). *Eating and growth disorders in infants and children.* Newbury Park London: Sage publications.

Wright, C., J. Loughridge, and G. Moore (2000). Failure to thrive in a population context: two contrasting studies of feeding and nutritional status. *Proceedings of the Nutrition Society 59*, 37–45.

Wright, C. M. (1996). *The Parkin Project - a study of screening and intervention in failure to thrive.* University of Newcastle-upon-Tyne. MD Thesis.

Wright, C. M. (1997). Who comes to be weighed: an exception to the inverse care law. *Lancet 350*, 642.

Wright, C. M. (1998). Girls weight monitoring chart for children with slow weight gain (birth to 60 months). Department of Child Health, University of Newcastle-upon-Tyne.

Wright, C. M. (2000). Identification and management of failure to thrive: A community perspective. *Archives of Disease in Childhood 82*(1), 5–9.

Wright, C. M., A. Avery, M. Epstein, E. Birks, and D. Croft (1998). New chart to evaluate weight faltering. *Archives of Disease in Childhood 78*, 40–43.

Wright, C. M., A. Aynsley-Green, P. Tomlinson, L. Ahmed, and J. A. Macfarlane (1992). A comparison of height, weight and head circumference of primary school children living in deprived and non-deprived circumstances. *Early Human Development 31*, 157–162.

Wright, C. M. and E. Birks (2000). Risk factors for failure to thrive: a population based survey. *Child: Care, Health and Development 26*(1), 5–16.

Wright, C. M., I. W. Booth, J. M. H. Buckler, N. Cameron, T. J. Cole, M. J. R. Healy, J. A. Hulse, M. A. Preece, J. J. Reilly, and A. F. Williams (2002). Growth reference charts for use in the United Kingdom. *Archives of Disease in childhood 86*(1), 11–14.

Wright, C. M., J. Callum, E. Birks, and S. Jarvis (1998). Effect of community based management in failure to thrive: randomised controlled trial. *British Medical Journal 317*(7158), 571–574.

Wright, C. M., J. Callum, and S. Jarvis (1999). The influence of slow weight gain in infancy on growth and blood pressure in childhood.

Wright, C. M. and T. Cheetham (1999). The strengths and limitations of parental heights as a predictor of attained height. *Archives of Disease in Childhood 81*(2), 257–260.

Wright, C. M., S. S. Corbett, and R. F. Drewett (1996). Sex differences in weight in infancy and the British 1990 standards. *British Medical Journal 33*, 513–514.

Wright, C. M., J. N. S. Matthews, A. Waterston, and A. Aynsley-Green (1994). What is the normal rate of weight gain in infancy? *Acta Paediatrica 83*, 351–356.

Wright, C. M., L. Parker, D. Lamont, and A. W. Craft (2001). Implications of childhood obesity for adult health. *British Medical Journal 323*, 1280–1284.

Wright, C. M. and E. Talbot (1996). Screening for failure to thrive - what are we looking for? *Child: Care, Health and Development 22*(4), 223–224.

Wright, C. M., A. Waterston, and A. Aynsley-Green (1993). Comparison of the use of Tanner and Whitehouse, NCHS and Cambridge standards in infancy. *Archives of Disease in childhood 69*, 420–422.

Wright, C. M., A. Waterston, and A. Aynsley-Green (1994). Effect of deprivation on weight gain in infancy. *Acta Paediatrica 83*, 357–359.

Wright, E. M. and P. Royston (1997). A comparison of statistical methods for age-related reference intervals. *Journal of the Royal Statistical Society (series A) 160*(1), 47–69.

Wright, J. A., C. A. Ashenburg, and R. C. Whitaker (1994). Comparison of methods to categorise undernutrition in children. *Journal of Pediatrics 124*, 944–946.

Ziebland, S., M. Thorogood, A. Fuller, and J. Muir (1996). Desire for the body normal: body image and discrepancies between self reported and measured height and weight in a British population. *Journal Epidemiology Community Health 50*, 105–106.

Zumrawi, F. Y., M. Yang, and T. Marshall (1992). The use of short term increments in weight to monitor growth in infancy. *Annals of Human Biology 19*, 165–175.