UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Human Development and Health

Associations of childhood head growth with health and human capital in adult life and in the next generation

by

Shivam Pandey

Thesis for the degree of Doctor of Philosophy

October 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE

Medical Statistics

Thesis for the degree of Doctor of Philosophy

ASSOCIATIONS OF CHILDHOOD HEAD GROWTH WITH HEALTH AND HUMAN CAPITAL IN ADULT LIFE AND IN THE NEXT GENERATION

Shivam Pandey

Abstract

Nature of the work undertaken: Most studies of the 'developmental origins of health and disease hypothesis' have related early weight and height measurements to later life outcomes, but few have considered head size. I study the association of early head size and growth with adult cognitive and cardiometabolic outcomes, and intergenerational outcomes, using the New Delhi Birth Cohort (NDBC). This was set up in 1969 in New Delhi, India and enrolled 20,755 married women in the reproductive age-group resulting in 9,169 pregnancies whose anthropometric data, including head circumference were collected from birth till early adulthood at defined time points. I develop and compare suitable statistical models and advise on the choice of method for analysis of such data.

Contribution to subject knowledge in the area: Head size and disproportion of head size relative to other body measurements at birth, and childhood head growth were unrelated to either educational attainment or blood pressure, and therefore early head size is not an indicator of early life programing in this population. Improving childhood nutrition and promoting linear growth up to age 2 years may be important for higher adult cognitive development. Contrastingly, becoming a heavier adolescent is associated with an increased risk of adult hypertension. Similar associations of early life maternal and paternal head growth with next generation birth weight suggest that they result from genetic factors which are non-modifiable or persisting environment between generations. Understanding the environmental factors influencing brain growth might help increase next-generation birth weight. Conditional and spline approaches provide similar goodness of fit in my data, and associations of head growth with the different adult outcomes were similar. Conditional growth modelling is suitable for studies with a small number of body measurements per individual, while spline models might be better for datasets with a larger number of measurements.

Contents Page

Table of Figures

Table of Tables

Academic Thesis: Declaration of Authorship

I, Shivam Pandey declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Associations of childhood head growth with health and human capital in adult life and in the next generation

I confirm that:

1. **This work was done wholly or mainly while in candidature for a research degree at this University;**
2. **Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;**
3. **Where I have consulted the published work of others, this is always clearly attributed;**
4. **Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;**
5. **I have acknowledged all main sources of help;**
6. **Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;**
7. **None of this work has been published before submission**

Signed: .....

Date: .....

Acknowledgements

Foremost, I would like to express my sincere gratitude to my chief supervisors Professors Clive Osmond and Caroline Fall for the continuous support of my Ph.D. study and research, and for their patience and continuous motivation throughout the duration of study. Their guidance has helped me in developing my research ability and skills. I would like to thank my co-supervisors Professors Sanjay Zodpey and N. Sreekumaran Nair for their guidance. I would like to thank Professors Santosh K Bhargava and HPS Sachdev, principal investigators of the New Delhi Birth Cohort for allowing me to use the study datasets. I would also like to thank Ms. Shikha Sinha for her help with the several data management queries and the overall understanding of the study datasets. My sincere thanks also goes to Dr. Niveditha Devasenapathy for her critical inputs on the thesis and on the subject as a whole. This work would not have been possible without her support. I would also like to express my gratitude to my examiners in the transfer viva Professor Hazel Inskip and Dr. Sarah Crozier for their critical comments during the transfer viva. Finally I would also like to thank my colleagues Dr. Sarah Kehoe and Ms. Chiara Di Gravio for their helpful comments on the work. I would also like to express my gratitude to Ms. Jane Pearce for her help with the formatting of the thesis and taking care of the several administrative issues that arose during the period of study.

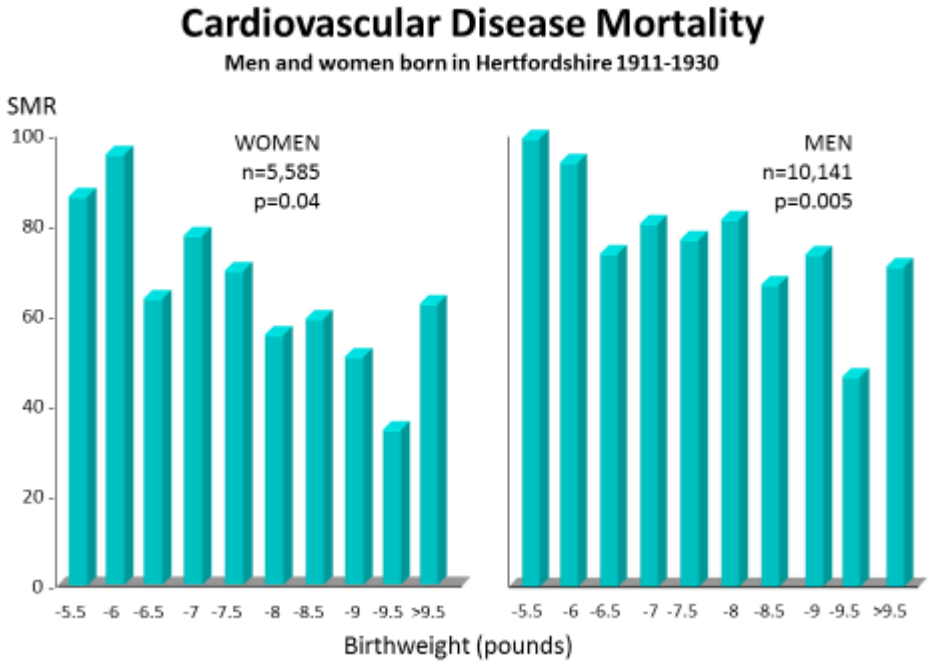
Chapter 1
Introduction

Growth during fetal life, infancy and childhood is important for health in future. I have studied growth, with a specific focus on head growth, which has been relatively less studied in relation to later outcomes. The first part of the chapter discusses the literature available in this research area and the second part discusses the various statistical approaches used to model growth in relation to outcomes in adulthood.

1.1 Early life characteristics and cardiovascular disease in later life

In 1989, David Barker showed that lower birth weight was associated with higher risk of death from cardiovascular disease (CVD). He noted that among both men and women, born in Hertfordshire, UK death rates from CVD fell progressively between the low and high birth weight groups (Figure 1.1) (1,2). Low birth weight was also strongly associated with type 2 diabetes and hypertension in men

Figure 1.1: Association of birth weight and cardiovascular disease mortality in Hertfordshire,UK (1,2)



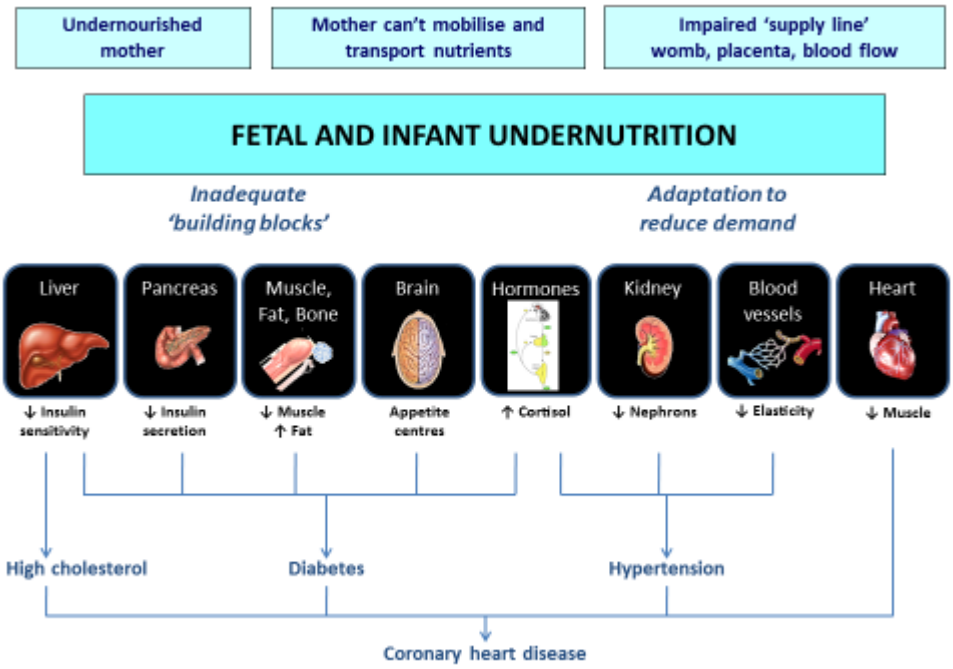
over 60 years of age in this population (3). A study conducted in India in Mysore in 1996 showed similar results to those obtained in Hertfordshire. The prevalence of coronary heart disease (CHD) fell from 11% in people whose birth weights were 5·5 lb

(2·5 kg) or less to 3% in those whose birth weights were more than 7 lb (3·1 kg), p for trend=0·09. The trends were stronger, and statistically significant, among people aged 45 years and over (p=0·03 for birth weight) (4). Studies in Finland further supported the findings; lower birth weight was associated with higher death rates from CHD, and a higher prevalence of type 2 diabetes (5,6). In the studies where gestational age was available, the associations appeared to be present with birth weight even after accounting for gestational age. These results suggested that the associations between low birth weight and adult CVD were related to poor fetal growth rather than premature birth. These were the early studies in the world linking birth size and early life growth to later life CVD and its risk factors in humans.

1.2 Developmental Origins of Health and Disease (DOHaD)

The interesting and consistent findings from these studies helped Barker to reach a very important common conclusion. He proposed that CVD in adulthood can be caused by poor nutrition of the mother at the time of her pregnancy (Figure 1.2). A mother who is undernourished does not have the necessary nutrition needed to sustain her body and as a result cannot provide the nutrition needed for growth and development of the fetus. The fetus uses the nutrients, which are scarce, for the development of the most important parts of the body, such as the head, and thus impairs the development of other important parts of the body such as the liver, pancreas, kidneys. He suggested that because these essential organs develop substantially during fetal life, this has permanent effects and leads to various metabolic problems in later life such as high cholesterol, diabetes and hypertension which ultimately lead to CHD. He named this phenomenon ‘fetal programming’ and this science became known as the ‘Fetal Origins of Adult Disease’ (7,8). It was later renamed ‘Developmental Origins of Health and Disease’ (DOHaD), as it gradually became clear that it was not only fetal growth that was important, but also postnatal growth (infant, child and adolescent) (9).

Figure 1.2: David Barker's concept of 'fetal programming'



1.3 Summary of systematic reviews of associations of birth weight and cardiovascular disease risk factors

Since Barker's initial work in the area, many studies in other parts of the world have confirmed these associations and tried to verify his hypothesis. A number of systematic reviews have also confirmed the associations derived from individual studies. Huxley and colleagues conducted a systematic review to examine the association between birth weight and CHD in later life. Seventeen published studies of birth weight and subsequent CHD were identified that included a total of 144,794 singletons. The combined adjusted relative risk for the association between birth weight and CHD events was 0.84 (95% CI: 0.81, 0.88) per kilogram of birth weight ($P <$ 0.0001) (10). Thus, 1kg higher birth weight is associated with a 10–20% lower risk of subsequent CHD. A review by Law and others in 1996 of 80 studies published since 1956, concluded that blood pressure fell with increasing birth weight (about 2 mmHg/kg) (11). Similar conclusions were found in a review of 20 studies conducted by Mu et. al. in 2012. The authors concluded that low birth weight compared with birth weight greater than 2500 g was associated with an increased risk of hypertension (odds ratio (OR):1.21; 95% confidence interval: 1.13, 1.30) (12). A meta-analysis in 2007, which pooled data from 14 studies and 132,180 persons, also showed an inverse association between birth weight (<2,500 g) and risk of type 2 diabetes in adulthood

(13). Low birth weight as compared to high birth weight was associated with an increased risk of type 2 diabetes (OR: 1.32 (95% CI: 1.06 to 1.64)) in adults. A systematic review in 2008 also showed an inverse association between birth weight and risk of type 2 diabetes in adults. Inverse associations were observed in 23 out of 30 populations for which data were available, 9 of which were statistically significant (14). In a systematic review of 11 studies in 2008, as compared to people with normal birth weight, in those with low birth weight, the OR for metabolic syndrome was 2.53 (95% CI: 1.57,4.08) in adulthood (15). A systematic review of 28 studies from 1990 onwards conducted in 2003 found a weak, but inverse association between birth weight and cholesterol levels (-0.06 mmol/L fall in total cholesterol per kg increase in birth weight (95% CI: -0.13 to 0.01 mmol/L per kg)), an association which was not affected after adjusting for current body size (16). A review of 58 relevant studies, conducted in 2004 also showed an inverse association between birth weight and subsequent cholesterol levels (17). The weighted estimate was -1.39 mg/dL(-0.036 mmol/L) per kilogram (95% CI, -1.81 to -0.97 mg/dl (-0.047 to -0.025 mmol/L). There was significant heterogeneity between their separate results ($P < 0.001$), with stronger associations in smaller studies and studies in infants. Thus, these systematic reviews confirmed the inverse association between birth weight and later life CVD and its risk factors shown in the earlier individual studies by Barker.

1.4 Infant growth and cardiovascular disease and its risk factors

Apart from size at birth, patterns of growth during infancy have also shown significant associations with later life CVD and its risk factors in many populations. In Hertfordshire, men with lower weight at the age of one year had increased cardiovascular mortality (2). A number of studies in Finland have shown lower weight gain during infancy to be associated with a higher risk of later life CVD and its risk factors. Low weight gain up to two years was associated with a higher risk of CHD at 30 years of age ($p<0.001$) (18). A 1 standard deviation (SD) increase in body mass index (BMI) at 2 years was associated with a reduced risk of stroke in later life (OR: 0.84 (95% CI: 0.77 to 0.92) ($p=0.0002$)) (19). Adults who had greater weight gain between birth and 2 years had a lower risk of diagnosed hypertension (19). The association remained even after adjusting for SES. Low weight gain during infancy was associated with an increased risk of type 2 diabetes (20). The highest ORs, 2.6 (95% CI

1.8–3.7), were among people whose birth weights were below 3,000 g and whose weights at 2 years were below 11.5 kg. These findings were consistent with a study in the New Delhi Birth Cohort (NDBC) in India, which showed that low weight gain in infancy was associated with diabetes at age 30 years (21).

In contrast, Singhal et. al. in 2004 showed that greater weight gain during infancy was associated with lower brachial artery flow-mediated endothelium-dependent dilation in adolescence (0.026 mm fall in mean arterial diameter per 100gm increase in weight, 95% CI: -0.040 to -0.012 mm) (22). A study in Stockholm showed that weight gain during infancy was associated with higher metabolic risk at age 17 years (standardized β:0.16 (95% CI:0.05 to 0.27), calculated by averaging the standardized values of waist circumference, blood pressure, fasting triglycerides, high-density lipoprotein cholesterol, glucose, and insulin level (23). The association was adjusted for birth weight, gestational age, current height, maternal fat mass and SES.

Thus, there does not seem to be a consistent direction of association between growth in infancy and risk of later life CVD risk factors.

1.5 Childhood growth and cardiovascular disease and its risk factors in later life

There is consistent evidence that rapid weight or BMI gain during later childhood and adolescence is associated with later life cardiometabolic diseases. A number of studies in Finland show this. Eriksson et. al in 1999, found that the highest death rates from CHD occurred in boys whose had above average BMI from ages 7 to 15 (5). Boys and girls who later developed hypertension and type 2 diabetes had above average weight at age 15 years ($p=0.008$ in boys and $p<0.001$ for girls) (24). Girls who developed CHD in later life had a rapid increase in BMI during childhood (hazard ratio: 1.17 (95% CI: 1.03 to 1.32)) ($p=0.02$)) (25). Similar findings were seen in the NDBC, where men and women having diabetes had a greater increase in BMI between 2 and 12 years. The OR for the disease with every SD increase in BMI was 1.36 (95% CI: 1.18 to 1.57) ($p<0.001$) (21) (Figure 1.4). A study by Baker et. al in 2007 confirmed that higher BMI during childhood is associated with an increased risk of CHD in adults between the ages of 25 to 60 years (26). The associations were stronger in boys than girls and increased with the age of the child in both the sexes. The authors also noted that compared to an average sized 13 year old boy, a boy of the same age and height

weighing 11.2 kg more had a 33% higher risk of having a CHD event in adulthood. Similar associations were also observed for girls. A study in 2013 which combined data of birth cohort studies from Brazil, Guatemala, India, Philippines and South Africa showed that higher than average weight gain between age 2 years and mid-childhood (4 years for Brazil, Guatemala, India and South Africa; 8 years for the Philippines) was associated with an increased risk of hypertension (OR: 1.22 (95% CI: 1.15 to 1.30)) and diabetes (OR:1.13 (95% CI:1.04–1.23)) in adulthood (27). Thus, rapid weight gain in childhood is consistently associated with risk of later life CVD and its risk factors. Studies that have plotted the childhood growth of adults with CVD, diabetes or hypertension have shown that they were small at birth and in infancy, but had rapid weight gain in childhood and adolescence (Figure 1.3 and 1.4). (6,21)

Figure 1.3: Body mass index, weight and height from birth to age 12 years of men and women in Helsinki, Finland who developed diabetes (6)

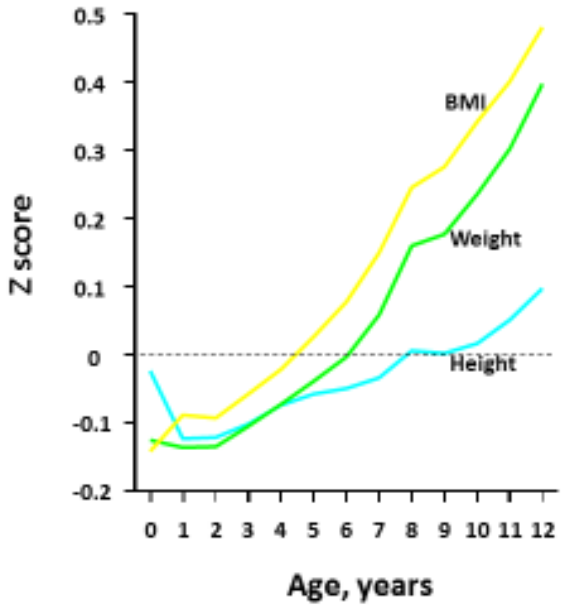
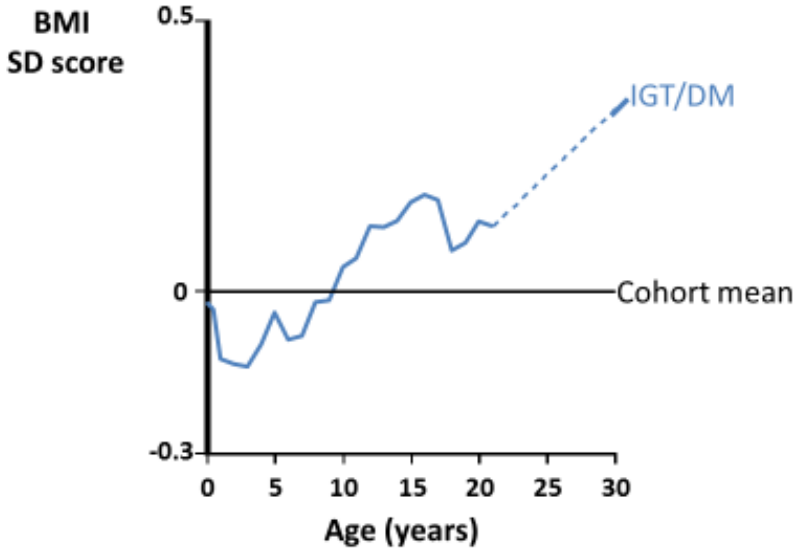


Figure 1.4: Body mass index from birth to adulthood of men and women in New Delhi who developed diabetes (DM) or impaired glucose tolerance (IGT), aged 30 years (21)



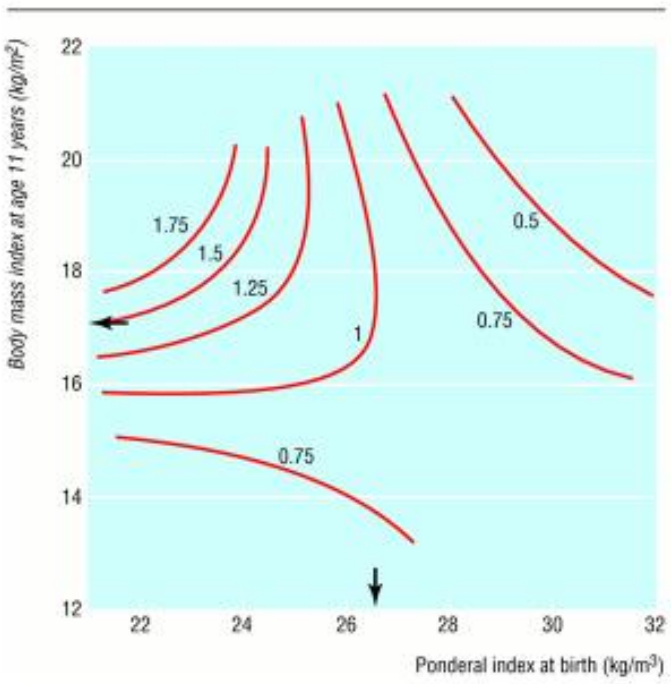
Footnote for both graphs: The 0 line refers to the average cohort values

Type 2 diabetes is strongly associated with obesity in adult life. After the age of 2 years, the BMI of young children decreases to a minimum at around 6 years of age before increasing again. The age when the minimum is reached is known as the age at adiposity rebound (28). An early age of adiposity rebound is associated with increased obesity in later childhood and adult life. In the Helsinki birth cohort early adiposity rebound is associated with an increased incidence of type 2 diabetes (6). The incidence decreased from 8.6% in subjects whose rebound occurred before the age of 5 years to 1.8% in those whose rebound occurred after 7 years ($p<0.001$).

1.6 Interactions between birth size and later growth

The Finnish cohort studies have shown interactions between birth size and childhood growth in relation to adult disease. Rapid weight gain after 1 year was associated with an increased risk of CHD but only among boys who were thin at birth (ponderal index (weight/height^3) <26) (Figure 1.5) (29). This means that individuals who are thin at birth have a higher risk of having CHD in later life if they gain more weight after 1 year. This is relevant for India, where most infants are thin at birth (30).

Figure 1.5: Interaction of ponderal index at birth and BMI with risk of coronary heart disease in later life (29)



1.7 Head size and cardiovascular disease and its risk factors

According to the fetal programming hypothesis described earlier, the undernourished fetus prioritizes the development of the brain, to the detriment of the rest of the body, such as the abdominal organs. Thus, it would be interesting to study not only head size and growth but also the ratio of head size to body size (length or weight). There are only a few studies which have associated head size and growth or these ratios with later life outcomes (Table 1.1). This may be because head circumference is not a common measure of body size at birth, and tends to be measured only in a research context. There are thus very few adult cohorts with head size available at birth.

The earliest findings of an association between head size and later life outcomes came from a study by Barker et. al in 1992 who showed that the blood pressure of 327 men and women aged 46 to 54 years in Preston, was inversely associated with head circumference at birth (31). In those with placental weights above 1.25 lb, mean blood pressure, and the risk of hypertension, rose as length decreased and as the ratio of head circumference to length increased (P = 0.02) (31). Mean systolic pressure rose by 14 mm Hg as the head circumference to length ratio increased from less than 0.65 to

greater than or equal to 0.7. In a study in Sheffield, men who had a small head circumference or were thin at birth, or both, had higher rates of mortality from CVD (32). These findings were confirmed in an Indian study in which small head circumference at birth was associated with an increased prevalence of CHD in people aged 45-60 years (4). In a population-based study in Norway, Risnes et. al. in 2009 showed an inverse association between head circumference at birth and death from CHD in later life (OR: 0.90 (95% CI: 0.83 to 0.98)) (p for trend=0.010) (33).

However subsequent studies in Finland have not shown associations between head size at birth and CHD (25) or stroke in later life (19). However smaller head size at birth was associated with earlier age at adiposity rebound, which in turn was associated with adult obesity and type 2 diabetes (34).

Thus, there is some evidence that a small head size at birth is associated with CVD and its risk factors in later life. However, the literature is limited and none could be found relating head to body size ratios at birth or infant and childhood head growth to CVD and its risk factors in later life.

Table 1.1: Association of head size at birth and cardiometabolic outcomes in later life

Author & Year	Cohort	Measure of head size or growth	Outcome	Result
Barker et. al, 1992 (31)	327 men and women aged 46 to 54 years in Preston, UK.	Head circumference at birth, ratio of head circumference to length	Systolic blood pressure (SBP), hypertension	Smaller head circumference at birth associated with higher SBP (p=0.04). Increased head/length ratio associated with increased risk of hypertension in persons with placental weights >1.25 lb (p=0.02)
Barker et. al, 1993 (32)	1,586 men born in Jessop Hospital, Sheffield, UK before 1925.	Head circumference at birth	Mortality ratios for CVD	Standardized mortality ratios decreased with increasing head circumference (chi-square=4.6, p=0.03).
Stein et. al, 1996 (4)	517 men and women born between 1934 and 1954 in Holsworth Memorial Hospital, Mysore, India.	Head circumference at birth	Prevalence of CHD.	Smaller head circumference at birth associated with higher prevalence of CHD (p=0.08). Associations were stronger for people aged 45 or older (p=0.02).
Forsen et. al, 2004 (25)	Helsinki, Finland. 4,130 girls born between 1934 and 1944 and still residing in Finland in 1971.	Head circumference at birth	CHD	No significant trends with CHD
Osmond et. al, 2007 (19)	12,439 people born in Helsinki between 1934 and 1944.	Head circumference at birth	Rates of hospitalization or death from stroke after 35 years of age	Head circumference at birth was not associated with risk of stroke in later life.
Eriksson et. al, 2014 (35)	2,877 children born at Helsinki University Central Hospital, Helsinki.	Head circumference at birth	Early age at adiposity rebound, which is associated with adult obesity.	Small head circumference at birth associated with an early age at adiposity rebound. The mean age at adiposity rebound rose from 5.8 years in babies with a head circumference of ≤33 cm to 6.2 in babies with a head circumference of >36 cm (p for trend = 0.007).
Risnes et. al, 2009 (33)	35,846 men and women born in St. Olav's University Hospital in Trondheim, Norway, from 1920 to 1959.	Head circumference at birth	Death from CHD	Head circumference at birth was inversely associated with deaths from CHD (Hazard ratio:0.90 (95% CI :0.83–0.97)) (p for trend<0.001).

1.8 Early life growth and human capital outcomes

In addition to cardiometabolic outcomes, research has shown associations between early life growth and human capital outcomes. Human capital refers to the knowledge, skill sets and motivation that people have, which provide economic value. Based on this, education can be taken as an investment which increases an individual's productivity. Education is a good summary measure for human capital because it helps individuals to acquire knowledge which can help create job opportunities, which can improve health conditions and reduce poverty.

A high quality review studied the association between birth weight and later childhood intelligence in individuals born at term (37-42 weeks) (36). The study included individuals who had completed valid cognitive tests and were between 5-16 years old. The studies showed a positive association between birth weight and cognitive function in adulthood. The authors were not able to do a meta-analysis as all the studies had different initiation dates, age of participants, tests used and definitions of the outcome measures. In other words, there was significant heterogeneity across studies. I therefore, describe two studies in this review to give an idea of the effect size. One study used data derived from 58,000 pregnancies in the USA (37). Unadjusted analysis showed that IQ increased from 91.9 points among those in the smallest birth weight category (95% CI: 1,500-2,500g) to 100.7 among those in the largest category (95% CI: 3,000-4,500 g). The associations became weaker but remained statistically significant after adjustment for potential confounders such as socioeconomic index, sex, birth order, maternal age, education and race. Another study in 2002 assessed cognitive ability at ages 7, 11 and 16 in 10,845 males and females born in the UK (38). For each kilogram increase in birth weight, cognitive ability SD score at age 7 increased by 0.15 (95% CI: 0.10 to 0.21) for males and 0.20 (95% CI: 0.14 to 0.25) for females.

Studies involving data from five birth cohorts in Brazil, Guatemala, India, Philippines and South Africa have shown an association between growth in height and weight gain up to age 2 years and increased number of years of attained schooling in later life (27,39-41). Similar associations were shown in other populations. It was found that in a Finnish population, slower growth and weight gain between birth and age 6 months and between ages 6 months and 2 years was associated with poorer later life cognitive function (42). A study in Guatemalan adults aged 25-42 years also showed

that growth in height up to 2 years was associated with increased levels of later life schooling (1 SD increase in height-for-age Z score was associated with 0.78 higher schooling scores (95% CI:0.25, 1.30) and also higher test scores for reading and non-verbal cognitive tests (41).

1.9 Head size /growth and human capital outcomes

Few studies have looked at the association of head size at birth and head growth at different time periods, with cognitive function in later life (Table 1.2), and these have inconsistent findings. An English study found no association between head size at birth, head/length and head/abdominal circumference ratios and intelligence (43). However, a study in a Finnish population showed a positive association between head size at birth and verbal (P=0.03), visuospatial (P=0.04) and arithmetic ability (P=0.002) at 20 years (38). Similar findings were shown in south India by Veena et. al in 2010 who found a positive association between head circumference at birth and learning ability at 9-10 years (44). The associations were reduced after further adjustment for current head circumference.

All studies which have looked at the association between head growth at different postnatal periods and cognitive function in later life have concluded that head growth early in life, up to infancy is positively associated with later life intelligence (45-49) (Table 1.2).

Table 1.2: Association of head size / growth and cognitive outcomes in later life

Author & Year	Cohort	Measure of head size or growth	Outcome	Type of growth approach used	Results
Gale et. al 2006 (47)	633, children Avon longitudinal study of parents and children, UK.	Head circumference at birth, head growth between birth-1yr, 1yr-4yr and 4yr-8yr.	Full scale IQ, Verbal IQ, Performance IQ.	Conditional body size approach	Full scale IQ increased an average of 2.41 points for each 1-SD head circumference at birth and 1.97 points for each 1-SD increase in head growth in infancy.
Lira et. al 2009 (49)	Cohort in Northeast Brazil. Two parallel cohorts-Low birth weight and appropriate birth weight. 202 low-birth weight individuals	Head size at birth, head growth between birth and 2 months, between 2 months and 6 months, between 6 months and 8 years.	Full-scale IQ, Verbal IQ, Performance IQ.	Conditional body size approach	Head growth between birth and 2 months was associated with an increase in full-scale IQ of 3.31 points (0.82 to 5.80) and 4.82 points (2.37 to 7.26) for every 1-SD increase between 2 months to 6 months. Every 1-SD increase in head growth between birth

	matched with 212 individuals of appropriate birth weight on sex and month of birth.				and 2 months and 2 months to 6 months associated with increased verbal IQ 3.23 (0.99 to 5.48) and 3.80 (1.54 to 6.06), and increased performance IQ at age 8 years (5.40 (2.53 to 8.27)).
Huang et. al 2013 (48)	8389 children aged 4-7 years of women from Hebei, Zhejiang and Jiangsu provinces in China	Postnatal head growth between ages 4 and 7 years.	IQ at 4-7 years	Two-stage least square (2-SLS) modelling	Every 1-SD increase in postnatal head growth between ages 4 years and 7 years was associated with increase in IQ after adjusting for sex 1.49 (1.22, 1.76).
Gale et. al 2004 (46)	599 Singleton children born to women aged 16 years or older at Princess Anne Hospital, Southampton, UK.	Head measured 18 weeks gestation, 48 hrs after birth, 9 months and 9 years.	Full-scale, verbal and performance IQ at 9 years.	Linear regression	head size at birth was associated with every 1-SD increase in full scale IQ at 9 months (2.59 (0.87 to 4.32)) and at 9 years (3.85 (1.96 to 5.73)). Head size at birth was associated with 1-SD increase in at Verbal IQ at 9 months (2.66 (0.49, 4.83)) and at 9 years (3.76 (1.81 to 5.72)). Head size at birth was positively associated with performance IQ at 9 months (2.88(0.66, 5.11)) and 9 years (3.16(1.16 to 5.16)).
Raikonnen et. al 2009 (42)	2786 Men born in Helsinki between 1934-44. Subjects had served in Finnish Defence Forces between 1952-72.	Head circumference at birth	Verbal, Visuospatial and arithmetic ability at age 20 years.	Conditional body size approach	Head size at birth was positively associated with verbal, visuospatial and arithmetic ability at age 20 years.
Veena et. al 2010 (44)	505 children born between June 1997 to August 1998 at Holdsworth Memorial Hospital, Mysore.	Head circumference measured 72 hrs after birth	Learning/long term ability and retrieval, Visuo-spatial ability	Linear regression	Increased head circumference at birth was positively associated with learning ability at 9-10 years (0.1 SD score (0.04, 0.23)). Associations were reduced after adjusting for current head circumference.
Martyn et. al 1996 (43)	1576 men and women born in Hertfordshire, Sheffield or Preston between 1920 and 1943.	Head circumference at birth, head/length and head/abdominal circumference ratios.	AH4 Intelligence test score.	Linear regression	No association between intelligence in later life and head circumference at birth. No association of head/length and head/abdominal circumference with intelligence in later life.

1.10 Intergenerational effects and the example of cardiovascular disease risk factors

Studies covered so far dealt with body measurements at birth and outcomes in later life in the same generation. However, the effect of early life conditions in individuals may extend beyond one generation (50). The possible explanations for intergenerational effects include

i. Genetic: Intergenerational programming can be caused by genetic factors. For example, the mother or father carries a gene in their DNA which leads to poor growth and low birth weight. The offspring also has the same genes and therefore will have a low birth weight. A genetic mechanism that could explain the association between low birth weight and high blood pressure is genetic linkage. Two genes, one causing low birth weight and the other causing high blood pressure could be located together on the same chromosome. These genes are inherited together by the offspring, which puts him at risk for developing these two outcomes. In this case either parents could contribute the gene associated with the subsequent outcomes in the next generation.

ii. Persisting environment: The risk for developing cardiometabolic outcomes in subsequent generations can also be transferred by persisting environment through generations. For example, if a family lives in a polluted environment they are more likely to develop disease and thus, have a low birth weight child. The child continues to live in the polluted environment and also develops disease.

iii. Learned behaviour: The risk for cardiometabolic diseases may be passed through several generations if the offspring mimic the behaviour of their parents. For example, if a mother smokes she is likely to have a low birth weight baby. The offspring copies the mother's behaviour and also starts smoking. She is again likely to give birth to a low birth weight offspring and is at risk of high blood pressure in later life.

iv. Fetal programming: Another way by which cardiometabolic diseases can be transmitted across several generations is by fetal programming. According to

this explanation, the mother is poorly nourished and gives birth to a low birth weight offspring. This permanently impairs the development of essential organs such as the kidney, heart, liver and pancreas leading to later life CVD and its risk factors.

With regards to the example of CVD risk factors, a number of recent studies have shown associations between parental risk of CVD and offspring birth weight (51). There is evidence that the association of maternal cardiovascular risk and offspring birth weight is stronger than that of the father. For example, lower maternal but not paternal birth weight is related to higher offspring blood pressure (52). Thus, although there are maternal and paternal effects on CVD risk factors in the offspring which may be genetic or due to the environment or both, there is also evidence for specific influences of maternal rather than paternal characteristics.

1.11 Factors influencing growth

Although my study is mainly about early life growth and later outcomes in the same or the next generation, I can also look at body growth as an outcome. The factors known to influence growth can be divided into non-modifiable factors such as genes, sex and race, and modifiable factors such as hormones, nutrition, socioeconomic conditions and presence of infections.

Non-modifiable factors

Genes are known to influence growth in height and weight as well as other characteristics such as hair colour and intelligence (53). There is a difference in the growth and development of boys and girls. Boys are in general are taller than girls. The body composition of boys and girls are also different from each other. Racial differences among the populations are responsible for height, weight, colour, features and body composition of the individual (53). For example a child of African race will have different physical features and the body structure will be different as compared to an individual from an Asian race. However, research has also shown that some racial differences are due to the environment. If the individuals from different races share the same environment, they will grow to be approximately the same (53).

Modifiable factors

Childhood growth is controlled by hormones, including insulin like growth factors (important in infancy), growth hormone (important from infancy until growth stops) and sex hormones (important during adolescence) (53). Growth and development also depends on nutrition (54). Optimal growth requires adequate intake of a range of nutrients including carbohydrates, fats and proteins and vitamins and minerals. In order to have a healthy diet that supports optimal development, the child requires a high quality diverse diet. Poverty and other factors mean that many children in low and middle income countries do not get enough nutrients for normal growth. Nutrition is therefore strongly related to socioeconomic status (SES). Engaging in excess physical work early in life impairs growth. SES can influence growth in other ways. People of low SES tend to live in poorer housing and be exposed to more environmental hazards, leading to infection and poor growth. They may have large families which limits each child's individual care and attention. Many studies have shown that people of higher economic status are taller than those of lower SES (55,56). People who live in low socioeconomic conditions are generally more prone to infections as they are more likely than others to live in overcrowded conditions (57). Living in close proximity to others makes it easier to transmit certain infectious diseases. Infections such as diarrhoea lead to rapid loss of nutrients from the body and also damage the gut lining, which leads to malabsorption of nutrients. All infections lead to an immune response which diverts nutrients and energy away from growth. Children with infection lose their appetite and stop eating, which further compromises their nutrition. There is not much research regarding factors affecting growth and development of the head. Only one research study conducted on the factors affecting head growth by Lira and others in 2009 in Brazil concluded that head growth in the first 6 months of life was related to maternal height and rate of infant weight gain (49).

Thus, there is a need for more research regarding factors affecting the growth of the head. Growth is a continuous process. When analysing the relationship between growth and later outcomes, we need statistical techniques that summarize the growth at specific time periods.

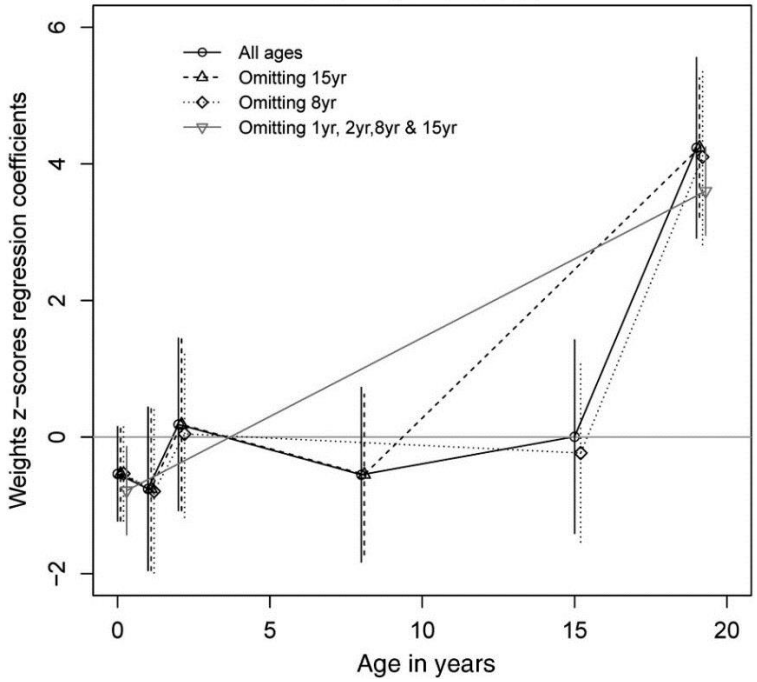
1.12 Statistical assessment of growth

There are several statistical methods which can be used to study growth of the individual at different ages. There are five broad statistical approaches which we can use to characterize growth i.e. changes in body size:

1. Approaches based on conditioning
2. Regression with change scores
3. Multilevel modelling e.g. the spline approach
4. Latent growth curve modelling
5. Growth mixture models
6. Superimposition by Translation and Rotation (SITAR)

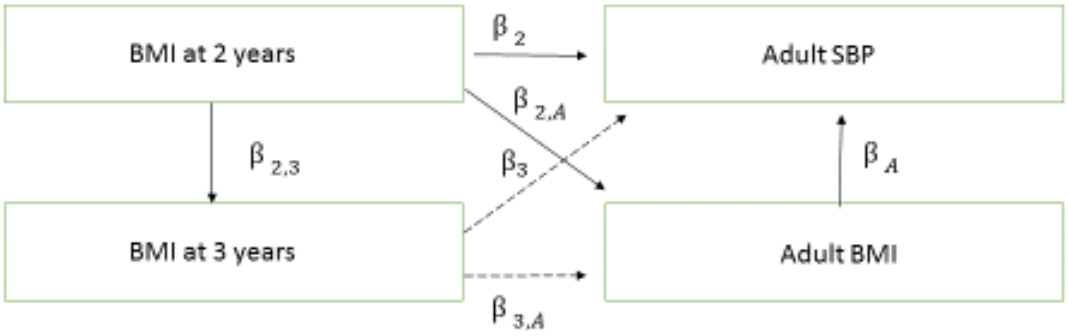
1. Conditional growth models
   (a) Life course plot: In the life course plot approach, the outcome variable is regressed on the body size measured at different ages using multiple regression (58). The regression coefficients obtained are plotted against age. Therefore, both the value of the coefficients and their change from one age to another are displayed together. Such a diagram is known as a life course plot. For example, an outcome such as blood pressure can be regressed on the combination of weight at birth and at different ages. The coefficients can be made comparable to each other by converting the weights at different ages to SD scores. A change in the direction of the multiple regression coefficients is interpreted as a critical phase of the association between growth and adult blood pressure.

Figure 1.6: Life course plot of association of early-life body growth and later life systolic blood pressure (derived from Tu et. al 2013) (60)



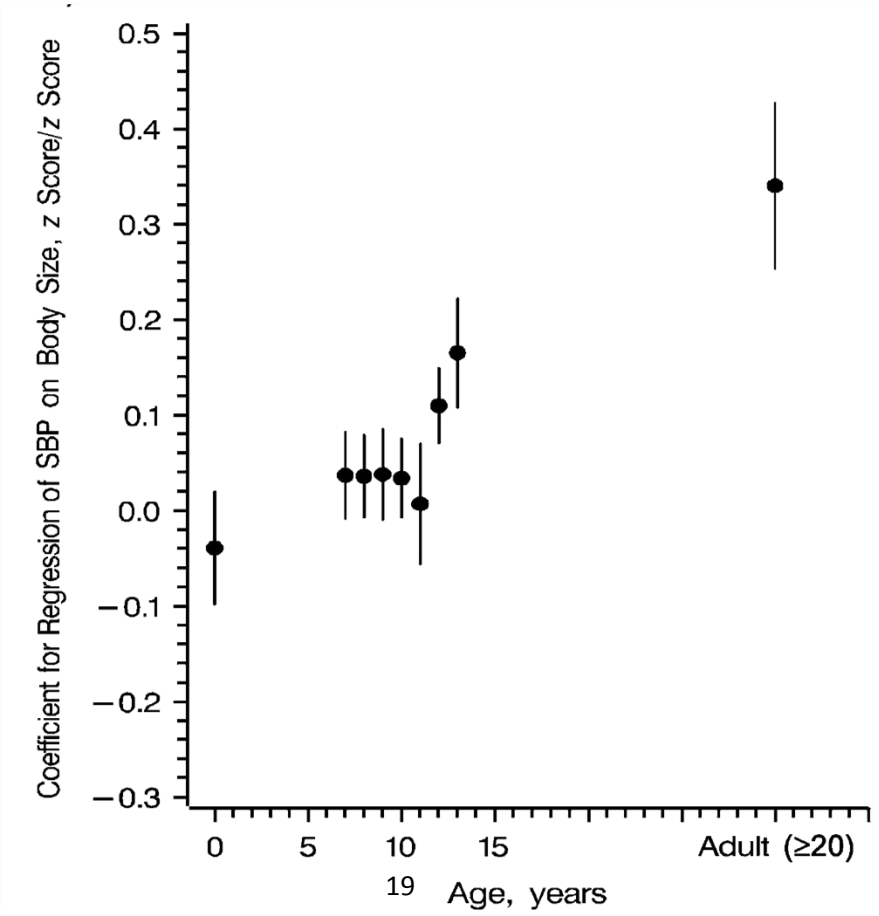
(b) Life course path analysis: The path analysis technique is an extension of multiple regression (59). It is an analysis which looks at the associations between body sizes at different time points and the associations between all body sizes and adult SBP (Figure 1.7). Using this technique, it is possible to assess direct and indirect effects and thus, the total effect of, size and change in body sizes at different ages on adult outcomes.

Figure 1.7: Life course path analysis diagram of association of early-life body mass index and later life systolic blood pressure



The direct effect of BMI at age 2 years on adult SBP is β_2 . The indirect effect of BMI at age 2 years can be calculated by adding the products of coefficients along all paths from BMI at age 2 years to adult SBP: $\beta_{2,A,}\beta_A + \beta_{2,3}\beta_3 + \beta_{2,3}\beta_{3,A}\ \beta_A$. The total effect is the sum of direct and indirect effects. When the total effects of change in body size on adult SBP, given past body size, are plotted in a graph it is known as a path analysis life course plot (Figure 1.8).The figure below shows the estimated total effects of body size on adult SBP given body size history from a sample of 1,284 Danish men born between 1936 and 1970. The total effects have been calculated. These include the potential indirect effect mediated through future body size. For example, the 9-year estimate is the total effect of BMI at age 9 years on adult SBP given birth weight and BMI at ages 7 and 8 years, including the effect mediated through BMI measured at ages 10 years to adulthood. The effect of change in relative body size on adult blood pressure was more significant after age 11 years than in earlier childhood. These results suggest that increases in body size prior to age 11 years are less harmful to adult blood pressure than increases occurring after this age.

Figure 1.8: Path analysis life course plot (derived from Gamborg et. al 2009 (58)).



(c) Conditional body size model: A common model for repeat exposures while analysing life course data is the conditional body size model (47). In this the idea is to remove the influence of growth at earlier time periods from the association between current growth and a later outcome and thus try to answer the question "what is the association of the exposures at the present time independent of the past time points?". This approach derives estimates of body size at defined time points which are uncorrelated with the conditional measures at all other time points. The conditional change scores are the residuals from the regression of the body size at each age on all earlier body size measurements. By construction, the conditional change variables at different time points are uncorrelated with each other. These scores can be interpreted as a change in body size which is above or below than expected in our cohort given earlier measures of the same body size. The conditional body size model is a regression of the later life outcome on the earliest measured body size and all conditional changes in body size over the time period. In these models, the coefficient for the outcome at age j represents the difference in outcome between two people with identical values for the outcome up to age (j-1), one of whom has a value of the body size which is one-SD higher than the other at age j. Such models can be used to obtain change above or below the expected value based on previous measurements and thus is a useful method for capturing accelerated or restricted growth. Using these models, we can explore sensitive periods in body development wherein changes in body size are associated with later life outcomes. This model is described in more detail in Chapter-3.

Wills et al. in 2016 has developed four further model parameterizations in accordance with their underlying growth pattern contrast (being bigger v/s being smaller, becoming bigger and staying bigger, growing faster v/s being bigger, becoming and staying bigger v/s being bigger) (60).

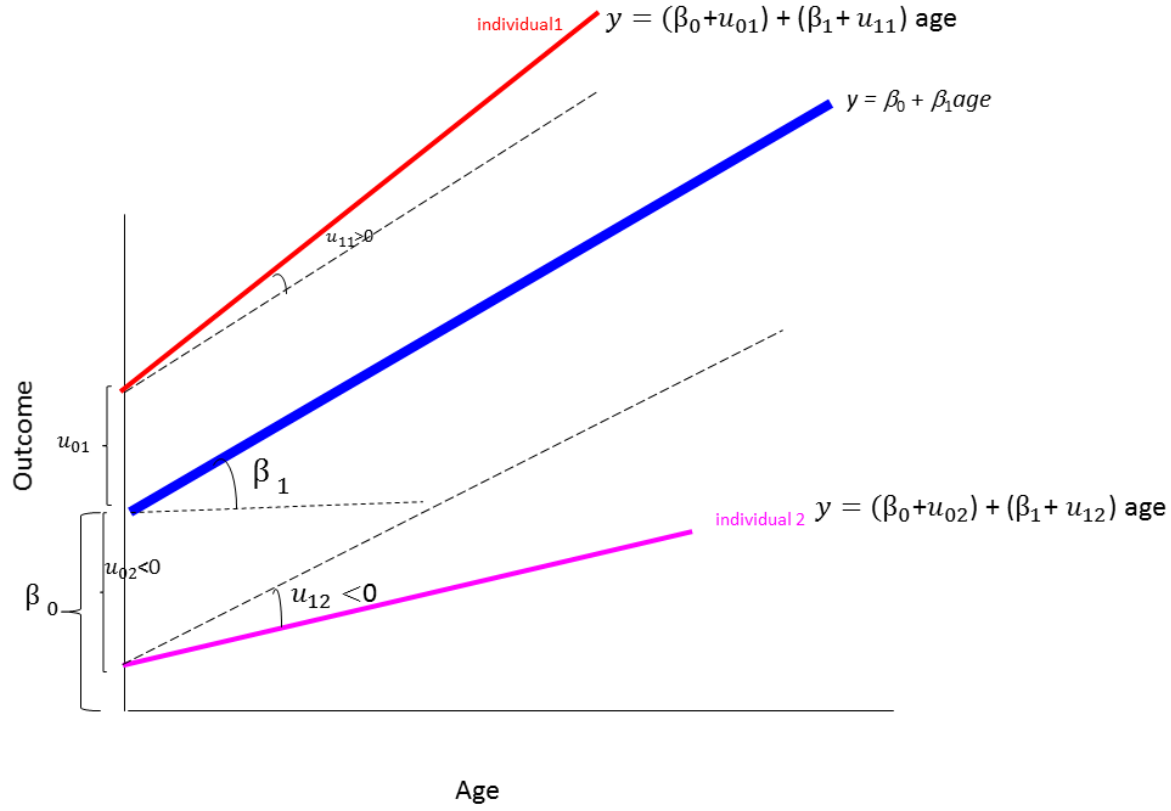
2. Regression with change scores: In this approach, the outcome is regressed on the growth in body size in different periods of time during the life course (61).

For p body weight measurements at different ages, there are (p-1) variables for increment changes in body size and the first body size or the (p-1) increment changes in body sizes and the last body size. We cannot regress the outcome on (p-1) incremental changes in body size and the first and last body size simultaneously, as these measures are linearly dependent.

The change-score body size model is a modified version of the life course plot model (58) and is better as compared to the life course plot model since changes in body size are less likely to be correlated to one another than in the life course plot. But, the problem of dependency among the repeated measurements still remains to a large extent.

3. Multilevel modelling: Multilevel modelling is one approach that can be used to model longitudinal data (62,63). In this method, individual-level random effects for the intercept capture the deviation of each individual from the average trajectory. In this method, there is no requirement for individuals to have been measured at the same ages, and varying numbers of measurements between individuals can be can be incorporated. In the case of missing data, these models make the assumption that the data are missing at random, that means that the probability of an observation being missing is associated with other observed variables for that individual, but is not associated with the value which is actually missing (64). A random slope model is displayed as:

Figure 1.9: Random Slope model



The bold line is the average value of the outcome at each age for the entire cohort. The coloured lines represent the trajectories for different individuals, which are assumed to vary for each individual (Figure1.9). These models assume that the error is normally distributed. Here β_0 and β_1 represent the average intercept and slope, respectively and u_{0j} and u_{1j} represent the deviation from the average intercept and slope, respectively, for individual j. The term e_{0ij}, represents the measurement error at the level of the occasion. These models can incorporate different correlation structures at the level of the occasion.

Multilevel models assume the growth in outcome is linear over time. This means the outcome grows by the same amount in each time period. But for most cases in health, the growth is nonlinear. One method by which this can be dealt with is by transforming the growth measurements or age, such that the association becomes linear. However, the interpretation of these results are very difficult. A second method is to include nonlinear terms in the model, such as nonlinear terms for age (63). The equation of a model with quadratic powers of age in a multilevel framework is as follows:

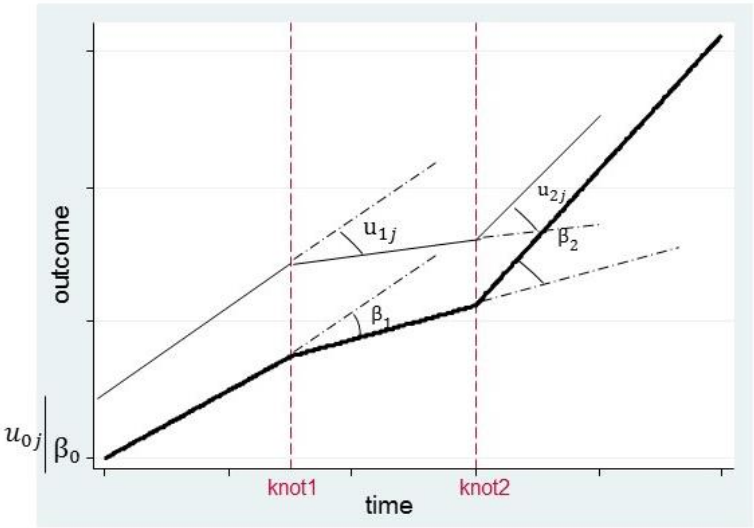
$$y_{ij} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})age_{ij} + (\beta_2 + u_{2j})age_{ij}^2 + e_{ij}$$

where $\beta_0, \beta_1 \, and \, \beta_2$ are the fixed coefficients describing the average shape of the trajectory, and u_{kj} describes the deviation of the individual's trajectory from the average.

The multilevel models that include nonlinear terms are not easy to interpret to explore associations between growth and later life outcomes. One approach that can yield more interpretable growth coefficients, is to use a series of linear splines, joined at different points, called knots to model the growth trajectory (63). These models have also been referred to as 'piecewise linear' or 'broken stick' models. Ages a_1, a_2, ..., a_{n-1} are "knots", a term that is used to convey that the model is a sequence of straight lines, one in each age interval, that are joined together at these ages to form a continuous curve, called a spline. For subject i we have n_i measurements of weight, denoted wt_{i_1}, ..., $wt_{i_{n_i}}$ measured at ages t_{i_1}, ..., $t_{i_{n_i}}$, which fall in the interval from a_0 to a_n inclusive (Figure 1.10).

Figure 1.10: Multilevel linear spline models

(derived from Corrie Macdonald-Wallis, University of Bristol)



The line fitted for subject i at age t is $\hat{wt_i}(t) = \sum_{k=0}^{n} (\beta_k + u_{ki}) s_{k(t)}$

where s0(t) = 1 and $s_{k(t)}$ = max(t-ak-1, 0) for k = 1, …, n are spline components; where β_0(intercept) and, β_1, …, β_n (slopes) are parameters that define the spline for the whole population; and where u_{0i}, u_{1i},, …, u_{ni}, are deviations that measure the differences

from the corresponding β term for subject i. The model assumes that all ε terms are independent random variables with mean zero and variance σ^2, and are independent of all u_{ki} terms. All u_{ki} terms are random variables with mean zero, and the covariance between u_{ki} and u_{li} is given by c_{kl}. .

I chose to use the conditional method in the first instance because it provides a description of growth in defined time windows in a particularly direct and interpretable way. It is most suited to situations such as this when data have been recorded in a regular and systematic way. Whilst the spline model shares some of the characteristics of the conditional model, it has several unattractive features. Firstly the piecewise linear function from which it depends is not a close approximation to the true curve. Second the variance and skewness of errors round the model can change markedly over an age range such as that spanned by the data used in our study. Third the slope estimates derived by the model can be heavily correlated, making it more difficult to interpret the results derived from the prediction of an outcome variable. This is particularly the case when, as here, some of the neighbouring intervals are relatively short and contain few data points. Nevertheless, it is possible to lessen some of the effects of these less attractive properties by analysing the standardized size measures. I do this in Chapter-6.

4. Latent growth curve modelling (LGCM): The conventional multilevel growth model assumes that the random effects vary across different individuals. In latent growth curve models the outcome does not vary across individuals but across groups of individuals (65). These models assume no variation within each group of individuals. These models are fitted within the help of structural equation models (SEM). The SEM incorporates the observed repeated measures on one or more latent factors to characterize the unobserved growth trajectories.

   The heterogeneity between the different trajectories is described by a finite set of unique polynomial functions, each corresponding to a different trajectory. When using LGCM, the researcher must specify the number of distinct trajectories to be extracted from the data. A priori knowledge regarding the number and shape of the trajectories can be obtained from the existing literature in the area of study. After obtaining parameter estimates from LGCM, posterior group membership probabilities are calculated for each distinct growth trajectories. A maximum probability assignment rule is then used to assign the

individuals to different growth trajectories. Trajectories having a higher group membership probability have more numbers of individuals assigned to them. Afterwards, average posterior probability of group membership is calculated for each trajectory identified in the data. The average posterior probability is a numerical approximation of the internal reliability for each trajectory. This is calculated by averaging the posterior probabilities of individuals having been assigned group membership to trajectory using the maximum probability assignment rule. Average probabilities greater than 0.7, a suggested rule of thumb indicates that the model groups individuals with similar patterns of change together and is able to distinguish between dissimilar patterns of change. Eg: Table 1.3 displays hypothetical data which is modelled by LGCM with three growth trajectories, resulting in three posterior probabilities for each individual. Using the maximum probability assignment rule, participants 1 and 5 would be assigned group membership to trajectory 3, participants 2 and 3 to trajectory 2, and participants 4 and 6 to trajectory 1.

Table 1.3 Hypothetical data which is modelled by latent growth curve modelling

Individual	**Trajectory 1**	**Trajectory 2**	**Trajectory 3**
1	0.11	0.11	0.81
2	0.02	0.88	0.02
3	0.05	0.87	0.11
4	0.94	0.03	0.06
5	0.21	0.04	0.77
6	0.96	0.02	0.02

5. Growth mixture models: The growth mixture model can be considered to be a more general case of the latent class growth curve models (65). The intercepts and slopes within each class have non-zero variances. If all variances in the growth factors are set to zero, growth mixture model gives the same results as a latent class growth model. The advantage of this model is that it allows variation across different groups of individuals and also allows variation in growth trajectories of individuals within these subpopulations.

6. Superimposition by Translation and Rotation (SITAR): This model uses a common spline function for all subjects whose growth trajectory can be modified by shifting two axes, body size and tempo of body growth, in order to

adapt this overall trajectory to the individual subject trajectories (66). A cubic spline function is fitted. The observed body growth of child i=1,2..n and t=1,2…t_i is expressed as:

$$Y_i(t) = \alpha_i + h\left(\frac{t-\beta_i}{e^{-\gamma_i}}\right) + \eta_{it}$$

where $Y_i(t)$ is the body size of subject i at age t, h(z) is the natural cubic spline curve of the growth variable regressed on z where z is the transformed age and $\alpha_i, \beta_i \; and \; \gamma_i$ are subject-specific random effects. These random effects are assumed to be drawn from a multivariate normal distribution with mean 0 and covariance matrix E. The error terms η_{it} are assumed to be independent and normally distributed random variables with mean 0 and constant variance τ^2, and independent of the random effects. The three parameters $\alpha_i, \beta_i \; and \; \gamma_i$ have a biological interpretation and are referred to as size, tempo and velocity respectively. α_i is a random body intercept that describes differences in mean body size, β_i is a random age intercept for differences in timing of the growth, and γ_i adjusts for the duration of the individual growth spurt. Three parameters are too few to be able to describe a growth curve observed across the complete lifecourse.

Discussion continues concerning the specification of growth models and their interpretation. Wills et al 2014, for example, correctly point out that the lifecourse plot, change scores model and the conditional model are simply reparameterizations of one another (67). The coefficients from each model are conditioned on different sets of transformed variables, and therefore they address different questions. Hence, different answers should be expected from these models, though their goodness of fit is identical. Tu et al. responded, pointing out that multilevel models, LGCM and growth mixture models are not reparameterizations of each other (68). They suggested that the different models addressed different research questions, and it is up to the researcher to choose the growth model which is appropriate for their research question.

Conclusion

The first part of the chapter gives an introduction to the earliest studies on fetal programming and the 'developmental origins of adult disease hypothesis'. This section also gives a summary of several important studies and reviews in this area and tries to

convey that there is very little literature specifically focusing on the association of head growth with later life outcomes.

The second part of the chapter gives an introduction to various statistical techniques used to model body growth and also discusses the conditional body size approach, which will be the main focus of my current research work. The main objectives of my research are as follows:

1. To study the associations of head size at birth and head growth at different ages during childhood with a human capital outcome (attained education) (Chapter 4).
2. To study the associations of head size at birth and head growth with adult SBP, a CVD risk factor (Chapter 4).
3. To study the associations of head size at birth and head growth with birth weight in the next generation (Chapter 5).
4. To compare the associations of head size at birth and head growth with adult CVD risk factors such as SBP, plasma triglyceride, plasma high density lipoprotein (HDL) cholesterol and fasting plasma glucose using two growth analysis techniques (Chapter 6).

I will carry out these analyses using data from the NDBC which is described in the next chapter.

Chapter 2

History of the New Delhi Birth Cohort

The NDBC study is one of the oldest and largest ongoing cohort studies in Asia. In this chapter, I have given a description of the cohort and described the data collected in the various generations and those applicable to my research work.

2.1 New Delhi Birth Cohort: Genesis and objectives

The NDBC was designed primarily to prospectively evaluate pregnancy outcomes with childhood growth and survival. The study was started in 1969 by Dr Shanti Ghosh, Dr I M Moriyama and Dr Santosh K Bhargava in a defined area of 12 sq. km in South Delhi, New Delhi, India. The goals of the study at inception were: (a) to establish norms for birth weight, and physical and neurological characteristics, at different gestational ages, in Indian newborns; (b) to document the morbidity and mortality patterns in small-for-dates and premature babies; (c) to study the growth and development of low birth weight babies; and (d) to assess the incidence of mental defects and congenital malformations among the small-for-dates and premature newborns. It was proposed to compare this group of babies with normal full term babies of higher birth weight. The proposed activities were to study an urban population of 100,000, to assess the socioeconomic background of each family along with per capita income. Further, it aimed to obtain a detailed obstetric history of every woman and to follow the women of child bearing age every two months to detect incident pregnancies by obtaining the menstrual history (last menstrual dates); Those who became pregnant were followed (a) every two months for assessing their nutrition; (b) to assess the outcome of pregnancy; and (c) to follow the baby for growth, development and morbidity pattern every three months; and (d) to measure neonatal, infant and childhood mortality rates.

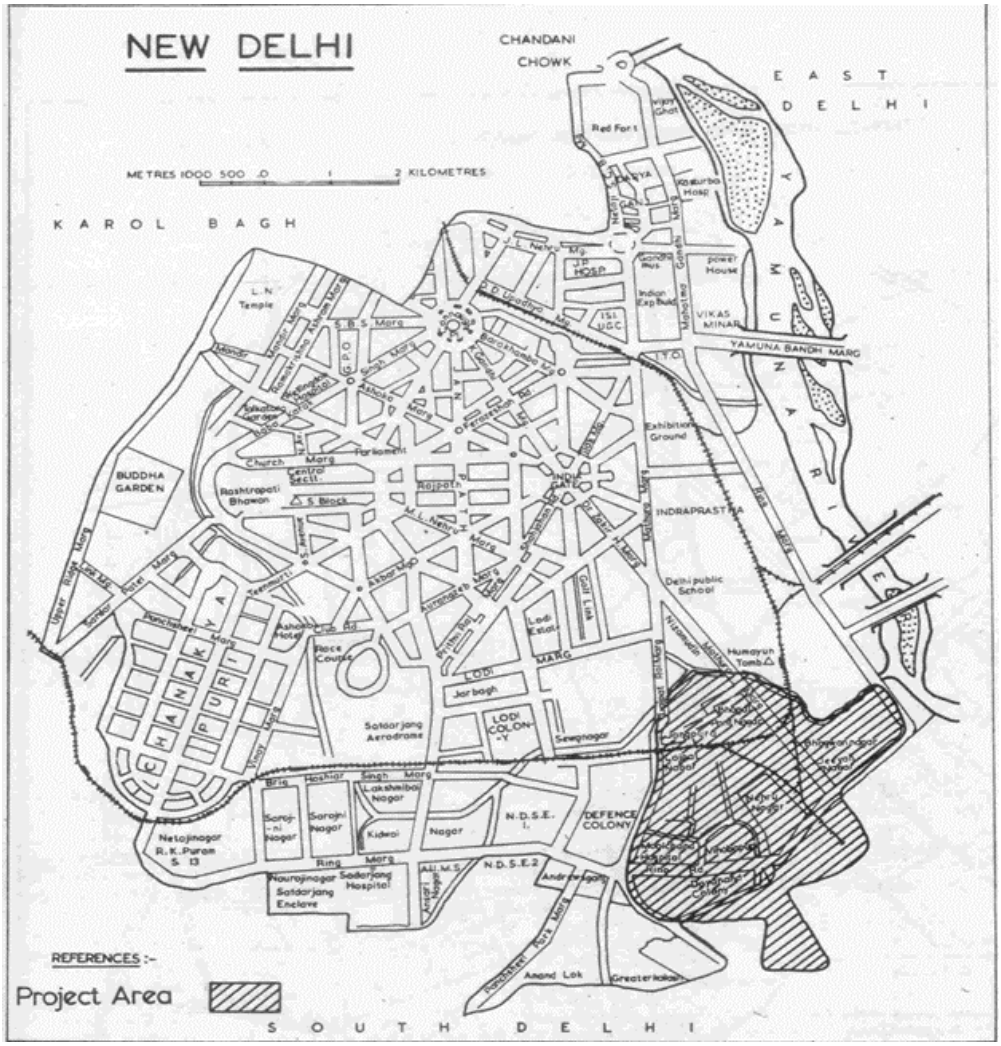
2.2 Description of the study area

The population of Lajpat Nagar (a colony situated in South Delhi) was initially chosen for the study. This population comprised families of different socioeconomic classes, varying from sophisticated middle class to people living in almost a rural setting in some places. Circulars in Hindi and English were sent to the house owners, giving them a brief introduction about the project and requesting them to cooperate with

the field workers who would visit their houses to obtain information about the families. A house-to-house survey was done by the field staff in October and November 1969. A diary was kept which had the name of the head of the household, address and number of members in each household. Since Lajpat Nagar was not able to fulfil the required population size, some additional colonies were also included within a total defined geographical area of 12 km^2 (shaded in Figure 2.1).

Figure 2.1: Map of New Delhi showing the project area (shaded)

(courtesy: New Delhi Birth Cohort study team)



2.3 Initiation of the cohort

Following the census carried out between October and November 1969, only families with a married woman were included. Information regarding the address and the number, names, ages and marital status of the family members was also recorded. Members of the household staying away permanently (e.g. married offspring, children living elsewhere) were excluded, but dependents living with the family were included. The relationship of each member of the family to the head of the household was also captured. No ethical clearance was obtained on initiation of the cohort as there were no ethical committees at that time. Institutional ethical committee clearance was obtained from 1998 onwards.

2.4 Training of staff and standardization of techniques

All staff involved in the study, which included the field interviewers and public health nurses, supervisors and anthropologists were aware of the objectives of the study. The field interviewers were then made to practice the techniques and form filling among hospital patients in order to give them practical experience and confidence. The public health nurses, senior supervisors and anthropologists were given special training on anthropometric techniques for infants, children and adults. The public health nurses also received training in recording blood pressure and measuring the haemoglobin content in blood.

2.5 Registration of ever-married women

At the time of initial visit, a card was made for each ever married woman with particulars of her current marital and menstruation status, occupation, present age, age at marriage and details of her past obstetric history. Women who were widowed, divorced, sterilized or post-menopausal were ineligible and were excluded from follow-up. Reproductively active women were visited every two months, and their menstrual dates were recorded. Some women became ineligible as the study proceeded, because either they or their husbands were sterilized. In the case of a missed menstrual period, the public health nurse contacted the woman within 10 days of the next expected period and if the woman had still not menstruated, booked her as a pregnant woman. Among a

population of 119,799 there were 20,755 married women in the reproductive age-group who were followed up, of whom 9,169 became pregnant

2.6 Follow up of the cohort

Nomenclature used for generations

The pregnant women followed up for pregnancy outcomes and their spouses will be referred to as the **F0** generation. Their children, followed up prospectively from birth until nearly 38 years comprises the main NDBC, the **F1** generation. The children born to F1 cohort members comprise the **F2** generation.

Follow up of pregnant woman (F0)

All pregnant women (n=9,169) were followed up by the public health nurses every two months to record morbidities. Blood pressure and haemoglobin were checked once during the third trimester. A record of the number of antenatal visits made, any hospitalization or use of drugs was also made. Around two weeks before the expected date of delivery, the public health nurse checked the weight of the pregnant women to assess weight gain during pregnancy. Frequent visits were made thereafter to record birth weight, other measurements at birth and gestational age. The type of delivery, whether live birth, twin, abortion or still birth was documented. There were 8,181 live births between December 1969 and November 1972, which included 8,030 singletons and 151 sets of twins. There were 202 stillbirth and 867 abortions.

Follow up of the infant (F1)

Newborns were examined by a paediatrician within 15 days of birth and the presence or absence of congenital malformations was noted. Trained field staff recorded the head circumference, weight and length of the newborn within 72 hours of birth, and at 3 months, 6 months, 9 months and 1 year (± 7days), and at 6-month intervals after 1 year (± 15 days). The weight of the babies was measured by infant weighing scales which were designed and manufactured locally. These scales had the capacity to weigh up to 10 kg. Later, scales with a capacity to weigh up to 16 kg became available. Aluminium infantometers, with 1 mm calibration, were used to measure the length of the babies. These were fairly sturdy instruments, able to withstand a certain amount of rough handling in the field. Aluminium was specially chosen for accuracy since

expansion with temperature variation is negligible. Steel measuring tapes were used to measure head and chest circumferences. Table 2.1 provides the number of measurements made at different ages stratified by sex. Children suffering from severe illness at the time of measurement were excluded. The field workers did not have access to the previous measurements when recording of new measurements. Measurements were entered in the database within a month of collection and incorrect values were reported. In this case the measurements were either discarded or repeated. In the case of the death of a baby or mother, attempts were made to assign the cause of death after obtaining information from the family, neighbours, hospital or medical attendants. If a family moved out of the area, attempts were made to find out the new address and the reasons for shifting. Similarly, new families moving in to the area until 30th November 1972 and babies born before 31st March 1973 were recruited into the cohort.

Table 2.1: Summary of head, height and body mass index at the different measurement occasions according to sex

	Males						Females					
	Head (cms)		Height (cms)		BMI (kg/m^2)		Head (cms)		Height (cms)		BMI (kg/m^2)	
Age (years)	N	Mean(SD)	N	Mean(SD)	N	Mean(SD)	N	Mean(SD)	N	Mean(SD)	N	Mean(SD)
0	3219	33.7 (1.3)	3260	48.7 (2.3)	3256	12.0 (1.3)	2979	33.3 (1.2)	3025	48.2 (2.2)	3023	12.0 (4.3)
0.25	2723	39.3 (1.3)	2778	59.5 (2.7)	2777	15.4 (1.6)	2522	38.4 (1.2)	2570	58.1 (2.5)	2568	14.7 (1.6)
0.50	2423	42.0 (1.3)	2457	65.3 (2.7)	2445	16.3 (1.6)	2253	40.9 (1.3)	2297	63.6 (2.6)	2291	15.7 (1.7)
0.75	2149	43.6 (1.3)	2185	68.9 (2.8)	2172	16.4 (1.6)	1989	42.2 (1.2)	2015	67.2 (2.7)	2004	15.9 (1.6)
1	1984	44.7 (1.3)	2003	72.0 (3.0)	1989	16.4 (1.5)	1791	43.5 (1.2)	1819	70.2 (3.1)	1804	15.9 (1.6)
1.5	1518	46.1 (1.3)	1548	77.1 (3.6)	1504	16.0 (1.4)	1346	44.8 (1.3)	1388	75.4 (3.8)	1336	15.5 (1.5)
2	1430	46.9 (1.3)	1442	81.3 (4.0)	1358	15.7 (1.4)	1307	45.7 (1.3)	1338	79.6 (4.0)	1262	15.4 (1.4)
2.5	1315	47.6 (1.3)	1337	85.3 (4.3)	1281	15.6 (1.4)	1183	46.5 (1.3)	1200	83.6 (4.4)	1153	15.3 (1.4)
3	1160	48.2 (1.3)	1170	89.0 (4.3)	1153	15.5 (1.3)	1132	47.1 (1.3)	1146	87.2 (4.6)	1127	15.2 (1.3)
3.5	1220	48.6 (1.3)	1221	92.3 (4.4)	1207	15.5 (1.3)	1183	47.7 (1.3)	1181	90.6 (4.7)	1169	15.2 (1.3)
4	1524	49.0 (1.3)	1531	95.8 (4.4)	1524	15.2 (1.2)	1434	48.1 (1.2)	1446	94.2 (4.7)	1435	15.0 (1.2)
4.5	1698	49.5 (1.3)	1702	99.1 (4.7)	1680	15.1 (1.1)	1563	48.5 (1.3)	1562	97.7 (4.7)	1546	11.4 (1.1)
5	1735	49.8 (1.3)	1733	102.2 (4.7)	1718	14.9 (1.1)	1628	49.0 (1.3)	1631	100.8 (4.9)	1618	14.6 (1.1)
5.5	1741	50.2 (1.3)	1738	105.5 (4.9)	1719	14.7 (1.1)	1644	49.3 (1.3)	1651	103.9 (4.9)	1641	14.4 (1.1)
6	1796	50.5 (1.4)	1789	108.5 (5.1)	1780	14.5 (1.1)	1686	49.7 (1.3)	1698	106.9 (4.9)	1691	14.3 (1.1)
6.5	1808	50.8 (1.4)	1810	111.5 (5.2)	1806	14.4 (1.1)	1678	50.0 (1.3)	1712	100.0 (5.2)	1709	14.2 (1.1)
7	1617	51.0 (1.4)	1621	114.4 (5.3)	1612	14.4 (1.2)	1508	50.3 (1.3)	1544	112.9 (5.3)	1540	14.2 (1.1)

7.5	1465	51.3 (1.4)	1465	117.1 (5.4)	1461	14.4 (1.2)	1307	50.6 (1.3)	1350	115.9 (5.5)	1348	14.2 (1.1)
8	1283	51.4 (1.4)	1290	120.0 (5.7)	1289	14.4 (1.3)	1197	50.8 (1.3)	1232	118.5 (5.6)	1231	14.2 (1.2)
8.5	944	51.7 (1.4)	951	122.7 (5.8)	949	14.6 (1.4)	854	51.0 (1.4)	864	121.3 (5.7)	863	14.4 (1.3)
9	720	51.8 (1.4)	726	125.1 (5.9)	724	14.7 (1.5)	649	51.3 (1.4)	660	123.9 (5.9)	660	14.5 (1.3)
9.5	351	51.9 (1.5)	353	127.7 (6.0)	350	14.6 (1.4)	323	51.5 (1.3)	327	126.9 (6.1)	326	14.6 (1.5)
10	305	52.2 (1.5)	311	130.3 (6.1)	310	15.0 (1.8)	264	51.6 (1.4)	273	129.4 (6.1)	273	14.7 (1.5)
10.5	125	52.4 (1.5)	127	131.7 (6.6)	126	15.2 (1.5)	126	51.9 (1.4)	125	131.0 (6.1)	125	14.9 (1.5)
11	220	52.0 (1.4)	224	135.7 (6.5)	224	15.5 (1.9)	208	52.1 (1.6)	214	136.5 (7.4)	214	15.4 (1.9)
11.5	438	52.1 (1.4)	441	138.9 (6.8)	441	15.5 (2.0)	408	52.6 (1.7)	410	139.2 (7.7)	410	15.7 (2.1)
12	726	52.3 (1.5)	732	141.5 (7.1)	732	15.8 (2.2)	709	52.7 (1.7)	721	141.9 (7.5)	720	16.1 (2.3)
12.5	908	52.4 (1.5)	918	144.1 (7.6)	918	15.9 (2.2)	904	53.2 (1.8)	916	145.3 (7.5)	914	16.7 (2.4)
13	1109	52.7 (1.6)	1139	147.1 (8.2)	1139	16.2 (2.4)	1114	53.4 (1.8)	1137	147.8 (7.0)	1136	17.2 (2.5)
13.5	1218	52.9 (1.6)	1243	150.5 (8.6)	1242	16.5 (2.4)	1247	53.7 (1.8)	1265	149.9 (6.6)	1261	17.6 (2.5)
14	1211	53.1 (1.7)	1251	150.5 (8.6)	1248	16.9 (2.5)	1284	53.7 (1.7)	1317	151.3 (6.3)	1312	18.2 (2.7)
14.5	1013	53.4 (1.6)	1036	157.5 (8.7)	1036	17.3 (2.6)	1095	53.9 (1.7)	1109	152.4 (6.1)	1105	18.6 (2.8)
15	810	53.8 (1.7)	828	160.4 (8.4)	825	17.7 (2.6)	847	53.9 (1.7)	871	153.4 (6.0)	859	18.9 (2.8)
15.5	547	54.0 (1.7)	559	162.8 (7.7)	557	18.1 (2.7)	597	53.9 (1.7)	603	153.9 (5.8)	600	19.4 (2.9)
16	424	54.2 (1.7)	434	164.6 (7.5)	424	18.4 (2.7)	452	53.7 (1.5)	454	154.1 (6.1)	446	19.6 (2.9)
16.5	422	54.1 (1.6)	430	166.3 (7.1)	397	18.7 (2.9)	400	53.4 (1.5)	405	154.3 (6.1)	387	19.6 (2.9)
17	409	54.3 (1.5)	411	167.5 (6.9)	393	19.1 (2.8)	432	53.3 (1.5)	435	154.9 (5.9)	428	19.7 (3.1)
17.5	403	54.6 (1.5)	403	168.7 (6.8)	398	19.4 (3.9)	406	53.2 (1.5)	408	154.9 (5.9)	404	19.8 (3.1)

18	418	54.8 (1.5)	419	169.2 (6.3)	419	19.6 (2.9)	393	53.3 (1.5)	396	154.9 (5.9)	394	20.0 (3.2)
18.5	361	54.9 (1.5)	363	169.7 (6.4)	363	19.9 (2.9)	419	53.3 (1.5)	424	155.4 (6.1)	423	20.1 (3.2)
19	285	55.1 (1.5)	287	170.3 (6.6)	287	20.0 (2.9)	409	53.3 (1.5)	413	155.5 (5.9)	413	20.3 (3.3)
19.5	276	55.3 (1.5)	282	170.0 (6.5)	281	20.1 (2.9)	402	53.3 (1.4)	406	155.7 (5.3)	406	20.2 (3.4)
20	246	55.4 (1.6)	250	170.1 (6.4)	250	20.4 (3.1)	297	53.3 (1.5)	297	155.7 (5.9)	297	20.2 (3.3)
20.5	85	55.6 (1.7)	88	169.7 (6.1)	88	20.3 (2.9)	131	53.3 (1.6)	131	155.9 (5.9)	131	20.2 (3.4)
21	26	55.8 (1.4)	26	170.6 (6.4)	26	21.1 (2.7)	11	53.2 (1.2)	11	156.5 (8.9)	11	19.9 (2.6)
Adult phase1 (26-32)	874	56.6 (1.7)	888	169.7 (6.3)	888	24.9 (4.3)	636	53.8 (1.7)	639	154.9 (5.7)	639	24.7 (5.1)

2.7 Follow up post-infancy until adulthood

There were several phases during the prospective follow up of the cohort (Figure 2.2).

Phase 1, 1969 to 1973: This phase included the recruitment of married women, their pregnancies and 8,181 live births. Of these 457 children died and 605 were lost to follow-up by the end of phase 1, leaving 7,119 infants and young children.

Phase 2, 1974 to 1980: There was a small gap between phase 1 and phase 2 because of a lack of continuity in funding. Phase 2 started with 7,119 infants and 2,414 infants, were lost to follow up in this phase. Some of these were deaths, but the major reason for this large lost to follow up was migration due to demolition of unauthorized housing and resettlement of the displaced population to outside the study area.

Phase 3, 1983 and 1987: Again, there was a gap in follow up because of funding. Of 4,104 children available at the beginning of this phase, 3,337 remained in follow up at the end.

Phase 4, **1987 to 1990**: In this phase the children born after 1970 were excluded, as one of the objectives was to study growth up to age 20 years. Thus, only 1,030 children were eligible in phase 4. Out of these, 836 were studied and 194 were lost to follow-up.

In 1995, 2,584 children from the original cohort could be traced back from the 3,337 participants in phase 3 and 1,526 of those gave consent for participation. This was the first time that the cohort had been contacted after phase 4. This was also the time when the decision was made to study the cohort from a DOHaD perspective. This was the first time that the NDBC group collaborated with the Medical Research Council Lifecourse Epidemiology Unit (MRCLEU) in Southampton. The original cohort data were stored on magnetic tapes, which were obsolete by 1995 and made the data difficult to access. Most of these tapes were translated into modern data formats using tape readers.

During the adult phase **(phase 5),** which was conducted between 1998 and 2002 at the ages of 26-32 years in which weight, height, head circumference waist and hip circumferences and skinfold thicknesses were measured. Blood pressure was also measured using standard equipment. Fasting plasma glucose, insulin, fibrinogen and plasminogen activator inhibitor-1 (PAI-1) and serum total, high density lipoprotein

(HDL) and low density lipoprotein (LDL) cholesterol, and triglyceride concentrations were also estimated. A 75-gram oral glucose tolerance test (OGTT) which measured plasma glucose fasting, and 30 and 120 minutes after an oral glucose load, was also carried out.

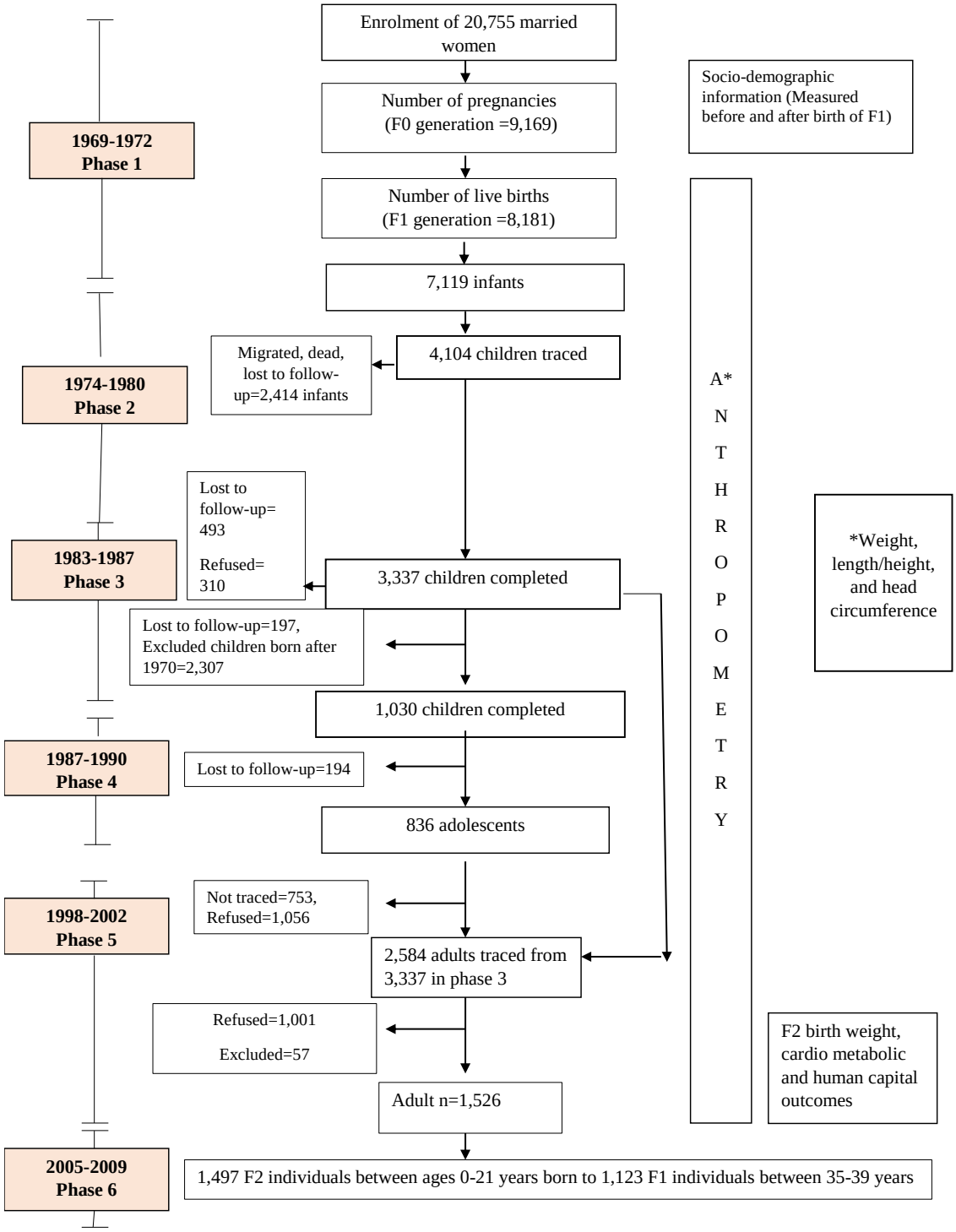
The protocol for the NDBC adult head measurements stipulated the landmarks to be used (just above the eyebrows or eyebrow ridges anteriorly and around the greatest bulge of the skull posteriorly, in order to obtain the widest circumference) and instructed the measurer to check that the tape was not slanted. If the participant's hairstyle impeded the correct placement of the tape (for example because of plaits or ponytails) measurers were instructed to untie hair, and if the subject was wearing headgear they were asked to remove it (including turbans in the case of Sikhs). Field staff were trained before the study started and their technique was supervised regularly. Inter and intra-observer variability studies were also carried out, approximately every 6-9 months during the study.

The most recent follow-up of the cohort **(phase 6)** was conducted between 2005 and 2009, when 1,123 of the phase 5 participants were evaluated between the ages 35 and 39 years. In this phase the investigators started to record information about the offspring of the cohort (F2 generation).

2.8 Variables measured in the F2 generation

Unlike the F1 generation, the F2 generation was not followed up prospectively in a systematic pre-designed manner. The variables were opportunistically measured when the cohort participants attended the clinics in phase 6. They were invited to bring their children for these visits. Anthropometry was recorded for all children above the age of 1 year. Blood pressure, hand grip strength, serum biochemistry, bioimpedance and a sleep questionnaire were recorded for all children above age 5 years. Information on birth events, immunization and school performance was elucidated from the parents by questionnaires. A series of small research grants enabled the follow up to continue. Up to 2012, a total of 1,492 F2 children were studied at ages ranging from 1-20 years.

Figure 2.2: Flow-Diagram of the New Delhi Birth Cohort since inception



2.9 Variables from the New Delhi Birth Cohort used in my study

Table 2.2 shows the variables from each generation that I have used in my research.

Table 2.2: Variables collected and used in my analyses

Generation	Variable type	Time point(s)	Variables used in my analyses
F0	Socioeconomic	Before birth, infancy, adulthood	Information on maternal education, assets (wealth), household income, paternal education, paternal occupation, access to health services, combination score of any health promotion or preventive health service utilization during antenatal and postnatal periods (e.g.: antenatal care, immunization), crowding index (people/rooms), child dependency ratio (children<18/adults). A combined socioeconomic score has been calculated.
F1	Anthropometry	At birth	Head circumference, birth weight, birth length, BMI , gestational age
	Growth	From birth until adolescence	Head circumference, weight, length/height, BMI
	Chronic diseases risk factors	Adult Phase	SBP, plasma triglycerides, plasma HDL cholesterol and fasting plasma glucose
	Human capital outcomes		Attained education was recoded as 0 years for illiterate, 3 years for primary education, 8 years for middle education, 12 for high school and 13.5 years for high school+ (senior secondary), 15 years for graduates and 17 years for a professional degree for the analyses.
F2	Anthropometry	Not followed up prospectively in a systematic pre-designed manner	Birth weight

Discussion

The NDBC has some favourable characteristics. Firstly, the location, Lajpat Nagar and its adjacent colonies, were large enough to fulfil the target of 100,000 population. There were four large hospitals nearby, which meant that women could deliver in or close to the study area. All the deliveries were either these hospitals or at home. A second advantage is the prospective longitudinal design. Such a design is appropriate to answer the research questions listed in chapter-1 that are exploring the associations of physical growth since birth with outcomes in adulthood.

There are two ways in which follow-up studies may be conducted over time, retrospectively and prospectively. The first approach is that one may identify a group with certain characteristics (such as birth weight) by means of historical records, at a certain defined time in the past, and then associate this characteristic with an outcome measured at a later age (such as adult CHD) (1). This type of study is called a historical or a retrospective cohort study. There are some advantages of this design. Retrospective cohort studies are typically constructed from healthcare databases that have already been collected, hence would save time and resources that are required for a prospective study with long follow-ups. This design would not have been possible in my study as there was no routine system of record keeping of size at birth or childhood growth in India at that time. The key disadvantage of retrospective cohort studies is that there can be selection bias if the cohort selected was not representative of Delhi as a whole. A second disadvantage is that the records may not have certain required risk factors, or the information may be incomplete. Prospective cohort studies can take care of these issues by having a sampling plan that is representative, collects information systematically and maintains a track of all included participants. However, prospective cohort studies are extremely resource intensive and can suffer from attrition bias affecting the internal and external validity of the study results. Attrition bias is more likely if it is differential between exposure groups and if the reasons for missingness are related to the outcomes. A study found no important bias with levels of loss to follow-up that varied from 5 to 60% when loss to follow-up was related to missing completely at random (MCAR) or missing at random (MAR) mechanisms (69). However, when observations were lost to follow-up based on a missing not at random (MNAR) mechanism, the authors found seriously biased estimates. The NDBC suffered from serious attrition of participants owing to long follow up. The main reasons were deaths in early childhood, migration

out of the locality in childhood because of demolition of unauthorized housing and migration out of the locality in adulthood due to greater mobility of the population in recent years. While mechanisms of missingness and the contribution of missingness to bias are difficult to establish, I will be able to test for systematic differences of exposure between individuals who remained in the study and those who were lost to follow-up. For example, in NDBC I will be able to compare size and SES at birth, and childhood growth, between cohort members with complete and incomplete follow-up. This comparison is described in Chapter-4 (Page 98).

A longitudinal study design usually involves obtaining repeated measurements of a particular exposure or outcome of interest. In the NDBC, the exposure is the head and the body measurements that were collected at several time points. A prospective design with the exposure measured at only one time point is more restrictive in terms of the scientific questions that can be answered. In DOHaD studies such as the present research, changes in body size are as important as body size at birth. Further, variables that have a time component may be more accurate than variables at a single time point because measurement errors can be identified. Another advantage of repeated measures data is that it can allow nonlinear growth to be modelled. This is of essence in the NDBC as the main exposure in the research, head circumference, follows a nonlinear growth trajectory.

In summary the NDBC, which has a prospective study design with repeated measurement of head and body measurements has several advantages. The data is suitable for exploring associations of birth size and growth with adult outcomes. However the influence of attrition on estimates should be explored and results should be cautiously interpreted.

Chapter 3
Methodology

The quality of data is very much dependent on adequate training of the interviewers in taking body measurements and on robust data entry methods. The primary aim of data cleaning is the detection and correction or removal of unrealistic values. Checking of the data is crucial to avoid incorrect results. In this chapter I have described the techniques undertaken for data cleaning of the head circumference data of the NDBC dataset. I have also described the mathematical technique of linear interpolation and how it was used to estimate the head circumference measurements at specified ages. Further I have described a conditional model that looks at the associations of body growth in specific time periods and adult outcomes. I have also summarized the method I used to compute the SES score using socioeconomic information collected from the F0 generation. Finally, I have described multilevel modelling, a statistical technique used in a subsequent chapter to assess associations of body growth measurements with outcomes in the next generation.

3.1 Data format

The head circumference data were provided in a "long" data file, wherein each individual was represented by multiple records in the data file, one per measurement. There were 7,522 individuals and 96,731 head measurements before data cleaning. Age was calculated for all the study participants from the date of measurement and the date of birth. I computed the number of individuals with head circumference recorded in each of the different measurement intervals (described in Chapter 2, table 2.1). The data cleaning procedure for head circumference measurements was initiated by deriving SD or z scores. I calculated the age-adjusted head circumference SD scores for every measurement in the cohort using two methods, the normal model and the LMS method which are described in sections in 3.2.1 and 3.2.3 below. A standardized score expresses how far the raw score is from the mean in terms of SD units. Standardized scores made the head circumference measurements comparable across the different ages.

3.2 Methods for creating age-adjusted standard deviation scores

3.2.1 Normal model

In the normal model, the assumption is that at each age of measurement, the body measurement has a normal distribution with a mean and SD that both vary smoothly with age (70). Least squares regression analysis by fitting splines is used to estimate both the mean and SD. A D^{th}-order spline is a string of concatenated polynomials of degree D joined at predefined data points called knots, such that the spline and its derivatives up to order (D-1) are all continuous. An approach to choose the degree of the spline is as follows (71). The spline of degree D with K knots, represented by $\xi_1, \xi_2 \ldots\ldots \xi_k$, is of the form:

$$y = \beta_0 + \sum_{d=1}^{D} \beta_d \, x^d + \sum_{k=1}^{K} b_k \, f(x, \xi_k)$$

where, $f(x, \xi_k) = 0$ if $x < \xi_k$ and $f(x, \xi_k) = (x - \xi_k)^D$ if $x \geq \xi_k$,

where y is the body measurement and x is age. Fitting polynomials of degree more than three is not recommended because it will be unrealistic (71).

The standardized scores are derived by fitting a spline through the longitudinal body measurements. The mean body measurements are estimated by the fitted values of the spline. The fitted values for SD are calculated by regressing the absolute residuals of the body measurement on age, using the following theory.

If X ~ N (0, 1)

then $E|X| = \int_{-\infty}^{0} -x \, \frac{1}{\sqrt{2\pi}} \, e^{\left\{\frac{-x^2}{2}\right\}} dx + \int_{0}^{\infty} x \, \frac{1}{\sqrt{2\pi}} \, e^{\left\{\frac{-x^2}{2}\right\}} dx$

Therefore, $E|X| = 2 \int_{0}^{\infty} x \, \frac{1}{\sqrt{2\pi}} \, e^{\left\{\frac{-x^2}{2}\right\}} dx$

Therefore, $E|X| = \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} x e^{\left[\frac{-x^2}{2}\right]} dx = \sqrt{\frac{2}{\pi}} \left[-e^{\left[\frac{-x^2}{2}\right]} \right]_0^{\infty} = \sqrt{\frac{2}{\pi}}$

So, if $E|residual| = Y$ $(obtained\ by\ fitting\ a\ spline\ to\ absolute\ residuals)$, then the SD

$$= Y \Big/ \sqrt{\frac{2}{\pi}}$$

So $sd\ at\ an\ age = fitted\ absresidual \Big/ \sqrt{\frac{2}{\pi}}$

Thus, the standardized scores are obtained by the formula:

$$z = \left(hc - mean\left(\widehat{hc}\right)\right) \Big/ \left(\sqrt{\frac{\pi}{2}}\ \left(|(\widehat{hc_{res}})|\right)\right.$$

Where, $mean\left(\widehat{hc}\right)$ gives the mean of the fitted value of the body measurements and $\left(\sqrt{\frac{\pi}{2}}\ \left(|(\widehat{hc_{res}})|\right)\right.$ gives the SD.

Extensions to the Normal model

3.2.2 Normal model with logarithmic transformation

Many body measurements tend to follow a right skewed non-normal distribution. This conflicts with the assumption that at each age, the data are normally distributed. Royston suggests an initial attempt to fit the Normal model without transformation (70). If the residuals from this model show a positive skew then a logarithmic transformation should be performed on the original values of the body measurements (y), and the model refitted on log(y). If the residuals from the refitted model are once again skewed, it is then recommended to try using a modified logarithmic transformation of the form log(y+C), where C is positive. C is varied until the residuals become normally distributed. If the effect of transformation on the body measurement is small, it is usually better to consider the original body measurements. If the coefficients of polynomial curves of degree higher than three are statistically significant, fractional polynomial curves are suggested as they may fit the data better (72). Fractional polynomials allow parameters to take fractional powers. While the conventional polynomial is of the form

$$a + b * age + c * age^2$$

where a,b and c are the coefficients

Fractional polynomials are of the form

$a + b * age^p + c * age^q$, where p and q are numbers selected from the set {-2, -1, -0.5, 0, 0.5, 1, 2, 3}. The p and q are placed in ascending order and power 0 is a natural

logarithmic transformation. As an example, a fractional polynomial with powers (0, 2) is of the form

$$a + b\,log(age) + cage^2$$

The advantage of the normal model is its simplicity. The estimation technique employed is least squares and it can be extended to use of simple data transformations. These are available in all statistical software. However, there are some cases that cannot be handled by this method. Skewed data may sometimes be corrected by transformation, but this is not always successful. Also age-varying skewness is difficult to adjust for using this method (72,73). Another disadvantage of this method is that even after transformation, kurtosis may remain in the data, which contradicts the assumption of normality. There are also other methods for calculating age-adjusted SD scores, one of them will be discussed in detail, the LMS method:

3.2.3 LMS Method

In this method, three parameters L (lambda (λ)), M (Mu (μ)) and S ((Sigma (σ)) are estimated for the body measurement (y) at every age of measurement (74,75). The M and S curves correspond to the median and coefficient of variation of the measurement at each age, whereas the L curve allows for skewness in the distribution of the body measurement. These three parameters are estimated simultaneously by maximum penalized likelihood and smoothed across age. The advantage of this approach is that the curve fitted across age is directly controlled by the values of three smoothing parameters, known as equivalent degrees of freedom (EDF). The LMS method uses the Box-Cox power transformation, which deals with the skewness present in the distribution of the body measurement and provides a way to normalize the measurement (76). This transformation can be defined as:

$$g_\lambda(y) = \left(y^\lambda - 1\right)/\lambda, \text{for } \lambda \neq 0$$

$$g_\lambda(y) = \ln(y), \text{for } \lambda = 0$$

The main assumption underlying the LMS method is that after Box-Cox power transformation, the body data at each age are normally distributed. If y is positively skewed, values of $\lambda < 1$ are needed. If y is negatively skewed, values of $\lambda > 1$ are needed. The z-score for a particular body measurement is calculated using these L, M

and S curves for a specific age and sex of the child. The formula for calculating SD score depends on the value of λ, μ and σ and is.

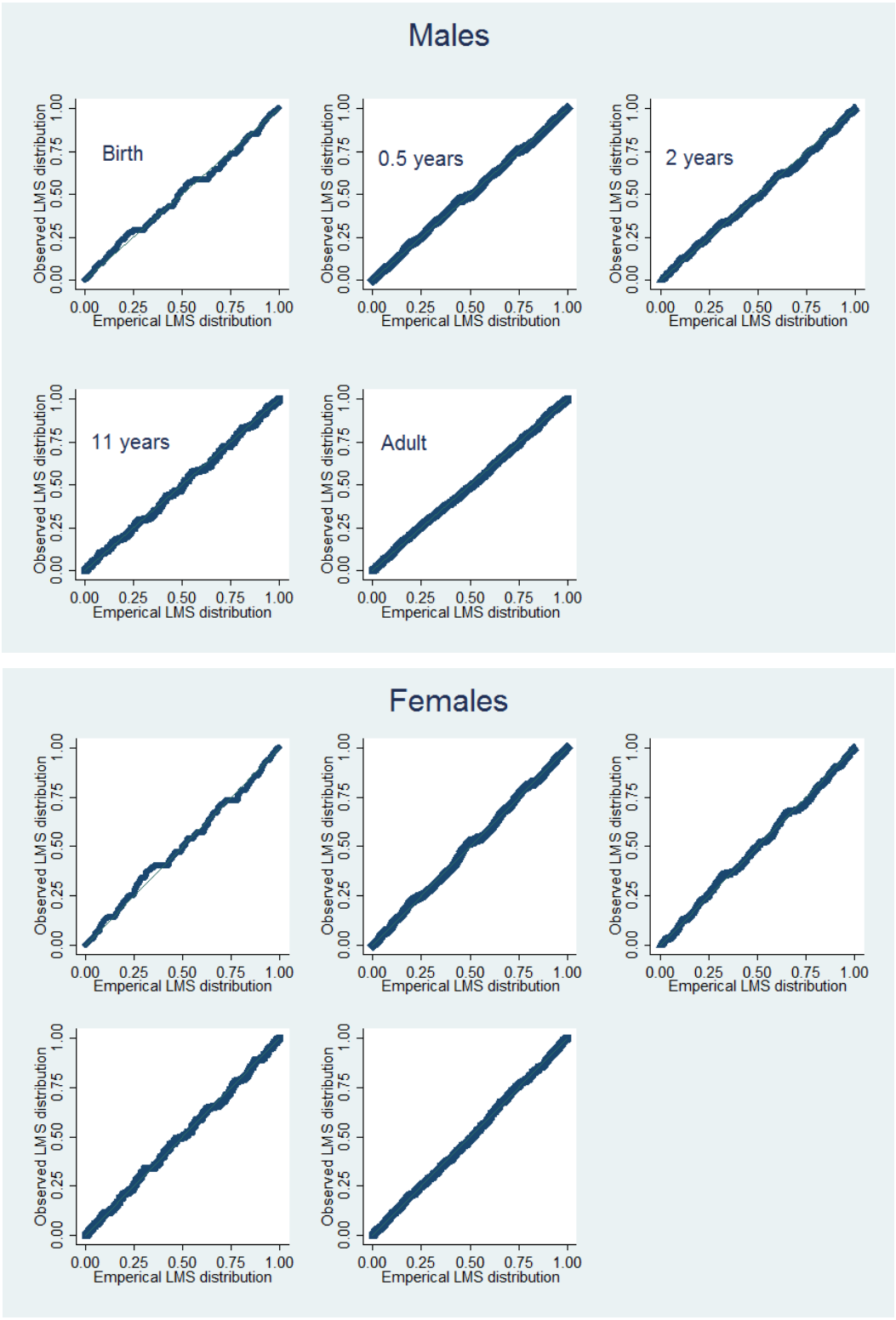
$$Z = \frac{(y/\mu)^{\lambda}-1}{\lambda\sigma}, \lambda \neq 0$$

Or,

$$Z = \frac{\log(y/\mu)}{\sigma}, \lambda = 0$$

This method is extremely flexible and is widely applicable (77). Body measurement SD scores can be easily calculated, even if the distribution is complex, and age-varying skewness is also handled. The L, M and S curves completely summarize the measurement's distribution over the age range. A potential disadvantage of this approach is that the choice of EDFs is subjective. A guideline has been suggested to evaluate the goodness of fit by comparing the difference in deviance (-2log (penalized likelihood)) between two models to a χ_e^2 distribution, e being the difference in the number of parameters being estimated in the old model as compared to the new model (74). However, in large datasets such as ours there is great power to detect small fluctuations in the L, M and S curves, some of which may be implausible. Therefore larger differences in deviance may be required. The probability-probability (P-P) plot can also be used to assess the goodness of fit (Figure 3.1). A slight deviation from normality can be observed for head size at birth for females. This may be because of small head sizes due to prematurity (53). The other periods seem normally distributed. There are other plots to assess goodness of fit. A detrended Q-Q plot displays the difference between the empirical and theoretical quantiles. These plots are made for every age of measurement and are also known as 'worm plots'. A model that fits the data well shows approximately a flat pattern of points at every age (78). Another disadvantage is that the normality assumption after applying the Box-Cox transformation may be violated if there is kurtosis, which is not handled by this method (74).

Figure 3.1: Sex-specific P-P plots of head circumference SD scores by the LMS method

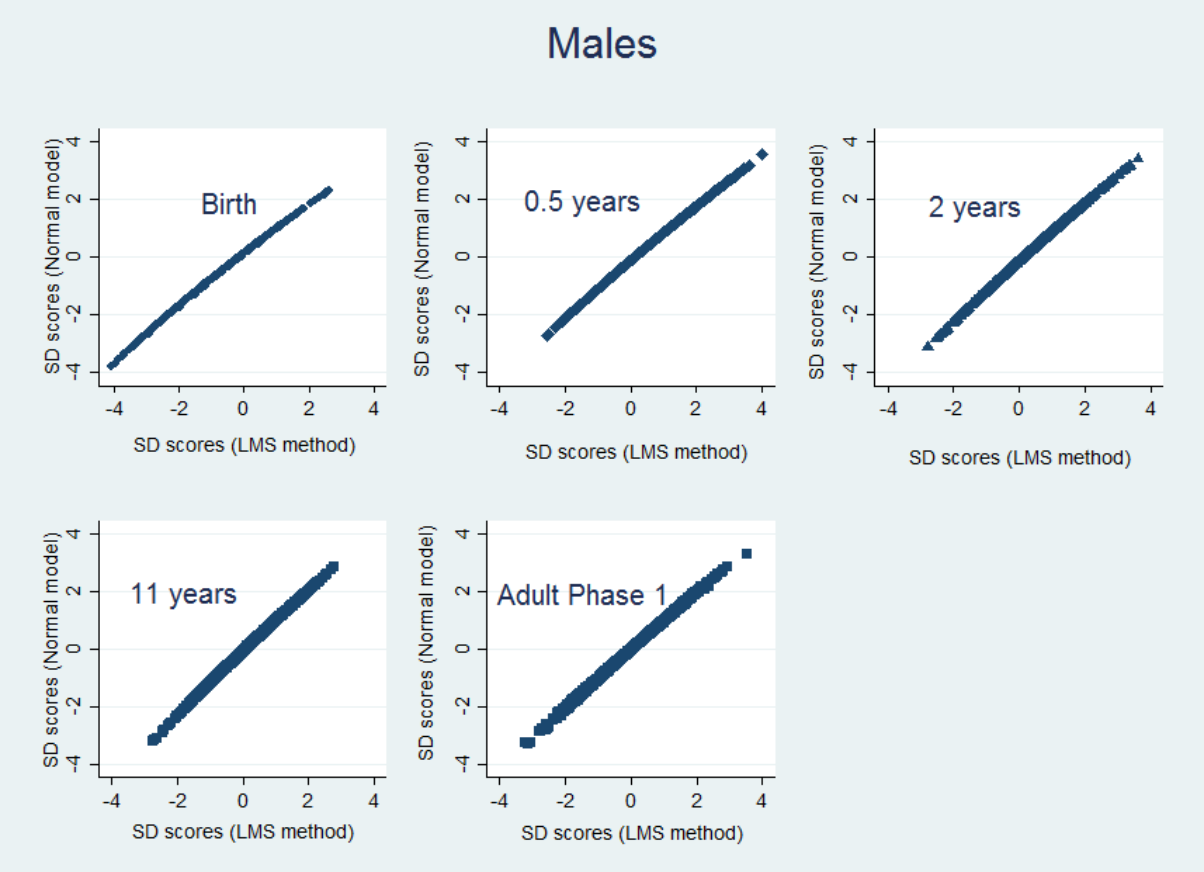


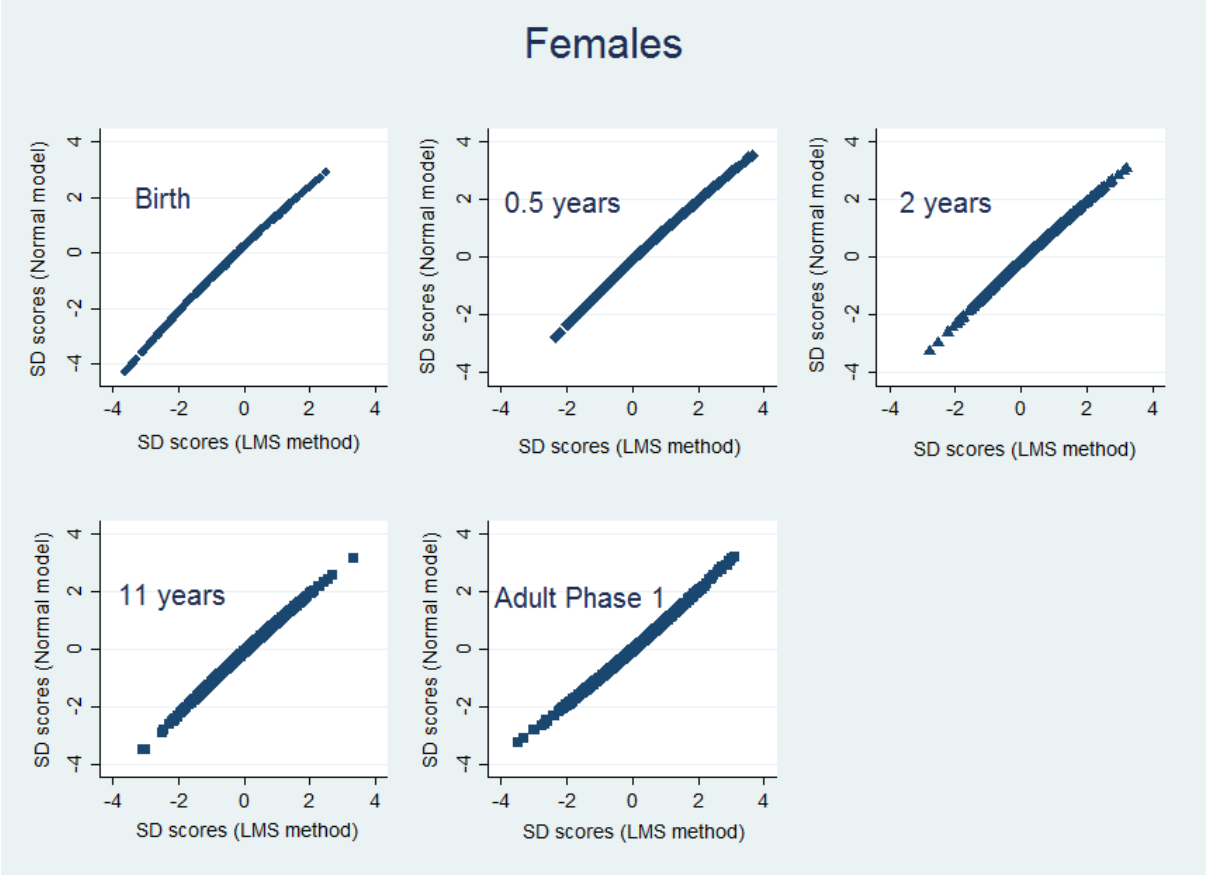
3.2.4 Comparison of the Normal model and the LMS method

I calculated the SD scores based on the normal model and the LMS method for every head measurement in the study. The correlation between the SD scores from the two

methods was (>0.99) in all the measurement time periods of interest for both the sexes (Figure 3.2). Also, the mean difference in the SD scores calculated by the two methods (LMS - Normal) was 0.0001 (SD: 0.12) for males and 0.0003 (SD: 0.08) for females, suggesting there was little evidence of bias in the SD scores calculated using either of the two methods. This suggested that I can use any one of the methods for calculation of the head circumference SD scores. I thus used the SD scores calculated using the normal model for the subsequent work.

Figure 3.2 Sex-specific correlation of head circumference standard deviation scores by the two methods





3.2.5 Data cleaning of head circumference data by creating age-adjusted standard deviation scores

Since the relationship between raw values of head circumference against age was clearly nonlinear with a period of rapid head growth during infancy, the standardized scores for the longitudinal head circumference measurements were derived by fitting a cubic spline through the set of head circumference measurements, through the "knots". For the present data, the knots were chosen at ages 0.25, 0.75, 2, 4, 5.5, 7, 8.5, 13 and 16 years as they divide the number of head circumference measurements evenly. Head circumference SD scores were calculated and values greater than 4 and less than -4 were considered implausible and were subsequently removed. For checking SD scores less than 4, "surprise" values were created by the mathematical technique of linear interpolation (described in section 3.3).

3.3 Linear Interpolation between two known points

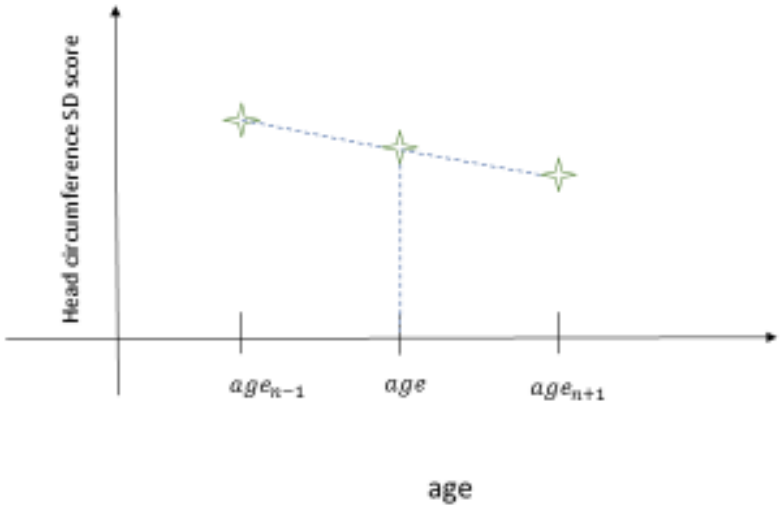
The following section describes the process of linear interpolation. Figure 3.3 shows two standardized head circumference scores (z (hc)) plotted against age. If two known points are given by the coordinates (age_{n-1}, z_{n-1}) and(age_{n+1}, z_{n+1}), the linear

interpolant is the straight line between these points (Figure 3.3). For a given value of age in the interval (age_{n-1}, age_{n+1}), the value of head circumference SD score at any age can be calculated from the equation:

$$\frac{\text{fitted}(z_n) - z_{n-1}}{age_n - age_{n-1}} = \frac{z_{n+1} - z_{n-1}}{age_{n+1} - age_{n-1}} \qquad (1)$$

$$\text{surprise} = |z_n - fitted(z_n)|$$

Figure 3.3: Linear interpolation of head circumference standard deviation score at specified ages (in years)



3.4 Data cleaning using surprise values

For each individual I plotted separately the age-adjusted standardized head circumference values against age and also the actual head circumference values against age. I calculated the absolute difference between an SD score and that which would have been predicted by interpolation between its neighbours. This measured the "surprisingness" of an observation. I called such values "surprise values". For the first value of an individual, "surprise" values were calculated by taking the absolute difference between the first SD score and the SD score at the next measurement for the individual. For the last value of an individual, "surprise" values were calculated by taking the absolute difference between the last SD score and the previous SD score for that individual. Starting with the individuals having the most surprising values for head circumference, I critically investigated the values at each time point and removed the head circumference values for those time points which had a surprisingly low or high value of head circumference at that age. After removing this suspicious head value for an individual, I recalculated the SD scores and the surprise values for all remaining

head circumference values. This was done iteratively for every value in the dataset. As an example, for individual ID=20165, (Figure 3.4), the graph of actual head circumference values against age was obtained (Figure 3.5). All points in all the subsequent graphs are labelled with their corresponding surprise values.

Figure 3.4: Plot of head circumference SD scores and age (ID=20165)

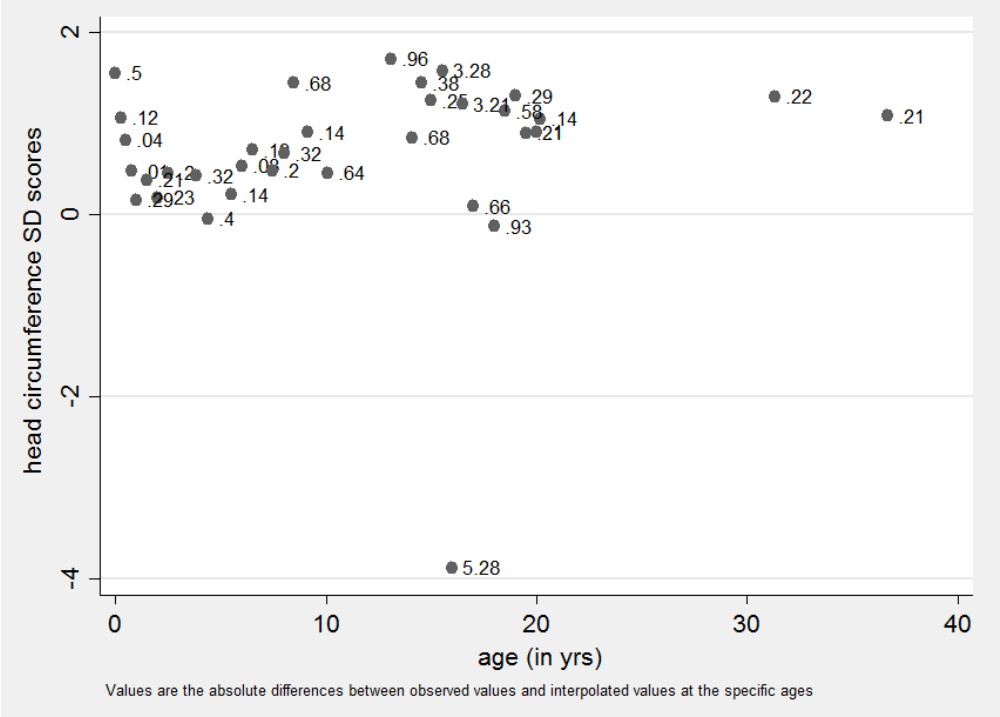
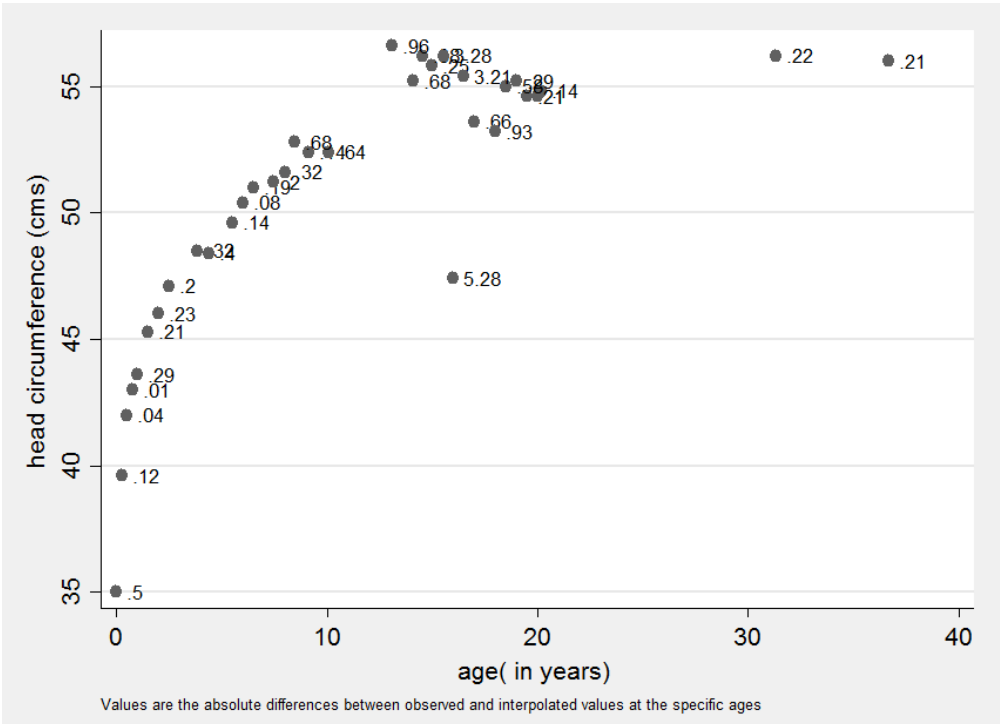
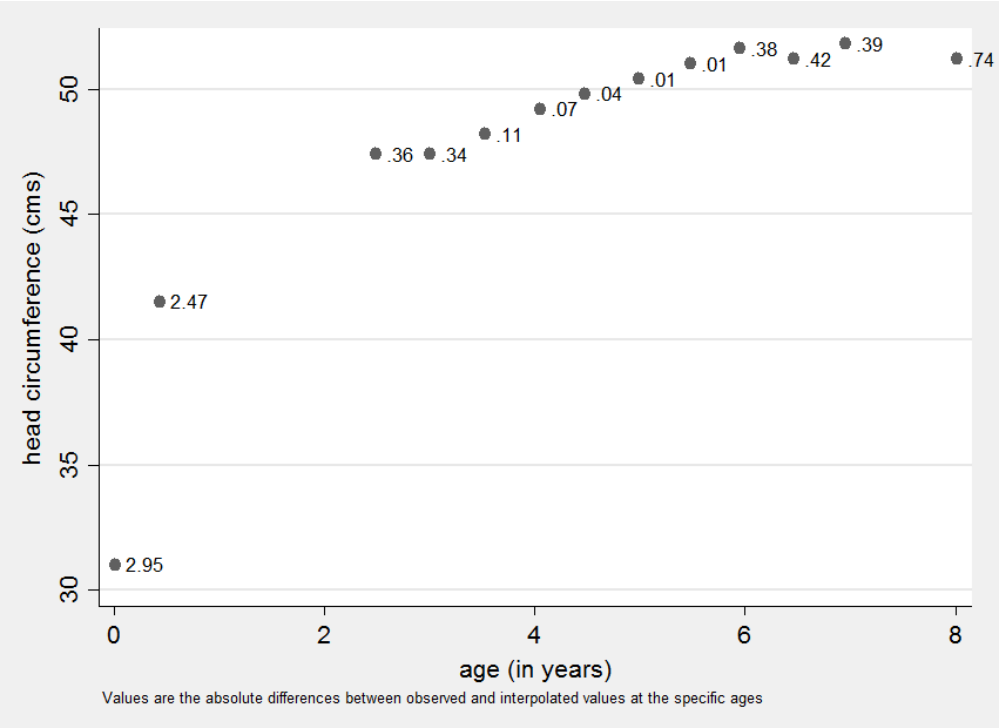


Figure 3.5: Plot of head circumference measurements and age (ID=20165)



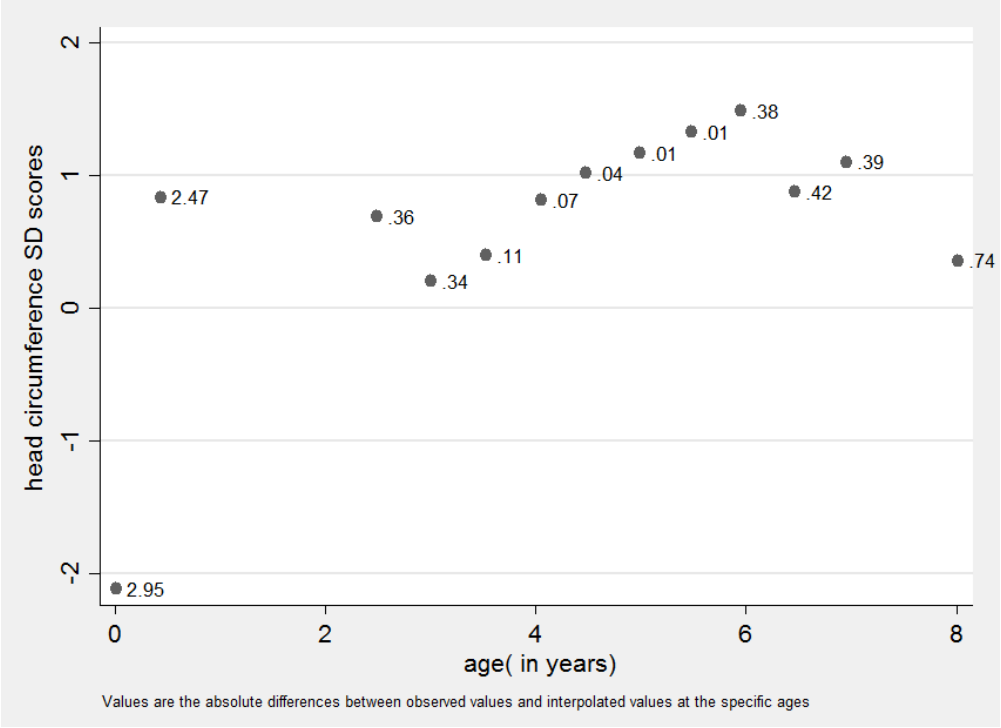
As seen from the above figures, the head circumference value with a surprise value of 5.28 looks very suspicious in both the plots. This observation was removed from the dataset and the same procedure was followed for other individuals.

Figure 3.6: Plot of head circumference measurements and age (ID=60775)



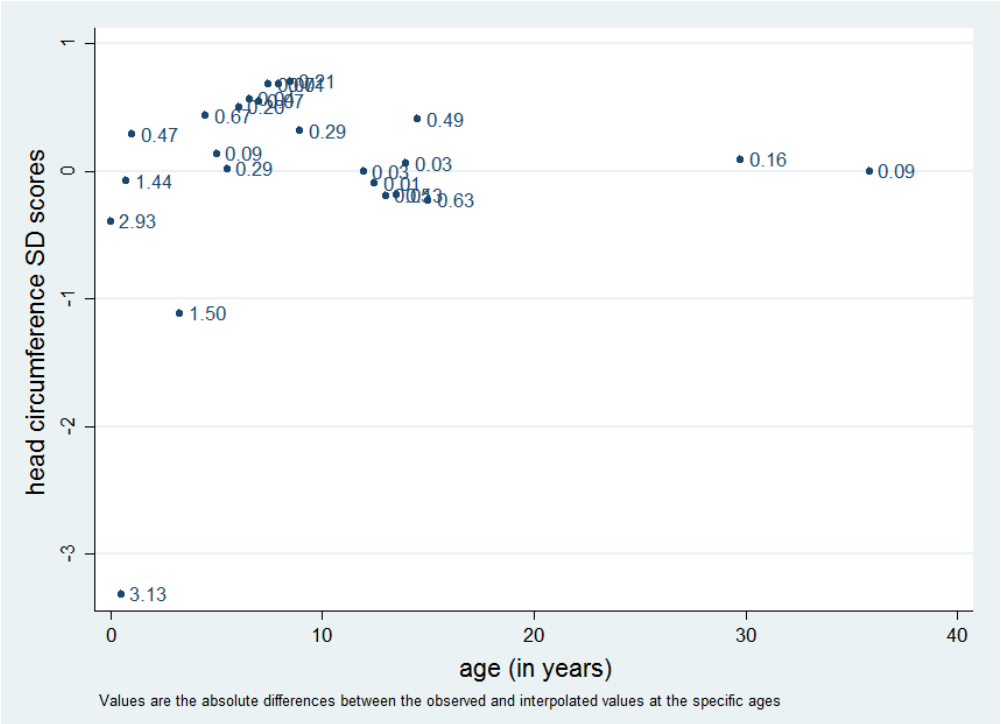
In the above example (Figure 3.6), the second reading for the given individual (ID=60775) has a surprisingly high value (surprise value is 2.47). But the reading has been made at the infancy stage for the particular individual and the head grows at a very rapid rate during the period of infancy and becomes stable after adolescence. Thus biologically, it would be wrong to remove the particular observation from the cohort even though mathematically, it can be considered suspicious. The plot of age-adjusted standardized head circumference scores against age for this individual was obtained as (Figure 3.7):

Figure 3.7: Plot of head circumference SD scores and age (ID=60775)



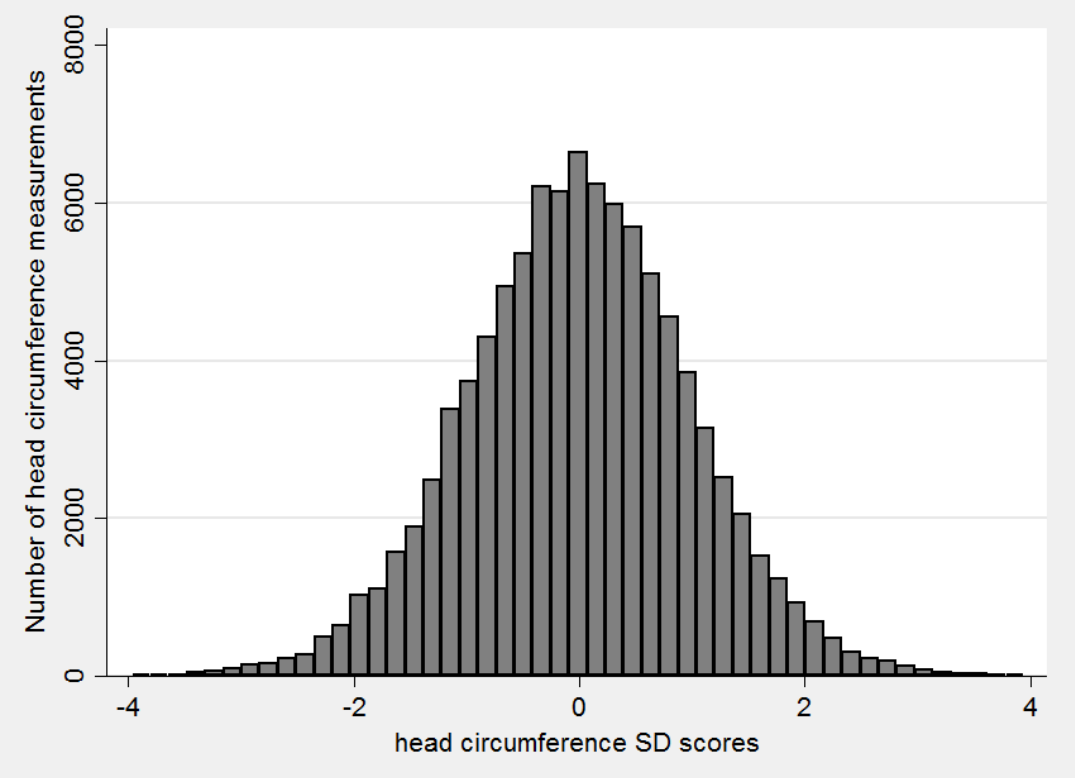
Taking another example, the birth measurement for ID= 80421 has a very high surprise value (2.93), but the individual is born preterm (gestational age is 36 weeks) (Figure 3.8). In the case of a preterm birth, the individual may have a smaller than expected birth head size, as the head has not developed completely.

Figure 3.8: Plot of head circumference SD scores and age (ID=80421)



On further inspection many other individuals below surprise value 2.93 had mathematically implausible values below this surprise value during infancy (graphs not shown) but which were biologically plausible, and there were no observations at or after age 2 years having a surprise value above 0.5, thus no other head circumference values were excluded from the dataset. At the end of the procedure, 605 values of head circumference were excluded out of 96,731 values (0.6%), leaving 96,126 values of head circumference in the dataset. The head circumference SD scores were recalculated separately for both males and females after removing implausible head circumference values. The distribution of standardized head scores after data cleaning was obtained as in Figure 3.9. Other anthropometric measurements such as height and BMI were also cleaned using this method.

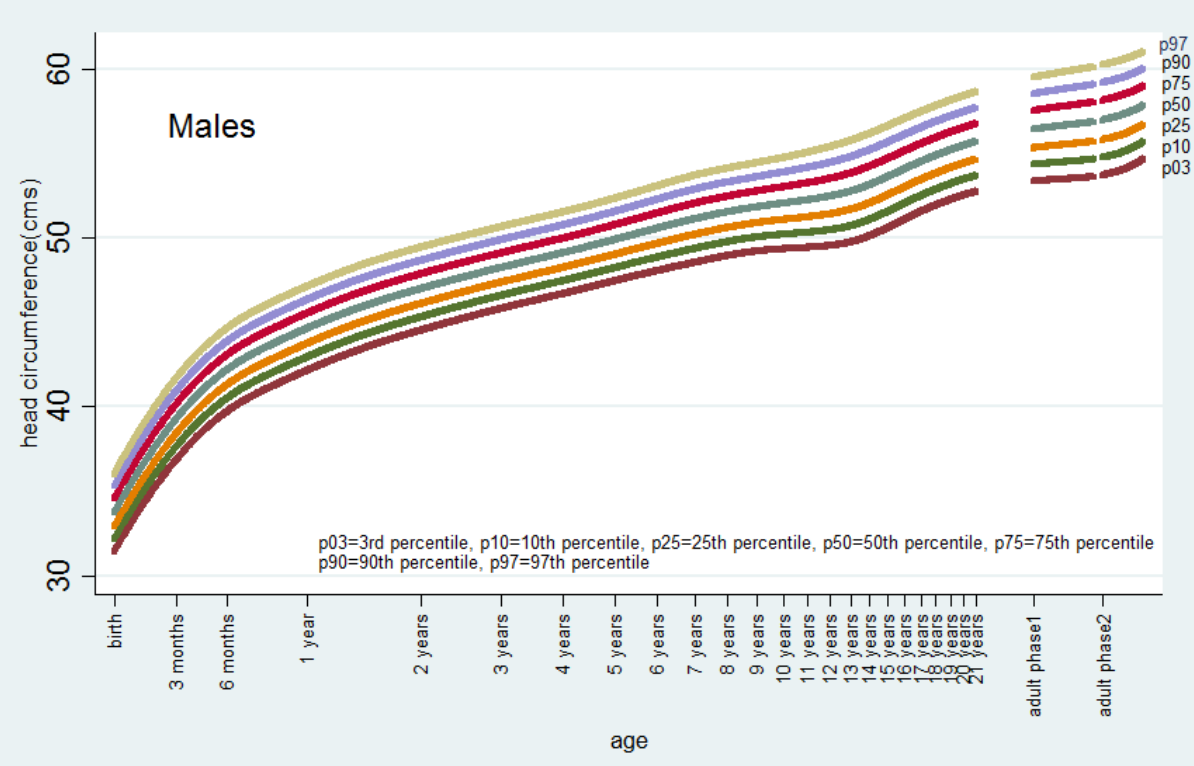
Figure 3.9: Distribution of head circumference standard deviation scores after data cleaning



I plotted sex-specific centile curves for the different ages showing the trajectory for head circumference growth at different stages of the follow-up. Log transformation of age was used to show clearly the period of rapid head growth during infancy. The sex-

specific centile curves are depicted in Figure 3.10 and the corresponding centile values displayed in Table 3.1. The adults in the study were followed up twice, at age 26-33 years and 34-39 years.

Figure 3.10: Centile curves for head circumference for males and females



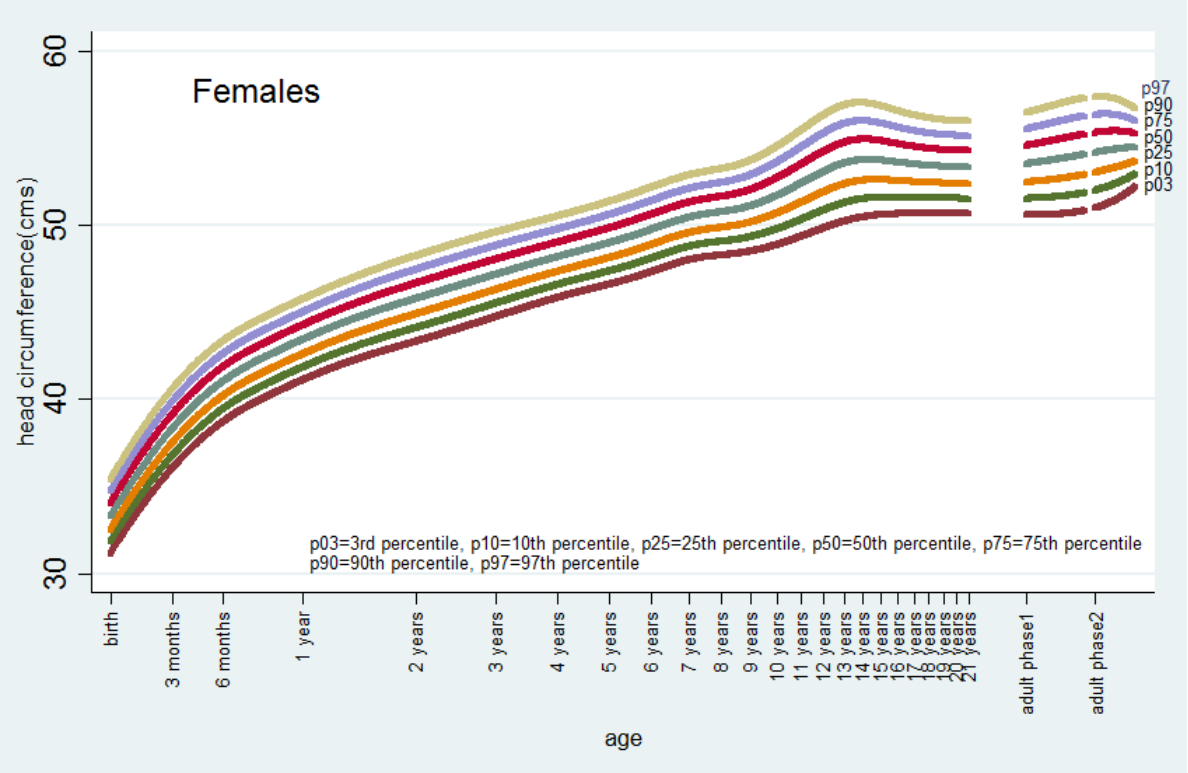


Table 3.1: Head circumference centile values according to age in males and females

Males

Age (years)	Centiles (cm)						
	3%	10%	25%	50%	75%	90%	97%
0	31.37	32.10	32.85	33.68	34.50	35.25	35.98
0.25	36.86	37.63	38.41	39.28	40.14	40.92	41.69
0.5	39.69	40.48	41.27	42.16	43.04	43.84	44.63
1	42.12	42.91	43.71	44.61	45.50	46.30	47.10
2	44.48	45.26	46.05	46.93	47.81	48.60	49.39
3	45.76	46.53	47.31	48.18	49.05	49.83	50.60
4	46.64	47.41	48.19	49.06	49.92	50.70	51.47
5	47.39	48.16	48.95	49.83	50.70	51.49	52.27
6	48.00	48.80	49.61	50.51	51.41	52.22	53.02
7	48.48	49.30	50.14	51.06	51.98	52.82	53.64
8	48.89	49.71	50.55	51.47	52.40	53.24	54.06
9	49.15	49.99	50.83	51.77	52.79	53.55	54.39
10	49.29	50.15	51.02	51.99	52.96	53.83	54.69
11	49.37	50.27	51.18	52.19	53.20	54.11	55.00
12	49.48	50.41	51.36	52.41	53.46	54.40	55.33

13	49.70	50.66	51.62	52.70	53.78	54.75	55.71
14	50.08	51.04	52.02	53.10	54.19	55.17	56.13
15	50.56	51.52	52.49	53.57	54.65	55.62	56.58
16	51.05	52.00	52.97	54.03	55.10	56.06	57.01
17	51.50	52.44	53.40	54.46	55.52	56.47	57.41
18	51.88	52.82	53.77	54.83	55.88	56.83	57.77
19	50.66	51.52	52.38	53.34	54.30	55.16	56.01
20	50.66	51.50	52.36	53.31	54.26	55.11	55.96
21	50.64	51.49	52.34	53.30	54.25	55.11	55.95
26-33	50.55	51.49	52.44	53.49	54.54	55.49	56.43
34-39	50.96	51.97	53.00	54.14	55.28	56.31	57.32

Females

Age (years)	**Centiles (cm)**						
	3%	**10%**	**25%**	**50%**	**75%**	**90%**	**97%**
0	31.13	31.81	32.49	33.25	34.00	34.69	35.36
0.25	36.05	36.78	37.51	38.33	39.14	39.88	40.60
0.5	38.70	39.44	40.18	41.02	41.85	42.59	43.33
1	41.09	41.83	42.58	43.42	44.25	45.00	45.74
2	43.29	44.07	44.86	45.75	46.63	47.42	48.20
3	44.71	45.48	46.26	47.13	48.00	48.79	49.56
4	45.81	46.55	47.31	48.15	48.99	49.74	50.49
5	46.56	47.32	48.08	48.94	49.79	50.55	51.31
6	47.29	48.06	48.85	49.72	50.59	51.37	52.14
7	47.96	48.74	49.52	50.40	51.27	52.06	52.83
8	48.24	49.03	49.83	50.72	51.61	52.41	53.20
9	48.46	49.29	50.14	51.07	52.01	52.85	53.68
10	48.85	49.75	50.66	51.67	52.68	53.59	54.49
11	49.32	50.29	51.28	52.37	53.47	54.46	55.43
12	49.80	50.83	51.88	53.04	54.21	55.25	56.29
13	50.19	51.25	52.33	53.52	54.72	55.80	56.86

14	50.43	51.48	52.54	53.71	54.89	55.95	57.00
15	50.56	51.56	52.57	53.69	54.82	55.83	56.83
16	50.61	51.56	52.52	53.58	54.64	55.60	56.54
17	50.65	51.55	52.46	53.47	54.49	55.40	56.30
18	50.66	51.53	52.41	53.39	54.37	55.25	56.12
19	50.66	51.52	52.38	53.34	54.30	55.16	56.01
20	50.66	51.50	52.36	53.31	54.26	55.11	55.96
21	50.64	51.49	52.34	53.30	54.25	55.11	55.95
26-33	50.55	51.49	52.44	53.49	54.54	55.49	56.43
34-39	50.96	51.97	53.00	54.14	55.28	56.31	57.32

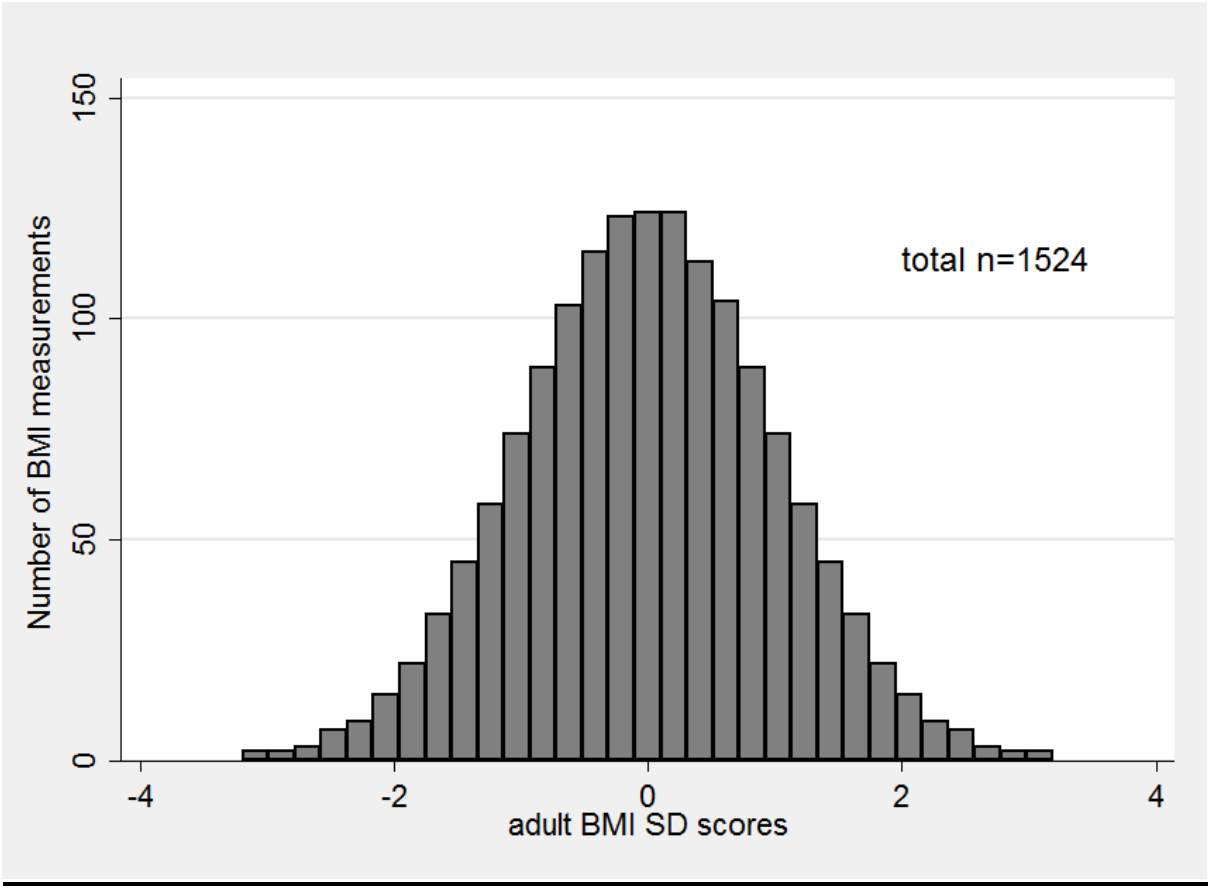
3.5 Creating standardized scores for adult anthropometry

Standardized scores for adult head circumference and height were calculated by subtracting the adult cohort mean and dividing by the cohort SD in adulthood, since both adult head circumference and height were symmetrically distributed. Standardized scores at the defined ages for adult BMI, which were skewed were calculated by using inverse normal transformations as proposed by Fisher and Yates (79), also known as the 'Fisher and Yates normal scores approach'. These scores were calculated separately for males and females. This process involved creating a modified rank variable and then computing a new transformed value for the i^{th} subject using the formula:-

$$Y_i' = \phi^{-1}\left(\frac{r_i - c}{N - 2c + 1}\right)$$

Where r_i is the ordinary rank for the i^{th} case among the N observations and ϕ^{-1} denotes the standard normal quantile function. Blom recommended the use of c=3/8 (80). The use of inverse normal transformation ensures that the resulting distribution of the variable is normally distributed. The distribution of the measurements after conversion is shown in Figure 3.11.

Figure 3.11: Distribution of adult body mass index standard deviation scores



After calculating the SD scores for head circumference and cleaning the data, using the technique of linear interpolation for each subject I interpolated values linearly between successive z-scores to estimate head circumference z-scores at 6 months and at birthdays from 1 to 21 years of age. The interpolated values were used if a measurement had been made within 6 months (up to 1 year), 1 year (age of 2 years), 1.5 years (age of 3 years), and 2 years (all ages after 3 years). Back transformation provided estimates of the head circumference measurements in centimetres at all these ages. A similar technique was also used for estimating height and BMI measurements at the specific ages. Table 3.2 provides the sex-specific distribution of body measurements at the interpolated ages.

Table 3.2: Sex-specific distribution of body size measurements across birthdays from birth to 21 years and adult life

	Males						Females					
	Head (cms)		Height (cms)		BMI (kg/m^2)		Head (cms)		Height (cms)		BMI (kg/m^2)	
Age (years)	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N
0	33.6 (1.3)	3666	48.6 (2.1)	3572	12.1 (1.2)	3568	33.2 (1.2)	3368	48.1 (1.9)	3368	11.9 (1.2)	3366
0.25	39.3 (1.3)	3763	59.4 (2.6)	3635	15.3 (1.6)	3611	38.3 (1.2)	3448	58.0 (2.5)	3377	14.7 (1.6)	3357
0.50	42.1 (1.3)	3806	65.3 (2.4)	3705	16.4 (1.6)	3666	40.9 (1.3)	3489	63.6 (2.4)	3420	15.7 (1.6)	3384
1	44.6 (1.3)	2762	72.0 (2.7)	2724	16.4 (1.5)	2656	43.4 (1.3)	2542	70.1 (2.9)	2523	15.9 (1.5)	2457
2	46.9 (1.3)	2779	81.1 (3.6)	2733	15.8 (1.2)	2575	45.7 (1.3)	2521	79.6 (3.6)	2469	15.4 (1.2)	2315
3	48.1 (1.3)	2663	88.7 (4.1)	2598	15.5 (1.2)	2477	47.1 (1.3)	2428	86.9 (4.3)	2366	15.2 (1.2)	2249
4	48.9 (1.3)	2501	95.5 (4.2)	2438	15.3 (1.1)	2350	48.1 (1.3)	2318	93.9 (4.4)	2254	15.1 (1.1)	2173
5	49.8 (1.3)	2212	101.9 (4.5)	2162	14.9 (1.1)	2111	48.9 (1.3)	2082	100.6 (4.4)	2036	14.7 (1.1)	1997
6	50.5 (1.3)	2130	108.1 (4.9)	2104	14.5 (1.1)	2082	49.7 (1.3)	2002	106.7 (4.7)	1976	14.3 (1.1)	1957
7	51.1 (1.4)	2051	113.9 (5.1)	2027	14.3 (1.2)	2012	50.3 (1.3)	1945	112.6 (5.1)	1928	14.1 (1.1)	1916
8	51.5 (1.4)	1989	119.7 (5.4)	1975	14.5 (1.3)	1968	50.8 (1.3)	1877	118.0 (5.4)	1862	14.2 (1.2)	1856
9	51.8 (1.4)	1855	124.9 (5.6)	1847	14.7 (1.4)	1839	51.2 (1.4)	1634	123.4 (5.7)	1727	14.2 (1.2)	1724
10	51.9 (1.4)	1750	130.1 (6.9)	1742	14.9 (1.6)	1732	51.7 (1.4)	1634	128.2 (5.3)	1627	14.5 (1.3)	1625
11	52.2 (1.5)	1674	135.9 (5.7)	1669	15.3 (1.7)	1661	52.2 (1.5)	1561	134.1 (7.4)	1564	15.2 (1.8)	1562
12	52.4 (1.5)	1666	140.3 (6.7)	1663	15.6 (2.0)	1657	52.8 (1.6)	1568	141.3 (7.6)	1569	15.9 (2.1)	1565
13	52.7 (1.6)	1658	145.6 (8.2)	1651	16.0 (2.2)	1646	53.4 (1.7)	1571	147.9 (6.8)	1569	17.1 (2.5)	1561
14	53.2 (1.6)	1637	153.0 (9.2)	1633	16.7 (2.4)	1629	53.8 (1.7)	1560	151.7 (5.7)	1551	18.2 (2.7)	1542

15	53.6	1558	160.3	1562	17.6	1555	53.8	1503	153.3	1500	19.1	1482
	(1.7)		(8.8)		(2.6)		(1.7)		(5.4)		(2.8)	
16	54.1	1342	165.2	1348	18.5	1320	53.3	1311	153.9	1318	19.7	1287
	(1.7)		(7.7)		(2.8)		(1.4)		(5.6)		(3.2)	
17	54.3	873	167.9	871	19.3	819	53.4	819	154.2	823	19.9	795
	(1.6)		(6.3)		(2.9)		(1.5)		(5.7)		(3.3)	
18	54.7	548	169.1	546	19.8	533	53.3	529	154.3	527	19.8	520
	(1.5)		(6.0)		(3.1)		(1.4)		(5.9)		(3.2)	
19	55.1	503	169.3	504	19.9	498	53.3	492	155.3	490	20.3	489
	(1.7)		(6.7)		(2.9)		(1.4)		(5.8)		(3.6)	
20	55.4	461	169.4	462	20.0	461	53.4	453	156.4	453	20.6	453
	(1.6)		(7.1)		(3.0)		(1.7)		(5.3)		(3.6)	
21	55.6	392	168.7	394	20.4	393	53.3	408	156.7	409	19.9	409
	(1.6)		(5.9)		(3.1)		(1.4)		(5.6)		(3.4)	
Adult	56.6	884	169.7	886	24.9	886	53.8	640	154.9	638	24.6	638
phase	(1.8)		(6.4)		(4.3)		(1.7)		(5.7)		(5.1)	

3.6 Conditional Models

I used conditional models to investigate how head growth during infancy and childhood are related to attained education in years and adult cardiometabolic outcomes. Conditional models help the researcher to assess growth during particular segments of early life in relation to an adult health outcome (28,41). In this approach, the health outcome in later life is regressed on the conditional body sizes at all preceding observation points plus the body size at the time of birth. This overcomes the problem of correlation among the body sizes at different points of time, called multicolinearity, in multiple linear regression. Conditional head size can be defined as the difference between the observed and predicted head size of the individuals at any defined age. The predicted head size at any age measured on occasion p can be obtained by regressing the observed head size on all preceding observed head sizes (1….. (p-1)) since birth and may also include other potential covariates.

E.g.: $Head\ circumference_{ip} = b_0 + \sum_{j=1}^{p-1} b_j \left(head\ circumference\ _{ij}\right) +\ e_{ip}$, where j = 1 means the body size at birth. In the above equation, the predicted head circumference on occasion p is $b_0 + \sum_{j=1}^{p-1} b_j \left(head\ circumference\ _{ij}\right)$. Thus the conditional head circumference on occasion p can be defined as the residual error (i.e. the difference between observed and predicted value). Due to the properties of linear

models, the conditional head circumference on occasion p is uncorrelated with the observed head sizes at all preceding points and also with all preceding conditional head sizes. The regression coefficients from a conditional body size analysis can be interpreted as the expected change in the adult health outcome, given a unit change in the predictor, where the predictor is body size at birth or conditional body size.

After the creation of the conditional variables, the health outcome in later life is regressed on all conditional body sizes plus the birth size.

$$Adult\ Outcome_i = \beta_0 + \sum_{j=1}^{W} \beta_j e_{ij} + t_i$$

Where e_{ij} is the jth conditional for subject i, and W is the number of time windows. The $\beta's$ are regression coefficients and t_i is the residual error.

The conditional variables represent the deviation in the child's size from the expected size at any particular age on the basis of the previous body size measurements. They represent faster or slower growth than would be expected at the particular age. E.g.: a child with a positive value for childhood (defined between 2 years to 11 years) conditional head circumference has a larger than expected head at age 11 years in view of his previous head circumference measurements up to age 2 years, and thus had a faster rate of head growth than expected between 2 and 11 years.

Because head circumference measurements are inevitably made with error, there will also be random error in the estimates of head growth. Such errors will lead to attenuation of effect size estimates. Sayers et al. gives an indication of how to model the effects of such errors (81). In simple situations, this can be done algebraically, and in more complicated situations it is possible to use simulation to address the issue.

Taking an example from the NDBC dataset, I studied the head growth trajectories of two individuals having similar conditional values from 6 months to 2 years but different head sizes at birth and 6months (Figures 3.12 and 3.13).

Figure 3.12: Head growth from birth to adulthood (ID=70060 and ID=40391)

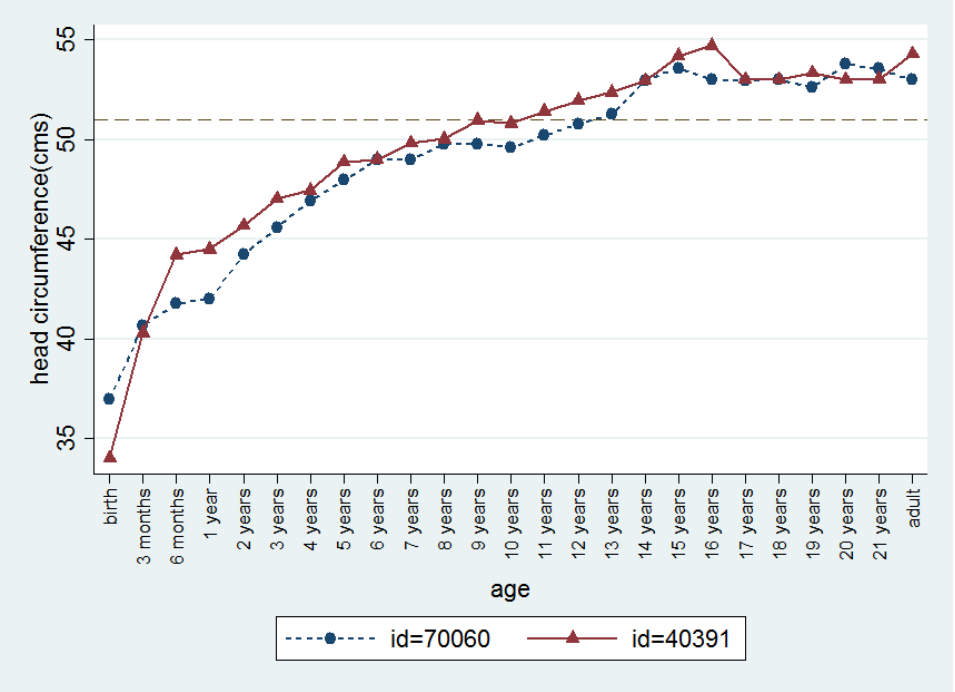
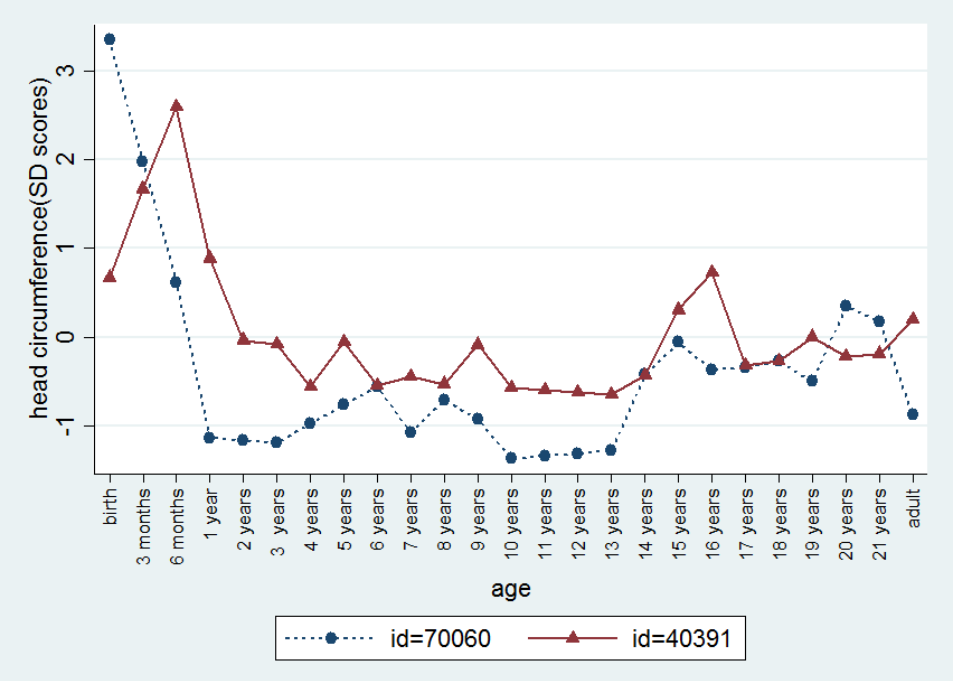


Figure 3.13: Head growth from birth to adulthood (ID=70060 and ID=40391) (standard deviation scores)



As can be seen from the above diagram, ID 70060 has a much bigger head size at birth as compared to ID 40391. The head size for ID 40391 grows at a faster rate than head size for ID 70060 during the period of early infancy and during the period of adolescence for ID 40391 as compared to ID 70060 (Table 3.3). The head growth for both IDs is similar in the other two growth periods. The adult head sizes for these two IDs are also similar.

Table 3.3: Head growth at clinically significant time periods for IDs 70060 and 40391

Head Conditionals (SD scores)	**ID=70060**	**ID=40391**
Birth	3.34	0.67
Birth-6m	-1.58	2.82
6m-2y	-2.60	-2.46
2y-11y	-1.17	-1.62
11y-adult	-0.20	0.63

3.7 Identifying crucial time periods for head growth

Based on the principle of parsimony, I would like to describe the head growth trajectory as accurately as possible but with as few parameters as possible. So, the time periods between which head growth is assessed need consideration. From a statistician/mathematician's perspective, the conditionals could be calculated at those ages where equal correlations are observed between the head circumferences. A clinician/paediatrician would recognize the particular period/periods which are crucial for a child's health and development and calculate the conditionals for those periods (53). In the first six months, the infant is exclusively on breast milk, which contains vital nutrients for his growth and development. At the end of infancy, the growth regulation changes from being controlled mainly by insulin-like growth factors to being regulated by growth hormone. The next phase of life in which growth patterns differ greatly between individuals is puberty, during which growth is influenced by sex hormones, and various physical and psychological changes start to occur.

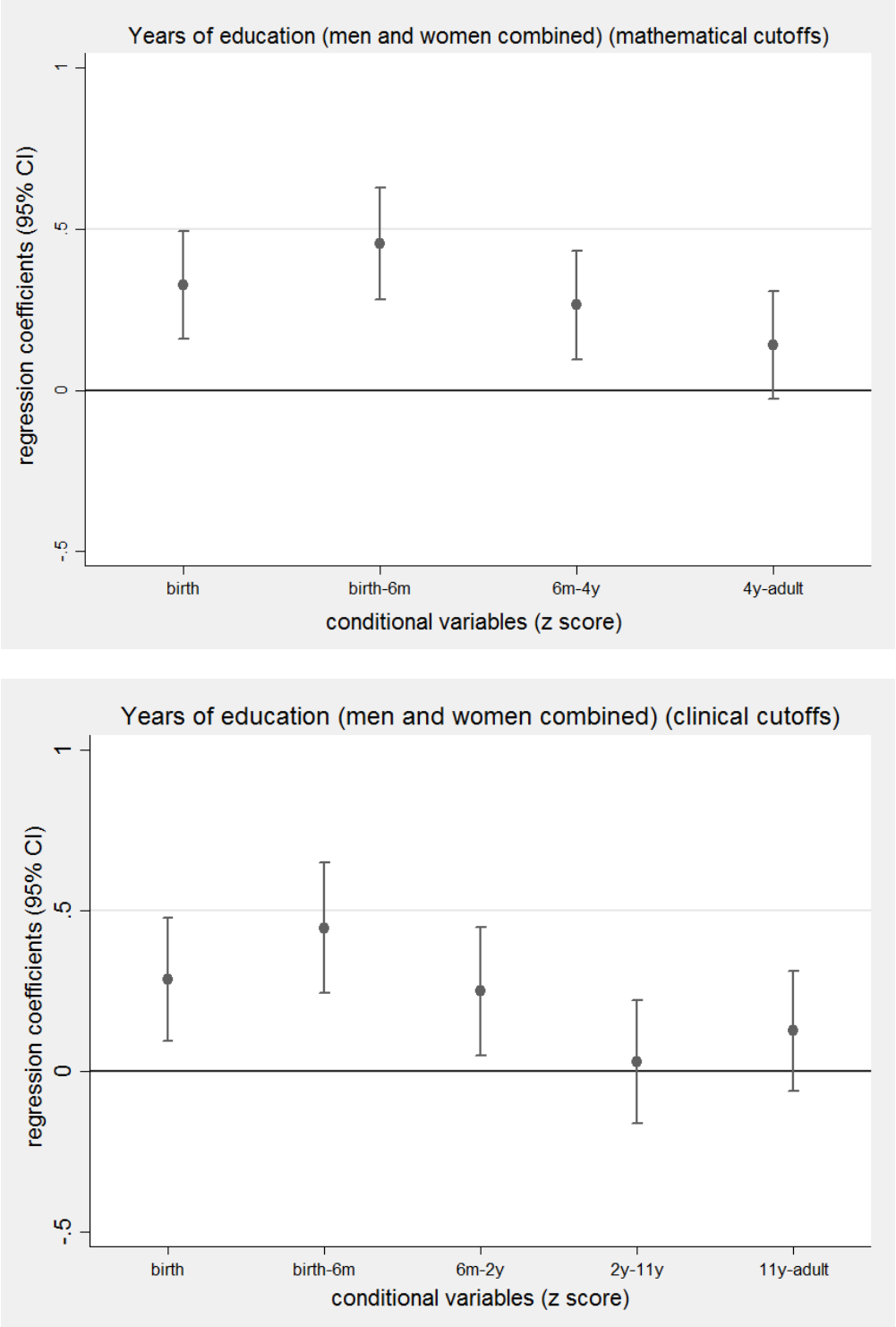
I evaluated the association between the conditional SD scores and years of adult education defined in chapter-2 (Table 2.2, page 37). This association was evaluated based on the different sets of divisions of the data (mathematical, clinical and more frequent divisions based on alternate measurements) separately for both the sexes as well as combined. I initially calculated sex-specific correlations. Since these were similar, I calculated partial correlations, controlling for sex. I present the partial correlations at the defined ages (Table 3.4). Graphs of the association of the conditionals with years of education based on these periods are also illustrated (Figure 3.14).

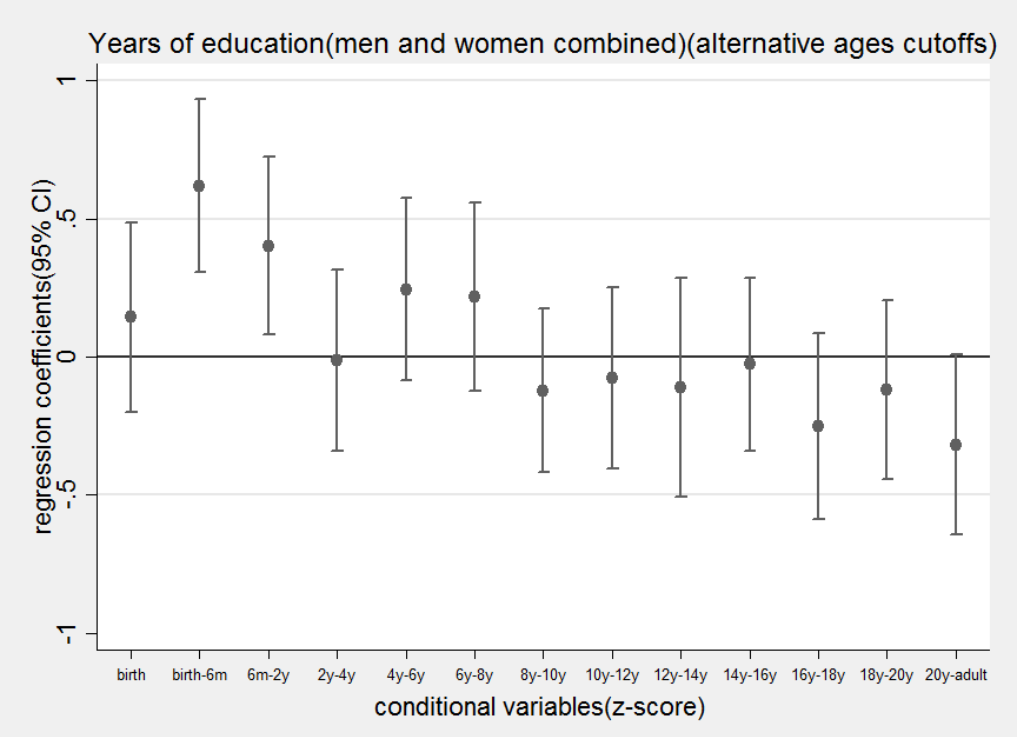
Table 3.4: Correlation between the head measurements at different measurement ages

Age (in years)	**0**	**0.25**	**0.5**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**adult**
0	1.0																								
0.25	0.70	1.0																							
0.5	0.61	0.84	1.0																						
1	0.46	0.66	0.80	1.0																					
2	0.44	0.60	0.70	0.84	1.0																				
3	0.43	0.57	0.66	0.76	0.90	1.0																			
4	0.41	0.55	0.64	0.74	0.83	0.91	1.0																		
5	0.41	0.56	0.65	0.74	0.81	0.86	0.91	1.0																	
6	0.42	0.56	0.64	0.75	0.80	0.83	0.87	0.92	1.0																
7	0.40	0.57	0.65	0.74	0.78	0.80	0.83	0.87	0.92	1.0															
8	0.40	0.57	0.64	0.72	0.76	0.78	0.81	0.86	0.90	0.94	1.0														
9	0.40	0.56	0.63	0.71	0.74	0.76	0.79	0.84	0.87	0.91	0.96	1.0													
10	0.40	0.56	0.64	0.72	0.74	0.76	0.78	0.83	0.87	0.90	0.94	0.99	1.0												
11	0.39	0.56	0.64	0.71	0.73	0.75	0.77	0.82	0.85	0.87	0.90	0.94	0.99	1.0											
12	0.39	0.54	0.61	0.70	0.71	0.73	0.76	0.80	0.82	0.85	0.87	0.89	0.94	0.96	1.0										
13	0.37	0.50	0.57	0.66	0.68	0.70	0.73	0.77	0.80	0.82	0.83	0.85	0.89	0.91	0.93	1.0									
14	0.36	0.49	0.55	0.64	0.65	0.67	0.70	0.74	0.77	0.79	0.81	0.82	0.85	0.86	0.88	0.89	1.0								
15	0.35	0.48	0.54	0.61	0.62	0.62	0.65	0.69	0.72	0.76	0.77	0.79	0.81	0.82	0.83	0.83	0.89	1.0							
16	0.33	0.49	0.55	0.62	0.64	0.63	0.64	0.69	0.72	0.75	0.76	0.78	0.80	0.81	0.82	0.80	0.84	0.92	1.0						
17	0.32	0.47	0.54	0.62	0.67	0.67	0.67	0.73	0.76	0.78	0.78	0.78	0.80	0.82	0.82	0.80	0.81	0.85	0.91	1.0					
18	0.30	0.43	0.49	0.59	0.66	0.69	0.68	0.74	0.78	0.81	0.79	0.80	0.81	0.82	0.82	0.81	0.81	0.80	0.85	0.92	1.0				
19	0.31	0.42	0.49	0.59	0.66	0.69	0.68	0.74	0.78	0.81	0.79	0.80	0.80	0.81	0.82	0.82	0.80	0.81	0.80	0.85	0.92	1.0			
20	0.30	0.44	0.49	0.59	0.65	0.68	0.68	0.73	0.78	0.80	0.80	0.79	0.80	0.81	0.80	0.78	0.78	0.79	0.81	0.85	0.89	0.92	1.0		
21	0.31	0.43	0.51	0.60	0.63	0.67	0.68	0.72	0.77	0.79	0.79	0.79	0.79	0.81	0.79	0.77	0.77	0.78	0.79	0.83	0.87	0.89	0.97	1.0	
Adult	0.35	0.43	0.50	0.60	0.61	0.62	0.65	0.66	0.70	0.72	0.73	0.73	0.75	0.75	0.74	0.73	0.74	0.75	0.74	0.77	0.78	0.76	0.73	0.74	1.0

By observation, the correlations at birth, birth-6months, 6months-4years and 4years-adulthood look similar. Graphs for clinically defined time periods and at alternative ages are also described below.

Figure 3.14: Association of head size at birth and head growth with attained years of education in adulthood



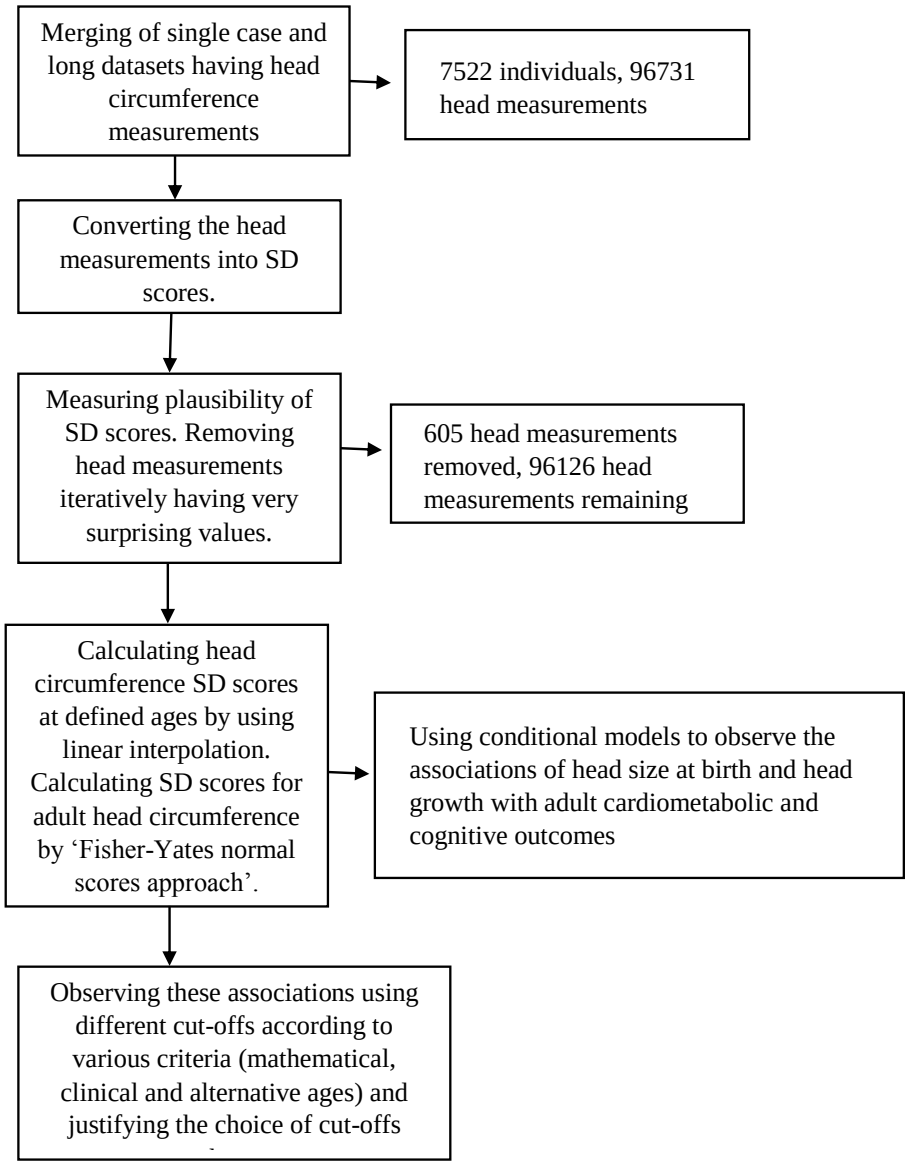


When mathematical periods were used, head growth during infancy and during 6 months to 4 years were positively associated with years of attained education. 1,442 head measurements were used in this analyses. In the case when clinical cutoffs were used, head size at birth and head growth in early and late infancy were positively associated with years of attained education. Head growth at later ages was unrelated to educational attainment. 1,052 head measurements were used in this analysis. When alternate ages were used as cutoffs, head growth up to late infancy was positively associated with years of education. 349 head measurements were used in this analyses. We can see from the above graphs that the cutoffs chosen on the basis of different criteria gave similar results. The cutoffs based on mathematical periods are decided only on the basis of the highest correlation between the ages, while the clinical cutoffs are decided on clinical significance. The cutoffs based on alternate ages require the estimation of a much larger number of parameters when compared to mathematically and clinically defined phases. The mathematical cutoffs combines the association of head growth in childhood and adult life, while the clinical cutoff separates the associations of periods of head growth in childhood and adult life and attained years of education in adulthood. Since I am dealing with medical data, it would be sensible to look at head growth in clinically important phases rather than based on any other

criteria. Thus, I will only look at growth in clinically important phases in all subsequent analyses.

Below is a diagrammatic representation summarizing the data management and methods used in data processing of the cohort data (Figure 3.15):

Figure 3.15: Illustration of the data management and methods used in data processing



3.8 Other outcome: Systolic blood pressure

In addition to years of attained education, I will examine the associations of head growth with SBP, a cardiometabolic outcome in adulthood.

3.9 Adjustment for socioeconomic status

SES is a common confounder in health research. Adverse childhood socioeconomic conditions impair growth (82) and cognitive development (39). Therefore it will be important to control for SES in my analyses. SES is multidimensional in nature and is a combination of several factors like education, income, dwelling characteristics and household assets. These components are correlated with one another. Correlation among the factors leads to the statistical issue of multicollinearity and also leads to statistically inefficient estimates. To avoid this problem, various factors representing SES can be reduced to a single composite variable and principal component analysis (PCA) is a data reduction technique commonly used for this purpose.

PCA is a multivariate statistical technique used to reduce the number of variables in data to a smaller number of variables, often called 'constructs' or 'dimensions' (83). PCA evaluates the most meaningful basis to express a large set of variables, which are often correlated strongly with one another. In a mathematical sense, PCA creates uncorrelated components from the correlated variables. PCA can be explained mathematically by the following equations.

$$PC_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1n}X_n$$
$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$
$$PC_m = a_{m1}X_1 + a_{m2}X_2 + \ldots + a_{mn}X_n$$

Where a_{mn} represents the weight for the m^{th} principal component and the n^{th} variable. PC_1 to PC_m are the m principal components, which are uncorrelated to each other. The components are ordered so that the first component (PC_1) explains the largest possible amount of variation in the original data. The variance of each principal component is given by its eigenvalue. Several studies have used this technique previously (84-86).

In the NDBC, a number of socioeconomic variables were collected from the F0 generation, which included both continuous as well as categorical variables (Table 3.5). For the purpose of PCA analyses, all the categorical socioeconomic variables were recoded into categories having sufficient observations in each category. For example, variables such as parental occupation which was initially a variable having six categories, was recoded into a new variable having 5 categories, the first two categories being combined together because of a small number of individuals in the first category (unemployed). These variables were linearly associated with years of education and

were included as continuous variables in the analysis. Variables such as crowding index and child dependency ratio, which were continuous were skewed and were made symmetric by log transformation. The dataset was then split into two sub-samples, one with the individuals having valid observations for the outcome years of education and another with the individuals having valid observations for both the outcome and conditional growth measurements. Table 3.5 shows the proportion of non-missing values of the socioeconomic variables in the full sample as well as the two subsamples. All the variables below have been considered for principal components analysis. Information like maternal education was collected at a time when the women in the cohort were being recruited for the study and variables such as paternal education were collected by recall when the cohort members were adults. Thus, there is a marked difference in the sample sizes for the two variables. Before conducting PCA, all the above socioeconomic variables were standardized to unify units and the missing values in each of these variables were imputed by their mean values. A binary indicator for missingness was generated for all the samples, coded 1 if any of the socioeconomic variables for an individual were missing and 0 otherwise.

Table 3.5: Number (%) of cohort members for whom data on various aspects of socioeconomic status are available

SES Variables	Full sample (N=7446)	Sub-sample with outcome years of education (N=1526)	Sub-sample with outcome years of education and all conditional growth variables (N=958)
Maternal years of education	5707 (76.6)	1366 (89.5)	910 (95.0)
Illiterate	1939 (34.0)	386 (28.3)	244 (26.8)
Primary	939 (16.5)	264 (19.3)	177 (19.5)
Middle	898 (15.7)	236 (17.3)	163 (17.9)
Matric	1186 (20.8)	293 (21.4)	190 (20.9)
College	745 (13.0)	187 (13.7)	136 (14.9)
Wealth(Assets)*	5401 (72.5)	1055 (69.1)	669 (69.8)
Not owned Thatched hut	58 (0.8)	5 (0.5)	1 (0.1)
Owned Thatched hut	32 (0.4)	4 (0.4)	1 (0.1)
Not owned Masonry build	1947 (26.1)	135 (12.8)	73 (10.9)
Owned Masonry build	1634 (21.9)	539 (51.1)	348 (52.0)
Not owned Flat	998 (13.4)	175 (16.6)	120 (17.9)
Owned Flat	486 (6.5)	165 (15.6)	107 (16.0)
Not owned Bungalow	148 (2.0)	8 (0.8)	5 (0.7)
Owned Bungalow	91 (1.2)	24 (2.3)	14 (2.1)

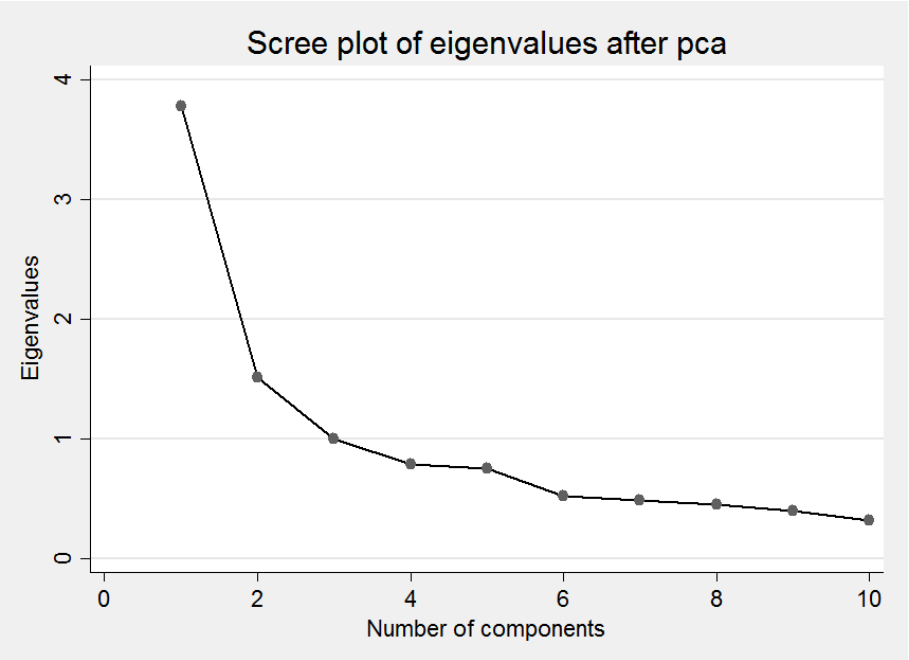
Individual Income (raw) Median (LQ, UQ)	5401 (72.5) 792 (480,1200)	1056 (69.2) 690 (459, 1200)	670 (69.9) 600 (456, 1200)
Parent's occupation	1532 (20.6)	1504 (98.6)	946 (98.7)
Unemployed	3 (0.2)	3 (0.2)	3 (0.3)
Unskilled manual labour	28 (1.8)	28 (1.9)	18 (1.9)
Semi-skilled manual labour	163 (10.6)	161 (10.7)	91 (9.6)
Skilled manual labour, small	333 (21.7)	328 (21.8)	211 (22.3)
Trainer, Clerical	761 (49.7)	741 (49.3)	476 (50.3)
Professional, big business	244 (15.9)	243 (16.2)	147 (15.5)
Utilities¥	5360 (72.0)	1057 (69.3)	671 (70.0)
No/low health services use	1559 (20.9)	336 (31.8)	208 (31.0)
Intermediate	1624 (30.3)	366 (34.6)	239 (35.6)
Highest use	2177 (40.6)	355 (33.6)	224 (33.4)
Sanitation	5403 (72.6)	1057 (69.3)	671 (70.0)
no toilet	1221 (22.6)	179 (16.9)	96 (14.3)
some excreta removal	2152 (39.8)	495 (46.8)	333 (49.6)
flush toilet	2030 (37.6)	383 (36.2)	242 (36.1)
Water supply	5404 (72.6)	1057 (69.3)	671 (70.0)
no piped water	953 (17.6)	124 (11.7)	68 (10.1)
shared piped water	2586 (47.9)	540 (51.1)	345 (51.4)
sole use piped water	1865 (34.5)	393 (37.2)	258 (38.5)
Crowding index (people/rooms)	5395 (72.5)	1055 (69.1)	669 (69.8)
Median (LQ, UQ)	4 (3,6)	4 (3,6)	4 (3,6)
Child dependency ratio (children<18 /adults)	5404 (72.6)	1057 (69.3)	671 (70.0)
Median (LQ, UQ)	1.0 (0.6, 1.7)	1.0 (0.7, 2.0)	1.2 (0.6, 2.0)
Paternal years of schooling	1455 (19.5)	1430 (93.7)	900 (93.9)
Illiterate	124 (8.5)	123 (8.6)	74 (8.2)
primary school	135 (9.3)	132 (9.2)	83 (9.2)
middle school	214 (14.7)	211 (14.8)	134 (14.9)
high school certificate	413 (28.4)	404 (28.3)	259 (28.8)
high school+	158 (10.9)	155 (10.8)	92 (10.2)
other graduate	285 (19.6)	280 (19.6)	186 (20.7)
professional degree	126 (8.7)	125 (8.7)	72 (8.0)

*assets (wealth); excluding water, sanitation or crowding variables

¥ Combination score of any health-promotion or preventative health service utilization during antenatal and postnatal period (e.g. antenatal care, immunization)

PCA was subsequently conducted in each of the three samples and only the first principal components score was used for analyses in all the cases (Figure 3.16), since the later components explained little additional variability. The number of components to be retained is usually decided by the point where the line flattens out.

Figure 3.16: Scree plot of the principal components



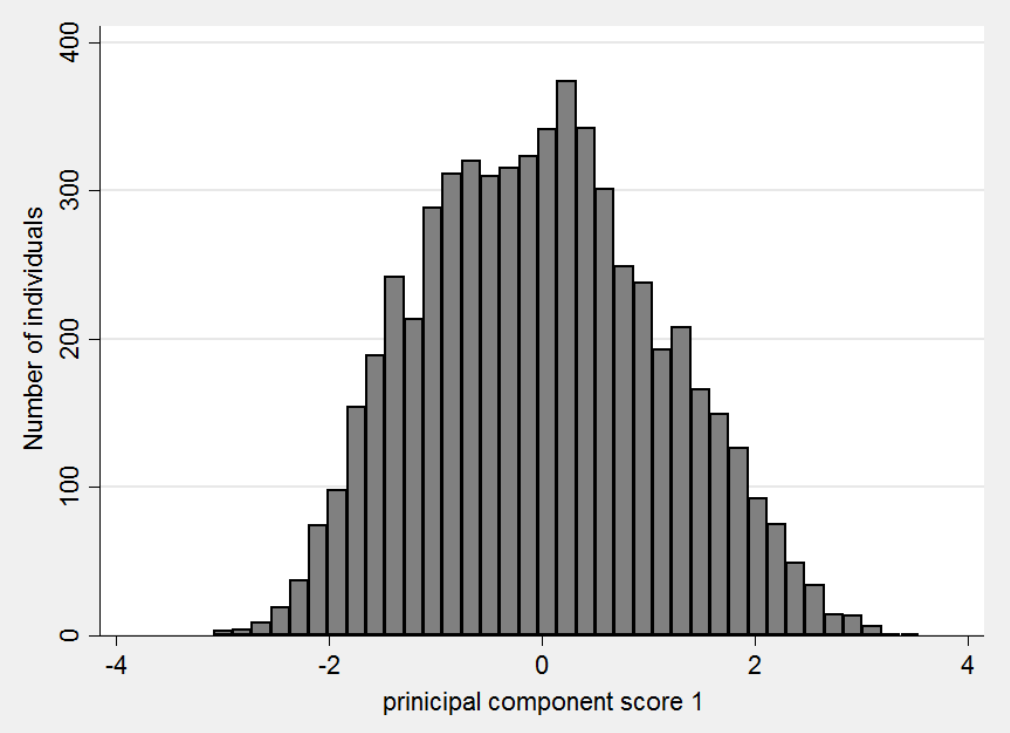
The principal components scores in the different samples were plotted against one another and their associations with each other were quantified using correlation. All the correlations were very strong (>0.99), which suggested any of the three samples could be used for the construction of the socioeconomic score. Thus, the full sample having 7,446 individuals was preferred in this case. The first principal component score for this sample explained 35% of the variation in these socioeconomic variables. The associations between the socioeconomic score of the first three components and the individual socioeconomic variables for the full sample are shown in (Table 3.6):

Table 3.6: Associations between principal components 1-3 (PC_1- PC_3) and the individual socioeconomic variables

SES Variables	PC_1	PC_2	PC_3
Maternal years of education	0.78	0.04	-0.03
assets (wealth)	0.47	-0.38	0.40
Individual income (raw)	0.79	0.11	-0.34
paternal years of schooling	0.31	0.71	0.42
social class – parental occupation	0.28	0.69	0.47
Utilities	0.59	-0.02	-0.09
Sanitation	0.63	-0.41	0.35
water supply	0.60	-0.42	0.40
crowding index (people/rooms)	-0.72	-0.10	0.41
child dependency ratio (children<18 /adults)	-0.56	-0.18	0.48

All variables were positively associated with PC_1, other than crowding index and child dependency ratio, which were inversely associated. Additionally, a histogram of PC_1 in the full sample was plotted (Figure 3.17). The curve was approximately normally distributed, and showed no evidence of clumping or truncation.

Figure 3.17: Distribution of first principal component score for the full sample (N=7446)



3.10 Forward Stepwise Regression

Multiple regression is used to assess how well one continuous dependent variable can be predicted from a set of independent (or predictor) variables. It determines how much variance in a continuous dependent variable can be explained by the predictors. Using the forward stepwise regression method I explore which SES components explain a significant proportion of variability in the composite SES variable created using PCA. This reduces the set of predictor variables to those that are necessary and account for nearly as much of the variance as is accounted for by the total set. The method helps to determine the level of importance of each predictor variable. This begins with an empty equation. Predictors are added one at a time beginning with the predictor with the highest correlation with the dependent variable. Once included, the variable remains in the equation thereafter. I will also discuss some of the disadvantages of this method in chapter-4.

3.11 Analysing outcomes in the next generation using multilevel modelling

In the NDBC study, data have been collected both on the cohort members and their offspring. Traditional methods assume that there is only one observation per individual. Thus there is a need to shift from traditional methods of analysing data to more sophisticated and complex methods which incorporate this dependency in the sample estimates. Individual observations that are clustered within a higher-level unit (e.g. siblings born to one parent) share a common environment and may be more similar than observations from individuals in different higher-level units (e.g. children born to different parents). A regression technique that has been developed to handle the correlation among these repeated measurements is known as linear mixed regression model. This technique is also known as ‘multilevel modelling’ or ‘hierarchical linear modelling’ (62).

In cases where there are more than one observation per individual, the average correlation measured between readings on individuals within the same unit (i.e. the intra-cluster correlation between the individual observations) is higher than the correlation measured between readings on individuals from different units. Mathematically, the intra cluster correlation coefficient (ICC) can be defined as (62):

ICC= (between cluster variability)/ (between + within cluster variability)

The process of building multilevel models is usually done in a stepped approach. First an ICC is calculated, and if it makes sense to use multilevel modelling given the nature of the data relative to the question, you advance step by step to a final model.

There are two types of multilevel models, the random intercept and the random slope model (62). The random-intercept model allows for variation in intercepts across the clusters without having to apply the traditional single level models to each level, whereas a more advanced model, the random-coefficients model accounts for variation in both intercepts and slopes across the levels. The dependent variable in all multilevel analyses can be quantitative, qualitative or time to event. In this thesis I consider a random intercept linear regression model with a quantitative outcome variable. The study has collected several variables in the F2 generation. For this thesis, I will only analyse birth weight.

3.12 Random Intercept Model

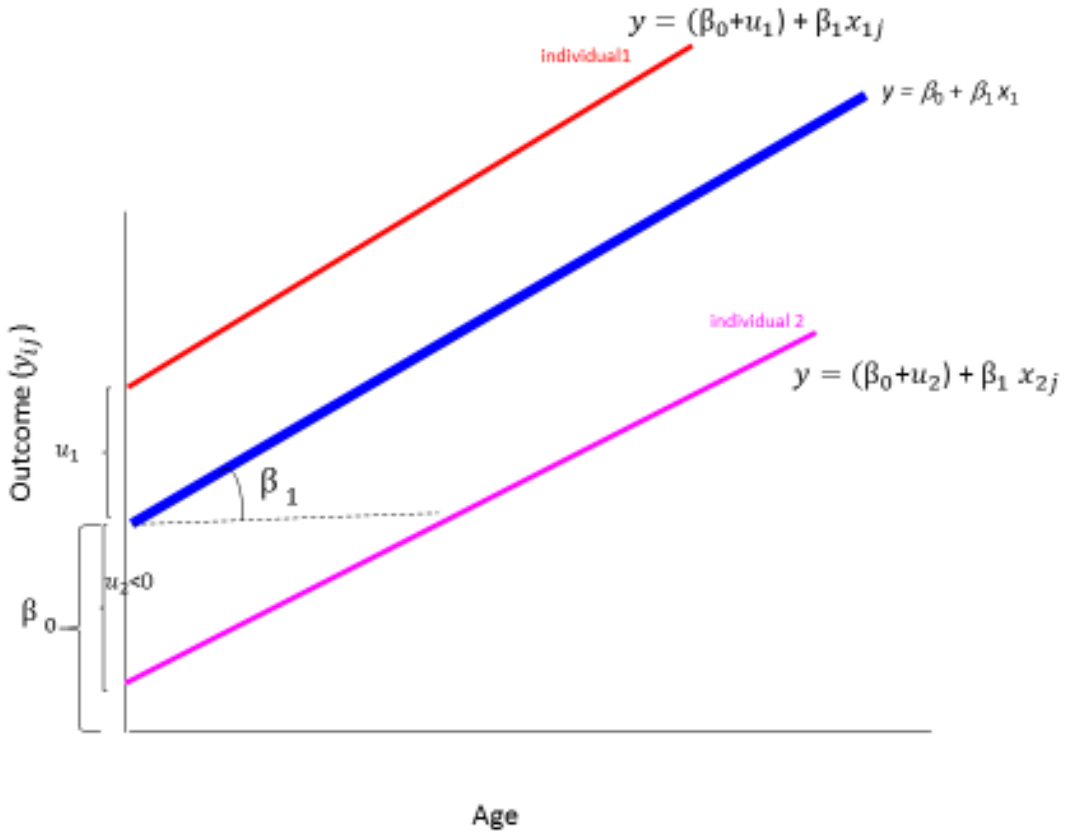
The general model for the response variable y_{ij} of parent i with child j can be specified as (Figure 3.18):

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + e_{ij}$$

Where $e_{ij} \sim N(0, \sigma_e^2)$,$u_i \sim N(0, \sigma_u^2)$

Here, u_i is a random parameter which is a parent-specific intercept, β_0 and β_1 are the intercept and slope respectively for the whole cohort and e_{ij} is a difference of a measurement for child j for parent i from the regression line for parent i. x_{ij} are the fixed effects in the model. To decide whether a random intercept or a random slope model might be suitable, we can compare the difference in likelihood ratios of the two models with a chi-square test (62).

Figure 3.18: Random Intercept model



Discussion

Need for data cleaning

Data cleaning is required to remove outliers in the data. An outlier is a point that is far outside the norm for a variable or a population (87). The presence of outliers can distort the different statistical estimates. Outliers can have other ill-effects on statistical analyses; their presence may increase the error variance and reduce the power of statistical tests. This can violate the assumption of normality required for many statistical tests of significance. Thus, they can increase the probabilities of Type I and Type 2 errors in the study.

Outliers can arise from several different causes. Anscombe et al. divides outliers into two categories (88). First are those that arise from errors in the data and second are those that arise from inherent variability in the data. The first category of outliers generally include outliers caused by human errors in data collection, recording or entry,

which is possible in this study. These errors are possible when a protocol or a standardized procedure for measuring the head is not followed perfectly. The study protocol has defined standardized methods for head measurement (see page 35, section 2.7). Therefore human error was minimised to the greatest extent possible, but human error due to inattention to these details can never be eliminated completely. There may also be types of points known as 'inliers' in the data. These are data points recorded in error, but falling within the expected measurement range (89). Longitudinal studies require checking the consistency of data across time points to detect inliers. Plots of individual values over time that show a recognizable pattern from which a discordant data point may be clearly identified. Inliers in the present study have been detected by the mathematical technique of linear interpolation and removed. There are still some surprising values even after data cleaning, but those are biologically plausible values.

Another way that outliers can be introduced in research studies is due to a biased sample. This type of error may be an issue in this study, as the population of Lajpat Nagar (a colony situated in South Delhi, India) was initially chosen for the study. At the time of recruitment, 59.9 percent of families had an income above 50 rupees per month (national average, 28.4). Only 14.9 percent of parents were illiterate (national average, 66.3) (21).

Methods of creating head circumference standard deviation scores and choice of method

Standardized scores made the head circumference measurements comparable across the different ages. Standardized head scores were derived using two methods, namely the 'normal model' (70) and the 'LMS method' (74). The SD scores were derived separately for the two sexes. The advantage of the normal model is its simplicity, but it is not suitable if the data are skewed. Skewed data may sometimes be normalised by applying suitable data transformation, but it is not always successful. Another disadvantage of this method is that, even after transformation, kurtosis may remain in the data, contradicting the normality assumption. Also age-varying skewness is not adjusted for by this method. The second method discussed in the chapter is the 'LMS method'. In this method three parameters, L (representing skewness), M (representing median) and S (coefficient of variation) are estimated for the body measurement at every age of measurement. These three parameters are estimated

simultaneously and smoothed across age. A potential disadvantage of this method is that the choice of the smoothing parameters may be subjective. A guideline that has been suggested is to compare the difference in deviances (75). Since the dataset used in my research has a large sample size, it has the power to detect even small changes in deviance, which may not be biologically plausible. Therefore, large differences in deviance should be considered for a study similar to mine. I estimated the LMS parameters in this study using the LMS chartmaker (90). Changes in deviance were observed sequentially. The best model chosen was the one having the minimum deviance and providing a smooth fit of the curves (90). LMS parameters can also be estimated using the GAMLSS package in R software (91). Another disadvantage is that the normality assumption after applying the Box-Cox transformation may be violated if there is kurtosis, which is not handled by this method (75).

In my study, the correlation between the SD scores from the two methods was (>0.99) in all the measurement time periods for both the sexes (Figure 3.2, page 46). Also, the mean difference in the SD scores calculated by the two methods (LMS-Normal) was 0.0001 (SD: 0.12) for males and 0.0003 (SD: 0.08) for females, suggesting there was little evidence of bias in the SD scores calculated using either of the two methods. This suggested that I can use either of the methods for calculation of the head circumference SD scores. Thus, I used the SD scores calculated using the normal model for all subsequent work. Since the head measurements were symmetrical at all ages except a slight skewness observed at birth, the Normal model is suitable for creating SD scores for head measurements.

Data cleaning using age-adjusted standard deviation scores

Data cleaning deals with data problems after they have occurred. Data cleaning for the present study was a three-stage process, which involved repeated cycles of screening, diagnosing and editing or removal of suspected data. It is efficient to detect these problems in a planned way. Data cleaning must be based on knowledge of technical errors and the expected ranges of the biological variables. The initial screening methods may not be statistical. Many outliers can be detected by the investigator's experience or knowledge from the literature. I removed head circumference values greater than 700 mm or less than 200 mm before creating the age-adjusted SD scores for head circumference. I

followed a graphical approach whereby I plotted the head measurements of every individual with respect to age, identifying surprising values at any particular time points, and iteratively removing those values. Some values like the head size at birth for many individuals were found to be very surprising. This was probably due to preterm births. Since a lot of growth takes place in the last few weeks of pregnancy, premature births results in a smaller size at birth (53). Thus these values which may be mathematically surprising, are biologically plausible. At the end of this procedure, 0.6% of the total head circumference values were excluded, which suggests a low error rate in collecting and recording the head measurements for this study, and good data quality.

Age-adjusted centile curves for head circumference: comparison with other populations

Using the cleaned age-adjusted SD scores, I plotted the head circumference centile curves separately for males and females. The overall pattern of head growth for both males and females is rapid growth in infancy, slower linear growth in childhood, and acceleration in puberty. The Delhi centile curves show an increasing head size during adulthood for both males and females, which is surprising because head growth is generally complete by age 21 years. In undernourished populations, growth can continue far longer than this, but is even then usually complete by the late 20's. One possibility for this that I explored was to see whether persons belonging to a higher social class (and therefore having larger heads) came later to clinics, therefore being older. This theory was not supported by my analysis and I could not find any other convincing explanation.

I compared the NDBC head data with data from other Indian populations. The head data of individuals from Delhi compares very closely with individuals from other parts of India. In a study by Bhalla et al in 1993, the authors described the pattern of head growth of one hundred and fifty four (86 male and 68 female) Punjabi infants in Chandigarh who were measured at monthly intervals up to age 1 year (92). The general pattern of growth that the authors noted was an initial sharp rise followed by slow gain during the second half of infancy. The males had larger head sizes than females at all ages after birth. The head circumference growth velocity deceased rapidly after birth up to 4 months, thereafter it decreased slowly. Sex differences in monthly growth rates at a few of the months were statistically significant in the study, which was also observed in my study. A study was conducted by Agarwal et al in 1993 on affluent urban Indian children from six cities (Bangalore, Calcutta, Delhi, Kota, Ludhaiana and Varanasi) (93). The

study was carried out for two years (1985-87) and consisted of two different cohorts (birth to 1 year, N=750 and 1-6 years, N=1,885). The centiles for head circumference in this study were larger than in the NDBC (maximum difference of 2 cms in median head size at any age for males and females). This may be due to the fact that there is a gap of 15 years between the early head sizes taken in the Delhi and Chandigarh studies and the nutritional capacity of India has improved substantially during this time (94).

The head sizes in this study were similar to other Asian populations. A study by Karabiber et al. in 2001 measured head circumference of 1,826 healthy Turkish children (945 males and 881 females) aged between 6 and 12 years (95). The 50^{th} centile values of head circumference for NDBC at the specified ages compared closely with data from this study. I also compared the head values of NDBC with that of English children (154 boys and 128 girls) born between 1970 and 1977, who had their heads measured from birth to 7 years (96). The babies were examined within 4 days of delivery and at the ages of 2, 6 , 12 and 18 months, and at 2,3,4 and 7 years. The head circumferences at the defined ages in the Delhi population were similar for the English boys but smaller than the English girls at the same ages (mean difference in head size at the 50^{th} centile of about 0.78 cms for boys and 2.3 for girls).

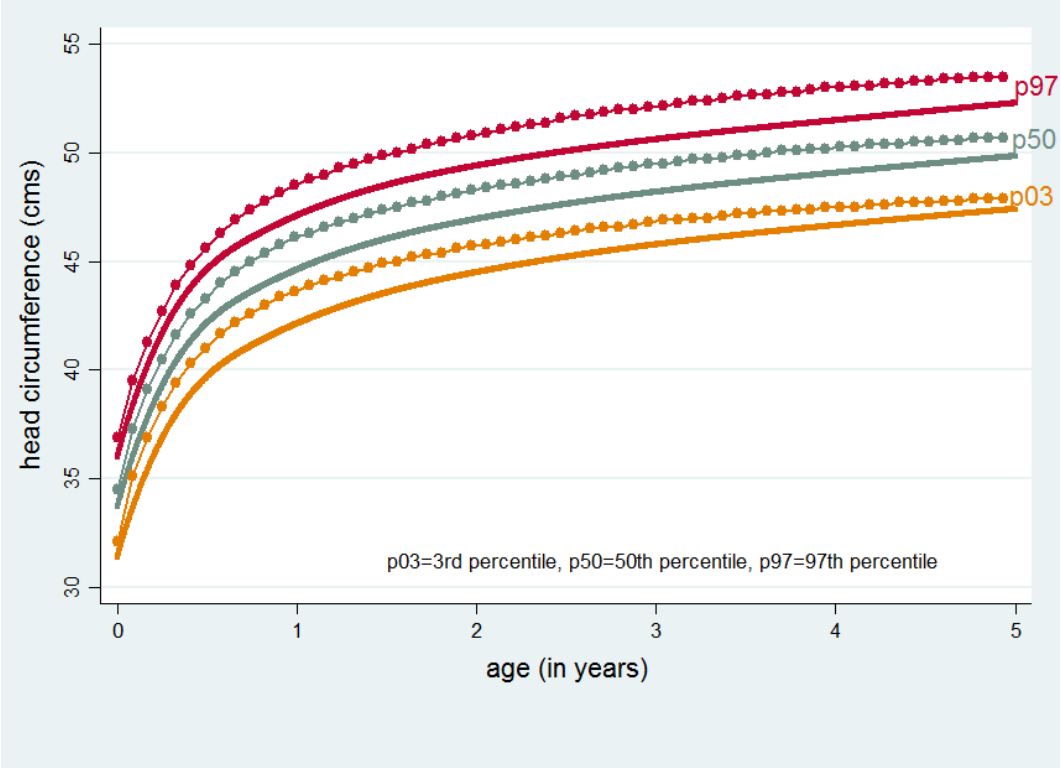
Other data suggest that head size in Asians is smaller than among white Caucasians. A study of Japanese children (97) observed that heads of children in both the sexes were smaller in comparison to individuals born in the United Kingdom in the same period (95). This observation was confirmed in a systematic review conducted by Natale et al. in 2014 which compared head circumference data from the World Health Organization (WHO)'s Multicenter Growth Reference Study (MGRS) with data from studies performed in 55 countries or ethnic groups in children up to age 5 years (98). The MGRS (July 1997 to December 2003) was a population-based study taking place in the cities of Davis, California, USA; Muscat, Oman; Oslo, Norway; and Pelotas, Brazil; and in selected affluent neighbourhoods of Accra, Ghana, and South Delhi, India (99). The study had two components: a longitudinal follow-up in which children were recruited at birth and followed up at home until they were 24 months old, and a cross-sectional survey involving children aged 18-71 months. Head circumference at birth was measured at all study sites. The MGRS authors concluded that all economically advantaged children who were breastfed as infants grow similarly. However, Natale et al. came to different conclusions. They made the point that head circumference was more variable across populations than height or weight. Indians in this study had smaller head sizes at the specified ages as

compared to other populations. This conclusion is consistent with my observations of the NDBC data, which indicated smaller head size than the WHO data (Figure 3.19). Natale et al. concluded that head size in breastfed children at any age examined was closer to local references than to the MGRS means and the use of a single international standard for head circumference is not justified.

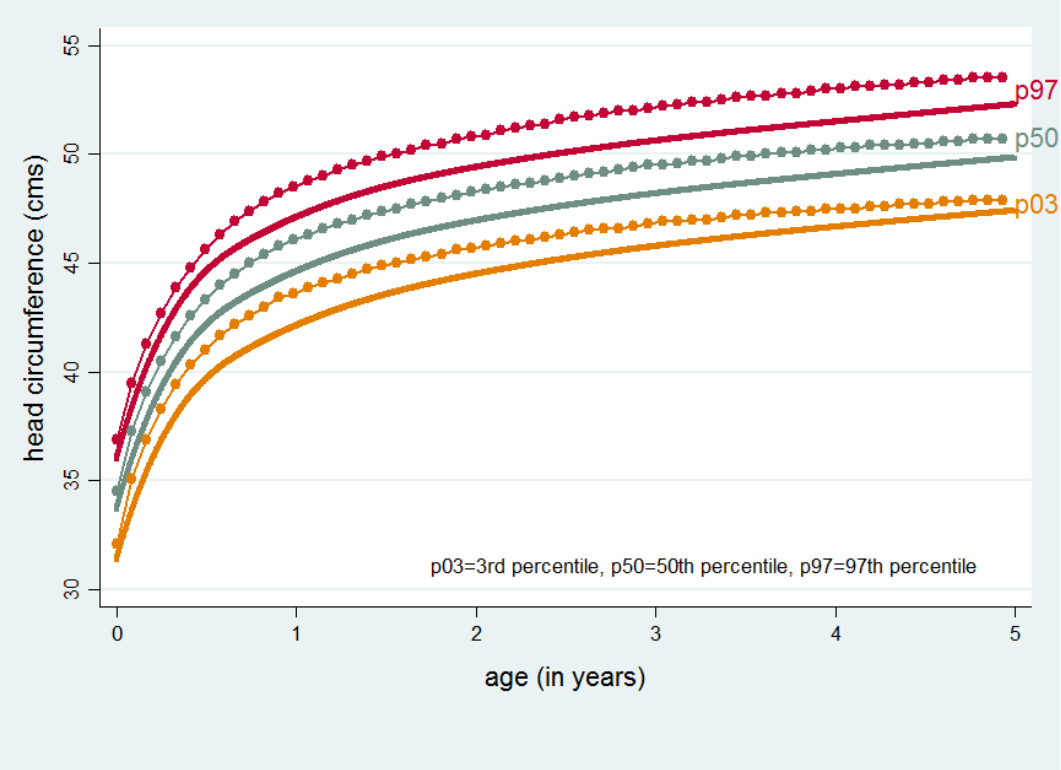
This raises the issue of which is more useful-an external or an internal growth standard. Head circumference varies considerably. The review by Natale et al. showed that the maximum variation between the extreme values in each age and sex group was 2.5 SD. Studies have found that head circumference in breastfed infants, for which the WHO growth reference was constructed is closer to the local standards than it is to the WHO charts. But, as growth is affected by early life nutrition, use of global standards can be useful for undernourished populations to aim to achieve the expected level of head growth based on the global standard.

Figure 3.19: Comparison of New Delhi Birth Cohort (smooth lines) and World Health Organization (dotted lines) growth curves for head circumference from birth to age 5 years

Males



Females



Conditional growth models

Conditional body size approach has been described in this chapter, which is used to observe the associations of head size at birth and head growth in the defined periods with adult outcomes. An advantage of the conditional body size approach is that the conditional body size measures are always statistically independent of each other, hence they help to deal with the issue of correlation among the predictors. Hence, these models resolve the issue of multicollinearity posed by other statistical methods of modelling growth data such as tracing z scores or the life course plot (58). The conditional body size measures have a simple and straightforward interpretation and the analysis is simple to implement. Disadvantages of conditional models are that they require all the body sizes to be measured, and at approximately the same age for the entire cohort. Missing data can therefore lead to a reduction in the analysis sample size. Another disadvantage is that these measures may be difficult to interpret for non-mathematicians, because change in body size is shown in SD units rather than units of measurement. Finally, due to the method used for the construction of the conditional variables, the comparison in this type

of analysis is between subjects within a population. Hence, this method cannot be used to characterize the growth of an external population. In chapter-6, I will explore how using a different growth modelling method influences my findings.

Chapter-4

Association of head growth with human capital and cardiometabolic outcomes in adulthood

Literature review indicates that associations between head growth during different periods in early life and later life cognitive and cardiometabolic outcomes have not been explored in an Indian population. Thus, using the F1 generation data from the NDBC I have examined associations of head growth with two adult outcomes using the conditional body size approach. These were years of education as a proxy for cognitive ability and SBP, a cardiometabolic outcome. I have further adjusted the associations of head growth with years of education for the SES of the mothers of the cohort members, because both postnatal growth and attained education reflect socioeconomic conditions in childhood, and so SES is a potential confounding factor (82). I have also examined whether disproportion between head and length at birth is associated with years of education in later life and SBP in adulthood. In the following section I present the distribution of the two outcome variables and the exposure variables used for the analyses.

4.1 Details of the outcome and exposure variables

Figure 4.1: Distribution of years of education

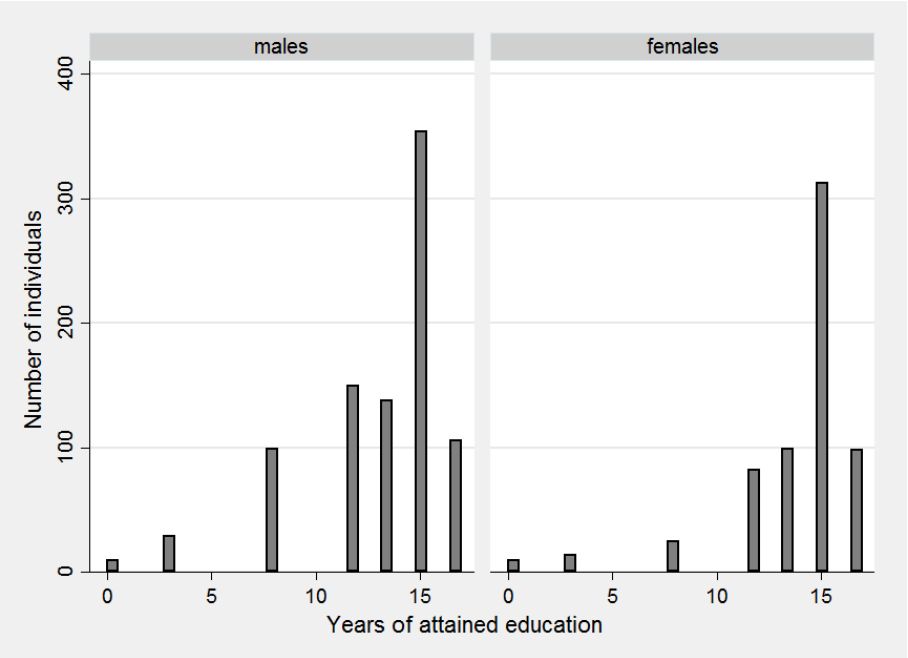
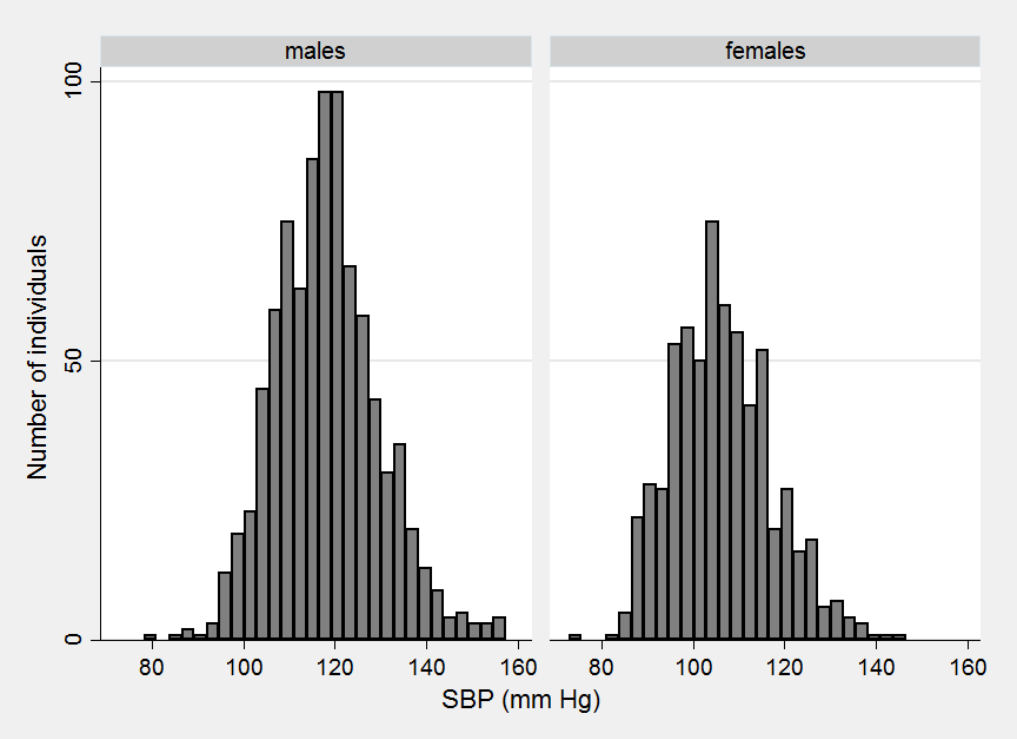


Figure 4.2: Distribution of systolic blood pressure



Systolic blood pressure was more symmetrically distributed than years of education, but I regarded them both suitable for use in linear regression untransformed (Figures 4.1, 4.2)).

Figure 4.3: Distribution of head circumference

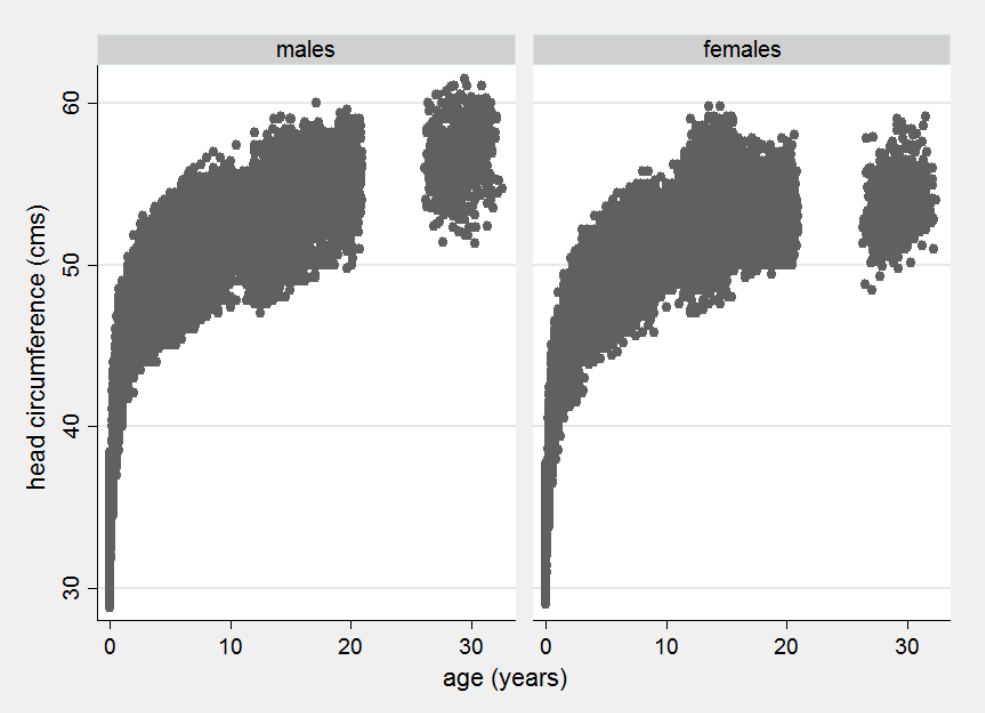


Figure 4.4: Distribution of height

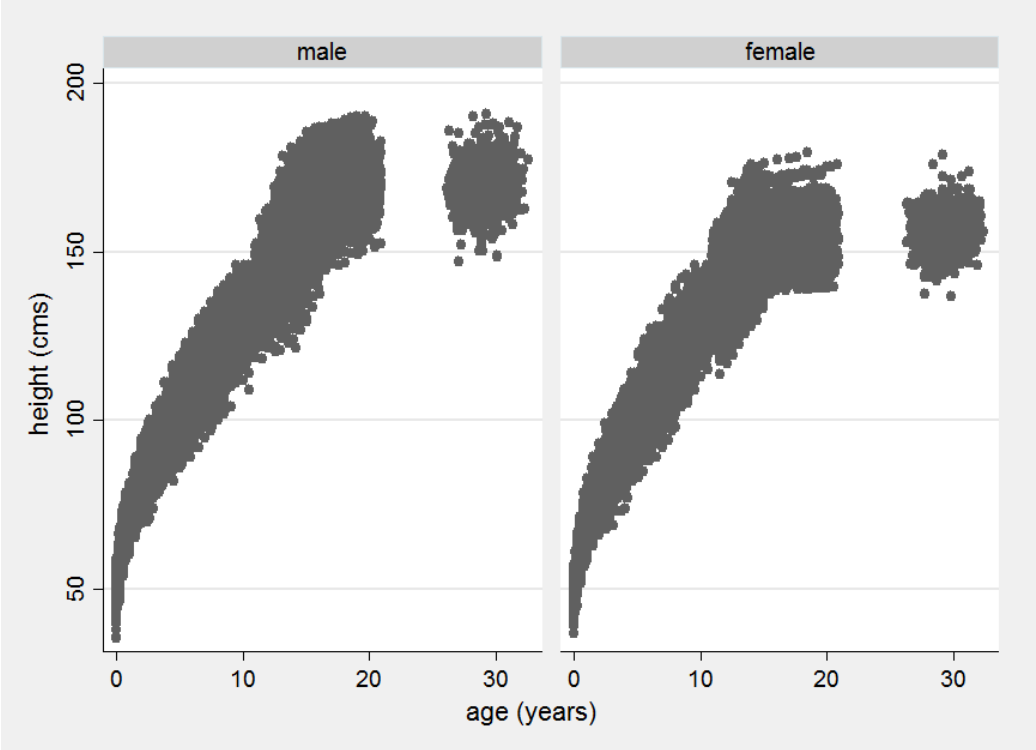
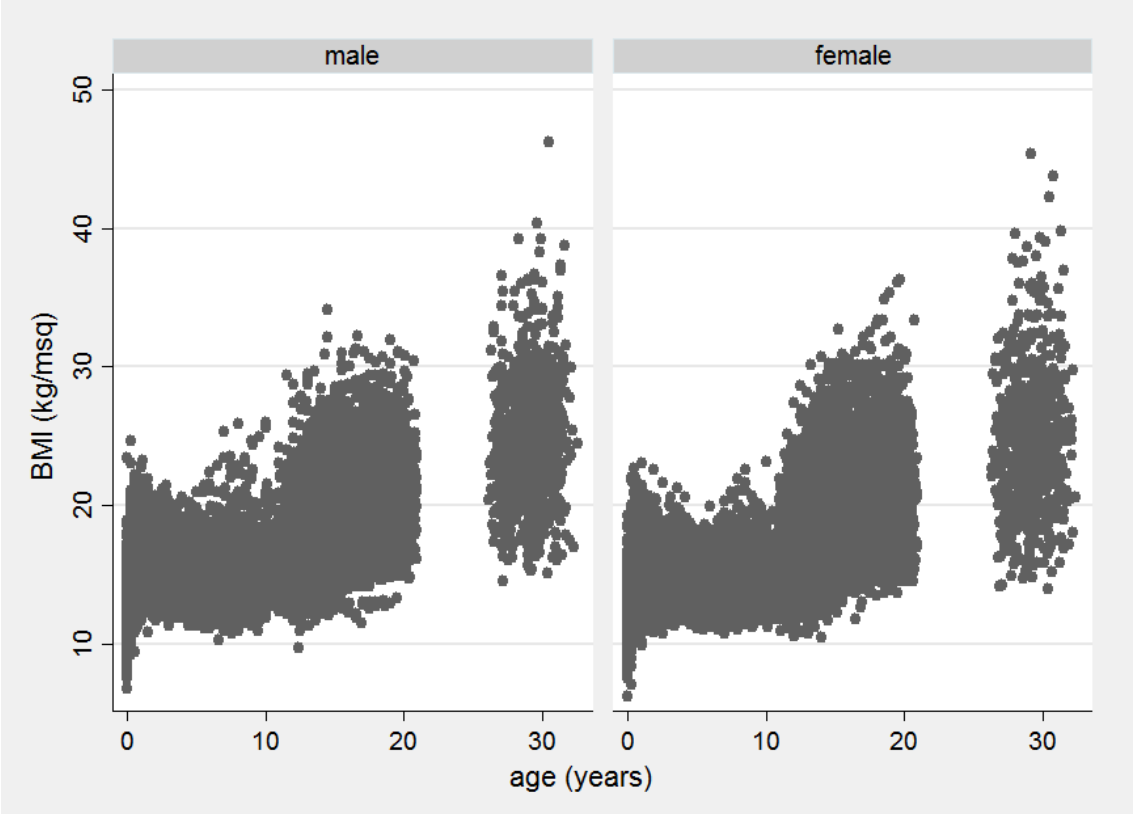


Figure 4.5: Distribution of body mass index



Graphs of head, height and BMI which are the exposures in the subsequent analyses, show a nonlinear association with respect to age (Figures 4.3-4.5). The graphs of head circumference follow a similar trajectory, a rapid increase during infancy and childhood and finally stopping during adolescence. BMI rose and fell during infancy and rose again during childhood and adolescence. This pattern was seen among both males and females.

4.2 Outcome: Years of education

4.2.1 Association of body growth at different ages with years of education

As an initial step, unadjusted associations of head growth at different time points (birth, birth to 6 months, 6 months to 2 years, 2 years to 11 years and 11 years to adult) with years of education, by sex, were analysed. This was followed by constructing formal interaction tests between sex and head growth measures. The same procedure was followed for determining associations of height growth measures and BMI growth measures at the different time points with years of education. Finally, sex stratified estimates of the association between head growth and education after adjusting for height and BMI growth were obtained. Interactions of sex with all three growth measures at different time points were tested using Wald's test. Since I did not find any statistically significant interactions, pooled results are presented for the adjusted models. The first adjusted model included head growth, height and BMI measured at the different time points. The second model included gestational age, and the third model further adjusted for SES.

Table 4.1 presents the unadjusted analysis showing the associations between years of education and the head growth variables considered simultaneously. In the pooled analysis, years of education was positively associated with head circumference at birth and with head growth during early and late infancy. Years of education increased by 0.29 years for each 1 SD increase in head circumference at birth, by 0.45 years for each 1 SD increase in head growth during early infancy and by 0.25 years for each 1 SD increase in head growth during late infancy. Head growth during childhood and adolescence were unrelated to years of education.

Table 4.1: Association of head growth and years of education in adulthood (unadjusted)

Outcome Variable: Years of education							
Predictor	Males (n=623)		Females (n=429)		Inter-action	Pooled (n=1052)	
	β (95% CI)	p-value	β (95% CI)	p-value	p-value	β (95% CI)	p-value
Head (z score)							
Birth	0.32 (0.06 to 0.58)	0.01	0.23 (-0.04 to 0.51)	0.09	0.6	0.29 (0.10 to 0.49)	0.003
birth-6m	0.38 (0.10 to 0.66)	0.007	0.54 (0.24 to 0.84)	<0.001	0.4	0.45 (0.24 to 0.65)	<0.001
6m-2y	0.24 (-0.03 to 0.51)	0.08	0.25 (-0.03 to 0.54)	0.07	0.9	0.25 (0.05 to 0.45)	0.01
2y-11y	-0.02 (-0.28 to 0.24)	0.8	0.10 (-0.19 to 0.38)	0.4	0.5	0.03 (-0.16 to 0.22)	0.7
11y-adult	0.21 (-0.05 to 0.47)	0.1	0.00 (-0.26 to 0.27)	0.9	0.2	0.13 (-0.06 to 0.32)	0.1

Table 4.2 shows the associations between years of education and the conditional height growth variables. In the unadjusted pooled analysis, years of education was positively associated with height at birth and with height growth during early and late infancy. Years of education increased by an average of 0.32 years for each 1 SD increase in height at birth, by an average of 0.46 years for each 1 SD increase in height growth during early infancy and by an average of 0.66 years per 1 SD increase in height growth during late infancy. Greater-than expected height growth during childhood showed a positive association of borderline significance with years of education, while height growth during adolescence was unrelated to attained education.

Table 4.2: Association of height growth and years of education in adulthood (unadjusted)

Outcome Variable: Years of education							
Predictor	Males (n=694)		Females (n=512)		Inter-action	Pooled (n=1206)	
	β (95% CI)	p-value	β (95% CI)	p-value	p-value	β (95% CI)	p-value
Height (z score)							
birth	0.29 (0.05 to 0.53)	0.03	0.41 (0.15 to 0.68)	0.003	0.3	0.32 (0.14 to 0.50)	<0.001
birth-6m	0.45 (0.22 to 0.69)	<0.001	0.47 (0.22 to 0.72)	<0.001	0.9	0.46 (0.29 to 0.63)	<0.001
6m-2y	0.62 (0.38 to 0.86)	<0.001	0.72 (0.46 to 0.97)	<0.001	0.5	0.66 (0.49 to 0.84)	<0.001
2y-11y	0.26 (0.03 to 0.50)	0.02	0.05 (-0.20 to 0.31)	0.6	0.2	0.17 (0.00 to 0.35)	0.05
11y-adult	-0.07 (-0.31 to 0.17)	0.5	0.04 (-0.29 to 0.22)	0.7	0.8	-0.05 (-0.23 to 0.12)	0.5

Table 4.3 shows the association between years of education and the conditional BMI growth variables, considered simultaneously. In the unadjusted pooled analysis, years of education was positively associated with BMI at birth and with increases in BMI during early and late infancy and childhood. Years of education increased by an average of 0.25, 0.44, 0.31 and 0.28 years for each 1 SD increase in BMI at birth, and conditional BMI gain during early and late infancy and childhood respectively. No statistically significant association was found between BMI growth during adolescence and attained education.

Table 4.3: Association of body mass index growth and years of education in adulthood (unadjusted)

Outcome Variable: Years of education							
Predictor	Males (n=688)		Females (n=507)		Inter-action	Pooled (n=1195)	
	β (95% CI)	p-value	β (95% CI)	p-value	p-value	β (95% CI)	p-value
BMI (z score)							
birth	0.35 (0.10 to 0.60)	0.006	0.09 (-0.20 to 0.37)	0.5	0.1	0.25 (0.06 to 0.43)	0.01
birth-6m	0.31 (0.07 to 0.55)	0.01	0.63 (0.37 to 0.89)	<0.001	0.07	0.44 (0.27 to 0.62)	<0.001
6m-2y	0.36 (0.12 to 0.60)	0.003	0.24 (-0.02 to 0.50)	0.07	0.5	0.31 (0.14 to 0.49)	0.001
2y-11y	0.33 (0.09 to 0.56)	0.008	0.21 (-0.05 to 0.47)	0.1	0.5	0.28 (0.10 to 0.45)	0.002
11y-adult	0.03 (-0.21 to 0.27)	0.8	0.05 (-0.22 to 0.31)	0.7	0.9	0.04 (-0.14 to 0.21)	0.6

I formally checked the interaction of sex with conditional measures of head, height and BMI at the defined time periods, which I now considered simultaneously. Only one out of the fifteen interactions was statistically significant, and the association was not strong, which could have been simply by chance. Therefore, I pooled both sexes (Table 4.4). After adjustment for height and BMI gain, head size at birth and conditional head growth ceased to be statistically significantly associated with years of education. Growth in height during early and late infancy, and BMI growth during late infancy remained positively associated with years of education. Years of education increased by an average of 0.35 and 0.74 years for each 1 SD increase in conditional height growth during early and late infancy and by an average of 0.54 years for every 1 SD increase in BMI during the period of late infancy (Figures 4.7 and 4.8). These associations changed little after further adjustment for gestational age.

Figure 4.6: Association between head growth and years of education (adjusted and unadjusted)

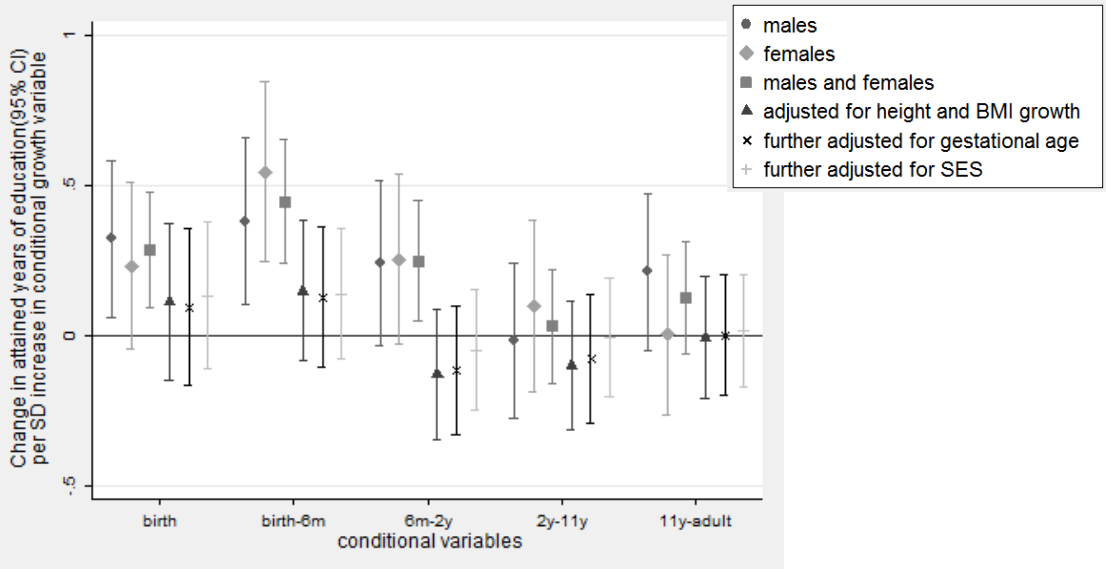


Figure 4.7: Association between height growth and years of education (adjusted and unadjusted)

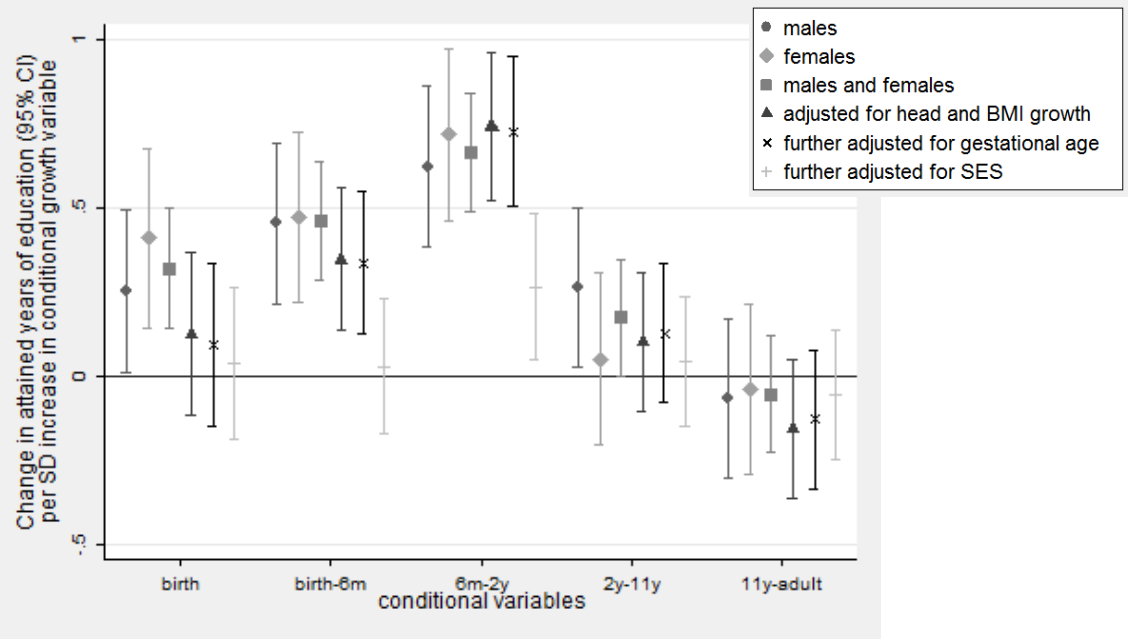


Figure 4.8: Association between body mass index growth and years of education (adjusted and unadjusted)

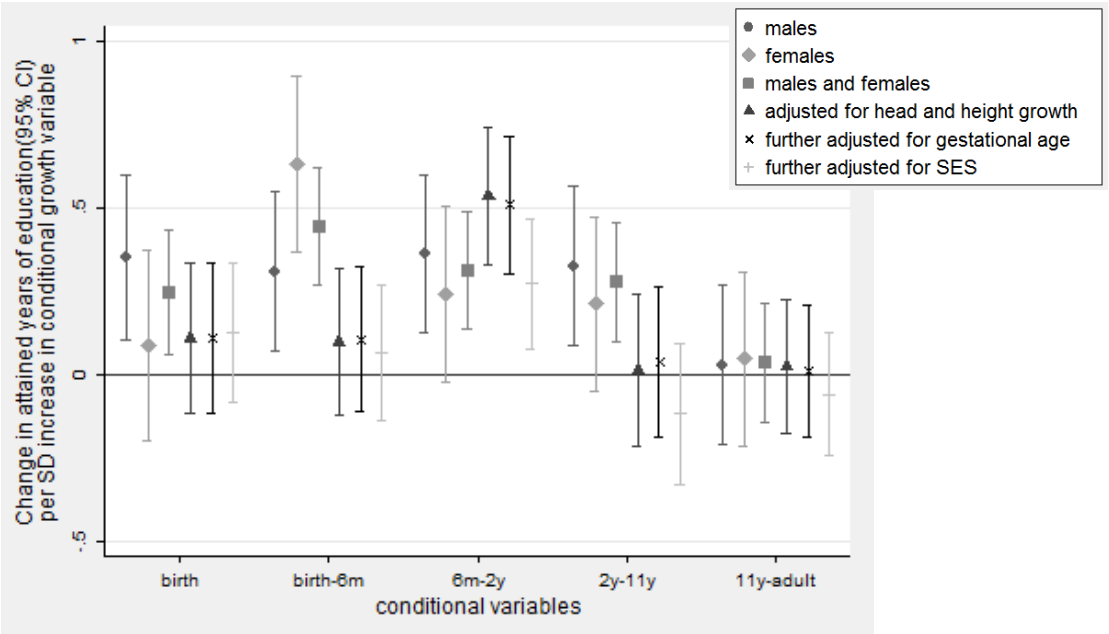


Table 4.4: Association of head growth and years of education in adulthood (mutually adjusted)

Predictor	Outcome Variable: Years of education			
	Pooled (n=958)		Mutually adjusted for gestational age (n=958)	
	β (95% CI)	p-value	β (95% CI)	p-value
Head (z score)				
Birth	0.11 (-0.15 to 0.38)	0.4	0.10 (-0.17 to 0.36)	0.4
birth-6m	0.15 (-0.09 to 0.38)	0.2	0.13 (-0.11 to 0.36)	0.2
6m-2y	-0.13 (-0.34 to 0.09)	0.2	-0.12 (-0.33 to 0.10)	0.2
2y-11y	-0.10 (-0.31 to 0.11)	0.3	-0.08 (-0.29 to 0.14)	0.4
11y-adult	-0.01 (-0.21 to 0.20)	0.9	0.00 (-0.20 to 0.20)	0.9
Height (z score)				
Birth	0.13 (-0.12 to 0.37)	0.3	0.09 (-0.15 to 0.34)	0.4
birth-6m	0.35 (0.14 to 0.56)	0.001	0.34 (0.12 to 0.55)	0.002
6m-2y	0.74 (0.52 to 0.96)	<0.001	0.73 (0.51 to 0.95)	<0.001
2y-11y	0.10 (-0.11 to 0.31)	0.3	0.13 (-0.08 to 0.33)	0.2
11y-adult	-0.16 (-0.36 to 0.05)	0.1	-0.13 (-0.33 to 0.08)	0.2
BMI (z score)				
Birth	0.11 (-0.12 to 0.34)	0.3	0.11 (-0.12 to 0.34)	0.3
birth-6m	0.10 (-0.12 to 0.32)	0.3	0.11 (-0.11 to 0.32)	0.3
6m-2y	0.54 (0.33 to 0.74)	<0.001	0.51 (0.30 to 0.71)	<0.001
2y-11y	0.01 (-0.21 to 0.24)	0.9	0.04 (-0.19 to 0.27)	0.7
11y-adult	0.03 (-0.18 to 0.23)	0.8	0.01 (-0.19 to 0.21)	0.9
Gestational age[#]	-	-	0.21 (-0.00 to 0.42)	0.05
Missing gestational age	-	-	-1.00 (-1.68 to -0.32)	0.004

#Gestational age known for 1379 (18.5%) individuals.

4.2.2 Adjusting for socioeconomic status: Adjusting for SES is important in this analysis as SES is independently associated with both head growth and years of education. Therefore, it might confound the association between head growth and years of education. As described in Chapter-3 (table 3.5, page 68) information was collected on ten socioeconomic variables of the parents of the cohort participants. I also described the construction of the combined variable using PCA. I considered the association between head growth measurements and years of attained education, adjusting for each of the socioeconomic variables using two approaches: one using PCA and one using a forward stepwise regression. All the SES variables were measured in different units and

were standardized before analysis. I replaced missing values with mean values and created a binary missing indicator which was coded 1 if the value was missing and 0 otherwise. I identified the individual socioeconomic variables which were associated with attenuation. For this, I determined the magnitude of association of the individual SES variables with years of education over several rounds. In the first round out of the ten SES variables, I determined the variable having the maximum correlation using the F-test. In the second round, I forced this variable in the model and determined the variable having the maximum correlation from the remaining nine variables. The importance of the variables was determined using the F-test. Eight rounds were conducted subsequently following a similar technique (Table 4.5). After including five SES variables, none of the remaining variables explained a significant amount of variability in years of education, and thus five variables were considered sufficient to be included in the final model (Table 4.6).

Table 4.5: Forward stepwise regression to identify factors associated with attained years of education

Step	Factors	R statistic	Increase in F-statistic	p-value
1	Paternal education	0.23	0	-
2	+ Paternal occupation	0.28	56.48	<0.001
3	+ Maternal education	0.30	18.0	<0.001
4	+ Assets (wealth)	0.30	5.77	0.003
5	+Crowding index	0.30	3.43	0.03
6	+Sanitation	0.30	1.29	0.2
7	+Individual income(raw)	0.30	1.21	0.2
8	+Child dependency ratio	0.30	1.26	0.2

9	+Water supply	0.30	0.49	0.6
10	+Utilities	0.30	0.46	0.6

The final model (Table 4.6) using the forward selection method included paternal education, paternal occupation, maternal education and assets (wealth) and crowding index as the individual SES variables.

Table 4.6: Socioeconomic status variables associated with attained years of education

SES variables(SD score)	Years of education (Coefficient (95% CI))	missing values n (%)	p-value
Paternal education	0.78(0.59 to 0.97)	-	<0.001
Missing(1=yes,0=no)	-1.61(-2.19 to -1.03)	5991 (80.5)	<0.001
Paternal occupation	0.82(0.65 to 0.99)		<0.001
Missing(1=yes,0=no)	0.54(-0.63 to 1.70)	5914 (79.4)	0.3
Maternal education	0.45(0.26 to 0.64)		<0.001
Missing(1=yes,0=no)	0.26(-0.28 to 0.80)	1739 (23.4)	0.3
Assets (wealth)	0.29(0.09 to 0.48)		0.004
Missing(1=yes,0=no)	-0.99(-3.71 to 1.73)	2052 (27.6)	0.4
Crowding index	-0.24(-0.43 to -0.05)		0.01
Missing(1=yes,0=no)	0.90(-1.82 to 3.63)	2051 (27.5)	0.5

Since these socioeconomic variables were correlated with one another I also calculated a socioeconomic score or 'construct' based on principal components analysis (See Chapter 3, page 67). The adjusted body growth coefficients based on these two approaches are presented in Table 4.7. I did not find a substantial difference in the associations using these two approaches and thus preferred the approach based on principal components analysis as it is based on estimating fewer parameters. The associations between years of education and the different body growth measurements were attenuated substantially after adjusting for SES (Figures 4.1, 4.2, 4.3), although height and BMI gain during late infancy remained significantly and positively associated with attained education.

Table 4.7: Association of head growth and years of education in adulthood

(mutually adjusted for socioeconomic status)

Predictor	Mutually adjusted for individual SES Factors (n=958)		Mutually adjusted for SES Construct (n=958)	
	β(95% CI)	p-value	β(95% CI)	p-value
Head (z score)				
Birth	0.15 (-0.09 to 0.38)	0.2	0.14 (-0.11 to 0.38)	0.2
birth-6m	0.12 (-0.09 to 0.33)	0.2	0.14 (-0.07 to 0.36)	0.1
6m-2y	-0.05 (-0.24 to 0.15)	0.6	-0.05 (-0.25 to 0.15)	0.6
2y-11y	-0.01 (-0.20 to 0.19)	0.9	-0.01 (-0.20 to 0.19)	0.9
11y-adult	0.04 (-0.14 to 0.22)	0.6	0.02 (-0.17 to 0.21)	0.8
Height (z score)				
Birth	-0.01 (-0.22 to 0.21)	0.9	0.04 (-0.19 to 0.26)	0.7
birth-6m	-0.03 (-0.22 to 0.17)	0.7	0.03 (-0.17 to 0.23)	0.7
6m-2y	0.25 (0.04 to 0.45)	0.02	0.26 (0.05 to 0.48)	0.01
2y-11y	0.06 (-0.13 to 0.24)	0.5	0.05 (-0.15 to 0.24)	0.6
11y-adult	-0.08 (-0.27 to 0.1)	0.3	0.06 (-0.25 to 0.14)	0.5
BMI (z score)				
Birth	0.08 (-0.13 to 0.28)	0.4	0.12 (-0.09 to 0.33)	0.2
birth-6m	0.06 (-0.13 to 0.25)	0.5	0.06 (-0.14 to 0.26)	0.5
6m-2y	0.26 (0.07 to 0.44)	0.007	0.27 (0.07 to 0.46)	0.008
2y-11y	-0.10 (-0.31 to 0.10)	0.3	-0.12 (-0.33 to 0.09)	0.2
11y-adult	-0.07 (-0.24 to 0.11)	0.4	-0.06 (-0.25 to 0.13)	0.5
Gestational age	0.10 (-0.09 to 0.29)	0.2	0.16 (-0.03 to 0.36)	0.1
Missing(1=yes,0=no)	-0.60 (-1.21 to 0.01)	0.05	-0.53 (-1.16 to 0.11)	0.1
Individual SES Factors(z score)				
Paternal occupation	0.74 (0.53 to 0.95)	<0.001	-	-
Missing(1=yes,0=no)	0.37 (-1.11 to 1.85)	0.6	-	-
Paternal education	0.77 (0.54 to 0.99)	<0.001	-	-
Missing(1=yes,0=no)	-1.30 (-2.00 to -0.60)	<0.001	-	-
Maternal education	0.30 (0.05 to 0.54)	0.01	-	-
Missing(1=yes,0=no)	0.03 (-0.79 to 0.85)	0.9	-	-
Assets(wealth)	0.24 (-0.01 to 0.50)	0.06	-	-
Missing(1=yes,0=no)	-1.59 (-5.21 to 2.03)	0.3	-	-
Toilet(sanitation)	0.13 (-0.13 to 0.39)	0.3	-	-
Missing(1=yes,0=no)	-0.00 (-0.23 to 0.22)	0.9	-	-
Utilities	-0.01 (-0.23 to 0.22)	0.9	-	-
Water	-0.08 (-0.35 to 0.19)	0.5	-	-
Crowding index	-0.09 (-0.35 to 0.17)	0.5	-	-
Missing(1=yes, 0=no)	-0.16 (-3.79 to 3.47)	0.9	-	-
SES Construct	-	-	1.23 (1.04 to 1.43)	<0.001
Missing(1=yes,0=no)	-	-	0.11 (-0.25 to 0.47)	0.5
R-square	0.31		0.24	

4.3 Outcome: Systolic blood pressure

4.3.1 Association of body growth at different ages with systolic blood pressure

As in the previous analyses, unadjusted associations of head growth at different ages with SBP were analysed. This was followed by formal interaction tests between sex and different head growth measures. The same procedure was followed for determining associations of height and BMI growth measures at different ages with SBP. Sex stratified estimates of association between head growth and SBP after adjusting for height and BMI growth were obtained. Interactions of sex with all three growth measures at different time points were tested using the Wald test. Since I did not find any statistically significant interactions, pooled results are presented. Bivariate associations were assessed between SBP and gestational age and SES. Since I did not find SES to be significantly associated with SBP, the models were not adjusted for SES.

Table 4.8 presents the unadjusted analysis, and shows that SBP was associated positively with conditional head growth during early infancy, childhood and adolescence. SBP increased by 0.84, 0.93 and 1.65 mmHg for each 1 SD increase in head growth at these ages. Head size at birth and head growth during late infancy were unrelated to adult SBP.

Table 4.8: Association of head growth and systolic blood pressure in adulthood (unadjusted)

Outcome Variable: SBP (mmHg)							
Predictor	**Males (n=618)**		**Females (n=424)**		**Inter-action**	**Pooled (n=1042)**	
	β (95% CI)	**p-value**	**β (95% CI)**	**p-value**	**p-value**	**β (95% CI)**	**p-value**
Head (z score)							
Birth	0.38 (-0.47 to 1.23)	0.3	0.27 (-0.84 to 1.38)	0.6	0.8	0.35 (-0.33 to 1.02)	0.3
birth-6m	0.99 (0.09 to 1.90)	0.03	0.62 (-0.58 to 1.82)	0.3	0.6	0.84 (0.12 to 1.57)	0.02
6m-2y	0.29 (-0.59 to 1.18)	0.05	-0.56 (-1.69 to 0.59)	0.3	0.2	-0.06 (-0.76 to 0.64)	0.8
2y-11y	1.30 (0.46 to 2.15)	0.003	0.46 (-0.68 to 1.59)	0.4	0.2	0.93 (0.25 to 1.61)	0.007
11y-adult	2.26 (1.41 to 3.12)	<0.001	0.78 (-0.30 to 1.85)	0.1	0.03	1.65 (0.99 to 2.32)	<0.001

SBP was positively associated with height growth during childhood, increasing by an average 0.93 mmHg for every 1 SD (Table 4.9). Length at birth and height growth during infancy and adolescence were not associated with SBP in adulthood.

Table 4.9: Association of height growth and systolic blood pressure in adulthood (unadjusted)

Outcome Variable: SBP (mmHg)							
Predictor	Males (n=688)		Females (n=505)		Inter-action	Pooled (n=1193)	
	β (95% CI)	p-value	β (95% CI)	p-value	p-value	β (95% CI)	p-value
Height (z score)							
Birth	0.48 (-0.39 to 1.34)	0.2	-0.54 (-1.56 to 0.48)	0.2	0.1	0.05 (-0.61 to 0.71)	0.8
birth-6m	0.75 (-0.12 to 1.61)	0.08	0.28 (-0.68 to 1.25)	0.5	0.4	0.55 (-0.10 to 1.19)	0.09
6m-2y	1.14 (0.28 to 1.99)	0.009	-0.27 (-1.24 to 0.71)	0.5	0.03	0.54 (-0.11 to 1.18)	0.1
2y-11y	1.13 (0.28 to 1.97)	0.009	0.64 (-0.34 to 1.61)	0.2	0.4	0.93 (0.28 to 1.57)	0.005
11y-adult	0.50 (-0.35 to 1.35)	0.2	-0.48 (-1.46 to 0.49)	0.3	0.1	0.09 (-0.55 to 0.73)	0.7

SBP was positively associated with growth in BMI in childhood and adolescence (Table 4.10). SBP increased by an average 1.27 and 3.51 mmHg for each 1 SD increase in BMI in childhood and adolescence respectively. BMI at birth, and growth in early and late infancy were not associated with SBP in adulthood.

Table 4.10: Association of body mass index growth and systolic blood pressure in adulthood (unadjusted)

Outcome Variable: SBP (mmHg)							
Predictor	Males (n=682)		Females (n=500)		Inter-action	Pooled (n=1182)	
	β (95% CI)	p-value	β (95% CI)	p-value	p-value	β (95% CI)	p-value
BMI (z score)							
Birth	-0.55 (-1.35 to 0.26)	0.1	0.26 (-0.69 to 1.21)	0.5	0.2	-0.21 (-0.82 to 0.41)	0.5
birth-6m	0.56 (-0.24 to 1.36)	0.1	0.11 (-0.83 to 1.06)	0.8	0.4	0.37 (-0.24 to 0.98)	0.2
6m-2y	0.57 (-0.22 to 1.37)	0.1	0.06 (-0.88 to 1.00)	0.8	0.4	0.36 (-0.25 to 0.96)	0.2
2y-11y	1.27 (0.47 to 2.07)	0.002	1.24 (0.28 to 2.21)	0.01	0.9	1.27 (0.65 to 1.88)	<0.001
11y-adult	3.90 (3.10 to 4.71)	<0.001	2.96 (2.00 to 3.91)	<0.001	0.1	3.51 (2.89 to 4.12)	<0.001

I checked the interaction of sex with conditional measures of head, height and BMI which I now considered simultaneously as in the previous analysis. Only two out of the fifteen interactions were statistically significant, and the associations were not strong. Therefore, I present only the pooled results.

After adjustment for height and BMI growth, head growth ceased to be statistically significantly associated with SBP. Greater-than-expected growth in BMI during childhood, which was of borderline significance and the statistically significant result could be because of multiple testing and adolescence remained associated with higher adult SBP. SBP increased by an average of 0.88 and 3.34 mmHg for each 1 SD increase in conditional BMI growth during childhood and adolescence (Table 4.11). There was little change in these associations after further adjustment for gestational age (Figure 4.11).

Figure 4.9: Association between head growth and systolic blood pressure (adjusted and unadjusted)

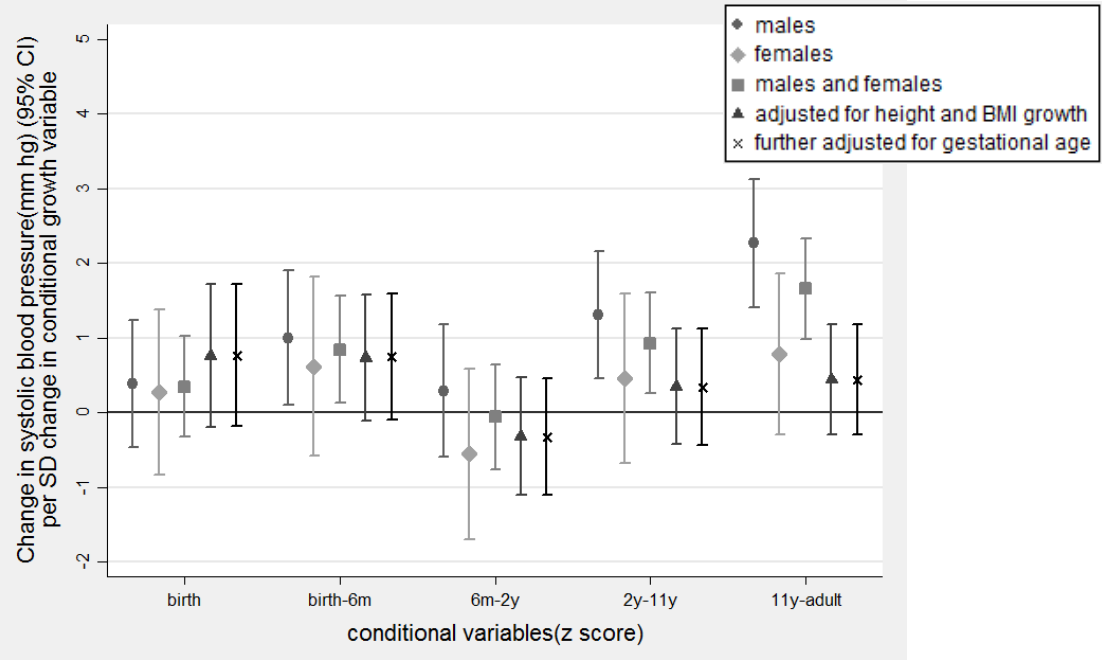


Figure 4.10: Association between height growth and systolic blood pressure (adjusted and unadjusted)

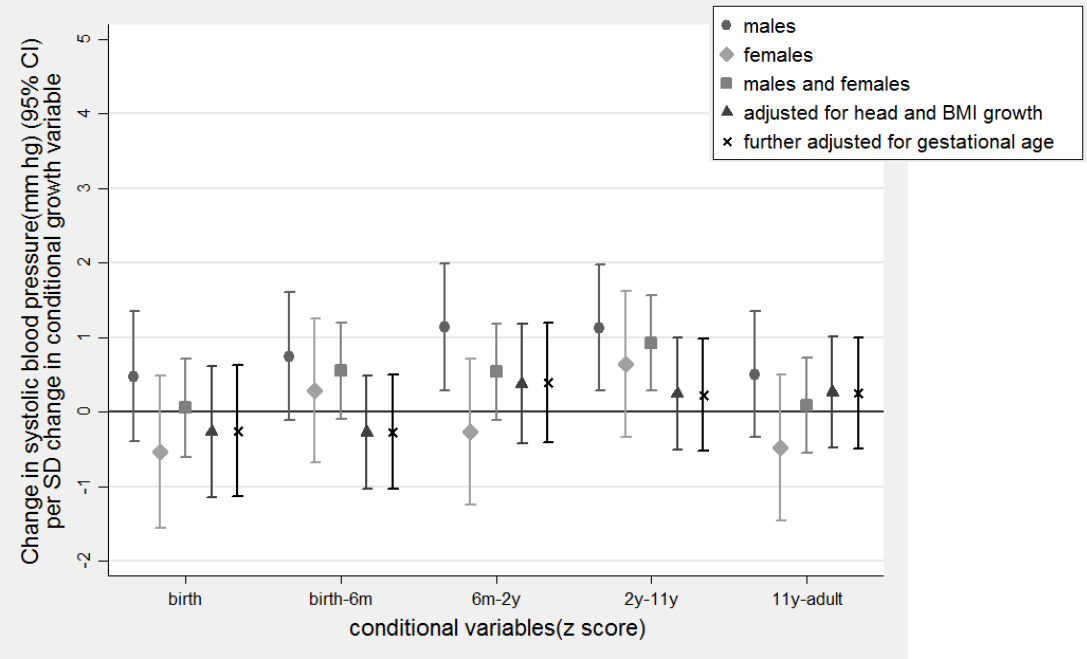
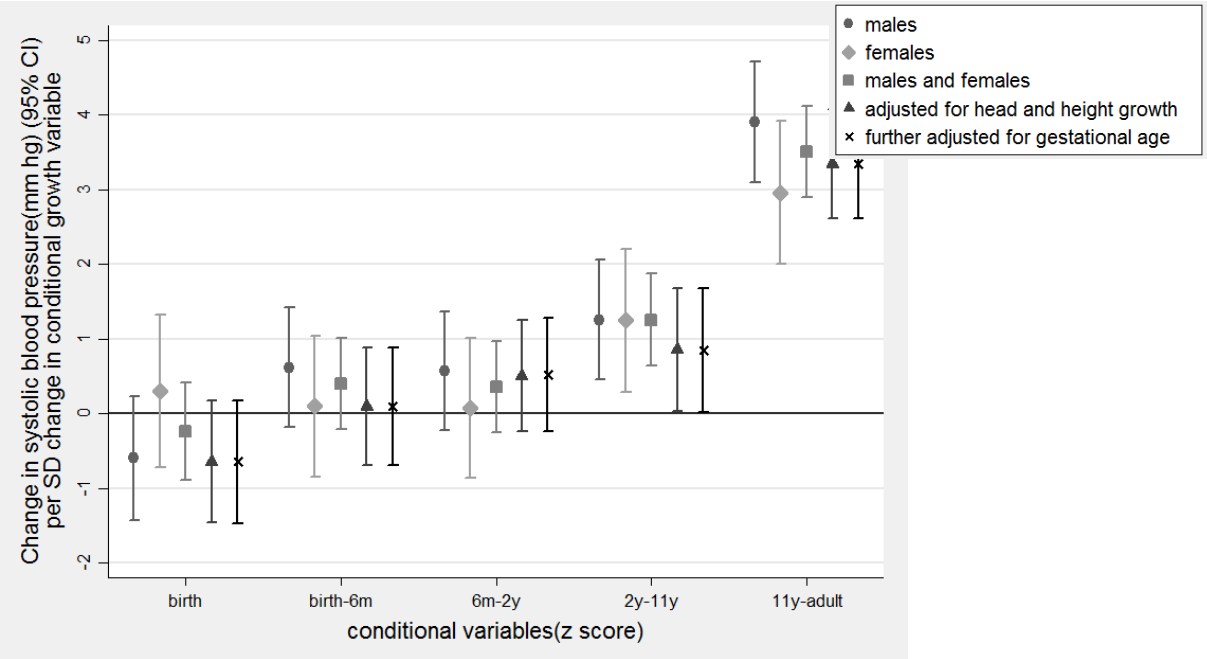


Figure 4.11: Association between body mass index growth and systolic blood pressure (adjusted and unadjusted)



There was a difference in the magnitude of association between head size at birth and later life SBP after adjusting for birth length and BMI. After adjustment, the strength of the positive association between head size at birth and later life SBP increased (from 0.35 to 0.75 mmHg), although it remained non-statistically significant.

Table 4.11: Association of head growth and systolic blood pressure in adulthood (mutually adjusted for gestational age)

	Outcome Variable: SBP (mmHg)			
Predictor	**Pooled (n=948)**		**Mutually adjusted for gestational age (n=948)**	
	β(95% CI)	**p-value**	**β(95% CI)**	**p-value**
Head (z score)				
Birth	0.75 (-0.20 to 1.70)	0.1	0.77 (-0.19 to 1.72)	0.1
birth-6m	0.74 (-0.11 to 1.58)	0.08	0.74 (-0.11 to 1.59)	0.08
6m-2y	-0.32 (-1.10 to 0.46)	0.4	-0.33 (-1.11 to 0.46)	0.4
2y-11y	0.35 (-0.43 to 1.12)	0.3	0.34 (-0.44 to 1.12)	0.3
11y-adult	0.44 (-0.30 to 1.18)	0.2	0.44 (-0.30 to 1.18)	0.2
Height (z score)				
Birth	-0.27 (-1.15 to 0.61)	0.5	-0.26 (-1.14 to 0.63)	0.5
birth-6m	-0.28 (-1.05 to 0.48)	0.4	-0.27 (-1.03 to 0.49)	0.4
6m-2y	0.38 (-0.42 to 1.18)	0.3	0.39 (-0.42 to 1.19)	0.3
2y-11y	0.24 (-0.51 to 0.99)	0.5	0.23 (-0.53 to 0.98)	0.5
11y-adult	0.27 (-0.48 to 1.01)	0.4	0.25 (-0.49 to 0.99)	0.5
BMI (z score)				

Birth	-0.59 (-1.36 to 0.19)	0.1	-0.65 (-1.47 to 0.18)	0.1
birth-6m	0.09 (-0.70 to 0.87)	0.8	0.09 (-0.70 to 0.88)	0.8
6m-2y	0.50 (-0.25 to 1.25)	0.1	0.52 (-0.24 to 1.27)	0.1
2y-11y	0.88 (0.05 to 1.70)	0.03	0.84 (0.02 to 1.67)	0.04
11y-adult	3.34 (2.61 to 4.07)	<0.001	3.35 (2.62 to 4.08)	<0.001
Gestational age	-	-	-0.10 (-0.87 to 0.68)	0.8
Missing	-	-	0.50 (-1.98 to 2.98)	0.6

4.3.2 Disproportion at birth and its association with blood pressure in adulthood

One of the aims of the analysis of association of body size with SBP and years of education in adulthood was to assess whether disproportion between head and height at birth was related to blood pressure and years of education. For assessing this specifically, I constructed five regression models all adjusted for sex, one including head size at birth only, the second including birth length only, the third including both together, the fourth including their interaction and the fifth model in which I considered the head/length ratio which is another method for assessing disproportion, with the outcome as adult SBP (Table 4.12).

Table 4.12: Associations of head and length at birth with adult systolic blood pressure and years of education

Predictor	**β (95% CI)**	**p-value**
Outcome: SBP		
Model 1 (n=1149)		
Female (1=y,0=n)	-11.66 (12.83 to -10.49)	<0.001
Birth head circumference (z)	0.43 (-0.15 to 1.00)	0.1
Model 2 (n=1397)		
Female	-11.57 (-12.76 to -10.38)	<0.001
Birth length (z)	0.15 (-0.46 to 0.76)	0.6
Model 3 (n=1384)		
Female	-11.62 (-12.82 to -10.42)	<0.001
Birth head circumference (z)	0.40 (-0.33 to 1.13)	0.2
Birth length (z)	-0.02 (-0.79 to 0.75)	0.9
Model 4 (n=1384)		

Female	-11.62 (-12.82 to -10.42)	<0.001
Birth head circumference (z)	0.40 (-0.33 to 1.13)	0.2
Birth length (z)	-0.02 (-0.79 to 0.75)	0.9
Birth head circumference*Birth length	0.01 (-0.47 to 0.49)	0.9
Model 5 (n=1384)		
Female	-11.62 (12.82 to 10.42)	<0.001
Birth Head/length ratio (z)	0.13 (-0.46 to 0.72)	0.6
Outcome: Years of education		
Model 6 (n=1463)		
Female	0.80 (0.46 to 1.14)	<0.001
Birth head circumference (z)	0.32 (0.15 to 0.49)	<0.001
Model 7 (n=1410)		
Female	0.80 (0.46 to 1.14)	<0.001
Birth length (z)	0.32 (0.14 to 0.50)	<0.001
Model 8 (n=1397)		
Female	0.73 (0.38 to 1.07)	<0.001
Birth head circumference (z)	0.18 (-0.03 to 0.39)	0.09
Birth length (z)	0.22 (-0.00 to 0.44)	0.05
Model 9 (n=1397)		
Female	0.73 (0.38 to 1.07)	<0.001
Birth head circumference (z)	0.18 (-0.03 to 0.39)	0.09
Birth length (z)	0.22 (-0.00 to 0.44)	0.05
Birth head circumference*Birth length	-0.03 (-0.17 to 0.11)	0.6
Model 10 (n=1397)		
Female	0.72 (0.38 to 1.07)	<0.001
Birth Head/length ratio (z)	-0.07 (-0.24 to 0.10)	0.4

If a disproportionately large head size relative to length at birth is associated with later elevated blood pressure, I would expect the head coefficient to increase in the height-adjusted analysis as compared to the unadjusted analysis. This was not the case; the head coefficient reduced in the adjusted analysis (Model 3). This was confirmed by an

alternative method of analysis which included the head/length ratio (Model 5). The interaction between head size at birth and birth length was also not statistically significant (p-value=0.9), strengthening the conclusion that disproportion at birth is not related to adult blood pressure in this population. There was also no association with years of education (Models 6-10) (Table 4.12).

Discussion

Years of education

Using the unadjusted conditional growth modelling approach, I found head size at birth and head growth up to 2 years to be positively associated with years of education. Similarly, I found positive associations of birth length and height growth during infancy, and BMI at birth and all ages up to 11 years, with years of education. However, after adjusting for the effects of height and BMI growth there were no significant associations between head size at birth and head growth at any age, and years of education. Greater height growth up to 2 years and greater BMI growth from 6 months to 2 years remained associated with increased years of education in adulthood. These associations became considerably attenuated after adjustment for SES, although both height and BMI growth from 6 months to 2 years remained significantly positively associated with attained education. SES was independently associated with years of education. Forward stepwise regression was used to identify the components of SES associated independently with educational attainment. Paternal occupation, paternal education, maternal education, assets (wealth) and crowding index were components of SES associated independently with years of education. Since height and BMI growth from 6 months to 2 years remained associated with education, independent of SES, we can say that infant growth is related to attained education, and that overall body growth was important rather than head growth specifically. There was no evidence that disproportion at birth was associated with years of education in later life.

The present analysis associated head size at birth and head growth during the life course with adult educational attainment, which I used as a proxy for cognitive ability. In the past, few studies have looked at the association of head size at birth with cognitive function in later life, and the results are inconsistent. Martyn and colleagues in their cohort of 1,576 men and women born in Hertfordshire, Preston or Sheffield between 1920 and 1943, found no association between head size at birth, head/length

and head/abdominal circumference ratios and intelligence in adulthood (43). These associations were adjusted for social class at birth. However, a study in a Finnish population showed a positive association between head size at birth and verbal ($p=0.03$), visuospatial ($p=0.04$) and arithmetic ability ($p=0.002$) at 20 years (44). These associations were adjusted for father's occupational status in childhood of the subjects. Similar findings were shown in Southern India by Veena et al in 2010 who found a positive association between head size at birth and learning ability at 9-10 years (45). Learning ability rose by 0.1 SD (95% CI: 0.04 to 0.23) per SD increase in head size at birth. This association was reduced after adjustment for current head circumference. These associations were adjusted for parent's educational level in completed years and current SES. However, studies which have looked at the association between head growth at different ages and cognitive function in later life have all concluded that head growth early in life, up to infancy is positively associated with later life intelligence (46-50). In a study by Gale et al in 2006 in the UK, head growth during infancy was positively associated with IQ in childhood. IQ increased by an average of 1.56 points (95% CI: 0.11 to 3.01 points) for every 1-SD increase in head growth between birth and 1 year (46). In another study by Gale et al in 2004, also conducted in the UK IQ at age 9 years increased by 1.98 points (95% CI:0.34 to 3.62) for every SD increase in head circumference at age 9 months (47). All associations of head growth were adjusted for mother's education and father's occupation, similar to the present study.

There could be several possible reasons why I did not find an independent association between head growth and years of education in adult life. One reason could be that I have used years of attained education as a proxy measure of cognitive function. This can be justified in one way, as previous literature has shown a positive association between education and cognitive function. A study of English children showed the correlation between intelligence at age 11 years and educational achievement in 25 academic subjects at age 16 years in 70,000 children to be 0.81 (100). However, years of education may not correlate well with cognitive function in the Indian context. This may be because education in India is very much dependent on SES and education status of the parents. If the parents are of high or middle SES, the children are more likely to be kept in schools and receive longer education. However, if the parents are from a low socioeconomic background the children are less likely to get removed from school early, even if they are intelligent. This may be because families are often large and the

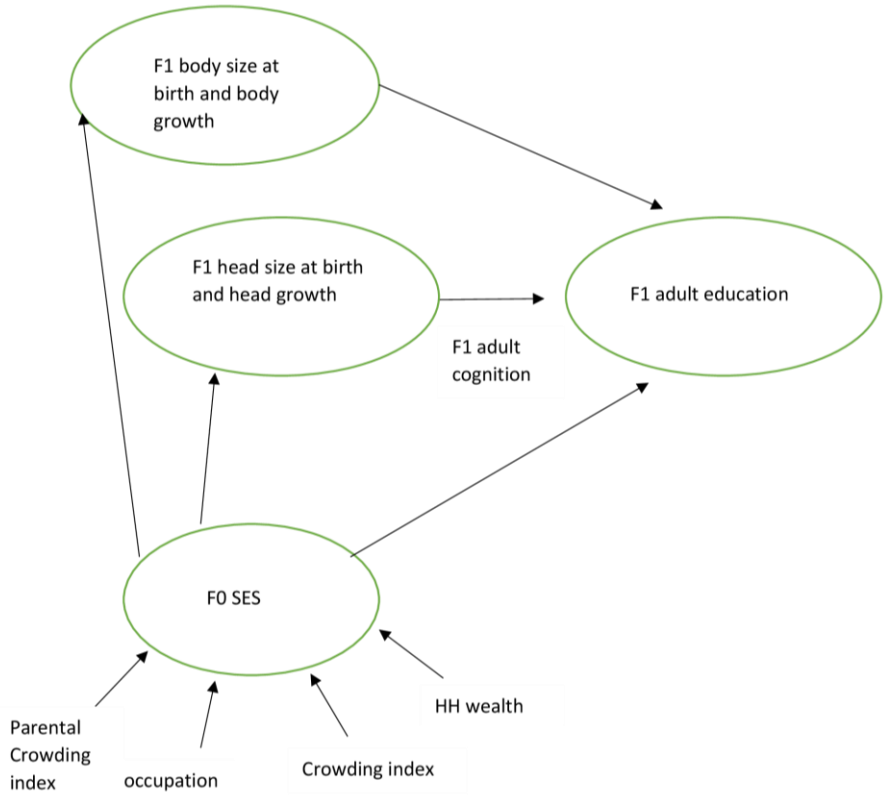
parents are usually engaged in low-paying jobs, as they themselves have received little or no education. Thus the household income is very limited and thus, the parents view the children as helping hands in the family who take care of the younger members in the family while the parents are away at work and do not educate them. Thus, education in India might be heavily dependent on socioeconomic conditions and may not be a good representation of an individual's cognitive ability. The studies mentioned earlier have used better measures of cognitive function. In Martyn's study, cognitive function was measured using the AH4 intelligence quotient (IQ test) and Mill Hill Vocabulary test, which has earlier been used in several community surveys of older people (43). In the Finland study, cognitive test scores were obtained from the Finnish forces basic ability test (42) The Indian study assessed cognitive function using standard neurophysiological tests applicable for use in school-aged children (44). The Avon longitudinal study of parents and children assessed cognitive function at age 4 years using the Wechsler Intelligence scale for children (47).

Another possible reason for the lack of association in my study between head size and attained education is that brain size is an inadequate proxy for brain development. Previous studies have shown a strong correlation between head circumference and brain volume. In newborns and children, head circumference is well correlated with brain size as measured by neuroimaging, eg :- MRI (101). Head circumference was an excellent predictor of brain volume in 1.7 to 6 years old children ($r = 0.93$). Since it is easy to measure head circumference, this is still the most frequently used surrogate measurement of overall brain development (102,103). Therefore head size is a good marker for brain size, but brain size may not be a good marker for brain function (103).

A further reason for my negative findings could be the high attrition rate in the cohort. Attrition can be associated with both bias and imprecision of the study results. The number of subjects on whom the outcome was measured was 1526, which is 19% of the original cohort. And the number with outcome and complete body measurements was 958 (13%). There were no statistically significant differences between head size at birth, 6 months, 2 years or 11 years between the final sample and the original cohort. Nor was there a difference in SES variables at birth (maternal education (p-value=0.7), assets (wealth) (p-value=0.9), utilities (p-value=0.1), sanitation (p-value=0.2), water supply (p-value=0.6) and income (raw) (p-value=0.1)) in these 958 individuals

measured till adulthood and the rest of the sample. Thus the sample selected appears representative of the entire set of body measurements, and does not appear a biased sample. Also, since the confidence intervals for the associations between head size at birth and head growth are narrow, the estimates are precise even after attrition. Another issue in these associations is that adjusting for SES might seem to be an overadjustment in the analysis. To be considered a confounder in an epidemiological association, a variable must satisfy certain conditions. Firstly, the confounding variable must have an association with the outcome and exposure. Previous literature confirms that SES is associated with both head size at birth and growth, which is the exposure and adult years of education, which is the outcome in this case (Figure 4.12).

Figure 4.12: Association of F1 head size at birth and growth with F1 adult cognition considering potential confounding by F0 socioeconomic status



The association between height growth up to 2 years and BMI growth between 6 months to 2 years support other studies involving five birth cohorts in Brazil, Guatemala, India, Philippines and South Africa which have shown an association between growth in height and weight gain up to age 2 years and increased number of

years of attained schooling in later life (27,39-41). An underlying reason could be better nutrition in early postnatal life, which might be associated with better brain development and therefore increased cognitive ability in adult life. A reason for the potential null findings with head growth after controlling for height and BMI could be a failure to adjust for other potential variables like mother's height, parity and year of intellect testing as they were not collected in the study. Previous research has cited these variables as independent risk factors for slower growth and/or poorer development of intellectual abilities (19). However, this could have also moved the results more towards the null value.

SES was found to be strongly related with attained education. This could be explained as better SES in childhood probably allows for more and better educational opportunities particularly with the Indian context. Further better SES of the household might mean greater exposure to cognitive stimulation in the home environment (through toys, play and language). A better cognitive environment in early life may be associated with later life educational success because it contributes to better reading ability and verbal skills in childhood and adolescence (104). These individuals are likely to also use these skills in adulthood. Early exposure to a cognitive simulating environment may enhance information processing skills and other cognitive abilities and may provide or increase the motivation to succeed in school and education (105). A limitation in the analysis is the use of stepwise regression methods for identifying the individual SES components associated with adult years of education. These methods involve developing a sequence of linear models in which the independent variables are entered one at a time. This approach contains some problems (106). Firstly, using this approach results in using incorrect degrees of freedom. Using incorrect degrees of freedom results in statistical significance becoming larger than expected. Most of the statistical software packages do not correctly calculate the degrees of freedom in a stepwise regression analysis. This is because the stepwise regression applies an F-test to the sum of squares at each stage in the regression models. Thus the analyst performs multiple statistical significance tests on the same dataset, ignoring the previous tests which can have consequences on the inferences. Multiple significance tests have been carried out in the present stepwise regression analysis which aimed at identifying the various individual variables associated with the attenuation in the associations of head size at birth and growth with adult cognitive ability. The issue discussed above may

have incorrectly chosen some variables purely based on inflated statistical significance. A second issue with the stepwise method is its procedure to identify the best predictor set. The best set of SES variables might not have been selected using this procedure in the present analyses as the results are dependent on the sampling error present in any given sample and thus can sometimes give incorrect conclusions. Changing the entry and exit probabilities may select a different set of variables. Therefore, the procedure of variable selection is purely based on statistical significance and does not consider the biological significance of the variables. A third issue in the use of stepwise regression methods is with regards to the replicability of the conclusions. In these methods, as variable selection are made at each step, there may be cases where one variable is chosen over another due to a small difference in their predictive abilities. This might not be the case in the present analysis as well as variables such as sanitation, individual income and child dependency ratio had similar predictive abilities, which can be seen with the similar changes in F-statistic in the stepwise model. These variables were not selected in the final model. Sampling errors may be of less concern when predictor variables are few and with a large sample size, which was the case in the present analysis. Another disadvantage is that multiple significance tests have been carried out in these analysis and thus some of the associations might have been statistically significant due to chance. Associations which were strongly statistically significant, still might not be of substantial clinical relevance since the effect sizes were small.

The hypothesis was that early life growth failure could be a marker of malnutrition that can affect neurological development leading to lowered adult educational attainment. Our analysis is suggesting towards that direction, as greater body growth and linear growth up to 2 years is associated with increased years of education in adulthood. Though further research is required in Indian population with better follow up data and improved measure of cognition, we can infer that improving childhood nutrition and promoting linear growth up to age 2 years may be important for higher cognitive development in adulthood. Improved SES conditions can also impact attainment of education significantly by enhancing information processing skills which may provide or increase the motivation to succeed in school and education.

Systolic blood pressure

In the unadjusted analysis I found that head size at birth was not associated with SBP in adult life. An increase in head growth during early infancy, childhood and adolescence was associated with higher SBP in adult life. I found positive associations of height growth in childhood, and BMI growth during childhood and adolescence, with SBP in adult life. The magnitude of the association for BMI growth during adolescence was considerably greater than that for growth at earlier ages. After adjusting for height and BMI growth, there were no statistically significant associations between head growth at any age and SBP. I did not find an independent association between SBP and SES (p-value=0.1), therefore SES was not considered in the adjusted model. There was no significant association between height growth at any age and SBP in the adjusted model. Greater than average BMI growth after 2 years was associated with higher SBP in adulthood. The strength and direction of these associations changed little after adjusting further for gestational age. There was no evidence that disproportion at birth was associated with higher adult blood pressure, since I would expect the head coefficient to increase after birth length adjustment. The interaction between head size at birth and birth length was also not statistically significant (p-value=0.9), strengthening this conclusion.

According to the fetal programming theory by Barker, the fetus uses the scarce nutrition provided for the development of the most important parts of the body such as the brain and the head, and impairs the development of the rest of the body, such as the trunk and muscles. This has been described in detail in Chapter-1. There are only a few studies which have associated head size and growth or ratios of head to overall body size, with respect to later life outcomes in individuals. This may be because head circumference is not a common measure of body size at birth, the most common being birth weight. The earliest findings of an association between head size and later life outcomes came from studies in Southampton, UK which showed that the blood pressure of 327 men and women aged 46 to 54 years in Preston, was inversely associated with head circumference at birth (31). Mean systolic pressure rose by 14 mm Hg as the head circumference to body length ratio increased from less than 0.65 to greater than or equal to 0.7. A limited number of studies have looked at the association of head size at birth and head growth with adult SBP, and the results are inconsistent. Very few studies have looked at the association between disproportion between head size at birth and SBP in

adulthood (107,108). These studies showed that SBP was higher in people who were short or who had small abdominal or head circumference at birth. Hypertension is a known risk factor for CVD (109). Studies exploring associations between head size at birth and later life CVD risk have also been inconsistent. There were no significant trends found with CHD in a Finnish population (25). However, a population-based study in Norway showed an inverse association between head circumference at birth and death from CHD in later life (33). The absence of relation between the birth measurements, in particular head circumference and SBP in adulthood provides no support for the possibility that any specific period of intrauterine growth is of importance in this population.

Our data showed that accelerated BMI gain in childhood and adolescence was associated with increased adult SBP. This is consistent with previous literature which has shown that greater than average BMI gain after age 2 is associated with increased adult SBP (24,110,111). These results suggest that efforts to prevent obesity-related diseases should start in childhood and should target not only children who are overweight or obese in infancy, but also children with a normal weight, but having rapid weight gain after infancy. I did not find an association between birth weight and birth length and blood pressure in adulthood. This finding is consistent with previous literature. Meta-analysis of the association of birth weight on blood pressure in adulthood have been conducted, concluding that there was no clear association (112). The association between birth length and later life SBP is also inconclusive. Several studies from high-income countries, have either failed to report an association (109,113, 114) or have found inverse associations (24,110,111). In this study blood pressure was recorded using standardized procedures by trained research team. Another strength of this analysis is that we are looking at the association of head size and growth after adjusting for height and BMI measurements which has been looked at in very few studies till date. A potential disadvantage of this analysis is that multiple significance tests have been carried out in these analysis and thus some of the associations might have been statistically significant due to chance. These associations which were strongly statistically significant, might not have substantial clinical relevance since their effect sizes were small.

The exploratory analysis suggests that overall body growth, rather than specifically head growth is related to elevated adult SBP. Becoming a larger child or adolescent might be a risk factor for developing higher adult blood pressure. Though further research is required in the Indian population in order to replicate the findings, we can infer that improved childhood nutrition and preventing adiposity during childhood and adolescence might be important for preventing higher blood pressure in adulthood

Chapter-5

Association of head growth with birth weight in the next generation

I found no studies of the association between head growth in early life and outcomes in the next generation. I used the conditional body size approach to measure associations between head growth in the F1 generation of the NDBC and birth weight in the F2 generation. I further adjusted these associations for SES of the parents of the cohort members (F0 generation) as slow postnatal growth in body measures such as head circumference reflects adverse economic conditions in childhood, and so SES is a potential confounding factor (82).

In the NDBC data was collected from three generations (F0, F1, and F2). Thus a particular cohort member will have one or more members in the subsequent generation (Table 5.1). Traditional methods assume that there is only one observation per individual. Thus there is a need to shift from traditional methods of analysing data to more sophisticated and complex methods which incorporate this dependency in the sample estimates. A regression technique that has been developed to handle the correlation among these repeated measurements is known as linear mixed regression model. This technique is also known as ‘multilevel modelling’ or ‘hierarchical linear modelling’ (62). The technique has been described previously in Chapter-3 (Page 72). Here I will use this technique to study associations between F1 head size and growth with F2 birth weight.

There were no marriages within the cohort, thus the mothers and fathers were from separate families. We also identified 15 pairs of twins having a valid birth weights. All subsequent analysis has been conducted only for singletons having a valid birth weight, since the birth weight of twins tends to be considerably lower than that of singletons (115). There were a total of 1,385 singletons born to 800 parents of the F1 generation (Table 5.1).

Table 5.1: Number of parents having specific combination of children

Pattern of F2 children	Number of F1 fathers (n=447)	Number of F1 mothers (n=353)	Number of F2 sons (n=774)	Number of F2 daughters (n=611)
1 boy	80	73	153	0
1 girl	61	51	0	112
2 boys	88	48	272	0
1 boy 1 girl	154	134	288	288
2 girls	41	28	0	138
3 boys	3	2	15	0
3 girls	2	1	0	9
2 boys 1 girl	3	3	12	6
1 boy 2 girls	12	10	22	44
2 boys 2 girls	1	2	6	6
1 boy 3 girls	0	1	1	3
2 boys 3 girls	1	0	2	3
3 boys 2 girls	1	0	3	2

5.1 Details of the exposure variables

Table 5.2 describes the data collected from each of the generation that will be discussed in this chapter.

Table 5.2: Details of outcomes and exposures collected in different generations

Generation	Outcome	Exposure/s
F0		SES*
F1		**Head, Height and BMI measurements** At birth, birth-6m,6m-2y,2y-11y,11y-adult, Sex
F2	Birth weight	Birth order, sex

*The SES variables have been described in detail in chapter-3.

Table 5.3 presents the parent-wise distribution of body growth variables which have been considered as exposure variables used in the analysis. Except at age 11 years, the body size measures for the fathers were greater than those for the mothers.

Birth order: A variable for birth order was created which ranked the F2 children in the ascending order of their birth. Mean F2 birth weight was tabulated for the different categories of birth order. Birth order was linearly and positively associated with F2 birth weight and was used as a continuous variable in all analyses.

Table 5.3 (a): Distribution of F1 measurements

Parent-level variables	**N**	**Fathers**	**N**	**Mothers**	**p-value**
Body growth variables		**Mean±SD**		**Mean±SD**	
At birth					
Head(cms)	431	33.8±1.3	338	32.3±1.1	<0.001
Height(cms)	427	48.2±2.2	339	48.2±1.9	0.008
BMI(kg/m^2)	400	12.1±1.3	289	11.9±1.2	0.05
At age 6 months					
Head(cms)	439	42.2±1.2	348	41.0±1.1	<0.001
Height(cms)	422	65.3±2.4	341	63.9±2.3	<0.001
BMI	411	16.5±1.5	296	15.8±1.6	<0.001
At age 2 years					
Head	425	46.9±1.3	334	45.7±1.2	<0.001
Height	411	81.2±3.6	328	79.7±3.6	<0.001
BMI	410	15.8±1.3	294	15.4±1.2	0.0001
At age 11 years					
Head	426	52.1±1.4	336	52.2±1.4	0.2
Height	425	135.2±6.3	335	135.9±6.4	0.1
BMI	409	15.3±1.6	295	15.1±1.7	0.1
At adult stage					
Head	429	56.6±1.8	308	54.0±1.7	<0.001
Height	431	169.7±6.3	307	155.3±5.5	<0.001
BMI	431	25.0±4.1	307	24.7±4.9	0.6

Table 5.3 (b): Distribution of F2 measurements

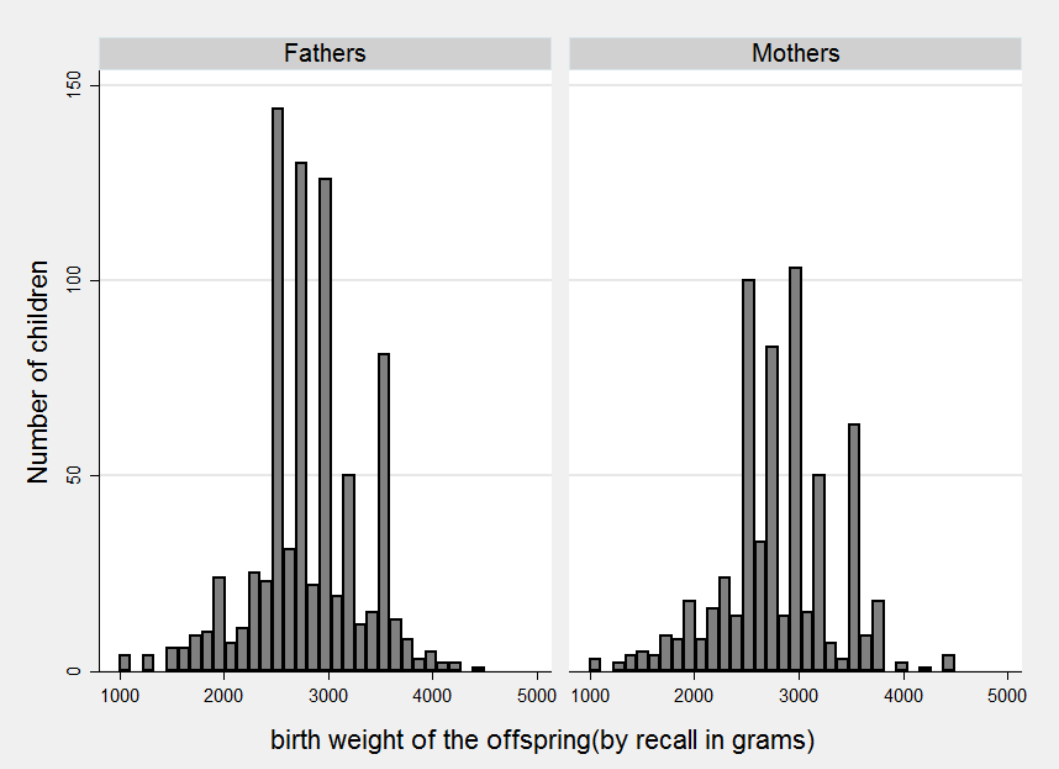
Child-level variables	N	Sex=Boys	n	Sex=Girls	p-value
Birth order					
Fathers n (%)					
1		233(52.5)		185(54.9)	0.2
2		186(41.9)		141(41.8)	
≥3		28(5.6)		11(3.3)	
Mothers n (%)					
1		168(50.9)		154(56.2)	0.06
2		139(42.1)		112(40.9)	
≥3		23(7.0)		8(2.9)	
Birth weight (gms)	774	2846±524	611	2786±499	0.003

SES score: Several socioeconomic variables were collected for the F0 generation in the study. In chapter 3, I described the principal component analysis method used for creating a composite socioeconomic score for every F1 mother and father. I will adjust for this in exploring the associations between body growth and birth weight in the next generation.

5.2 Details of the outcome variable

F2 birth weight was symmetrically distributed for offspring from both F1 mothers and fathers, so could be used as an outcome measure in the subsequent multilevel analyses in its original form (Figure 5.1). Both the histograms also show evidence of rounding which could be due to the birth weights being collected by recall.

Figure 5.1: Distribution of F2 birth weight



5.3 Analytic approach

The first step in the multilevel analysis was to consider if the data justified the decision to use multilevel models for the analysis. For this I estimated the clustering of the data. The ICC accounts for clustering in the data (62). The ICC was similar for the null model for the fathers and mothers 0.36 (95% CI: 0.27 to 0.45) and 0.36 (95% CI: 0.26 to 0.47).

As an initial step, unadjusted associations of head growth of the F1 generation at different time points (birth, birth to 6 months, 6 months to 2 years, 2 years to 11 years and 11 years to adult) and birth weight by sex of the F2 generation were analysed. This was done separately for both sexes of the F1 generation, since associations of body growth with outcomes in the next generation are different for mothers and fathers, as has been described in detail in Chapter-1 (page 13).

These models for head circumference were also tested for the presence of random slopes. Two out of 20 likelihood ratio tests were statistically significant, which were association of head growth between 6m-2y for F1 fathers and head growth

between 11y-adult for F2 sons with F2 birth weight. These could be due to multiple testing.

This was followed by constructing formal interaction tests between sex of the F2 generation and different head growth measures of the F1 generation. A similar procedure was followed for both the sexes of the F2 generation. Since, I did not find any statistically significant interactions of the body measures with the F1 or F2 sex, the body associations were pooled for both F1 and F2 sexes in a final unadjusted model for head growth. Models for studying associations between F1 height and BMI growth in the different time periods and F2 birth weight were constructed using the same method. Finally, an adjusted model was constructed which pooled data from both F1 and F2 generation, and which also adjusted for height and BMI growth simultaneously. These associations were further adjusted for F1 gestational age, F2 birth order and SES of the F0 generation.

5.4 Association of the critical periods of body growth with birth weight in the next generation

Table 5.4 presents the unadjusted analysis showing the associations between birth weight in the F2 generation and F1 head growth variables considered simultaneously. In the unadjusted final pooled model, which combined the data for both F1 and F2 generation, birth weight in the F2 generation was positively associated with head growth during early infancy (Figure 5.2). F2 birth weight was positively associated with F1 head size at birth (32.7 gms (95% CI: 0.3 to 65.0)). Birth weight increased by 66.7 grams for each 1 SD increase in head growth during early infancy. Greater-than expected head growth during childhood and adolescence were not associated with the birth weight in the F2 generation.

Figure 5.2: Association between F1 head growth and F2 birth weight (adjusted and unadjusted)

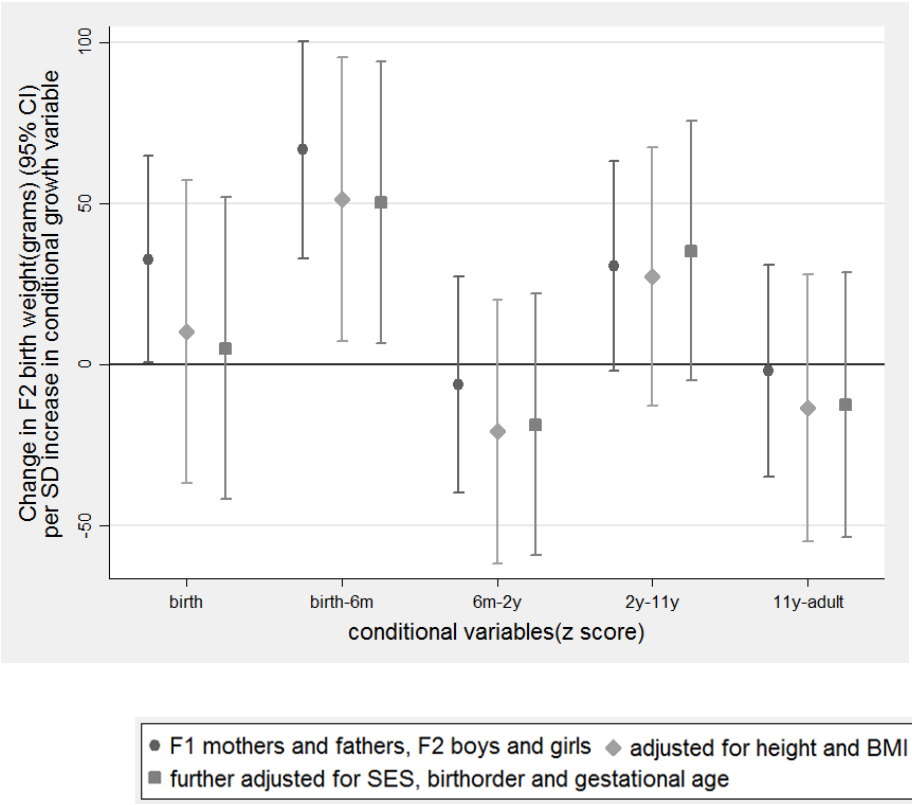


Table 5.4: Association of parental head growth and birth weight of children (unadjusted)

Outcome Variable: Birth weight (grams)							
Predictor	**Boys (n=n_2)**		**Girls (n=n_3)**		**Inter-action**	**Pooled (n=n_5)**	
	β(95% CI)	**p-value**	**β(95% CI)**	**p-valu**	**p-value**	**β(95% CI)**	**p-value**
Fathers (n=n_1) (n_1=286, n_2=366)			**(n_1=236, n_3=288)**			**(n_1=375, n_5=654)**	
Head (z score)							
Birth	27.1 (-23.3 to 77.4)	0.2	20.2 (-36.6 to 77.0)	0.4	0.9	22.3 (-18.3 to 62.9)	0.2
birth-6m	125.8 (71.7 to 179.9)	<0.001	26.6 (-35.5 to 88.8)	0.3	0.008	87.2 (43.4 to 131.0)	<0.001
6m-2y	-20.9 (-77.7 to 35.9)	0.4	-31.9 (-96.6 to 32.9)	0.3	0.5	-31.2 (-77.1 to 14.8)	0.1
2y-11y	4.6 (-50.7 to 59.8)	0.8	4.6 (-50.7 to 59.8)	0.8	0.09	25.8 (-17.4 to 69.0)	0.2
11y-adult	-20.6 (-73.5 to 32.2)	0.4	-20.6 (-73.5 to 32.2)	0.4	0.9	-27.2 (-70.2 to 15.8)	0.2
Mothers (n=n_4)(n_4=209, n_2=255)			**(n_4=172, n_3=206)**			**(n_4=269, n_3=461)**	
Head (z score)							
Birth	86.6 (9.2 to 164.1)	0.02	12.7 (-47.8 to 73.1)	0.6	0.1	51.1 (-2.8 to 104.9)	0.06
birth-6m	53.0 (-19.7 to 125.7)	0.1	21.2 (-46.8 to 89.3)	0.5	0.2	39.1 (-15.7 to 93.8)	0.1
6m-2y	18.2 (-51.8 to 88.2)	0.6	20.3 (-41.0 to 81.6)	0.5	0.9	21.5 (-29.2 to 72.1)	0.4
2y-11y	71.8 (1.2 to 142.5)	0.04	2.0 (-58.1 to 62.2)	0.9	0.1	36.5 (-14.2 to 87.2)	0.1
11y-adult	66.1 (-6.6 to 138.8)	0.07	-5.7 (-66.8 to 55.4)	0.8	0.07	35.6 (-17.0 to 88.2)	0.1
Interaction							
Birth	0.1		0.8			0.3	
birth-6m	0.1		0.9			0.1	
6m-2y	0.3		0.2			0.1	
2y-11y	0.1		0.1			0.7	
11y-adult	0.05		0.5			0.05	
Pooled (n=n_6) (n_2=495, n_6=621)			**(n_3=408, n_6=494)**			**(n_5=644, n_6=1115)**	
Birth	45.8 (2.9 to 88.7)	0.03	17.2 (-24.2 to 58.7)	0.4	0.3	32.7 (0.3 to 65.0)	0.04
birth-6m	95.4 (51.6 to 139.2)	<0.001	23.5 (-22.3 to 69.4)	0.3	0.005	66.7 (32.6 to 100.8)	<0.001
6m-2y	-1.4 (-45.7 to 43.0)	0.9	-7.8 (-52.7 to 37.1)	0.7	0.5	-6.3 (-40.1 to 27.6)	0.7
2y-11y	34.3 (-9.5 to 78.2)	0.1	35.2 (-6.7 to 77.0)	0.1	0.6	30.7 (-2.1 to 63.6)	0.06
11y-adult	13.8 (-29.5 to 57.0)	0.5	-17.6 (-61.1 to 25.8)	0.4	0.2	-1.9 (-35.1 to 31.4)	0.9

n_1= number of fathers, n_2= number of boys, n_3= number of girls, n_4=number of mothers, n_5=number of boys and girls, n_6=number of fathers and mothers

Table 5.5 presents the unadjusted analysis showing the associations between birth weight in the F2 generation and F1 height growth variables considered simultaneously (Figure 5.3). In the final pooled model, birth weight in the F2 generation was positively associated with birth length and height growth up to 2 years, conditional on height at all preceding time points. Birth weight increased by 37.9 grams for each 1SD increase in birth length and by 43.4 and 44.1 grams for every SD increase in height growth during early and late infancy. Greater-than expected height growth during childhood and

adolescence were not associated with the birth weight in the F2 generation individual. The association of F1 head growth between 2y-11y with F2 birth weight has a p-value of 0.06. This could be regarded as close to statistical significance, but it also could be due to multiple testing.

Figure 5.3: Association between F1 height growth and F2 birth weight (adjusted and unadjusted)

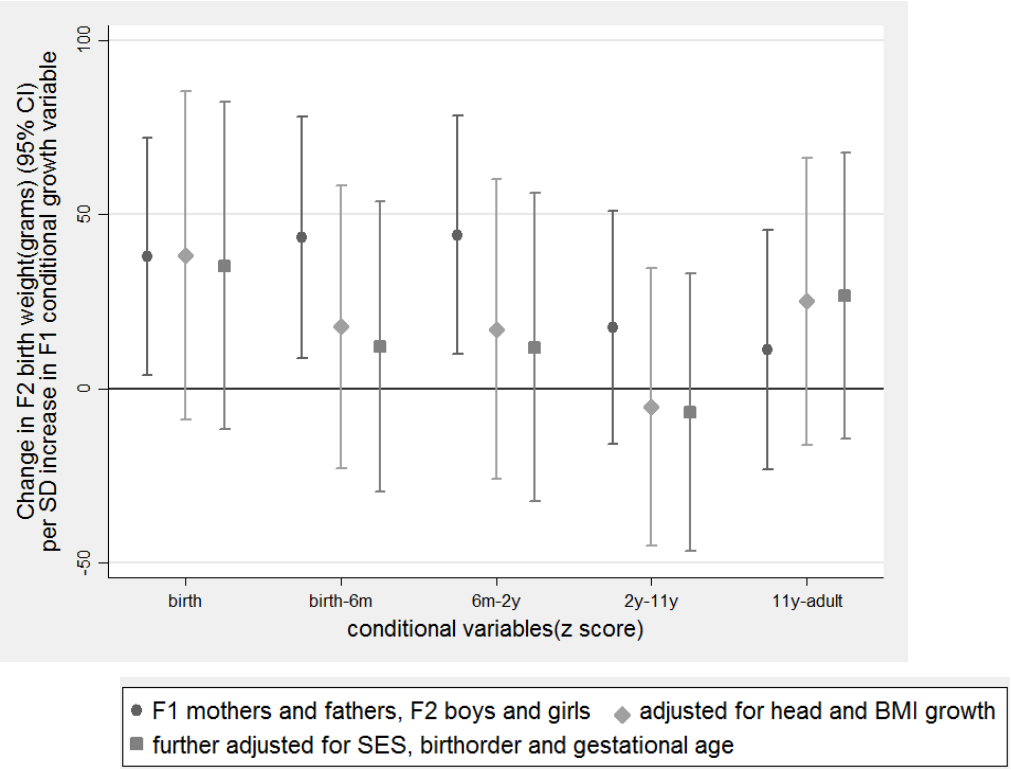


Figure 5.4: Association between F1 body mass index growth and F2 birth weight (adjusted and unadjusted)

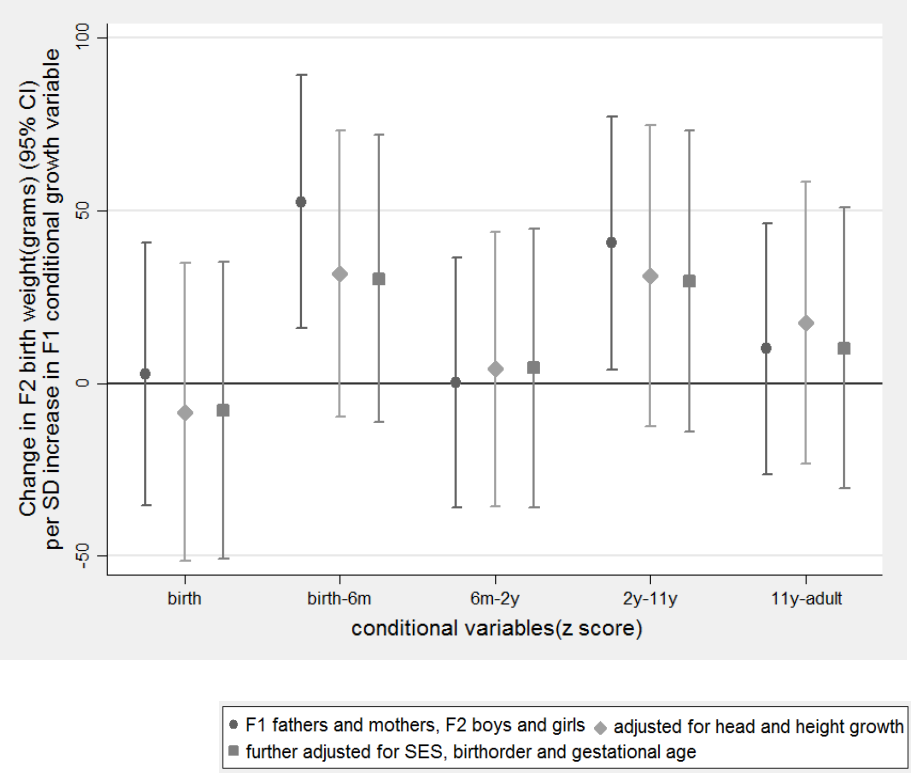


Table 5.5: Association of parental height growth and birth weight of children (unadjusted)

Predictor	Outcome Variable: Birth weight (grams)						
	Boys (n=n_2)		Girls (n=n_3)		Inter-action	Pooled (n=n_5)	
	β(95% CI)	p-value	β(95% CI)	p-value	p-value	β(95% CI)	p-value
Fathers (n=n_1)	**(n_1=276, n_2=356)**		**(n_1=222, n_3=270)**			**(n_1=358, n_5=626)**	
Height (z score)							
Birth	48.7 (-9.3 to 106.6)	0.09	81.0 (20.0 to 142.1)	0.01	0.3	61.3 (16.7 to 106.0)	0.007
birth-6m	62.7 (5.8 to 119.5)	0.03	16.2 (-47.5 to 79.8)	0.6	0.3	55.8 (10.2 to 101.5)	0.01
6m-2y	29.5 (-26.9 to 85.9)	0.3	31.4 (-29.4 to 92.2)	0.3	0.8	29.1 (-15.5 to 73.7)	0.2
2y-11y	-0.5 (-57.4 to 56.3)	0.9	56.1 (-3.3 to 115.6)	0.06	0.09	24.1 (-20.1 to 68.2)	0.2
11y-adult	5.9 (-50.0 to 61.8)	0.8	-31.1 (-94.2 to 32.0)	0.3	0.5	-9.4 (-54.4 to 35.6)	0.6
Mothers (n=n_4)	**(n_4=204, n_2=249)**		**(n_4=167, n_3=201)**			**(n_4=263, n_3=450)**	
Height (z score)							
Birth	12.1 (-63.7 to 87.8)	0.7	15.8 (-46.6 to 78.2)	0.6	0.7	5.1 (-49.0 to 59.2)	0.8
birth-6m	43.1 (-29.9 to 116.2)	0.2	4.4 (-63.0 to 71.8)	0.8	0.3	27.4 (-27.0 to 81.8)	0.3
6m-2y	102.8 (28.0 to177.6)	0.007	25.4 (-42.0 to 92.8)	0.4	0.06	65.9 (11.4 to 120.3)	0.01
2y-11y	45.5 (-30.0 to 120.9)	0.2	8.3 (-50.2 to 66.9)	0.7	0.5	6.8 (-45.7 to 59.3)	0.7
11y-adult	-8.6 (-86.1 to 68.9)	0.8	27.6 (-36.4 to 91.6)	0.3	0.5	38.7 (-15.9 to 93.2)	0.1
Interaction							

Birth	0.4		0.1		0.09		
birth-6m	0.6		0.8		0.4		
6m-2y	0.1		0.8		0.3		
2y-11y	0.8		0.2		0.5		
11y-adult	0.4		0.7		0.1		
Pooled (n=n_6)	**(n_2=480, n_6=605)**		**(n_3=389, n_6=471)**			**(n_5=621, n_6=1076)**	
Birth	34.2 (-11.9 to 80.2)	0.1	51.1 (7.3 to 94.8)	0.02	0.3	37.9 (3.6 to 72.3)	0.03
birth-6m	53.6 (8.8 to 98.5)	0.01	12.9 (-33.5 to 59.2)	0.5	0.2	43.4 (8.5 to 78.2)	0.01
6m-2y	60.3 (15.3 to 105.3)	0.009	27.6 (-17.4 to 72.7)	0.2	0.2	44.1 (9.7 to 78.5)	0.01
2y-11y	-1.4 (-47.3 to 44.5)	0.9	35.0 (-6.9 to 76.9)	0.1	0.1	17.6 (-16.1 to 51.3)	0.3
11y-adult	23.3 (-21.8 to 68.3)	0.3	-4.9 (-49.8 to 40.1)	0.8	0.4	11.2 (-23.3 to 45.8)	0.5

n_1= number of fathers, n_2= number of boys, n_3= number of girls, n_4=number of mothers, n_5=number of boys and girls, n_6=number of fathers and mothers

Table 5.6 presents the unadjusted analysis showing the associations between birth weight in the next generation and parental BMI at birth and growth variables considered simultaneously. In the final pooled model for both F1 and F2 generation, birth weight in the F2 generation was positively associated with BMI growth during early infancy and in childhood, conditional on BMI growth at all preceding time points. Birth weight in the F2 generation increased by 52.5 and 40.6 grams for each 1 SD increase in BMI growth during early infancy and childhood. No statistically significant associations were found for greater than expected BMI at birth and BMI growth in late infancy and adolescence and birth weight in the F2 generation.

Table 5.6: Association of parental body mass index growth and birth weight of children (unadjusted)

Outcome Variable: Birth weight (grams)							
Predictor	**Boys (n=n_2)**		**Girls (n=n_3)**		**Inter-action**	**Pooled (n=n_5)**	
	β(95% CI)	**p-value**	**β(95% CI)**	**p-value**	**p-value**	**β(95% CI)**	**p-value**
Fathers (n=n_1)	**(n_1=247, n_2=318)**		**(n_1=202, n_3=246)**			**(n_1=323, n_5=564)**	
BMI (z score)							
Birth	-4.9 (-67.8 to 58.0)	0.8	-48.4 (-122.5 to 25.8)	0.2	0.7	-21.0 (-71.8 to 29.8)	0.4
birth-6m	64.2 (1.9 to 126.4)	0.04	83.9 (17.4 to 150.3)	0.01	0.4	68.9 (20.3 to 117.4)	0.006
6m-2y	-7.1 (-67.3 to 53.2)	0.8	0.4 (-67.5 to 68.3)	0.9	0.6	-5.4 (-52.9 to 42.1)	0.8
2y-11y	-0.9 (-59.5 to 57.6)	0.9	6.6 (-67.1 to 80.3)	0.8	1.0	0.1 (-48.5 to 48.7)	0.9
11y-adult	-20.3 (-82.6 to 41.9)	0.5	-2.1 (-67.4 to 63.1)	0.9	0.3	-3.8 (-51.9 to 44.3)	0.8
Mothers (n=n_4)	**(n_4=187, n_2=229)**		**(n_4=153, n_3=183)**			**(n_4=242, n_3=412)**	
BMI (z score)							
Birth	60.6 (-24.3 to 145.5)	0.1	5.8 (-60.1 to 71.7)	0.8	0.2	32.4 (-26.2 to 91.0)	0.2
birth-6m	70.9 (-4.2 to 146.0)	0.06	-8.1 (-78.8 to 62.5)	0.8	0.04	32.7 (-24.2 to 89.5)	0.2
6m-2y	-4.0 (-81.1 to 73.1)	0.9	18.7 (-49.3 to 86.7)	0.5	0.8	6.2 (-50.4 to 62.8)	0.8
2y-11y	119.5 (37.9 to201.0)	0.004	89.1 (24.7 to 153.5)	0.007	0.7	94.4 (37.4 to 151.4)	0.001

11y-adult	43.4 (-33.3 to 120.1)	0.2	-10.1 (-78.8 to 58.6)	0.7	0.4	29.7 (-26.8 to 86.2)	0.3
Interaction							
Birth	0.7		0.2		0.1		
birth-6m	0.4		0.04		0.3		
6m-2y	0.6		0.8		0.7		
2y-11y	1.0		0.7		0.01		
11y-adult	0.3		0.4		0.3		
Pooled (n=n_6)	**(n_4=434, n_6=547)**		**(n_3=355, n_6=429)**			**(n_5=565, n_6=976)**	
Birth	22.3 (-28.6 to 73.1)	0.3	-22.8 (-73.0 to 27.3)	0.3	0.3	2.7 (-35.7 to 41.2)	0.8
birth-6m	64.3 (16.3 to 112.2)	0.009	50.6 (2.1 to 99.0)	0.04	0.5	52.5 (15.6 to 89.5)	0.005
6m-2y	-7.0 (-54.6 to 40.6)	0.7	6.8 (-41.6 to 55.1)	0.7	0.7	0.1 (-36.2 to 36.5)	0.9
2y-11y	45.1 (-2.8 to 93.0)	0.06	43.8 (-5.6 to 93.2)	0.08	0.7	40.6 (3.6 to 77.6)	0.03
11y-adult	7.3 (-41.1 to 55.7)	0.7	-4.4 (-52.0 to 43.1)	0.8	0.8	9.9 (-26.6 to 46.6)	0.5

n_1= number of fathers, n_2= number of boys, n_3= number of girls, n_4=number of mothers, n_5=number of boys and girls, n_6=number of fathers and mothers

Table 5.7 Association of parental head growth and birth weight of children

(mutually adjusted for height, body mass index, gestational age, birth order and socioeconomic status)

Predictor (n_1)	Outcome Variable: Birth weight (grams) (n_2)			
	Pooled (n_1=551, n_2=954)		Mutually adjusted for gestational age, birth order and SES (n_1=551, n_2=954)	
	β(95% CI)	p-value	β(95% CI)	p-value
Head (z score)				
Birth	10.3 (-37.6 to 58.2)	0.6	5.3 (-42.6 to 53.1)	0.8
birth-6m	51.2 (6.3 to 96.1)	0.02	50.4 (5.6 to 95.2)	0.02
6m-2y	-20.8 (-62.6 to 21.0)	0.3	-18.6 (-60.2 to 23.1)	0.3
2y-11y	27.2 (-13.7 to 68.2)	0.1	35.1 (-6.0 to 76.3)	0.09
11y-adult	-13.6 (-55.8 to 28.6)	0.5	-12.7 (-54.7 to 29.4)	0.5
Height (z score)				
Birth	38.0 (-9.9 to 85.9)	0.1	35.1 (-13.0 to 83.1)	0.1
birth-6m	17.7 (-23.8 to 59.1)	0.4	12.1 (-30.6 to 54.7)	0.5
6m-2y	16.9 (-26.9 to 60.7)	0.4	11.9 (-33.4 to 57.2)	0.6
2y-11y	-5.3 (-45.8 to 35.2)	0.7	-6.7 (-47.3 to 33.8)	0.7
11y-adult	25.1 (-16.9 to 67.1)	0.2	26.9 (-15.1 to 68.8)	0.2
BMI (z score)				
Birth	-8.4 (-52.2 to 35.5)	0.7	-8.0 (-51.9 to 35.9)	0.7
birth-6m	31.7 (-10.6 to 74.0)	0.1	30.2 (-12.3 to 72.8)	0.1
6m-2y	4.0 (-36.5 to 44.5)	0.8	4.3 (-37.0 to 45.6)	0.8
2y-11y	31.1 (-13.4 to 75.5)	0.1	29.6 (-15.0 to 74.3)	0.1
11y-adult	17.7 (-23.8 to 59.3)	0.4	10.6 (-31.1 to 52.2)	0.6
Gestational age (z)*	-	-	34.7 (-7.2 to 76.5)	0.1
Missing (gestational age)	-	-	2.9 (-122.3 to 128.1)	0.9
birth order	-	-	62.4 (14.8 to 110.0)	0.01
SES*	-	-	10.1 (-31.6 to 51.7)	0.6
Missing (SES)	-	-	76.0 (-1.2 to 153.3)	0.05

*Gestational age available for 252 (18%) subjects, SES calculated for 819 (59%) subjects.

Table 5.7 presents the results of the adjusted analysis after multivariable adjustment for the effects of height and BMI growth on the association between F1 head growth and F2 birth weight. Except head growth in early infancy, none of the body size at birth or body growth measures in the defined time periods were associated with F2 birth weight.

Birth weight in the F2 generation increased by 51.2 grams for each 1 SD increase in head growth during early infancy. These associations changed little after further adjustment for gestational age, birth order and SES. Birth order was positively associated with F2 birth weight in the final adjusted model.

Discussion

This chapter described an analysis of F1 head size and growth with F2 birth weight. The objective of this analysis was to assess 1) whether there was evidence of intergenerational effects; 2) whether they are specific to F1 head (as opposed to weight, length or BMI) and 3) whether the associations differ between mothers and fathers, which would be informative about mechanisms of intergenerational effects.

The analysis was based on 441 F1 fathers and 353 F1 mothers having 1,385 F2 singleton offspring. The main statistical analysis technique used is multilevel modelling. This is essential because F1 parents had more than one child in this study and the F2 birth weights will tend to be correlated. I started the analysis by looking at the unadjusted associations of F1 head size at birth and growth with F2 birth weight. I found no significant differences in the associations between mothers and fathers, or between boys and girls, and therefore pooled the sexes.

I also tested for the presence of random slopes of head size at birth and head growth in the different time periods in the unadjusted models for head circumference for mothers and fathers, and boys and girls using a likelihood ratio test. Two out of 20 tests were significant (6m-2y for fathers and 11y-adult for sons), which could be due to multiple testing. I found no presence of a random slope in either of those models and thus used random intercept models for all subsequent analyses. In the pooled analyses, F1 head size at birth was positively associated with birth weight in the F2 offspring ($p=0.04$) (Table 5.4). F1 head growth in early infancy was also positively associated with birth weight in the next generation. In the unadjusted analyses, I also found positive associations of F1 birth length and height growth up to 2 years, and of BMI gain in early infancy and childhood with next generation birth weight. After adjusting for the effects of height and BMI growth there were no statistically significant associations of F1 head, height or BMI size at birth and growth with F2 birth weight, except F1 head growth in early infancy which was positively associated. F2 birth order remained positively associated with F2 birth weight in the final model. This is the first

analysis in the world to our knowledge looking at the associations of parental head size at birth and head growth with offspring birth weight. The associations of head size at birth and growth with offspring birth weight were similar among mothers and fathers, which suggests that these may be due to genetic influences or persisting shared environment between the mothers and fathers, rather than specific effects of the environment provided by the mother.

Mechanisms of intergenerational inheritance

Many epidemiological studies in different populations have demonstrated an association between low birth weight and subsequent disease in adulthood. This evidence has given rise to the fetal origins hypothesis, which suggests that if the fetus is exposed to an adverse environment in the utero it is associated with CVD and its risk factors in adulthood (7,8). There are studies which show that this phenomenon may not be limited to a single generation (52). There is a comprehensive review on the mechanisms underlying intergenerational effects of fetal programming (50). According to this review, there are several potential explanations for the transmission of programming effects to the subsequent generations. These are (i) adverse environmental conditions that persist across generations. Eg: Kuzawa and Sweet recently showed that the marked disparities between African-Americans and Whites in the USA in outcomes like low birth weight reflect the effect of adverse social influences including discrimination and social disadvantage, which are shared by both parents and which could persist across generations (116). The authors argued that these influences could be associated with the burden of CVD in African-Americans; (ii) adverse in-utero experiences that permanently affect growth and development, thus resulting in changes in physiology, size or behaviour. Here, the 'abnormal environment' is the altered maternal physiology, behaviour or size that have the potential to be associated with fetal development. These changes then induce programming effects in the next generation; (iii) effects which are transmissible through genes; (iv) intergenerational changes through learned behaviour: In humans there is evidence for a similar behaviour programming effect which can be transmitted across generations Eg: a child exposed to family violence may continue the same type of behaviour towards his/her spouse and children (117), and parenting styles and conduct disorders may remain the same across generations (118,119). If the intergenerational effects are similar for both mothers and fathers, they are usually due to genetic effects or shared environment. If maternal

effects are greater than paternal effects, these may be due to learned behaviour or the adverse in utero experiences. Since in this analyses there were no statistically significant differences between maternal and paternal effects, I can conclude that these may be genetic or due to adverse environmental conditions shared by the parents and which persist across generations.

Comparison with previous literature

This analysis is unique in the sense that this is the first in the world to my knowledge studying the association between parental head size at birth and head growth with offspring birth weight. Previous studies have looked at the association of parental birth weight with offspring birth weight in mothers and fathers. Addo et al. in 2015 in a pooled analyses involving four birth cohorts in low and middle income countries, which included the NDBC showed that maternal and paternal birth weight were positively associated with birth weight in the offspring (120). A 1 SD increase in maternal birth weight was associated with an increase in offspring birth weight of 102 gms (95% CI: 79 to 125). For fathers, a 1 SD increase in birth weight was associated with an increase of 57 gms (95% CI: 26 to 89) in offspring birth weight. In an Indian study, Agnihotri et al. studied two birth cohorts of successive generations in Vellore in India (121). The parental cohort comprised 472 fathers and 422 mothers, and data were obtained on 1525 offspring. In this study, a low birth weight (LBW) mother had 2.8 times risk (95 % CI: 1.2 to 6.4) of delivering a LBW baby (p=0.02) and a LBW father was twice as likely to be associated with a LBW offspring (OR: 2.2 (95% CI: 1.0 to 4.8)) after adjusting for adult BMI, adult height, SES score, parity and sex of the infant (p=0.05). Every 100 gms increase in maternal birth weight was associated with an increase in offspring birth weight of 13.9 gms (95% CI: 6.4 to 21.4). For fathers birth weight increased by 18.1 gms (95% CI: 10.0 to 26.2) ($p<0.01$ for both). In a similar study, Veena et al analysed 468 mother-offspring pairs and 341 father-offspring pairs born in Mysore, India (122). The study showed that the birth weight of both parents is equally associated with the offspring birth weight. However, some studies have shown a specific association of the mother's birth weight with birth weight in the offspring. Kuzawa et al looked at the associations of parental birth weight with offspring birth weight in the Philippines (123). Data included birth weight of offspring (n=1,101) born to female members

(n=382) and spouses of male members (n=275). There was an interaction between sex of the parent and parental birth weight which was of borderline significance (p=0.068), showing that maternal and paternal birth weight are associated with offspring birth weight differently. All subsequent models were stratified on sex of the parent for whom birth weight was measured. Each kg in mother's and father's birth weight predicted a 271±53g (p<0.001) and 132±55g (p=0.017) increase in offspring birth weight respectively. A similar study was conducted by Horta et al in 2009 wherein separate analyses were carried out for men and women (124). The postnatal growth in males was not expected to affect offspring birth weight. Their inclusion in the analyses was aimed at testing the specificity of findings with regards to the mothers. The data included 848 women who had delivered a child and 525 fathers. Maternal birth weight, and weight and length-for-age z score at age 20 months were positively associated with next generation birth weight, whereas paternal variables were not related to the outcome. Conditional growth modelling analyses showed that women who had greater weight gain in the first 20 months of life had heavier babies, whereas paternal weight gain was not associated.

Thus, some studies show a greater maternal association while others show a similar association from both parents. From the above studies, equal associations are observed in developing countries, and unequal associations are observed in developed countries. This is possibly because in developing countries the persistence of adverse environmental conditions could be a major explanation for the observed intergenerational effects. In developing countries, there is a high percentage of the population living in poverty and disadvantage which may persist across several generations, which might not be the case in developed countries. Also, there could be a genetic effect, due to which equal associations were observed from both parents, an issue which I did not investigate further in the current analysis. Where there is a specific maternal effect, a potential mechanism is that programmed changes in the mother can influence the development of her offspring during pregnancy (47). The abnormal environment in this case is the altered maternal phenotype, and this can influence the development of the next generation. This can be due to the inadequate maternal nutrition which might be having a predominant effect in these associations in the developed countries.

I did not find any associations of F1 height or BMI at birth, or growth in any postnatal time period, with F2 birth weight after adjusting for head size at birth and head growth. A recent Brazilian study that analysed 2,226 mother-child pairs found that maternal height in the lower quartile was related to an increased risk for having children with low birth weight when compared to mothers in the upper quartile for height (125). However, a systematic review by Han et al. concludes that only studies with unadjusted data express associations between women of short stature and an increased risk of low birth weight in the offspring (126). Previous studies have shown an association between parental body size at birth and growth with offspring birth weight. Ramkrishnan et al. showed that the offspring's birth weight increased by 53 gms for every 1 cm increase in mother's birth length (127). The study by Addo et al showed that a 1 SD increase in maternal length growth between 0-2 years was associated with a 46 gms (95% CI: 21 to 70 gms), $p<0.01$ increase in offspring birth weight (120). A 1 SD increase in maternal height growth in childhood was associated with 27 gms (95% CI: 4 to 51 gms) increase in offspring birth weight. A study by Martin et al in 2003 showed a significant association of both maternal and paternal growth measures with offspring birth weight (128). This study showed that mother's height in childhood was positively associated with her offspring's birth weight. Leg length, but not trunk length was the component of maternal height associated with offspring birth weight. For each unit increase in SD-score for maternal childhood leg length, there was a 96g (95% CI: 6-186 gms) increase in offspring birth weight in the adjusted model. The study showed weaker positive associations of paternal height and leg length in childhood with offspring birth weight. For paternal childhood height, leg length and trunk length there were respectively a 47 gms(-14 to 107 gms), 47 gms(-19 to 113 gms) and 13 gms(-48 to 73 gms) increase in offspring birth weight per unit increase in SD score. These associations did not differ according to the sex of the offspring. In the study by Veena et al in 2004, the mother's BMI in adulthood had a stronger association than paternal BMI in adulthood with offspring birth weight (mother:18 g/kg/m^2 (95% CI: 9 to 27), $p<0.001$; father:15 g/kg/m^2 (95% CI: 1 to 30), $p=0.04$) (122).

Genetic effects and birth weight

Previous studies have shown associations between fetal genes and birth weight. Freathy et al in 2009 genotyped SNPs at five identified Type 2 diabetes loci in 7,986 mothers and 19,200 offspring from four studies of white Europeans (129). The authors tested the

association between maternal or fetal genotype at each locus and birth weight of the offspring. The authors found that type 2 diabetes risk alleles were associated with reduced birth weight when inherited by the fetus (21g (95% CI:11-31), $p=2*10^{-5}$ and 14g (4-23), p=0.004 lower birth weight per risk allele respectively). The 4% of offspring carrying four risk alleles at two loci were 80 g (95% CI: 39-120) lighter at birth than 8% carrying none (p for trend=$5*10^{-7}$). There were no associations between birth weight and fetal genotypes at the other three loci or maternal genotypes at any locus. These findings were confirmed in a meta-analysis of six genome-wide association (GWA) studies (n=10,623 Europeans from birth cohorts) (130). The authors found that 9% of Europeans carrying four birth weight lowering alleles were on average 113g (95% CI;89-137g) lighter at birth than the 24% with zero or one alleles (p-value for trend=$7*10^{-30}$).

These studies show that genetic factors have an influence on next-generation outcomes, such as birth weight. Both mother and father can theoretically influence the offspring birth weight both through the genes that are passed on by them, but it appears to be a small effect. I did not look at such associations in my research.

Intergenerational effects and F1 postnatal growth

In my analysis, a specific association of head growth from birth-6m was found in the adjusted analysis, which was similar for both parents. There can be several possible reasons for this association. Firstly, it is known that increased head growth between birth-6m is positively associated with adult cognition (47). Increased education is associated with increased birth spacing when that individual becomes a parent. Educated parents are more likely to have increased spacing between births. Increased birth spacing is associated with a higher offspring birth weight (131,132).

In my analysis, first born had a lower birth weight than subsequently born children. A potential explanation for this finding is that physiologic studies have suggested possible anatomic suggestions for the reduced birth weight of first born. During a mother's first pregnancy, structural changes take place in the arteries, thereby increasing the blood flow (133-137). These changes do not disappear completely following the pregnancy. Therefore, the subsequent offspring are from start of the pregnancy exposed to reduced resistance of the blood vessels and hence, greater blood flow as compared to the first born, thus this promotes fetal growth (138).

Appropriateness of the analytical approach

I used conditional growth variables, which control for prior measures. The use of uncorrelated measures of growth for four life course periods is a strength of this study that allowed me to assess the relative importance of parental body growth during specific important developmental periods. Another strength is the use of mixed linear regression models that controlled for any potential correlation between the birth weights of the offspring from the same parent (62). This method has been used in a similar previous study in a pooled analysis of four birth cohorts, which included the NDBC (120). The authors used mixed regression modelling as it can assign to each parent a parent-specific random effect. The main statistical models used in a study by Hypponen et al. were random effects regression models to account for the dependence of subsequent births to one parent, something which has been done in this study as well (139). As in this study, all analyses in my study were adjusted for the gestational age of the offspring (139). Gestational age was adjusted for in this analysis to distinguish prematurity from restricted fetal growth. In this analysis, the main outcome is the child's birth weight, which is clustered at the level of the parent. When subjects are clustered, it creates several problems. Observations within a cluster tend to be more alike than observations selected at random. Children from the same parents will be similar to one another because they share genes and environment with their siblings. Standard statistical methods assume that the observations are independent. If observations are clustered, the analysis must be modified to take the clustering into account. When cluster designs are used, there are two sources of variance in the observations. The first is the variability of birth weights within a cluster, and the second is variability between clusters. These two sources combine to produce the overall variance. Both the between and within variance must be taken into account while analysing the data. The effect of the increased variance due to a cluster design is that it increases the size of the standard errors and widens the confidence intervals, thus increasing the p-values, as compared to a study without any clustering of data. The effect of this is that the sample size is reduced, power is lost and the type-1 error rate is increased. Therefore adjusting for correlation within the cluster is necessary. Linear mixed models have been used to adjust for the correlations in the present study.

There are several other statistical techniques to adjust for the within-cluster correlation, apart from linear mixed models. One is a clustered robust standard error

method, in which model applied is a linear model but the standard errors are adjusted for clustering using the Huber-White (also called Sandwich) standard error approach (140). This method only adjusts the standard error related to the confidence interval; the point estimates are left unchanged. Another method is the Generalized Estimating Equations (GEE) approach, in which a working correlation matrix is specified to adjust the within cluster correlation (141). Unlike the random effects method which is used in this analysis, the GEE approach estimates the regression parameters averaging over the clusters. This method is also known as a 'population average approach'. I will expect the results for these two approaches to be very similar to the ones obtained by using mixed models, which has currently been used.

Making a choice between the different methods as well as interpreting associations from the two techniques is an important issue. The GEE approach treats the clustering as a nuisance factor which has to be adjusted for. The dependence between repeated observations is taken into account by robust variance estimation. Also, missing values are handled differently by GEE and the linear mixed models method. GEE requires the missing completely at random (MCAR) assumption (62). Under MCAR, missingness does not dependent on any individual characteristics (64). Mixed models assume the data to be missing at random (MAR) (64). This allows missingness to depend on an individual's previously observed values of the dependent variable. This assumption is more realistic in practice, which is an advantage of this approach over GEE. An advantage of the mixed model method is that the data can be highly unbalanced. This is relevant in the present study as every F1 parent has different number of F2 children.

Other strengths and limitations

The fact that F2 birth weights were obtained by recalled rather than measured may be a potential weakness, as it may introduce inaccuracy. However, several previous studies have shown good agreement between recalled and measured birth weight (142-145). These results have been confirmed in a recent systematic review and meta-analysis by Shenkin et al. in 2017 (146). In total, 40 studies were available (n=78,997 births from 78,196 parents). Agreement between recalled and recorded birth weight was high in the meta-analysis. Pooled estimate of correlation in 23 samples from 19 studies (n=7406) was 0.90 (95% CI: 0.87-0.93). The difference between recalled and recorded birth weight in 29 samples from

26 studies (n=29,293) was small (range:-86 to -129g: random effects estimate 1.4g (95% CI:-4.0 to -6.9g)). Studies were heterogeneous, with no evidence for an effect of time, person reporting, recall bias or birth order. However, recalled birth weight may not be so useful in developing countries. There are few studies comparing the accuracy of recorded and registered birth weight in developing countries.

There are no studies from an Indian population comparing recalled and recorded birth weight. Recall may be less accurate in India because Indian mothers do not give much importance to the birth weight of the child, and therefore may not remember it. Lower levels of literacy and numeracy in India may also increase inaccuracies. As the data on all children from one parent were collected at one time point, there is a chance that mothers providing data on more than one child might report similar or even identical birth weight for all their children. Therefore, in my study, the issue of recalled birth weight being completely different from actual birth weight cannot be completely ruled out.

Another disadvantage of my study was that the mothers and fathers in the cohorts were from different families. Hence, it was not possible to compare both paternal and maternal effects in determining an offspring's birth weight. For example, people tend to marry people having a similar height and social and economic background (147). The associations of these type of effects on the birth weight of the offspring could not be investigated in this analysis. Another disadvantage was the small sample size of offspring birth weight in the study. This could be because birth weight was ascertained opportunistically for the F2 generation when the F1 generation attended the research clinic, leading to variable recall periods. This could have biased the results towards the null values, due to an increase in error in the estimates, thus reducing the power of the study. A statistical issue that arises with a small sample size is that this generally leads to a loss in the precision of the estimates, thus bringing them towards the null values. Another limitation in the study is the significant loss to follow-up in the F2 generation. Unlike the F1 generation, the F2 generation was not followed up prospectively in a systematic pre-designed manner. The variables were measured opportunistically during the adult phases of F1 generation, when the subjects were invited to bring their children also (F2 generation). A bias that arises due to unsystematic follow-up is selection bias. There could have been selection bias in this case, as the subjects in phase 5 were invited to the clinics. For example people of lower SES might not have come to the clinics, possibly because of lack of

transportation. The critical question then becomes, “does the association between head size/growth and the outcome variable differ according to the level of SES?” There is no evidence that this is likely to be the case. A further disadvantage is that multiple significance tests have been carried out in these analysis and thus some of the associations might have been statistically significant due to chance. Associations which were strongly statistically significant, still might not be of substantial clinical relevance since the effect sizes were small.

To conclude, increased F1 head growth in early infancy was positively associated with birth weight in the next generation. The similar associations of both parents head growth with birth weight in the next generation are probably due to shared genetic factors or persisting environment between generations. Genetic influences cannot probably be modified. However, an understanding of which environmental factors influence brain growth may help in increasing birth weight in the next generation.

Chapter 6: Comparison of conditional and spline approaches for modelling head size at birth and head growth

The assessment of human growth requires the assessor to make serial observations on a particular body measurement or body measurements, thereby resulting in longitudinal data. The analysis of such a dataset can be challenging due to lack of independence of observations between time points. Several statistical approaches for modelling growth data have been proposed, to identify sensitive growth periods throughout the lifecourse of an individual that may be related to higher risk of diseases in later life (61). Two such strategies are the multilevel linear spline approach and the conditional body size approach. The conditional body size approach has been described in detail in chapter-3 (Page 58). In brief in the conditional body size approach the health outcome in later life is regressed on all conditional body sizes, which are defined as the difference between observed and predicted body sizes at any age plus the size at birth. In chapter 4, I presented the relationship between head growth and an adult cardiometabolic and a cognitive outcome in the NDBC dataset. In this Chapter, I looked at the same associations using an alternative statistical approach, the spline approach. The comparison between the two growth approaches will be extended to other adult cardiometabolic outcomes such as plasma triglyceride, plasma HDL cholesterol and fasting plasma glucose. My objective of performing the above two methods of analysis is (a) to assess the robustness of the previous findings (b) to explore reasons for any difference in results are encountered. This exercise would help in providing suggestions to future analysts working in longitudinal data as the most appropriate method given the nature of their data.

6.1 Multilevel linear spline models

Growth is a nonlinear process. Nonlinear growth trajectories can be modelled well by using regression models containing polynomial terms. However, interpretation of such models is challenging. An approach that can yield more interpretable growth coefficients is to use a series of linear splines, joined at ‘knot points’, to model the growth trajectory (148). These models have also been referred to as ‘piecewise linear’ or ‘broken stick’. As an example, a multilevel linear spline model for weight with knots at 3 and 12 months would allow different linear slopes from 0 to 3 months, 3 to 12

months and beyond 12 months. The slopes for these models will be different for every individual.

6.1.1 Choosing the knot points

A method is to place knot points at the centiles of the age distribution (148). This approach involves dividing the distribution into quantiles with respect to age and placing the knot points at the quantiles. Another method is to decide the position of the knots based on knowledge of the underlying biology of growth patterns. I considered this approach for selecting the knot points at 6 months, 2 years and 11 years. The underlying biological reasoning for choosing the critical time points has been described in detail in Chapter-3. These periods are important for a child's health and development, because during each of these phases, growth is regulated by a different set of hormones. For both the conditional and the multilevel spine approach the same time points will be considered, which will help in the comparison between the two growth analytic approaches. In the next section a statistical method to quantify time period-specific deviations from the subject-specific fitted line will be discussed.

6.2 Quantifying time period-specific deviations from the subject-specific fitted line

Multilevel linear spline models estimate the slope (growth velocity) in the different growth periods for every individual separately. Hence, growth period-specific random effects were derived for every subject in the multilevel dataset. These are called BLUPs (149). The acronym 'BLUP' stands for Best Linear Unbiased Predictor. It is a method of estimating random effects. BLUPs are linear functions of the data. These random effects have the minimum mean squared error within the class of linear unbiased estimators and are sometimes called 'predictors' to distinguish them from the fixed effect estimates. BLUPs were generated for head circumference measured in cms and also a SD score.

Table 6.1 shows an example of an individual (ID=50037) for whom deviations from the population line have been generated (BLUPs). Using these, I generate estimated slopes for the subject. The fitted lines are given in figure 6.1.

Table 6.1 (a): Subject-specific deviations of head size at birth and in the different growth periods

Time period (years)	Full population estimate (birth in cms, growth periods in cms/year)	ID=50037 estimate (birth in cms, growth periods in cms/year)	Deviation of ID=50037 from population line (BLUP)
0	34.08	33.14	-0.94
0-0.5	17.25	16.54	-0.71
0.5-2	3.47	3.24	-0.23
2-11	0.55	0.60	0.05
11-adult	0.21	0.23	0.02

Table 6.1 (b): Fitted values of head size at birth and at specific ages

Age (years)	Full population estimate (cms)	ID=50037 estimate	Deviation of ID=50037 from fitted line
0	34.08	33.14	-0.94
0.5	43.43	41.41	-2.02
2	48.64	46.27	-2.37
11	53.69	51.67	-2.02
37	59.05	57.65	-1.40

Figure 6.1: A linear regression fitted spline for ID=50037 with population spline

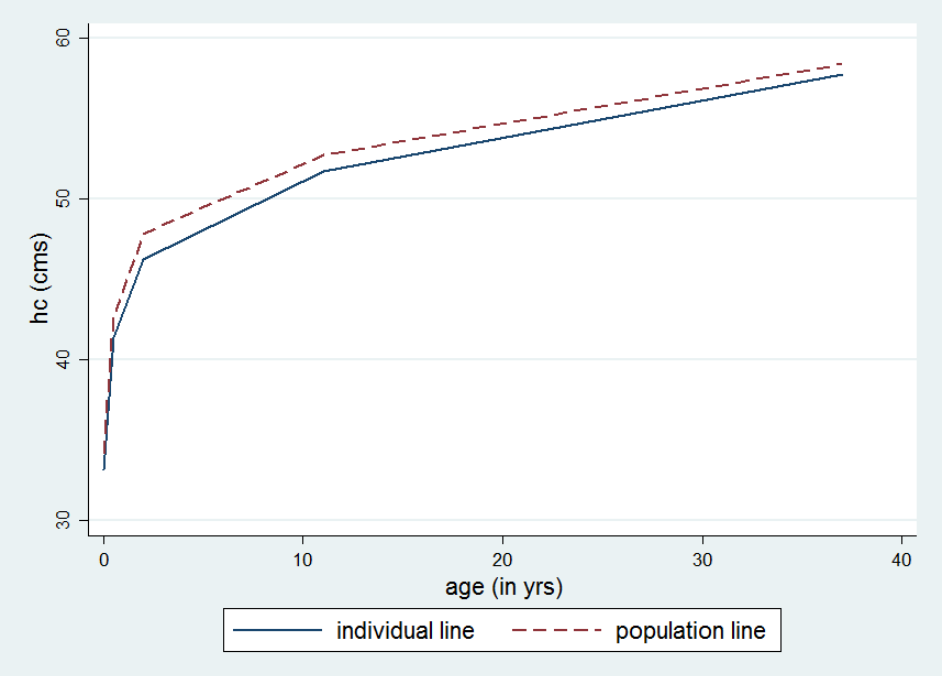


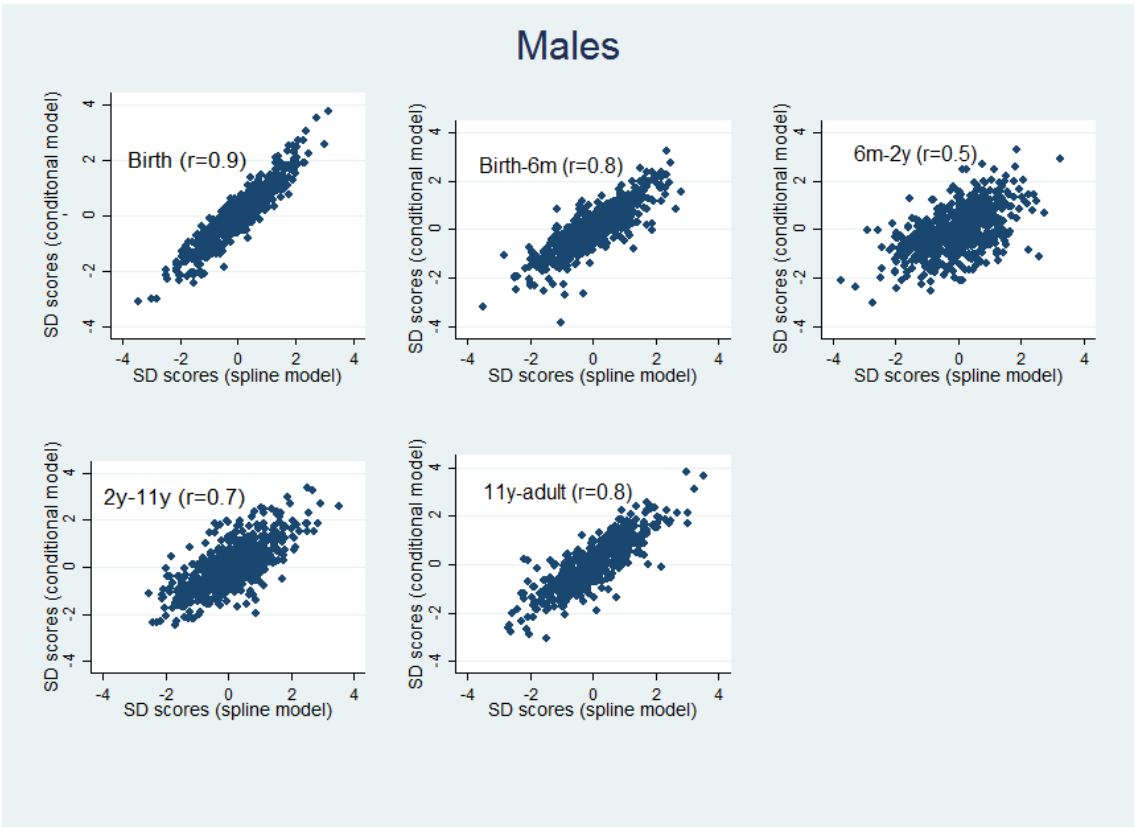
Table 6.2 gives the distribution of the estimated slopes of head circumference at the different ages of interest in the cohort. These are the velocities of head growth generated in the different time periods by fitting linear splines for every subject in the study

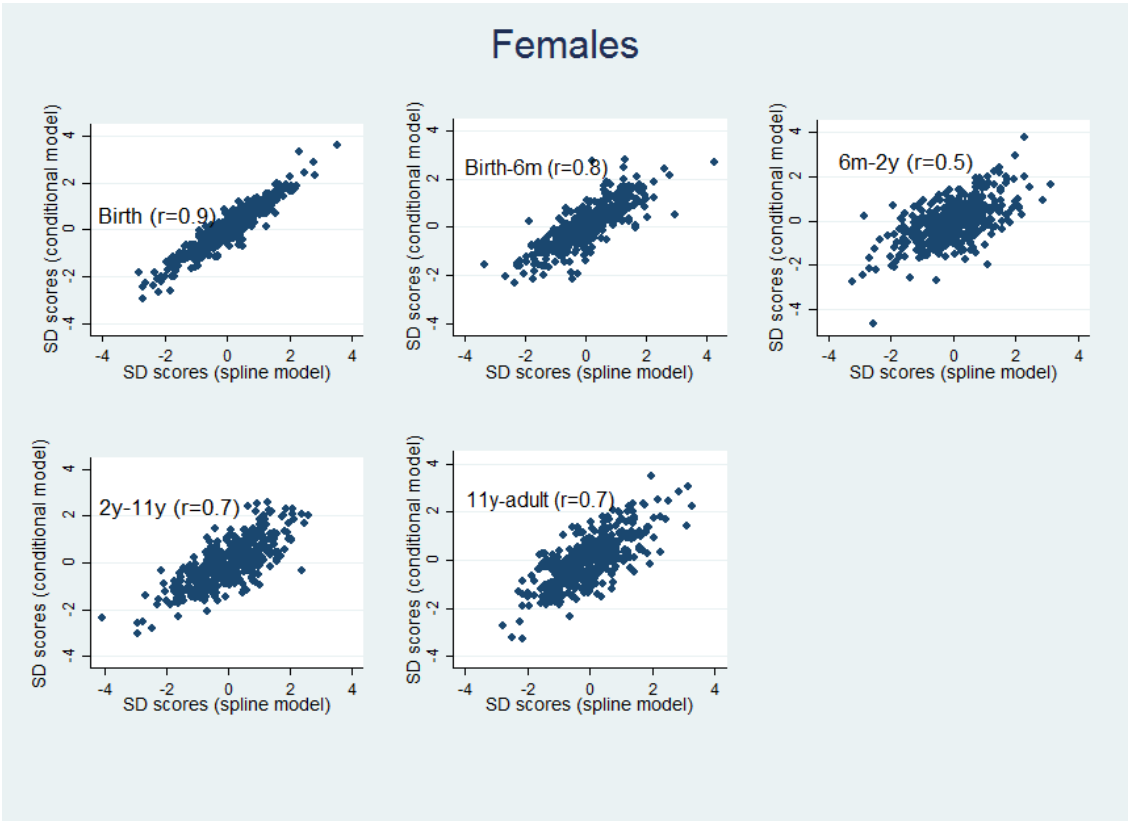
Table 6.2: Head size at birth and average estimated slopes in the different time periods

Linear spline slopes	**N**	**Males**	**N**	**Females**
Body growth variables		**Mean±SD**		**Mean±SD**
At birth				
Head (cms)	3526	34.7±2.4	3235	34.1±2.3
Head (SD score)	3526	0.02±0.85	3235	0.02±0.86
From birth-6months				
Head (cms/year)	3251	17.22±1.55	2996	16.15±1.48
Head (SD score)	3251	-0.02±1.48	2996	-0.04±1.49
From 6months-2years				
Head (cms/year)	3209	3.47±1.13	2922	3.36±1.05
Head (SD score)	3209	-0.02±0.35	2922	-0.02±0.38
From 2years-11years				
Head (cms/year)	2519	0.59±0.17	2339	0.73±0.16
Head (SD score)	2519	-0.00±0.06	2339	0.00±0.07
From 11 years-adulthood				
Head (cms/year)	1675	0.18±0.05	1579	0.08±0.06
Head (SD score)	1675	-0.00±0.02	1579	-0.00±0.02

Since both the estimated slopes and the conditionals are derived within the same age intervals, they may be plotted and compared. In Figure 6.2, they are expressed as standardized scores.

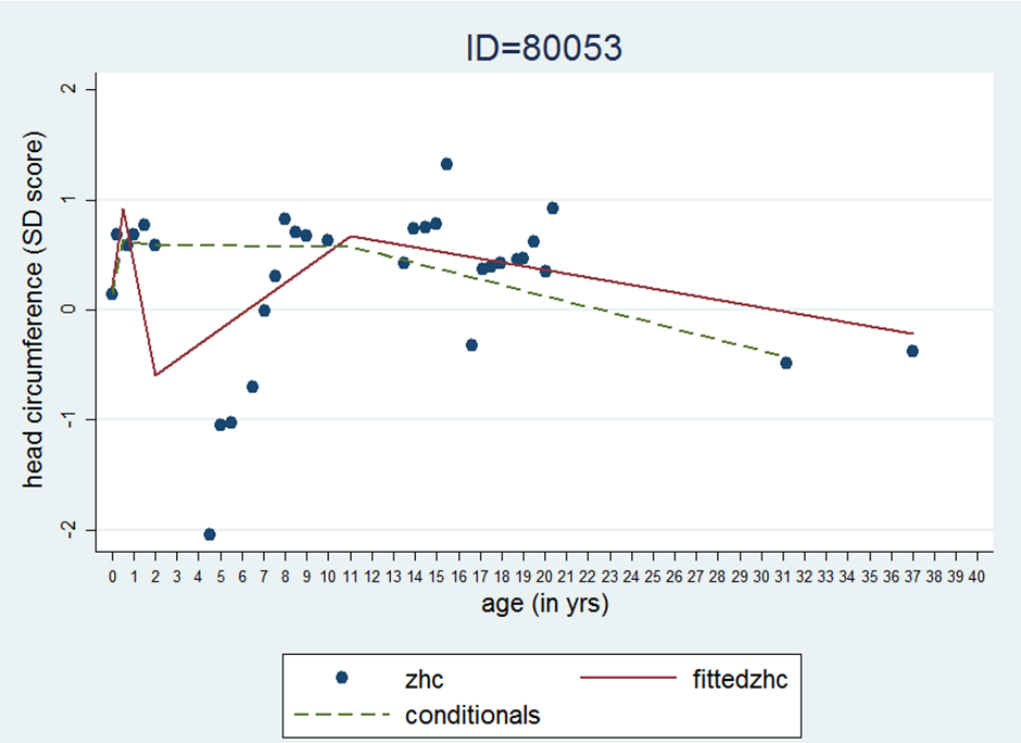
Figure 6.2: Sex-specific correlation of the growth variables using the two approaches (using head circumference measured in cms and transformed into standard deviation scores)





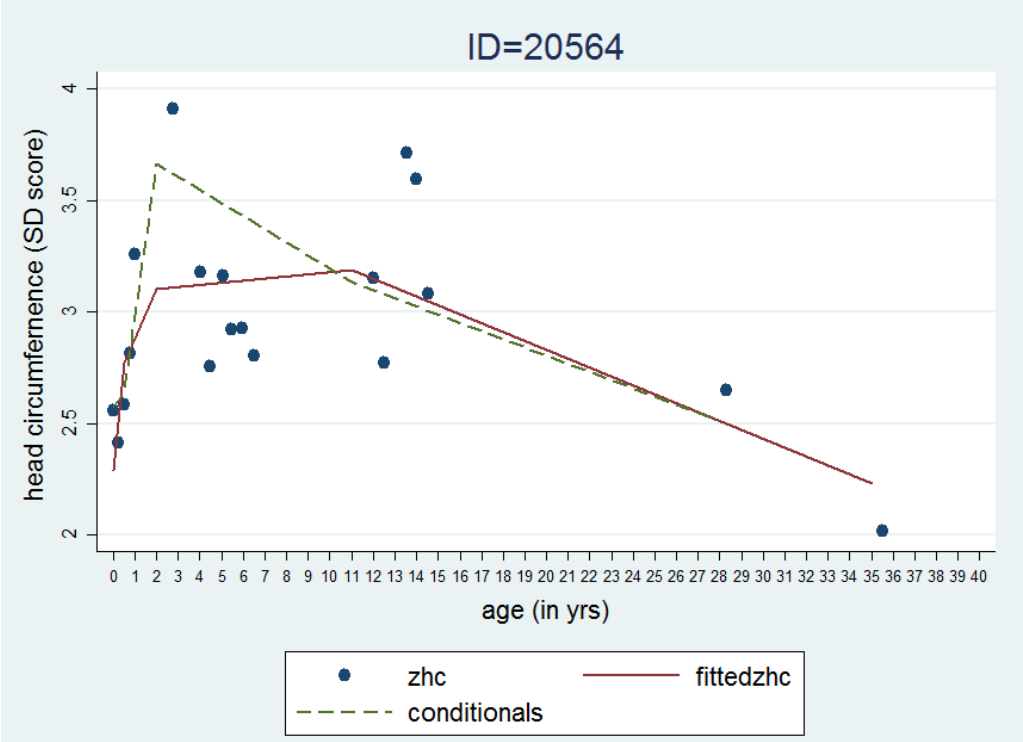
The two values were strongly correlated for both head size at birth and all growth periods, except in the period from 6 months-2years (6m-2y). To determine the cause of this difference, I identified individuals for whom both splines and conditionals could be calculated. This will help to visualize the trajectory of the fitted line from 6m-2y for the two methods, the spline and the conditional approach. Plots for some surprising individuals are displayed below (Figures 6.3 and 6.4):

Figure 6.3: Differences in fitted individual-specific lines using the multilevel linear spline (solid line) and the conditional model (dotted line)



From the above plot (Figure 6.3), we can see the fitted line from the conditional model uses only the head sizes close to the significant ages for plotting the fitted line. The fitted line for the spline model is influenced by outlying observations, and is directed towards the outlying observations as can be seen for the observation having a SD score less than -2. From this plot, we conclude that the spline fitted line can be influenced by outlying observations, whereas the conditional fitted line is less affected by such observations.

Figure 6.4: Differences in fitted individual-specific lines using the multilevel linear spline (solid line) and the conditional model (dotted line)



Taking ID=20564, the spline method identifies the trajectory of head growth and plots the fitted line (Figure 6.4), whereas the conditional model fits the regression line only through the points at the significant ages.

Thus, there is a difference between the head growth in the 6m-2y period for these methods. Hence, there is a low correlation in the head growth in the two approaches for this period.

6.3 Comparison of the conditional and spline approaches for different outcomes

I then evaluated the unadjusted associations of head size at birth and head growth in the different growth periods for individuals having non-missing conditionals and spline slopes with years of education in adulthood (Table 6.4) and selected adult cardiometabolic outcomes (Figures 6.5-6.9). Below is a table showing the number of individuals contributing data to each of the outcomes in the different analyses (Table 6.3). As can be seen from the table, the splines use more data than the conditional approach.

Table 6.3: Number of individuals in the different analyses

Growth Variables	Years of education	SBP (mmHg)	Triglycerides (mg/dl)	HDL (mg/dl)	Fasting blood glucose (mg/dl)
Males					
Splines alone	835	829	817	818	818
Conditionals alone	623	618	612	613	612
Splines + conditionals	615	610	604	605	604
Females					
Splines alone	602	594	586	584	586
Conditionals alone	429	424	425	423	425
Splines + conditionals	426	421	422	420	422

The spline method is able to fit a line, because the spline method does some smoothing, and is able to calculate deviations based on the average estimates at the ages which have no data, but these would have quite a high variability and be unreliable. The conditional method, on the other hand calculates deviations based on the observed data only. Therefore, the deviations from the spline method are more unreliable than the conditional method. All subsequent models were constructed only for individuals having both splines and conditionals.

6.3.1 Outcome: Years of education

Initially I analysed unadjusted associations of head size (cm) at birth and head growth (cm/year) in the previous age intervals with years of adult education, separately for males and females. The same procedure was followed to derive estimates using age-adjusted head circumference SD scores and then using conditionals. I constructed formal interaction tests between sex and head size at birth and growth measures. (Table 6.4 and Figure 6.5).

Table 6.4: Association of head size at birth and growth with years of education in adulthood using the two growth modelling approaches

Outcome Variable: Years of education							
Predictor	**Males (n=615)**		**Females (n=426)**		**Inter-action**	**Pooled (n=1041)**	
	β(95% CI)	**p-value**	**β(95% CI)**	**p-value**	**p-value**	**β(95% CI)**	**p-value**
Spline approach (using head circumference in cm)							
Head parameter (standardized)							
Birth	0.38(0.10 to 0.66)	0.008	0.29(0.02 to 0.57)	0.03	0.6	0.34(0.14 to 0.54)	0.001
birth-6m	0.40(0.13 to 0.66)	0.003	0.52(0.25 to 0.79)	<0.001	0.5	0.45(0.26 to 0.64)	<0.001
6m-2y	0.02(-0.29 to 0.33)	0.8	-0.03(-0.32 to 0.27)	0.8	0.8	0.00(-0.22 to 0.22)	0.9
2y-11y	0.14(-0.18 to 0.45)	0.3	0.02(-0.31 to 0.34)	0.9	0.6	0.09(-0.14 to 0.32)	0.4
11y-adult	0.12(-0.17 to 0.40)	0.4	0.02(-0.29 to 0.32)	0.9	0.6	0.08(-0.13 to 0.28)	0.4
R^2	0.0290		0.0437			0.0332	
Spline approach (using head circumference in SD score)							
Head parameter (standardized)							
Birth	0.57(0.26 to 0.88)	<0.001	0.50(0.19 to 0.81)	0.002	0.7	0.54(0.32 to 0.76)	<0.001
birth-6m	0.49(0.19 to 0.80)	0.002	0.68(0.37 to 0.99)	<0.001	0.4	0.57(0.35 to 0.79)	<0.001
6m-2y	0.12(-0.18 to 0.42)	0.4	0.03(-0.28 to 0.34)	0.8	0.6	0.08(-0.14 to 0.30)	0.4
2y-11y	0.08(-0.22 to 0.39)	0.5	0.02(-0.31 to 0.35)	0.9	0.7	0.06(-0.16 to 0.29)	0.5
11y-adult	0.14(-0.15 to 0.42)	0.3	0.05(-0.24 to 0.34)	0.7	0.6	0.1(-0.11 to 0.30)	0.3
R^2	0.0262		0.0509			0.0328	
Conditional approach (using head circumference in SD score)							
Head parameter (standardized)							
Birth	0.32(0.05 to 0.58)	0.02	0.24(-0.04 to 0.51)	0.09	0.6	0.29(0.09 to 0.48)	0.004
birth-6m	0.39(0.11 to 0.66)	0.007	0.55(0.25 to 0.85)	<0.001	0.4	0.45(0.25 to 0.66)	<0.001
6m-2y	0.26(-0.01 to 0.54)	0.06	0.26(-0.03 to 0.54)	0.07	0.9	0.26(0.06 to 0.46)	0.01
2y-11y	-0.00(-0.26 to 0.26)	0.9	0.11(-0.18 to 0.40)	0.4	0.5	0.04(-0.15 to 0.24)	0.6
11y-adult	0.21(-0.05 to 0.47)	0.1	0.00(-0.26 to 0.27)	0.9	0.2	0.13(-0.06 to 0.32)	0.1
R^2	0.0299		0.0418			0.0318	

Figure 6.5: Association of head size at birth and growth with years of adult education (n=1041)

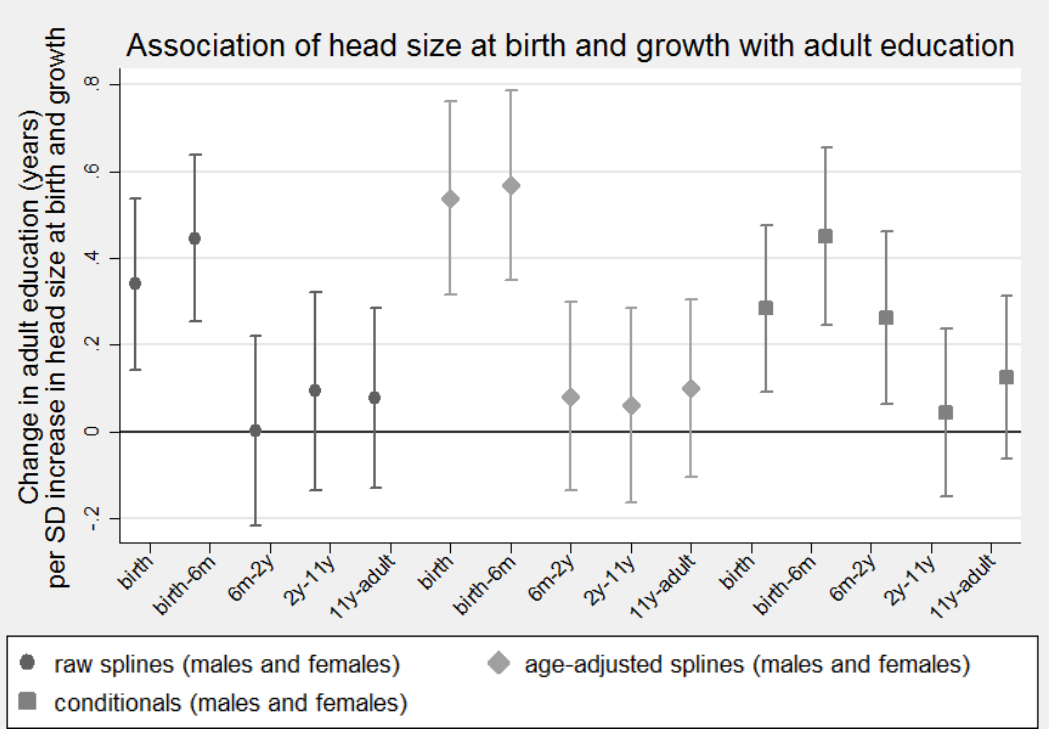
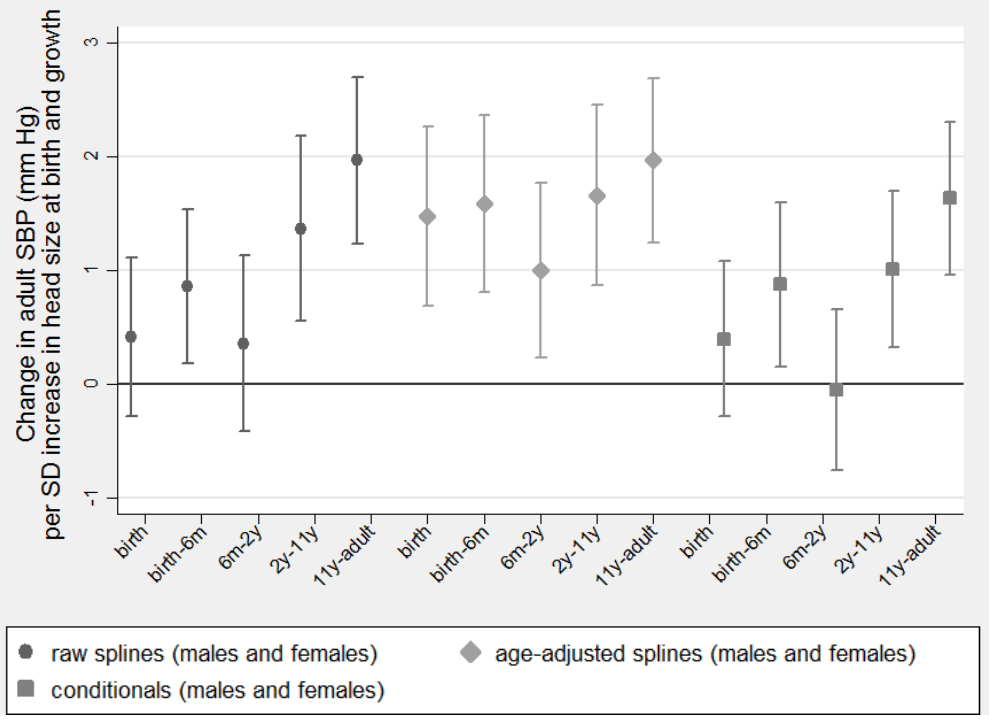


Figure 6.5 shows the associations from the three models. Both spline approaches gave similar results. Greater than average head size at birth, and greater than average head growth between birth and 6 months, were positively associated with years of adult schooling in all three models. However, results from the spline and conditional approaches differed for head growth between 6m-2y, being statistically significant using the conditional method. The difference in head growth between the two methods is greatest at 6m-2y (Figure 6.2). This may be because the spline method identifies the full trajectory of head growth and plots the fitted line, whereas the conditional model only uses data at the knots, as discussed following Figures 6.3 and 6.4. The spline and conditional methods show similar associations in the remaining time periods.

6.3.2 Adult cardiometabolic outcomes

Outcome: Systolic blood pressure

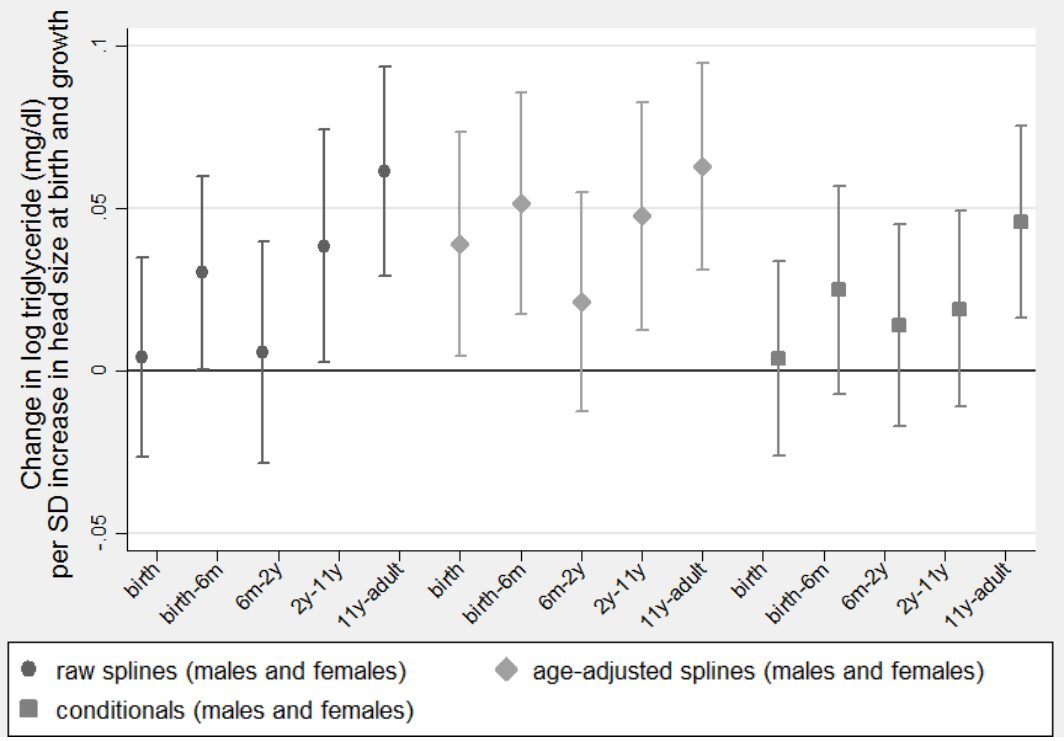
Figure 6.6: Association of head size at birth and growth with adult systolic blood pressure (n=1031)



In the case of SBP for each method (Figure 6.6), the regression coefficients became increasingly positive from birth to adulthood with the exception of 6m-2y, which was always the least positive. However, none of the fifteen regression coefficients comparisons for the same age intervals were statistically significant.

Outcome: Plasma triglycerides

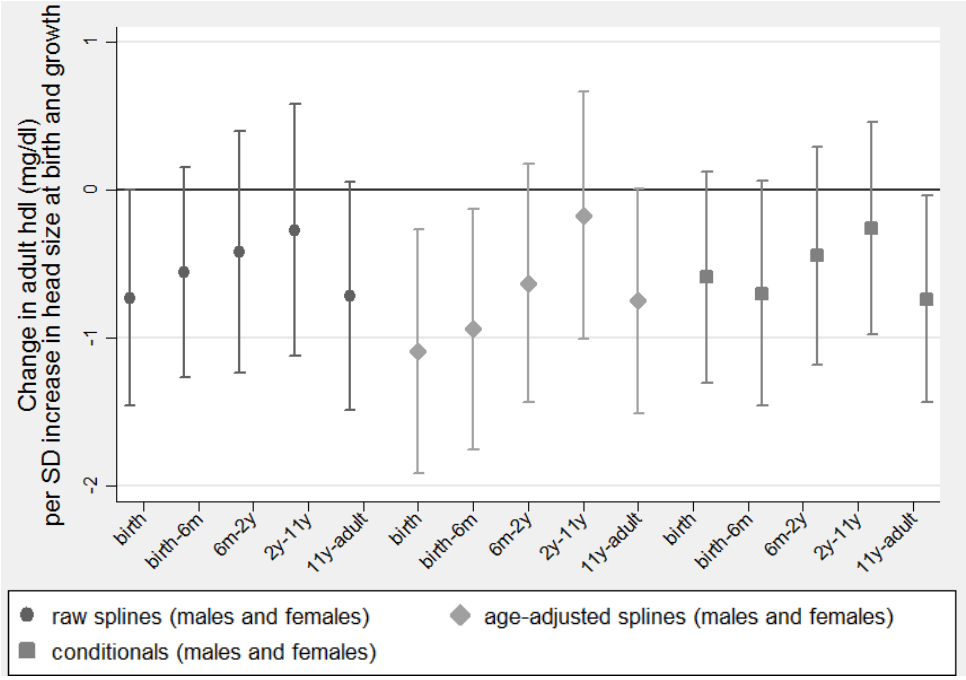
Figure 6.7: Association of head size at birth and growth with adult triglycerides (n=1026)



In the case of triglyceride for each method (Figure 6.7), the regression coefficients became increasingly positive from birth to 6m, then fell before continuing to rise between 2 years to adulthood. However, none of the fifteen regression coefficients comparisons for the same age intervals were statistically significant.

Outcome: Plasma HDL cholesterol

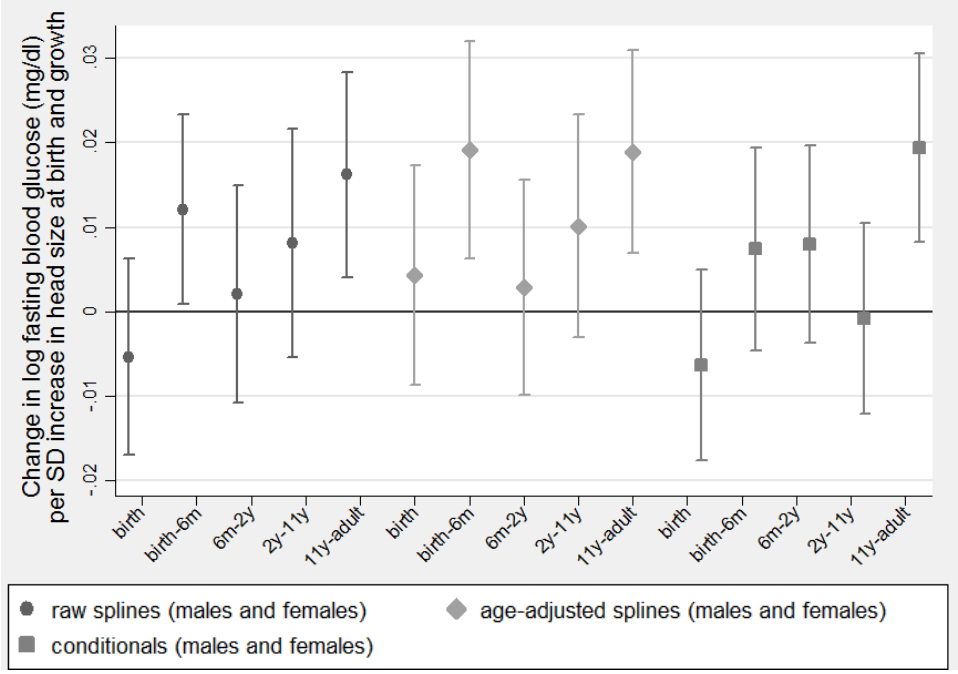
Figure 6.8: Association of head size at birth and growth with adult high density lipoprotein (n=1025)



In the case of HDL for each method (Figure 6.8), the regression coefficients were negative, but increased from birth to 11 years and then fell. However, none of the fifteen regression coefficients comparisons for the same age intervals were statistically significant.

Outcome: Fasting plasma glucose

Figure 6.9: Association of head size at birth and growth with adult fasting glucose (n=1026)



For fasting plasma glucose (Figure 6.7) the pattern of associations was similar to that seen for triglycerides (Figure 6.9), though the association for the conditional model between 2y-11y was a bit weaker.

6.4 Goodness of fit of conditional and spline approaches

I now compared the model fit in the two approaches in order to assess whether the spline model provides a significantly better fit as compared to the conditional approach. There are several ways to assess the goodness of fit (150). These include calculating the mean square error (MSE) and the R-square (R^2) statistic (150). If it can be assumed that the correct model gives an unbiased estimate of the variance, the values for MSE can be compared between different models. R^2 is the proportion of variation in the dependent variable explained by the model fit using the independent variables. This proportion can give an idea of how well the predicted equation fits the data. I have used R^2 as a goodness of fit measure to compare between the three modelling approaches. This was done for the different adult outcomes (Table 6.5).

Table 6.5: Goodness of fit of the two approaches for the various adult outcomes

Outcome	R^2	p-value for extra prediction beyond that		
		Model 1	Model 2	Model 3
1. Years of Education				
Males (n=615)				
Model 1: Spline using raw data	0.0290	X	0.083	0.265
Model 2: Spline using z-scores	0.0262	0.042	X	0.524
Model 3: Conditional	0.0299	0.318	0.873	X
Females (n=426)				
Model 1: Spline using raw data	0.0437	X	0.105	0.926
Model 2: Spline using z-scores	0.0509	0.310	X	0.901
Model 3: Conditional	0.0418	0.818	0.350	X
Males and Females				
Model 1: Spline using raw data	0.0489	X	0.425	0.588
Model 2: Spline using z-scores	0.0487	0.372	X	0.710
Model 3: Conditional	0.0467	0.396	0.560	X
2. SBP (mmHg)				
Males (n=610)				
Model 1: Spline using raw data	0.0634	X	0.077	0.430
Model 2: Spline using z-scores	0.0711	0.421	X	0.555
Model 3: Conditional	0.0615	0.300	0.072	X
Females (n=421)				
Model 1: Spline using raw data	0.0122	X	0.401	0.308
Model 2: Spline using z-scores	0.0109	0.340	X	0.310
Model 3: Conditional	0.0135	0.364	0.431	X
Males and Females				
Model 1: Spline using raw data	0.0382	X	0.092	0.440
Model 2: Spline using z-scores	0.0406	0.222	X	0.476
Model 3: Conditional	0.0359	0.204	0.093	X
3. Triglyceride (mg/dl)				
Males (n=604)				
Model 1: Spline using raw data	0.0314	X	0.467	0.659
Model 2: Spline using z-scores	0.0336	0.660	X	0.634
Model 3: Conditional	0.0265	0.278	0.167	X
Females (n=422)				
Model 1: Spline using raw data	0.0165	X	0.043	0.517
Model 2: Spline using z-scores	0.0162	0.041	X	0.063
Model 3: Conditional	0.0127	0.323	0.036	X

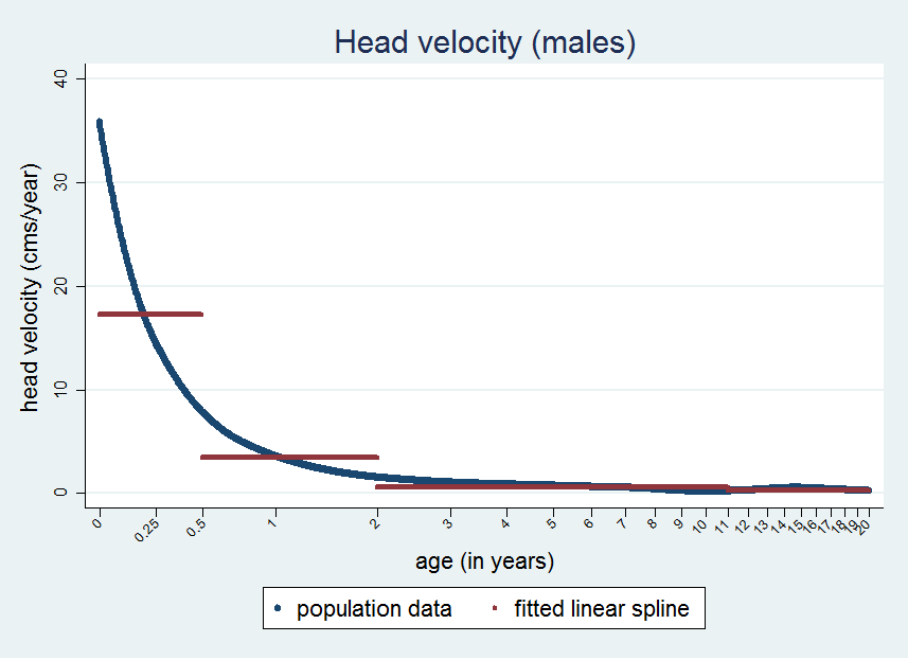
Males and Females				
Model 1: Spline using raw data	0.0193	X	0.336	0.996
Model 2: Spline using z-scores	0.0202	0.454	X	0.924
Model 3: Conditional	0.0137	0.293	0.146	X
4. HDL (mg/dl)				
Males (n=605)				
Model 1: Spline using raw data	0.0221	X	0.803	0.891
Model 2: Spline using z-scores	0.0228	0.862	X	0.890
Model 3: Conditional	0.0217	0.852	0.792	X
Females (n=420)				
Model 1: Spline using raw data	0.0032	X	0.015	0.677
Model 2: Spline using z-scores	0.0041	0.018	X	0.634
Model 3: Conditional	0.0062	0.865	0.769	X
Males and Females				
Model 1: Spline using raw data	0.0119	X	0.167	0.518
Model 2: Spline using z-scores	0.0133	0.269	X	0.720
Model 3: Conditional	0.0131	0.703	0.695	X
5. Fasting blood glucose (mg/dl)				
Males (n=604)				
Model 1: Spline using raw data	0.0133	X	0.020	0.285
Model 2: Spline using z-scores	0.0172	0.051	X	0.373
Model 3: Conditional	0.0162	0.484	0.309	X
Females (n=422)				
Model 1: Spline using raw data	0.0165	X	0.826	0.398
Model 2: Spline using z-scores	0.0176	0.886	X	0.248
Model 3: Conditional	0.0178	0.465	0.256	X
Males and Females				
Model 1: Spline using raw data	0.0128	X	0.121	0.373
Model 2: Spline using z-scores	0.0151	0.276	X	0.359
Model 3: Conditional	0.0161	0.854	0.482	X

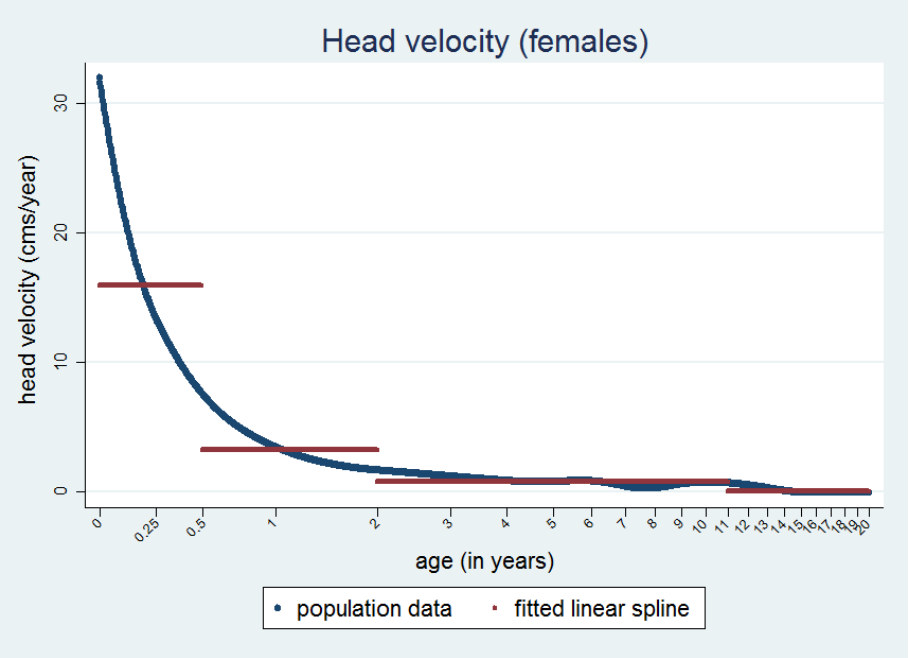
As there were no significant differences in the associations between males and females, I will describe only the pooled associations for the different outcomes. There were no statistically significant differences in the variability explained between the spline model and the conditional models for the different outcomes. The percentage of variability explained for the first two outcomes (years of education and SBP (4%)) is greater than that explained for the rest of the outcomes (Triglyceride, HDL and fasting blood glucose (1%)).

Growth velocity

Here I consider how well the linear spline model fits the head circumference data. The panels of Figure 6.10 shows two lines; the first is the head velocity obtained by differentiating the cubic spline fitted through all the head circumference measurements; the second is the head velocity implied by the linear spline fit.

Figure 6.10: Sex-specific velocity curves by cubic (population data) and linear splines





As can be seen from the above graph, the linear spline is discontinuous at the knots and between the knots it is a constant, while the cubic spline assumes a continuous head velocity, which is more realistic and biologically plausible. However, the number of parameters to be estimated in a cubic spline is greater than that in a linear spline.

Discussion

A lot of literature already exists on the modelling of child growth data using the spline approach (151,152) but none is focussed on head growth. The multilevel linear spline method uses a linear spline approach for describing head growth. The linear spline approach might not be appropriate for the current research because higher order regression splines may better capture the nonlinear trajectory between the knots, and might be more suitable for modelling child growth immediately after birth, when acceleration and velocity of head growth are the greatest (53). Also, higher order regression splines such as quadratic and cubic assume that the velocity of growth in children is continuous and smooth, whereas the linear splines assume that the velocity is constant in a small number of intervals which is an assumption that might be biologically implausible for body growth. But the higher order splines require the estimation of many more parameters as compared to the linear spline approach, and thus requires a larger number of head measurements per individual, which is a prerequisite that was not met in the current analyses. Also, since the approach based on the age-adjusted SD splines constructs the growth on a standardized scale, this approach might be more suitable than the approach based on the raw splines, as the variability in early head growth is more than head growth in later life, such as in adolescence.

Advantages of the linear spline approach as compared to the conditional approach

There is one major advantage of the spline approach over the conditional approach. The head circumference data in NDBC did not have equal measurements for every subject, and was therefore unbalanced. Linear mixed models are more suitable for the analysis of unbalanced data because the conditional models require an estimate at specific ages which might not be possible in a longitudinal study, whereas splines can calculate a deviation with no such assumption (63). This can be observed in the larger

number of head size at birth and growth measures calculated for the spline approach as compared to the conditional approach (Table 6.3).

Disadvantages of the linear spline approach as compared to the conditional approach

Although there can be individuals for whom it is possible to calculate a spline deviation but not a conditional deviation, but these deviations may be unreliable if there are few data points, when they may be shrunk towards the population slope. The spline method calculates the estimated deviations based on the population values in the interval for such individuals, and thus assigns average values in such cases ('shrinkage') (62). These deviations will have high variability, since they are based on the estimated population values. Also, while the conditionals are constructed to be uncorrelated to each other, the spline measurements are moderately correlated with each other (correlation between 6m-2y and 2y-11y for raw splines is r = -0.4). This can be seen by the slightly wider confidence intervals for the spline approach as compared to the conditional approach when analysing the individuals having both splines and conditionals as predictors of adult outcomes such as years of education.

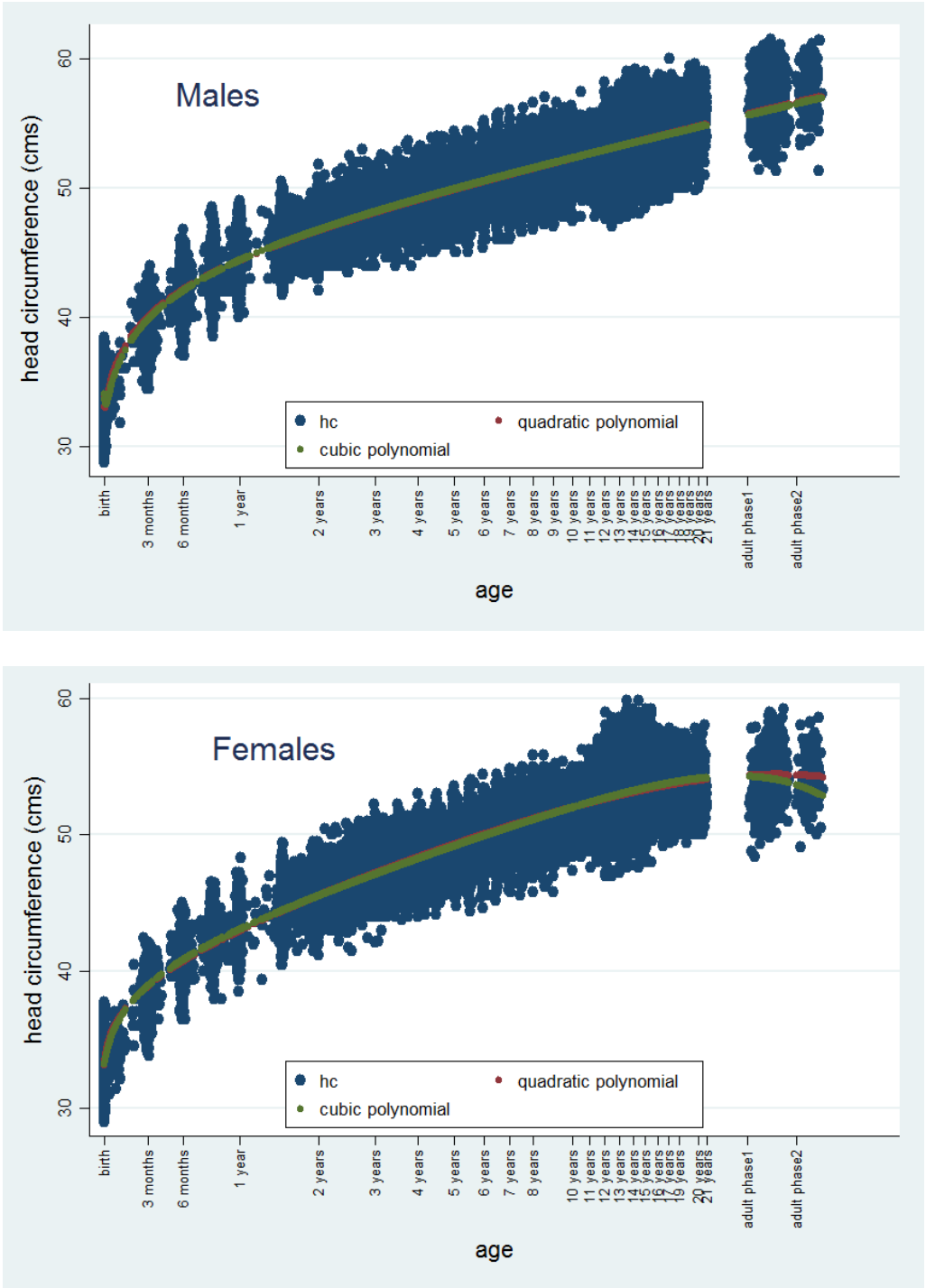
The model fit was compared between the three models (raw splines, SD splines and conditional model) using the R^2. Values for the three models were similar, and the difference in predictive ability between the models was not statistically significant (all p-values >0.05). Therefore, the estimated deviations from the spline approach were less precise than the conditional approach, with little difference in model fit. This suggests that the spline models predicts the outcomes as well as the conditional models because they have less bias.

Other approaches for modelling body growth

Other approaches are available for modelling body growth, which do not assume the linear growth assumed by the spline model. Fractional polynomials are useful for modelling the growth structure (75,152). The issue is, however, these models derive the powers from a predefined set of numbers and hence may be unsuitable. These were described in chapter-3 (section 3.2.2, Page 42). A three degree fractional polynomial was fit to the head circumference data. The curve clearly does not fit the data well for the females as the fitted line for a quadratic fractional polynomial looks different from the fitted line for a cubic fractional polynomial in the adult age period (Figure 6.11).

The line for the fitted value using the cubic fractional polynomial is a poorer fit as compared to a quadratic fractional polynomial passing above the average values in infancy, and the AIC for this model is lesser. The age axis of the graphs are in log scale to depict the rapid head growth until infancy.

Figure 6.11: Sex-specific average head growth characterized by a quadratic and cubic fractional polynomial model



Conclusion

Thus, in choosing between the conditional and the spline approach, the choice of approach depends on the number of measurements available per individual in the dataset and their distribution across the age range. When a study has measurements at all of a small number of defined time points the conditional approach is suitable. This chapter shows that for a large dataset in the sense of multiple measurements at different time points, both the spline method and the conditional method give similar results. For the current dataset both the spline and the conditional approach seem to give similar results, with little difference in model fit or in the precision of the associations at birth and in the different growth periods.

Conclusions and proposed future work

According to the fetal programming theory described by Barker, an undernourished fetus uses the scarce nutrition provided for the development of the most important parts of the body such as the brain, and restricts the development of the rest of the body, such as the trunk and muscles. There are only a few previous studies which have looked at the associations of head size and growth, or the ratios of the head to the rest of the body, with later life outcomes. This may be because head is not commonly measured at birth or in childhood. Thus, the current thesis aimed to study the associations between head size and growth, and also the ratio of head size to body size (length) at birth, with several adult outcomes and also outcomes in the subsequent generation in the NDBC study.

My first objective was to study the associations of head size at birth and head growth with human capital outcomes in adulthood in the same generation. I used educational attainment as a proxy for cognitive ability. In this analysis, I found head size at birth and head growth up to 2 years to be positively associated with years of education. However, after adjusting for height and BMI growth there were no significant associations between head size at birth or head growth at any age and attained education. Greater height growth up to 2 years and greater BMI growth from 6 months to 2 years remained positively associated with years of education. These associations became considerably attenuated after adjustment for parental SES, although both height and BMI growth from 6 months to 2 years remained significantly positively associated with attained education. SES was independently positively associated with years of education. Paternal occupation, paternal education, maternal education, assets (wealth) and crowding index were components of SES associated independently with educational attainment. Since height and BMI growth from 6 months to 2 years remained associated with education, independent of SES, we can say that overall body growth was more important than specifically head growth. This result suggests that early life growth failure could be a marker of malnutrition that can affect neurological development leading to lowered educational attainment. Also, improved socioeconomic conditions could impact attainment of education significantly.

My second objective was to look at the associations of head size at birth and head growth with adult SBP, which is a CVD risk factor. This analysis showed

contrasting results to those observed with years of education. In the unadjusted analysis I found that head size at birth was not associated with SBP in adult life. An increase in head growth during early infancy, childhood and adolescence was associated with higher SBP in adult life. After adjusting for height and BMI growth, there were no statistically significant associations between head growth at any age and SBP. Greater than average BMI growth after 2 years was associated with higher SBP in adulthood. The strength and direction of these associations changed little after adjusting further for gestational age. There was no evidence that disproportion between the size of the head and the rest of the body at birth was associated with higher adult blood pressure. Therefore, there was no evidence that a 'brain-sparing' effect, in which the fetus adapts to preserve oxygen and nutrition supply to the brain, causes high blood pressure in this population. This analysis suggests that, consistent with many other studies, greater weight gain in childhood is associated with higher adult SBP. Improved childhood nutrition and preventing adiposity during childhood and adolescence may be important for preventing higher blood pressure in adulthood.

My third objective was to look at the associations of parental head size at birth and head growth with offspring birth weight. The statistical analysis technique used in this case is multilevel modelling. This is essential because many parents had more than one child in this study and their birth weights will be correlated, violating the standard statistical assumptions. There were no statistically significant differences in the associations between mothers and fathers, or between boys and girls, and therefore I considered them together. In the pooled analyses, parental head size at birth and growth in early infancy were positively associated with offspring birth weight. After adjusting for the effects of height and BMI growth parental head growth in early infancy remained positively associated. This is the first analysis looking at the associations of parental head size at birth and head growth with offspring birth weight. The associations of head size at birth and head growth with offspring birth weight were similar among mothers and fathers, which suggests that these may be due to genetic influences or persisting shared environment between the parents, rather than specific effects of the environment provided by the mother. Genetic influences cannot probably be modified. However, an understanding of which environmental factors influence brain growth may help in increasing birth weight in the next generation.

My fourth objective was to compare the conditional body size approach, which is the primary statistical model used in this thesis, to an alternative approach i.e. splines. Both raw and age-adjusted splines were considered. The choice of approach depends on the study design. In my dataset both the spline method and the conditional method gave similar results.

The present work has only evaluated the associations of head size at birth and head growth with later life cardiometabolic and cognitive outcomes in the same generation and birth weight in the next generation. It would be interesting to evaluate these results in other birth cohorts. Some recently established birth cohorts have collected longitudinal head size data, for example the Mysore Parthenon Cohort and Pune Maternal Nutrition Study (153,154). Associations of head size at birth and head growth can also be evaluated for other cardiometabolic outcomes such as HDL, LDL, fasting plasma glucose and plasma triglycerides. Associations of head size at birth and head growth with other outcomes, such as age at adiposity rebound, could also be explored. Head size at birth and head growth can also be seen as outcome variables and factors associated with them can be evaluated. These factors can be divided into non-modifiable such as heredity, sex and race, and modifiable such as exercise, socioeconomic conditions and presence of infections during early life. Alternative growth methods for modelling head circumference such as group-based trajectory modelling can also be studied (155). The outcome in these models will be repeated cardiometabolic outcomes, and the exposures will be head size at birth and the conditional head measures. Another analysis that can be considered is growth mixture modelling, which can be used to identify individuals with similar growth trajectories, which are then associated with a later life outcome which is measured at a single time point (65).

Appendix

I have presented the following three seminars:

1. "Association of childhood head growth with adult systolic blood pressure: findings from the New Delhi Birth Cohort" at the International Society for Clinical Biostatistics conference in Birmingham, UK in August 2016 (oral).

2. "Association of head growth with adult human capital: findings from the New Delhi birth cohort study" at the annual conference of Society for Natal Effects on Health in Adults (SNEHA) conference in September 2014 (poster).

3. "Associations of childhood (F1) head growth with birth weight in the next generation (F2): findings from the New Delhi Birth Cohort" at the annual conference of Society for Natal Effects on Health in Adults (SNEHA) conference in February 2016 (poster).

List of References

1. Barker DJ, Winter PD, Osmond C, Margetts B, Simmonds SJ. Weight in infancy and death from ischaemic heart disease. Lancet. 1989 Sept; 2(8663):577-80.

2. Osmond C, Barker DJP, Winter PD, Fall CHD, Simmonds SJ. Early growth and death from cardiovascular disease in women. BMJ. 1993 December; 307(6918):1519-24.

3. Hales CN, Barker DJ, Clark PM, Cox LJ, Fall C, Osmond C, Winter PD. Fetal and infant growth and impaired glucose tolerance at age 64. BMJ. 1991 Oct; 303(6809):1019-22.

4. Stein CE, Fall CH, Kumaran K, Osmond C, Cox V, Barker DJ. Fetal growth and coronary heart disease in South India. Lancet. 1996 Nov; 348(9037):1269-73.

5. Eriksson JG, Forsen T, Tuomilehto J, Winter PD, Osmond C, Barker DJ. Catch-up growth in childhood and death from coronary heart disease: longitudinal study. BMJ. 1999 Feb; 318(17181):427-31.

6. Eriksson JG, Forsen T, Tuomilehto J, Osmond C, Barker DJ. Early adiposity rebound in childhood and risk of type 2 diabetes in adult life. Diabetologia. 2003 Feb; 46(2):190-4.

7. Barker DJ. Fetal origins of coronary heart disease. BMJ. 1995; 311 (6998):171-4.

8. Barker DJP. Mothers, Babies and Health in Later life. 2nd ed. London: BMJ Publication group; 1998.

9. Eriksson JG. Developmental Origins of Health and Disease- from a small body size at birth to epigenetics. Ann Med. 2016; 48(6):456-67.

10. Huxley R, Owen CG, Whincup PH, Cook DG, Rich-Edwards J, Smith GD, et al. Is birth weight a risk factor for ischemic heart disease in later life?. Am J Clin Nutr. 2007 May; 85(5):1244-50.

11. Law CM, Sheill AW. Is blood pressure inversely related to birth weight? The strength of evidence from the systematic review of the literature. J Hypertens. 1996 Aug; 14(8):935-41.

12. Mu M, Wang SF, Sheng J, Zhao Y, Li HZ, Hu CL, et al. Birth weight and subsequent blood pressure: a meta-analysis. Arch Cardiovasc Dis. 2012 Feb; 105(2):99-113.

13. Harder T, Rodekamp E, Schellong K, Dudenhausen JW, Plagemann A. Birth weight and subsequent risk of type 2 diabetes: a meta-analysis. Am J Epidemiol. 2007 Apr; 165(8):849-57.

14. Whincup PH, Kaye SJ, Owen CG, Huxley R, Cook DG, Anazawa S, et al. Birth weight and risk of type 2 diabetes: a systematic review. JAMA. 2008 Dec; 300(24):286-97.

15. Silveria VM, Horta BL. Birth weight and metabolic syndrome in adults: meta-analysis. Rev Saude Publica. 2008 Feb; 42(1):10-8.

16. Owen CG, Whincup PH, Odoki K, Gilg JA, Cook DG. Birth weight and blood cholesterol level: a study in adolescents and systematic review. Paediatrics. 2003 May; 111(5 Pt 1):1081-9.

17. Huxley R, Owen CG, Whincup PH, Cook DG, Colman S, Collins R. Birth weight and subsequent cholesterol levels: exploration of the “fetal origins” hypothesis. JAMA. 2004 Dec; 292(22):2755-64.

18. Barker DJ, Osmond C, Forsen TJ, Kajantie E, Eriksson JG. Trajectories of growth among children who later have coronary events as adults. N Eng J Med. 2005 Oct; 353(17):1802-9.

19. Osmond C, Kajantie E, Forsen TJ, Eriksson JG, Barker DJ. Infant growth and stroke in later life: the Helsinki birth cohort study. Stroke. 2007 Feb; 38(2):264-70.

20. Eriksson JG, Forsen TJ, Kajantie E, Osmond C, Barker DJ. Childhood growth and hypertension in later life. Hypertension. 2007 Jun; 49(6):1415-21.

21. Bhargava SK, Sachdev HS, Fall CH, Osmond C, Lakshmy R, Barker DJ et al. Relation of serial changes in childhood body-mass index to impaired glucose tolerance in young adulthood. N Eng J Med. 2004 Feb; 350(9):865-75.

22. Singhal A, Cole TJ, Fewtrell M, Deanfield J, Lucas A. Is slower early growth beneficial for long-term cardiovascular health? Circulation. 2004 Mar; 109(9):1108-13.

23. Ekelund U, Ong KK, Lime Y, Neovius M, Brage S, Dunger DB et al. Association of weight gain in infancy and early childhood with metabolic risk in young adults. J Clin Endocrinol Metab. 2007 Jan; 92(1):98-103.

24. Eriksson J, Forsen T, Tuomilehto J, Osmond C, Barker D. Fetal and childhood growth and hypertension in adult life. Hypertension. 2000 Nov;36(5):790-4.

25. Forsen T, Osmond C, Eriksson JG, Barker DJ. Growth of girls who later develop coronary heart disease. Heart. 2004 Jan; 90(1):20-4.

26. Baker JL, Olsen LW, Sørensen TI. Childhood body-mass index and risk of coronary heart disease in adulthood. N Eng J Med 2007 Dec; 357(23):2329-37.

27. Adair LS, Fall CH, Osmond C, Stein AD, Martorell R, Ramirez-Zea M et al. Associations of linear growth and relative weight gain during early life with adult health and human capital in countries of low and middle income: findings from five birth cohort studies. Lancet. 2013 Aug; 382(9891):525-34.

28. Cambell MW, Williams J, Carlin JB, Wake M. Is the adiposity rebound a rebound in adiposity? Int J Pediatr Obes. 2011; 6:207-15.

29. Eriksson JG, Forsen T, Tuomilehto J, Osmond C, Barker DJ. Early growth and coronary heart disease in later life: longitudinal study. BMJ. 2001 Apr; 322(7292):949-53.

30. Yajnik CS, Fall CH, Coyaji KJ, Hirve SS, Rao S, Barker DJ et al. Neonatal anthropometry: the thin-fat baby. The Pune Maternal Nutrition Study. Int J Obes Relat Metab Disord. 2003; 27(2):173-80.

31. Barker DJ, Godfrey KM, Osmond C, Bull A. The relation of fetal length and ponderal index and head circumference to blood pressure and risk of hypertension in adult life. Peadiatr Perinat Epidemiol. 1992 Jan; 6(1):35-44.

32. Barker DJ, Osmond C, Simmonds SJ, Wield GA. The relation of small head circumference and thinness at birth to death from cardiovascular disease in adult life. BMJ. 1993 Feb; 306(6875): 422-6.

33. Risnes KR, Nilsen TI, Romundstad PR, Vatten LJ. Head size at birth and long-term mortality from coronary heart disease. Int J Epidemiol. 2009 Aug; 38(4):955-62.

34. Eriksson JG, Osmond C, Kajantie E, Forsen TJ, Barker DJ. Patterns of growth among children who later develop type-2 diabetes or its risk factors. Diabetologia. 2006 Dec; 49(12):2853-8.

35. Eriksson JG, Kajantie E, Lampl M, Osmond C, Barker DJ. Small head circumference at birth and early age at adiposity rebound. Acta Physiol (Oxf). 2014 Jan; 210(1):154-60.

36. Shenkin SD, Starr JM, Deary IJ. Birth weight and cognitive ability in childhood: a systematic review. Psychol Bull. 2004 Nov; 130(6):989-1013.

37. Mattle TD, Breshnan M, Begg MD, Susser E. Influence of variation in birth weight within normal range and within sibships on IQ at 7 years: Cohort Study. BMJ. 2001; 323(7308):310-4.

38. Jefferis BJ, Power C, Hertzman C. Birth weight, childhood socioeconomic environment, and cognitive development in the 1958 British birth cohort study. BMJ. 2002 Aug;325(7359):305.

39. Victora CG, Adair L, Fall C, Hallal PC, Martorell R, Richter T et al. Maternal and child under nutrition: consequences for adult health and human capital. Lancet. 2008; 371(9609):340-57.

40. Martorell R, Horta BL, Adair LS, Stein AD, Ritcher L, Fall CH et al. Weight gain in the first two years of life is an important predictor of schooling outcomes in pooled analysis of five birth cohorts from low and middle income countries. J Nutr. 2010 Feb; 140(2):348-54.

41. Hodinott J, Behrmann JR, Maluccio JA, Melgan P, Quisumbing AR, Ramirez-Zea M et al. Adult consequences of growth failure in early childhood. Am J Cli Nutr. 2013 Nov; 98(5):1170-8.

42. Raikonnen K, Forsen T, Henriksson M, Kajantie E, Heinonen K, Pesonnen AK et al. Growth trajectories and intellectual abilities in young adulthood: the Helsinki Birth Cohort Study. Am J Epidemiol. 2009 Aug; 170(4):447-55.

43. Martyn CN, Gale CR, Sayer AA, Fall C. Growth in utero and cognitive function in adult life: follow-up study of people born between 1920 and 1943. BMJ. 1996 Jun; 312(7043):1391-6.

44. Veena SR, Krishnaveni GV, Wills AK, Kurpad AV, Muthayya S, Hill JC et al. Association of birth weight and head circumference at birth to cognitive function in 9 to 10-year-old children in South India: prospective birth cohort study. Pediatr Res. 2010 Apr; 67(4):424-9.

45. Gale CR, Walton S, Martyn CN. Fetal and postnatal head growth and risk of cognitive decline in old age. Brain. 2003 Oct; 126(10):2273-8.

46. Gale CR, O'Callaghan FJ, Godfrey KM, Law CM, Martyn CN. Critical periods of brain growth and cognitive function in children. Brain. 2004 Feb; 127(2):321-9.

47. Gale CR, O'Callaghan FJ, Bredow M, Martyn CN. The influence of head growth in fetal life, infancy and childhood on intelligence at the ages of 4 and 8 years. Paediatrics. 2006 Oct; 118(4):1486-92.

48. Huang C, Martorell R, Ren A, Li Z. Cognition and behavioural development in early childhood: the role of birth weight and postnatal growth. Int J Epidemiol. 2013 Feb; 42(1):160-71.

49. Lira PI, Eickmann SH, Lima MC, Amorim RJ, Emond AM, Ashworth A. Early head growth: relation with IQ at 8 years and determinants in term infants of low and appropriate birth weight. Dev Med Child Neurol. 2010 Jan; 52(1):40-6.

50. Drake AJ, Walker BR. The intergenerational effects of fetal programming: non-genomic mechanisms for the inheritance of low birth weight and cardiovascular risk. J Endocrinol. 2004 Jan; 180(1):1-16.

51. Davey Smith G, Hart C, Ferrell C, Upton M, Hole D, Hawthorne V et al. Birth weight of offspring and mortality in the Renfrew and Paisley study: prospective observational study. BMJ. 1997 Nov; 315(7117):1189-93.

52. Barker DJ, Sheill AW, Barker ME, Law CM. Growth in utero and blood pressure levels in the next generation. J Hypertens. 2000 Jul;18(7):843-6.

53. Tanner JM. Fetus into Man- Physical growth from conception to Maturity. 2nd ed. USA: Harvard University Press; 1990.

54. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet. 2013;382(9890):427-51.

55. O. G. Eiben, A. Barabbas, G. Kontra, and E. Panto, "Differences in growth and physical fitness of Hungarian urban and rural boys and girls," HOMO, vol. 47, no. 1–3, pp. 191–205, 1996.

56. Silva PA, Birbeck J, Williams S. Some factors influencing the stature of Dunedin 7 year old children: a report from the Dunedin Multidisciplinary

Health and Development Research Unit. Aust Pediatr J. 1985 Feb: 21(1):27-30.

57. Krieger J, Higgins DL. Housing and health: Time again for public health action. Am J Public Health. 2002; 92(5):758-68.

58. Cole TJ. Modelling postnatal exposures and their interactions with birth size. J Nutr. 2004 Jan; 134(1):201-4.

59. Gamborg M, Andersen PK, Baker JL, Budtz-Jørgensen E, Jørgensen T, Jansen G et al. Life course path analysis of birth weight, childhood growth and adult systolic blood pressure. Am J Epidemiol. 2009 May; 169(10):167-78.

60. Wills, A.K., Strand, B.H., Glavin, K., Silverwood, R.J., Hovengen, R. Regression models for linking patterns of growth to a later outcome: infant growth and childhood overweight. BMC Medical Research Methodology. 2016; 16:41.

61. Tu YK, Tilling K, Sterne JA, Gilthorpe MS. A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. Int J Epidemiol. 2013;42(5):1327-39.

62. Goldstein H. Multilevel statistical models. 4th ed. United Kingdom: Wiley Series in Probability and Statistics; 2011.

63. Howe LD, Tilling K, Matiajasevich A, Petherick ES, Santos AC, Fairley L. Linear spline multilevel models for summarizing childhood growth trajectories: a guide to their application using examples from five birth cohorts. Stat Methods Med Res. 2013 Oct; 0(0):1-21.

64. Little RJA, Rubin DB. Statistical analysis with missing data; 2nd ed. New Jersey: John Wiley and Sons; 2002.

65. Muthen B, Muthen LK. Integrating person-centred and variable-centred analysis: growth mixture-modelling with latent trajectory classes. Alcohol Clin Exp Res. 2000 Jun; 24(6):882-91.

66. Cole TJ, Donaldson MD, Ben-Shlomo Y. SITAR- a useful instrument for growth curve analysis. Int J Epidemiol. 2010; 39(6):1558-66.

67. Wills, A.K., Silverwood, R.J. and De Stavola, B.L. Comment on Tu et al. 2013. A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. Int J Epidemiol. 2014;43 (5), 1662-64.

68. Tu, Y.K., Tilling, K., Sterne, J.A. and Gilthorpe, M.S. Authors' reply to the letter to the editor by Wills et al. Int J Epidemiol. 2014; 43 (5), 1664-1665.

69. Kristman V, Manno M, Cote P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004; 19(8):751-60.

70. Royston P. Constructing time-specific reference ranges. Stat Med. 1991 May; 10(5):675-90.

71. Royston P, Wright EM. How to construct 'normal ranges' for fetal variables. Ultrasound Obstet Gynecol. 1998; 11(1):30-8.

72. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Appl Statis. 1994; 43:429-67.

73. Silverwood RJ, Cole TJ. Statistical methods for constructing gestational age-related intervals and centile charts for fetal size. Ultrasound Obstet Gynecol. 2007;29(1):6-13.

74. Cole TJ. Fitting smoothed centile curves to reference data(with discussion). J.R. Statis. Soc A. 151:385-418.

75. Cole TJ. Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med. 1992; 11(10):1305-19.

76. Box GEP, Cox DR. An analysis of transformations(with discussion). J.R. Statist. Soc B. 1964.26:211-52.

77. Wright EM, Royston P. A comparison of statistical methods for age-related reference intervals. J.R. Statis Soc A. 1997;160:47-69.

78. van Burren S, Fredriks M. Worm plot: a simple diagnostic device for modelling growth reference curves. Stat Med. 2001; 20(8):1259-77.

79. Fisher RA, Yates F. Statistical tables for biological, agricultural, and medical research. 1st edn. Oliver & Boyd; Edinburgh: 1938.

80. Blom G. Statistical estimates and transformed beta-variables. New York: Wiley; 1958.

81. Sayers A, Heron J, Smith A, et al. Joint modelling compared with two stage methods for analysing longitudinal data and prospective outcomes: A simulation study of childhood growth and BP. Statistical Methods in Medical Research. 2016;26(1):437-452.

82. World Health Organization. Physical Status: The Use and Interpretation of Anthropometry. Report of a WHO Expert Committee. (WHO Technical Report Series no. 854). Geneva, Switzerland: World Health Organization; 1995.

83. Vyas S, Kumaranayake L. Constructing socioeconomic status indices: how to use principal components analyses. Health Policy Plan. 2006 Nov; 21(6):459-68.

84. Balen J, Mc Manus DP, Li YS, Zhao ZY, Yuan LP, Utzinger J et al. Comparison of two approaches for measuring household wealth via an asset based index in rural and peri-urban settings of Hunan province, China. Emerg Themes Epidemiol. 2010 Sep; 7(1):7.

85. Gunnsteinsson S, Labrique AB, West KP Jr, Christian P, Mehra S, Shamim AA et al. Constructing indices of rural living standards in Northwestern Bangladesh. J Health Popul Nutr. 2010 Oct; 28(5):509-19.

86. Howe LD, Hargreaves JR, Huttly SR. Issues in the construction of wealth indices for the measurement of socioeconomic position of low-income countries. Emerg Themes Epidemiol. 2008 Jan; 5:3.

87. Jarell MG. A comparison of two procedures, the Mahalnobis Distance and Andrew-Pregibon statistic, for identifying multivariate outliers. Research in the schools. 1994; 1:49-58.

88. Anscombe FJ. Rejection of outliers. Technometrics. 1960; 2; 123-47.

89. Winkler WE. Problems with inliers. Washington (DC) : Census Bureau Research Reports Series RR 98/05. Available: http://www.census.gov/srd/papers/pdf/rr9805.pdf. Accessed 1 August 2016.

90. Pan H, Cole T. LMS Chartmaker. A program to construct growth references using the LMS method. Version 2.54. 2011.

91. Stasinopoulos M, Rigby B, Akantziliotou C. 2nd edn. 2008. p. 206. Instructions on How to Use the gamlss Package in R.

92. Bhalla AK, Walia BN. Centile growth charts for head circumference in Punjabi infants. Indian Pediatr. 1993; 30(1):41-6.

93. Agarwal DK, Agarwal KN. Physical growth in Indian affluent children (birth-6 years). Indian Pediatr. 1994; 31(4):377-413.

94. Sachdev HPS. Nutritional status of children and women in India: Recent Trends. NFI Bull. 1997;18(3):2-5.

95. Karabiber H, Durmaz Y, Yakinci C, Katto O, Gumusalan Y, Yologu S et al. Head circumference measurement of urban children aged between 6 and 12 in Malatya, Turkey. Brain Dev. 2001; 23(8):801-4.

96. Ounsted M, Moar VA, Scott A. Head circumference charts updated. Arch Dis Child. 1985; 60(10):936-9.

97. Ishikawa T, Furuyama M, Ishikawa M, Ogawa J, Wada Y. Growth in head circumference from birth to fifteen years of age in Japan. Acta Pediatr Scand. 1987; 76(5):824-8.

98. Natale V, Rajagopalan A. Worldwide variation in human growth and the World Health Organization growth standards: a systematic review. BMJ Open. 2014; 4(1):e003735.

99. de Onis M, Garza C, Victora CG, Onyango AW, Frongilio EA, Martinez J. The WHO Multicenter Growth Reference Study: planning, study design and methodology. Food Nutr Bull. 2004; 25(1 Suppl):315-26.

100. Deary IJ, Strand S, Smith P, Fernandes C. Intelligence and educational achievement. Intelligence. 2007 Feb; 35(1):13-21.

101. Cooke RW, Lucas A, Yudkin PL, Pryse-Davies J. Head circumference as an index of brain weight in the fetus and newborn. Early Hum Dev. 1977 Oct;1(2):145-9.

102. Bartholomeusz HH, Courchesene E, Karns CM. Relationship between head circumference and brain volume in healthy normal toddlers, children and adults. Neuropediatrics. 2002;33(5):239-41.

103. Wickett JC, Vernon PA, Lee DH. Relationships between factors of intelligence and brain volume. Personality Individ Diff. 2000; 29: 1095-122.

104. Ceci SJ, Williams WM. Schooling, intelligence, and income. Am Psychol 1997; 52:1051–8.

105. Bradley RH, Caldwell BM, Rock SL. Home environment and school performance: a ten-year follow-up and examination of three models of environmental action. Child Dev. 1988 Aug; 59(4):852-67.

106. Walter S, Teimeier H. Variable selection: current practice in epidemiological studies. Eur J Epidemiol. 2009; 24(12):733-6.

107. Martyn CN, Barker DJ, Jespersen S, Greenwald S, Osmond C, Berry C. Growth in utero, adult blood pressure and arterial compliance. Br Heart J. 1995 Feb; 73(2):116-21.

108. Kumaran K, Fall CH, Martyn CN, Vijaykumar M, Stein C, Shier R. Blood pressure, arterial compliance and left ventricular mass: no relation to small size at birth in south Indian adults. Heart. 2000 Mar;83(3):272-7.

109. He J, Welton PK. Elevated systolic blood pressure as a risk factor for cardiovascular and renal disease. J Hypertens Suppl. 1999 Jun;17(2):s7-13.

110. Hardy R, Wardsworth ME, Langenberg C, Kuh D. Birth weight, childhood growth, and blood pressure at 43 years in a British birth cohort. Int J Epidemiol. 2004 Feb; 33(1):121-9.

111. Zhao M, Shu XO, Jin F, Yang G, Li HL, Liu DK et al. Birth weight, childhood growth and hypertension in adulthood. Int J Epidemiol. 2002 Oct. 31(5):1043-51.

112. Huxley R, Neill A, Collins R. Unravelling the fetal origins hypothesis: Is there really an inverse association between birth weight and subsequent blood pressure? Lancet. 2002;360:659-65.

113. Sorensen HT, Thulstrup AM, Norgdard B et al. Fetal growth and blood pressure in a Danish population aged 31-51 years. Scand Cardiovascular J. 2000;34:390-5.

114. Whincup P, Cook D, Papacosta O, Walker M. Birth weight and blood pressure: cross-sectional and longitudinal relations in childhood. BMJ. 1995; 311:773-6.

115. Chaudhari S, Bhaleraa MR, Vaidya U, Pandit A, Nene U. Growth and development of twins compared with singletons at ages one and four years. Indian Pediatr. 1997 Dec;34(2):1081-6.

116. Kuzawa CW, Sweet E. Epigenetics and the embodiment of race: developmental origins of US disparities in cardiovascular health. Am J Hum Biol. 2009; 21(1):2-15.

117. Doumas D et al. The intergenerational transmission of aggression across 3 generations. J Fam Violence. 1994. 9:157-75.

118. Kerr DC, Capaldi DM, Pears KC, Owen LD. A prospective three generational study of father's constructive parenting: influences from family of origin, adolescent adjustment and offspring temperament. Dev Psychol. 2009. 45(5):1257-75.

119. Smith CA, Farrington DP. Continuities in antisocial behaviour and parenting across three generations. J Child Psychol Psychiatr. 2004;45(2):230-47.

120. Addo OY, Stein AD, Fall CH, Gigante DP, Gutupalli AM, Horta BL et al. Parental childhood growth and offspring birth weight: pooled analyses from four birth cohorts in low and middle income countries. Am J Hum Biol. 2015; 27(1):99-105.

121. Agnihotri B, Antonioswamy B, Priya G, Fall CH, Ragupathy P. Trends in human birth weight across two successive generations. Indian J Pediatr. 2008; 75(2):111-7.

122. Veena SR, Kumaran K, Swarnagowri MN, Jayakumar MN, Leary SD, Stein CE et al. Intergenerational effects on size at birth in South India. Pediatr. Perinat. Epidemiol.2004;18(5):361-70.

123. Kuzawa CW, Eisenberg DT. Intergenerational predictors of birth weight in the Philippines : correlations with mother's and father's birth weight and test of maternal constraint. PLoS One. 2012;7(7):e40905.

124. Horta BL, Gigante DP, Osmond C, Barros FC, Victora CG. Intergenerational effect of weight gain in childhood on offspring birth weight. Int J Epidemiol. 2009; 38(3):724-32.

125. Costa e Silva LI, Gomes FM, Valente MH, Escobar AM, Brentani AV, Grisi SJ. The Intergenerational effects on birth weight and its relations to maternal conditions, Sao Paulo, Brazil. Biomed Res Int. 2015:615034.

126. Han Z, Lutsir O, Mulla S, Mc Donald SD. Maternal height and the risk of preterm birth and low birth weight: a systematic review and meta-analyses. J Obstet Gynacol Can. 2012; 34(8):721-46.

127. Ramakrishnan U, Martorell R, Schroeder DG, Flores R. Role of Intergenerational effects on linear growth. J Nutr. 1999;129(25 Suppl):544S-549S.

128. Martin RM, Smith GD, Frankel S, Gunnell D. Parent's growth in childhood and the birth weight of their offspring. Epidemiology. 2004; 15(3):308-16.

129. Freathy RM, Bennet AJ, Ring SM, Shields B, Groves CJ, Timpson NJ et al. Type 2 diabetes risk alleles are associated with reduced size at birth. Diabetes. 2009;58(6):1428-33.

130. Freathy RM, Mook-Kanamori DO, Sovio U, Prokopenko I, Timpson NJ, Berry DG et al. Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. Nat Genet. 2010; 42(5):430-5.

131. Das JC, Paul N. Association of maternal socio-biological factors with birth weight of new born baby. JCMCTA.2008;19(1):37-42.

132. Kaur S, Upadhyay AK, Srivastava DK, Srivastava R, Pandey ON. Maternal correlates of birth weight of a new born: A hospital based study. Ind. J. Comm Health. 2014;26(2):187-91.

133. Karn MN, Penrose LS. Birth weight and gestation time in relation to maternal age, parity and infant survival. Ann Eugen, 1951; 16(2):147-164.

134. Roberts DF, Tanner RE. Effects of parity on birth weight and other variables in a Tanganyika Bantu sample. Br J Prev Soc Med,1963;17(4):209-215.

135. Wilcox MA, Chang AM, Johnson IR. The effects of parity on birth weight using successive pregnancies. Acta Obstet Gynecol Scand. 1996; 75(5);459-463.

136. Ong KK, Ahmed ML, Emmett PM et al. Association between postnatal catch-up growth and obesity in childhood; prospective cohort study. BMJ. 2000; 320(7240):967-971.

137. Khong TY, Adema ED, Erwich JJ. On an anatomical basis for the increase in birth weight in second and subsequent born children. Placenta. 2003;24(4):348-353.

138. Chattopadhyay A. Men in maternal care: evidence from India. J Biosoc Sci. 2012; 44(2):129-53.

139. Hypponen E, Power C, Smith GD. Parental growth at different life stages and offspring birth weight: an intergenerational cohort study. Peidatr. Perinat. Epidemiol. 2004;18(3):168-77.

140. Huber PJ. Robust statistics. New York: Wiley , 1981.

141. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrka. 1986; 73:13-22.

142. Agegboye AR, Heitmann B. Accuracy and correlates of maternal recall of birth weight and gestational age. BJOG. 2008; 115(7):886-93.

143. Tate AR, Dezateux C, Cole TJ, Davidson L. Factors affecting a mother's recall of her baby's birth weight. Int J Epidemiol. 2005;34(3):668-95.

144. O'Sullivan JJ, Pearce MS, Parker L. Parental recall of birth weight: how accurate is it? Arch Dis Child. 2000; 82(3):202-3.

145. Walton KA, Murray LJ, Gallagher AM, Cran GW, Savage MJ, Boraham C. Parental recall of birth weight: a good proxy for recorded birth weight? Eur J Epidemiol. 2000; 16(9):793-6.

146. Shenkin SD, Zhang MG, Der G, Mathur S, Mina TH, Reynolds RM. Validity of recalled v. recorded birth weight: a systematic review and meta-analysis. J Dev Orig Health Dis. 2017.8(2):137-148.

147. Bengston VL, Acock AC, Allen KR, Dilworth-Anderson P, Klein DM. Sourcebook of family theory and research. Sage Publications, Inc, 2005.

148. Frank E Harrell Jr. Regression Modelling Strategies. 2nd ed. Springer; 2006.

149. Robinson GK. That BLUP is a good thing: The estimation of Random Effects. Statistical Sc.1991;6(1):15-51.

150. Kutner MH, Naschtsheim CJ, Neter J, Li W. Applied linear statistical models. 5th ed. New York: Tata McGraw-Hill; 2005.

151. Fairley L, Patherick ES, Howe LD, Tilling K, Cameron N, Lawlor DA et al. Describing differences in weight and length growth trajectories between white and Pakistani infants in the UK: analysis of the Born in Bradford birth cohort study using multilevel linear spline models. Arch Dis Child. 2013; 98(4):274-9.

152. Howe LD, Tilling K, Matijasevich A, Petherick ES, Santos AC, Fairley L et al. Linear spline multilevel models for summarizing childhood growth trajectories: a guide to their application using examples from five birth cohorts. Stat Methods Med Res. 2016;25(5):1854-74.

153. Krishnaveni GV, Veena SR, Hill JC, Karat SC, Fall C. Cohort profile: Mysore parthenon birth cohort. Int J Epidemiol. 2015; 44(1):28-36.

154. Rao S, Yajnik CS, Kanade A, Fall CH, Margetts BM, Jackson AA et al. Intake of micronutrient-rich foods in rural Indian mothers is associated with the size of their babies at birth: Pune Maternal Nutrition Study. J Nutr. 2001; 131(4):1217-24.

155. Kagura J, Adair LS, Munthali RJ, Pettifor JM, Norris SA. Association between Early Life Growth and Blood Pressure Trajectories in Black South African Children. Hypertension. 2016; 68(5):1123-31.