Early life risk factors and epigenetic biomarkers of obesity across the life course

Natassia Robinson

Submitted to Newcastle University for the degree of Doctor of Philosophy

Institute of Health & Society

September 2019

Word count 78,257

Abstract

Obesity prevalence continues to rise and can be partially attributed to the obesogenic environment. However, there is increasing evidence that environmental exposures in early development can influence later-life disease, known as the Developmental Origins of Health and Disease hypothesis. Whilst some early life exposures have been associated with later-life adiposity, the underlying mechanism is less understood. One hypothesised mechanism is through epigenetic changes, such as DNA methylation.

Multiple longitudinal cohorts were used to investigate the hypothesis; DNA methylation is a mediating mechanism between early life events and subsequent obesity. The Newcastle Thousand Families (NTFS) and Gateshead Millennium (GMS) studies (established 1947 and 2000 respectively), were used to investigate the impact of early life exposures (i.e. socioeconomic status, growth, adversity) on childhood (GMS) and adult (NTFS) obesity. Using both cohorts provided an opportunity to investigate regional temporal changes on childhood obesity, and the impact of obesogenic environments. The Avon Longitudinal study of Parents and Children (ALSPAC), which has methylation data (Illumina 450K array), was used to investigate associations between early life exposures and DNA methylation (in childhood and late adolescence) at CpG loci.

Early life rapid weight gain (RWG) was consistently associated with childhood body composition in both local cohorts over time. In ALSPAC, RWG was significantly associated with a 1% increase in childhood methylation (age 7, *n*=116) at an individual CpG locus (CG11531579). Furthermore, the highest levels of methylation (+2%) were in those with RWG who were subsequently overweight/obese (OWOB, age 17).

The CG11531579 loci was investigated further in NTFS adults (age 50, *n*=134) to examine whether the epigenetic marks persist. RWG was also associated with methylation changes in adults, although this was a decrease in methylation (-2%, age 50). These findings suggest that RWG in infancy is associated with small, dynamic variations in methylation at this locus.

Acknowledgements

I would like to take the opportunity to thank everyone who has helped me get through the last 4 years. Firstly, I would like to thank my talented supervisory team: Jill M^{c}Kay, Heather Brown, Mark Pearce, Viviana Albani and Hyang-Min Byun, for all of your continued support and for the opportunities you have afforded me. You have been patient mentors and role models, allowing me intellectual freedom and independence, which I truly appreciate. I have learnt an incredible amount from all of you, and would not have been able to complete this project without your help and guidance.

Having worked across 3 institutes and 4 buildings throughout the course of my PhD, I'd like to acknowledge those who made me feel welcome, and thank Hyang-Min for welcoming me into her lab group.

The project, training courses and internship would not have been possible without the funding from BBSRC for which I am very grateful. I would also like to thank all the cohort participants and administrators who make this research possible. In particular, I would like to give special thanks to Kay Mann and Katharine Kirton for facilitating access to the NTFS data; Angela Jones, Wendy Wrieden, Kathryn Parkinson and Laura Basterfield for access to GMS data; and Caroline Relton, Gemma Sharp and Sian Crosweller for access to the ALSPAC data. Thanks to Kay Mann for patiently passing on your knowledge of path analysis. I would also like to thank Lotte Houtepen and Caroline Relton for their guidance on the analysis of the ALSPAC data, and Charlotte Wright for the advice regarding analysis of GMS data. I'd like to thank Yuri for his patient responses to my many R-related queries, and Jess for addressing my general queries and concerns, and for being a supportive friend from the start.

I'd like to thank my best pals, especially the occupants of Cardy T, for keeping me sane throughout. To the PhD girls, I'm glad we got to share this experience and have formed some solid friendships that could probably only spawn from a collective experience of the utter turmoil of doing a PhD. I'd like to thank Oli for his wisdom, wine, and words of encouragement, and for creating a peaceful home environment and literary escape. Finally I would like to thank my family for their support, and to my nana, Emma Rose Robinson, for her unwavering belief that I would even go to university (little did she know I would still be here 8 years later), and for being the reason I got into research in the first place.
Thank-you all so much!

Abbreviations

ACE	Adverse childhood experience
ALSPAC	Avon Longitudinal study of Parents and Children
ANOVA	Analysis of Variance
APS	Adenosine 5' phosphosulfate
ARIES	Accessible resource for integrated epigenomics studies
ATP	Adenosine triphosphate
BDNF	Brain-derived neurotrophic factor
BF%	Body fat percentage
BIA	Bioelectrical Impedance
BIC	Bayesian information criterion
BMI	Body mass index
BMR	Basal metabolic rate
BSGS	Brisbane systems genetic study
BWT	Birthweight
BWTz	Birthweight z-score
CFI	Comparative fit index
CG	CpG site
CI	Confidence interval
CRH	Corticotropin-releasing hormone
CVD	cardiovascular disease
DMR	Differentially methylated region
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DOHAD	Developmental origins of health and disease
DXA	Dual-energy X-ray absorptiometry
EI	Energy intake
EWAS	Epigenome-wide association study
FDR	False discovery rate
FETA	FFQ EPIC Tool for Analysis
FFQ	Food frequency questionnaire
FMI	Fat mass index
FP	Fractional polynomial
FTO	fat mass and obesity-associated gene
GCSE	General Certificate of Secondary Education
GDM	Gestational diabetes mellitus
GFI	Goodness-of-fit index
GMS	Gateshead Millennium Study
GR	Glucocorticoid receptor
GWAS	Genome-wide association study
GWG	Gestational weight gain
HBW	High birthweight
HIV	Human immunodeficiency virus
HPA	hypothalamic–pituitary–adrenal
HZ	Heterozygous
IDT	Integrated DNA Technologies
IGF	Insulin-like growth factor
IPW	Inverse probability weighting

IQR	Interquartile range
ISVA	Independent surrogate variable analysis
IUGR	Intrauterine growth restriction
IUPAC	International Union of Pure and Applied Chemistry
KW	Kruskal-Wallis
LBW	Low birthweight
LGA	Large for gestational age
MAF	Minor allele frequency
MVPA	Moderate–vigorous physical activity
NCMP	National child measurement programme
NEST	New-born Epigenetic Study
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NK	Natural killer
NSHD	National survey of Health and Development
NTFS	Newcastle Thousand Families study
OB	Obese
OLS	Ordinary least squares
ONS	Office of National Statistics
OPCS	Office of Population and Census Statistics
OR	Odds ratio
OW	Overweight
OWOB	Overweight/obese
PACE	Pregnancy And Childhood Epigenetics
PAL	Physical activity level
PCR	Polymerase chain reaction
PHE	Public Health England
PROBIT	The Promotion of Breastfeeding Intervention Trial
B-PROOF	B-Vitamins for the PRevention Of Osteoporotic Fractures
RCT	Randomised controlled trial
RNA	Ribonucleic acid
RT	Rapid thrive
RWG	Rapid weight gain
SD	Standard deviation
SES	Socioeconomic status
SGA	Small for gestational age
SNP	Single nucleotide polymorphism
SVA	Surrogate variable analysis
TDEE	Total daily energy expenditure
TI	Thrive index
UOJ	Ultra-orthodox Jewish
WC	Waist circumference
WHO	World Health Organisation
WHR	Waist-to-hip ratio

Table of Contents

Table of tables

Table of figures

Table of boxes

Chapter 1. Introduction

Overweight and obesity, particularly in children, have increased at alarming rates in recent times. Whilst genetic factors contribute to overweight/obesity (OWOB), the recent surge in incidence and younger age of onset suggest early life and environmental factors are key to this phenomenon.

The developmental origins of health and disease (DOHaD) hypothesis proposes that exposures in early life predispose an individual to diseases, such as obesity, in later life. The literature suggests that early life factors related to birthweight, early growth, maternal body composition and lifestyle factors, infant feeding, and adversity and sleep in infancy may be linked to childhood body composition. It is unclear if each of these early life factors has an independent influence on childhood overweight and obesity, beyond socioeconomic status (SES) and obesogenic environments. Therefore, this thesis examines the relationship between early life factors, lifestyle and SES and the development of obesity in children and adults, and how these factors and their influence on obesity outcomes have changed over time with the emergence of modern obesogenic environments.

The mechanisms behind the DOHaD hypothesis are largely unknown, but DNA methylation (DNAm), an epigenetic mechanism with the capacity to regulate gene expression, has been proposed to be involved. Accumulating evidence suggests that early life factors are molecularly programmed. These epigenetic changes could reveal more about the underlying biological mechanisms of how early life factors increase susceptibility to later obesity. Epigenetic changes could also be used as predictive biomarkers of future risk, in order to identify those who would benefit from early intervention.

Therefore, underlying epigenetic mechanisms of theoretically relevant early life exposures on child and adult obesity were investigated. Differential methylation has been associated with some early life exposures (such as birthweight), however this is the first study that attempts to link to DNAm with both early life exposures and adiposity outcomes. Furthermore, many of these changes have only been identified in cross-sectional studies, however due to the latency between the exposure and outcome there is the need to examine these relationships using longitudinal cohorts.

The thesis is organised as follows: chapter 1 provides the background literature and rationale for this study and presents the aims of the study. Chapter 2 details the data and the statistical methods used in analyses. Chapter 3 and chapter 4 investigate early life and socioeconomic factors influencing obesity in adults and children, using data from two longitudinal cohorts from the same region in the UK born 50 years apart. Chapter 5 uses these same cohorts to examine the role of the changing socio-economic environment on childhood obesity. The potential role of DNAm in mediating obesity-associated exposures is examined using data from a third UK longitudinal study in chapter 6. Validation of these findings was carried out using samples from the original cohort of adults in chapter 7. Finally, chapter 8 discusses the findings, summarises strengths and limitations of the study and concludes with future research directions.

1.1 Obesity: An overview

1.1.1 Epidemiology of obesity

Obesity was first highlighted as a major global concern by the World health Organization (WHO) in 1997 and since then has become an ever-increasing issue. The Health Survey for England has been measuring the nation's height and weight since the early 1990's and has determined that and has shown that the prevalence of OWOB has been increasing steadily: over the last three decades adult obesity prevalence has increased from less than 10% to almost 25% of the population in England (noo, 2015). Of particular concern is the 10 fold increase in childhood obesity worldwide (Butland et al., 2007, Abarca-Gómez et al.), that children are becoming obese at younger ages (Johnson et al., 2015), and the upwards trends in prevalence of severe obesity (PHE Publishing, 2018). The government's Foresight report predicts that obesity will affect 60% of men, 50% of women and 25% of children by 2050 (Butland et al., 2007).

Routine measurement of children's weight and height in England has been achieved via the National Child Measurement Programme (NCMP) which commenced in 2006 (Ridler et al., 2009). The programme successfully measures over 95% of eligible children, with measures taken at both the start (reception, age 4/5) and the end (year 6, age 10-11) of primary school. The 2016/17 NCMP data show that prevalence of obesity in reception was 9.6%, and 20% in year 6 children (NHS Digital, 2017). Prevalence was highest in the North East (10.7% in reception, 22.5% in year 6) compared to other regions in England (NHS Digital, 2017).

There are distinct social inequalities in childhood obesity; the prevalence of obesity is 2 fold higher in the most deprived decile compared to the least (PHE Publishing, 2018).

1.1.2 The relationship between child and adult obesity

OWOB in childhood are of great concern due to the 'tracking' into adulthood, which is the persistence and relative stability of overweight over time (Twisk, 2003). A systematic review by Singh et al,. determined that all high-quality studies reported at least 2x increased risk of OWOB in adults in those that were OWOB in childhood (Singh et al., 2008). However, most of the studies were from cohort studies from 20 years prior to the year 2000, and all of the studies were from high-income countries, therefore as prevalence of obesity has increased over time these estimates may be conservative for current populations. Although there have been few studies outside of North America and Europe, there is evidence that BMI tracking is influenced by ethnicity (Bayer et al., 2011a), and that BMI increases with age may also be higher in black children compared to white children (Freedman et al., 2005a).

The risk of the tracking of obesity is much higher than for OWOB, and multiple cohort studies have demonstrated that obese children are around 5 times more likely to be obese as adults (Simmonds et al., 2015b). In terms of when to intervene, obesity appears resistant to change, as by age 5, most excess weight prior to puberty has been laid down (Gardner et al., 2009) and children are more likely to remain obese (Mostazir et al., 2015, Buscot et al., 2018). Childhood obesity is also related to adult morbidity (Llewellyn et al., 2016). Therefore, understanding the key early life factors driving the upwards prevalence of childhood obesity will be important for prevention and for designing effective interventions.

However, this tracking of obesity is a relatively recent phenomenon. Examining the trends using several British birth cohorts with over 56,000 participants, found that cohorts born more recently had greater probabilities of overweight or obesity at younger ages (Johnson et al., 2015). However, even in the cohorts born between 1946 and 1970, tracking was consistently stronger at the higher quantiles of the BMI distribution (Norris et al., 2019). Bayer et al., (2011) conducted a meta-analysis using multiple cohorts from all ages and time periods and found strong evidence of tracking of weight over time, and a low probability of spontaneous weight changes (without intervention) (Bayer et al., 2011b). These findings suggest that a high proportion of children obese today are likely maintain an obese body weight throughout life, which poses a significant public health crisis.

Regarding the mechanisms underlying the tracking of body composition, there is evidence of the tracking of both physical activity and of diet behaviours between childhood and adulthood, each to similar degrees (Craigie et al., 2011), whilst follow-up from adolescence to adulthood seems to also support that physical activity is pertinent (Kvaavik et al., 2003).

However, an important feature of tracking is that those at high risk for diseases later in life can be identified at an early age, and therefore this feature could be utilised and those at 'high-risk' could be the target for early risk reduction interventions (Twisk, 2003). Crucially, the majority of obese adults (70%) were not obese in childhood (Simmonds et al., 2016), suggesting factors impacting and accumulating over the life course. This also emphasises that further understanding of the many factors and how they interact to influence obesity across the life course is necessary. Furthermore, this highlights the need for biomarkers that could identify those who are at risk of subsequent obesity, so that interventions can be targeted.

1.1.3 The consequences of obesity

On average, obesity decreases life expectancy by three years, and severe obesity by up to 10 years (Prospective Studies, 2009). Obesity can cause ill health in childhood and premature mortality and physical morbidity in later life (Reilly and Kelly, 2010), drastically impacting on the individual's physical and psychological health as well as socioeconomic effects (Reilly et al., 2003). The clinical consequences can range from type 2 diabetes (T2DM), metabolic complications, risk of some cancers, cardiovascular disease (CVD), asthma, sleep apnoea, and menstrual cycle abnormalities (in females)(Reilly et al., 2003, Lee, 2009, Reilly and Kelly, 2010, Umer et al., 2017). Furthermore, life course body mass index (BMI) gains and earlier obesity onset are associated with poorer physical functioning in middle age (age 50), stressing the importance of prevention and delaying onset of obesity (Rogers et al., 2019).

Obesity also represents a global economic problem, as obesity-related diseases could cost the National Health Service (NHS) an extra £2.51 billion by 2035 (The UK Health Forum and Cancer Research UK, 2015).

1.1.4 Definition and measurement of body composition in adults

Obesity is defined by the WHO as “abnormal or excessive fat accumulation that may impair health” (World Health Organization, 2019). Therefore, definitions of obesity in children or adults should meet these two criteria of: diagnosing high body fat and increased risk of health outcomes.

There is no accepted ‘gold standard’ of measuring obesity. There are many methods for determining body composition, which each have acknowledged advantages and drawbacks (

Table 1.2). No single measure is error-free, and the choice of measurement is dependent on cost, availability, the population under study and the outcomes of interest (i.e. disease risk).

BMI, a measure of weight related to height, is the measure most frequently used to categorise individuals into weight categories (Table 1.1). Obesity is defined according to the WHO criteria as a BMI greater than 30 kg/m^2, with well-established risks of all-cause mortality (Aune et al., 2016). Overall, BMI is a quick, easy, inexpensive measure to determine weight status. Generally, BMI is practical on a population level, however it represents a proxy, rather than a direct measure of body fat, and there is some disagreement as to whether BMI is the best diagnostic measure (Adab et al., 2018). Furthermore, BMI does not adjust for sex or age, or consider variation by pubertal status or ethnicity (although different cut-offs do exist). BMI also does not provide information on fat distribution or proportions of lean and fat mass (World Health Organization, 2011c).

Alternative measures to BMI that are viable in a clinical setting include measures of abdominal fat (i.e. waist-to-hip ratio or waist circumference) or body fat (i.e. skinfold measurements or bio-electrical impedance).

Waist circumference (WC) is an indicator of visceral fat (the fat stored around internal organs). Cut-off points are used to identify individuals at increased risk of metabolic complications (Table 1.1). Waist circumference-based measures have been shown to more accurately reflect obesity prevalence than BMI (O'Neill, 2015), and are independently associated with cardiovascular risk (Huxley et al., 2009) and all-cause mortality (Pischon et al., 2008, Sahakyan et al., 2015). Whilst a WC greater than the cut-off is purported to increase risk, it will not accurately capture all of those at risk due to differences in body composition (small frame), however it does represent a quick method of identifying individuals with central obesity. An additional measure including the hip circumference, the waist-to-hip ratio (WHR), can reveal more about fat distribution. There is often not much difference between the measures of central obesity, either WC or WHR (Seidell, 2009). WHR is arguably a more informative measure as it describes fat distribution around the waist in proportion to the hips, and is usually a strong predictor of cardiovascular risk (Dalton et al., 2003, Myint et al., 2014, Egeland et al., 2016).

Table 1.1. Categorisation of body composition according to the proxy measures BMI, waist circumference and waist-hip ratio

Measure	Ranges	Category
BMI	<18.5 kg/m^2	Risk of underweight
	18.5-24.9 kg/m^2	Normal weight
	25-29.9 kg/m^2	Overweight
	30-34.9 kg/m^2	Class I obesity
	35.0–39.9 kg/m^2	Class II (severe) obesity
	≥ 40.0 kg/m^2	Class III (morbid) obesity
Waist circumference	>94cm (M), >80cm (F)	Increased risk of metabolic complications
	>102cm (M), >99cm (F)	Substantially increased risk of metabolic complications
Waist-hip ratio	≥0.90cm (M), ≥0.85cm (F)	

M, Male; Female. WHO cut-off points and risk of metabolic complications for Caucasian populations *(World Health Organization, 2011c).*

Alternatively, body fat can be measured directly. Estimates of lean and fat mass can be made from routine measures such as skinfolds or bioelectrical impedance. Advanced imagining techniques can provide more accurate measures of body composition (Lee and Gallagher, 2008), although these are costly and more difficult to perform in routine settings. Examples of these are dual energy x-ray absorptiometry (DXA), developed to measure bone mineral mass, or air displacement plethysmography, which measures the volume of air displaced by the subject in a confined space (i.e. the BOD POD). Body fat determined by bioelectrical impedance (BIA), estimates body composition based on the conductive properties of lean and adipose tissue. However, BIA replies on the assumption that fat free mass is composed of 73% water, and is therefore affected by hydration level through food and drink consumption, medications, and stage of the menstrual cycle (Dehghan and Merchant, 2008). BIA has the advantage of being fast, simple and low susceptibility to inconsistencies from operator technique (Prentice and Jebb, 2001). The choice of prediction equation for BIA is a great source of variability (Reilly et al., 1996), and needs to be appropriate for the ethnicity of study participants (Dehghan and Merchant, 2008).

Frankenfield *et al*,. found that 30% of men and 46% of women with a non-obese BMI (below 30 kg/m^2) had 'obese levels' of body fat (Frankenfield et al., 2001), indicating that BMI will misclassify some of those at risk. BMI measures both lean and fat mass and does not

distinguish between the ratios, which can change over time. For example, the loss of lean tissue and hormonal changes (i.e. growth hormone and Insulin-like growth factor (IGF) decreases with age) are important considerations in middle age.

An advantage of using BF% over BMI is in that two individuals could have the same height and weight and therefore the same BMI whilst one has a higher proportion of body fat and the other higher muscle mass. Similarly, the same could be said for individuals with the same BMI but one apple-shaped, with a high central fat distribution and the other pear shaped with less visceral fat. These two individuals would have different health risks associated with their body composition and therefore it is important taking into account different measures where data are available.

Table 1.2 Advantages and disadvantages of different methods of measuring body composition routinely used in cohort studies

Measure	Advantages	Disadvantages
BMI	▪ Simple, inexpensive measure ▪ Easy to obtain, therefore useful screening tool and population measure ▪ Well-characterised with standard cut-off points ▪ BMI categories are associated with health risks ▪ Strongly correlated with sophisticated measures of body fat	▪ Questionable validity for non-Caucasian populations ▪ Not an accurate measure in athletes as it does not differentiate fat from muscle mass ▪ Does not account for age or sex or age ▪ Provides no information on fat distribution
Percent body fat using BIA	▪ Fast and simple ▪ Less prone to error from operator technique ▪ Portable and convenient	▪ BIA is less accurate than advanced imaging techniques ▪ Hard to calibrate ▪ Body water could be affected by dehydration or illness
Waist circumference or WHR	▪ A measure of central fat distribution and visceral fat ▪ An independent risk factor for disease, and specifically a marker of CVD ▪ Simple, straightforward and inexpensive ▪ Strongly correlated with more sophisticated measures of body fat	▪ Perhaps less accurate than a measure of waist and height ▪ WHR could be prone to error, as requires accuracy of two measurements (waist and hip)

1.1.5 Measurement of body composition in children

When choosing a measure in children, there are similar considerations as in adults, based on cost, acceptability, and ease of application (

Table 1.2). There is no 'gold standard' in body composition measurement in children, however BMI is the most common means of categorising weight. The main difference in children is that measurements need to take into account that they are still growing and hence require growth charts. For example, growth references derived from an appropriate reference population can be used to transform absolute BMI values (standard deviation z-scores). Each growth curve has its own set of recommended thresholds to classify overweight or obesity. The UK90 growth reference is most commonly used in UK populations, and either centiles or standard deviation can be used as BMI cut-offs (Weng et al., 2012, Bammann et al., 2014, Woo Baidal et al., 2016). BMI cut-offs in children have high specificity and moderate sensitivity (Javed et al., 2015), meaning that some obese children may be missed, but healthy weight children are not likely to be wrongly classified as OWOB (Mast et al., 2002, Bedogni et al., 2003). In children, BMI has good acceptability as a measure, compared to more intrusive measures, or those that require clothing to be removed, which may not be appropriate in some settings or populations (Simmonds et al., 2015a).

Similar to body composition measures in adults, growth charts have been developed for use in children's for measures: BMI (Lindgren et al., 1995, Wells et al., 2012, Weber et al., 2013), BIA (Chumlea et al., 2002), skinfolds (Tanner and Whitehouse, 1975), DXA (Van der Sluis et al., 2002) and more recently for 4-component models (divides body weight into fat, water, mineral, and protein)(Wells et al., 2012).

Whilst waist measurements of obesity have been well characterised as risk factors in adults, there is less evidence to suggest that WHtR is also associated with cardio-metabolic risk in children (Kuba *et al*., 2013). Generally fat distribution patterns, either android ('apple-shaped') or gynoid ('pear-shaped'), generally start to emerge in puberty (Lobstein et al., 2004), although some studies have found evidence of fat patterning in pre-pubertal children (Mast et al., 1998). Waist-to-height ratio is a measure of whether the amount of upper body fat is appropriate relative to height (McCarthy and Ashwell, 2006). It is calculated as the waist measurement divided by height measurement in cm. A simple cut-off of a WHtR >0.5 is the commonly accepted value (regardless of sex or age) determining 'waist obesity', which translates to a weight that is over half height. In children WC sensitivity ranges from 35-100% and specificity from 81-100%, however waist-to-height ratio (WhtR) has the best diagnostic accuracy, although examined in fewer studies (Simmonds et al., 2015a).

Whether it is a better indicator than BMIz is unclear, however increased waist obesity (using WHtR) has been found in children even when BMIz remains the same (Fredriksen *et al*., 2018).

Direct measures of fat have advantages over proxy measures. Fat mass can be measured in similar ways as in adults and carries the same limitations (section 1.1.4). Skinfold measurements offer an alternative, however compliance is low and there are issues in ensuring correct technique when measuring young children (Kehoe et al., 2011). Alternatively, fat mass can be measured using BIA, and in systematic reviews has been found to have good reliability (Talma et al., 2013), reproducibility (Chula de Castro et al., 2018), and performs well in measuring change in body composition (Meredith-Jones et al., 2015) in children. However, BIA may be prone to measurement error (Talma et al., 2013), underestimate fat mass (Chula de Castro et al., 2018), and in terms of outcomes is not superior to BMI in predicting cardiovascular risk factors in OWOB children (Bohn et al., 2015). When examining fat mass in children it is important that measures take into account height. Fat mass index (FMI) is calculated as fat mass (kg) divided by height (m) squared. Findings from the Gateshead Millennium study determined that FMI was a more sensitive measure than BMI (Basterfield et al., 2012a) and proxy measures of adiposity were inferior (Basterfield et al., 2012b).

Whilst BMI is imperfect, there are a lack of validated reference values for alternative measures of adiposity in children (Javed et al., 2015). BMI is the only measure recommended for use of determining obesity in children in the UK (Simmonds et al., 2015a), and will therefore likely remain the most frequently used measure for now. However, there is support for using additional measures such as WHtR or FMI which may be able to detect larger waist circumferences and higher fat mass in children.

1.2 Aetiology of obesity

1.2.1 Current concepts

The convincing factors that increase risk of obesity determined by WHO include; high intakes of energy dense, nutrient poor foods and a sedentary lifestyle (World Health Organization, 2003). However, obesity aetiology extends beyond the archaic notion that weight gain simply results from intake of more calories than required for whole-body metabolism (energy expenditure). Energy expenditure is a combination of the basal metabolic rate, the

thermic effect of food, and activity related expenditure (Levine, 2005). From a physical perspective, the laws of thermodynamics regarding energy conservation in living organisms are often applied as an explanation, namely calories in vs. calories out, or dietary intake vs. physical activity.

This reductionist approach disregards the multitude of factors that influence food intake, energy expenditure, and whole-body metabolism, which are regulated by complex feedback mechanisms. Regulatory processes in the human body ensure that, for the most part, body mass and energy intake/expenditure remain fairly stable (Jéquier and Tappy, 1999). Therefore, excess weight gain must be driven by other factors. The environmental basis for obesity was also addressed in the Governments Foresight report, which acknowledged the multiple societal influences and the need for a multifaceted approach (Butland et al., 2007). Addressing the underlying causes of obesity requires an understanding of the socio-economic, psychological, behavioural, and socio-cultural factors, including the drivers of food choices, eating patterns and participation in physical activity. However, environmental factors are ubiquitous and not everyone has obesity, suggesting underlying factors that increase susceptibility for some.

It is important to distinguish between risk factors and causes. Causality is the study of the relationship between an event (the cause) and an outcome (the effect) which occurs consequently. The presence (or absence) of a causal factor can lead to illness or disease, therefore when a cause is removed the outcome should cease. Whereas a risk factor is something that increases the statistical risk of a disease, but a risk factor is not necessarily a cause and could be a surrogate for the underlying cause.

In terms of the genetics of human obesity, there is evidence for genetic mutations and genetic variation. Specific gene mutations, such as those related to leptin, have been implicated in development of monogenic obesity, these however are very rare (Montague et al., 1997). Distinguished genetic mutations include those involved in food intake control (Neuropeptide Y, leptin, Pro-opiomelanocortin), energy regulation (β2-adrenergic and β3-receptors), thermogenesis (uncoupling proteins 1–3), adipogenesis, signalling (peroxisome proliferator-activated receptor), and leptin and the leptin receptor (Farooqi and O'Rahilly, 2004). There is also the capacity for small genetic changes, such as single nucleotide polymorphisms (SNPs), to influence susceptibility across the life course (Thorleifsson et al., 2009, Willer et al., 2009). A genome-wide association study, which combined multiple

polymorphisms into a risk score using 97 genetic loci, explained 2.7% of the variation in BMI (Locke et al., 2015). Although important, genetic variants remain fairly stable over time and cannot explain the extent of the rise in obesity (Yang et al., 2007).

Gene-environment interactions refer to phenotypic changes in response to environmental cues, dependent on genotype. These dictate individual responses to environment and downstream effects on weight. However, those genetically susceptible would be more likely to gain weight in a weight-promoting environment, and hence genetic factors may exert more of an effect when colluding with obesity risk factors. For example, in the case of the FTO genotype, which is associated with an increased risk of obesity in children and adults (Frayling et al., 2007), the risk can be attenuated by physical activity (Kilpeläinen et al., 2011), or diet composition (Sonestedt et al., 2009).

In summary, probable aetiological factors in obesity are likely a mixture of genetics, individual lifestyle factors, and perhaps more importantly, gene-environment interactions.

1.2.2 Obesogenic environments

Increases in childhood obesity in the last 20 years have been somewhat attributed to social change and a changing environment. Over this time, there has been widening social inequalities in childhood overweight (NHS Digital, 2016, Bann et al., 2018), and changes in the social patterning of obesity (Knai et al., 2012). All of the environmental influences and conditions of life that encourage overweight and obesity have been termed the 'obesogenic environment' (Lake et al., 2011).

The ever-increasing 'obesogenic' macro and microenvironments have developed over the last 30 to 40 years. Diets have changed; food is widely available and easily accessible, in particular energy-dense convenience foods. Decreasing levels of physical activity and increasingly sedentary lifestyles have been facilitated by labour-saving devices, improvements in transport and shifts to more sedentary jobs, highlighting another dimension to the problem. Additionally, marketing of food has become aggressive and advertising ever-present.

Health behaviours are often targeted for interventions at the individual level. Social characteristics are a more difficult avenue for intervention as they require large-scale changes, but are important determinants of health (Tarlov, 1999). Minimising the exposure to harmful environments may even-out the distribution of health, although there are

obvious health inequalities between social groups that need to be considered in public health interventions.

1.2.3 Obesity and inequalities

In the UK, socioeconomic status (SES) and obesity are related across all ages (PHE, 2013). Current statistics from Public Health England show an increased risk of obesity with increasing levels of deprivation across all indicators for women, whilst the relationship is less clear for men (Marmot et al., 2010). Although trends in UK childhood obesity prevalence have begun to stabilise in the past decade, this is not the case for children in the lower socioeconomic groups (Stamatakis et al., 2009).

The relationship between SES and obesity is thought to be due to a number of factors, which have been explored by Cutler et al. (2008). Generally, lower SES leads to greater exposure to health-compromising issues. It also creates vulnerability (i.e. lower income→ lower nutrient intake→more vulnerable to infection→more time off work→lower earnings) thereby facilitating the cycle of disadvantage. This is similar to Riley's life course model, which theorises the accumulation of risk, which is the idea that exposures accumulate throughout life, for example from periods of illness, adverse conditions and detrimental health behaviours (Riley, 1989). Lower SES can also contribute to psychosocial issues, for example, uncertainty of future access to resources can lead to increased stress levels.

Research consistently shows that adverse early life socioeconomic conditions are associated with poorer adult health (Smith et al., 1998, Power et al., 1999, Poulton et al., 2002). Socioeconomic factors are considered important risk factors in childhood obesity, but also a potential source of confounding. There is a cycle of obesity in families, and parental obesity can be considered one of the greatest risk factors for childhood obesity (Parsons et al., 1999). This is likely related to genetics and SES, but also further exacerbated by maternal overweight as an *in utero* exposure. Obesity risk is lower in adoption studies but still remains, suggesting the shared socioeconomic environment has an effect (Stunkard and Sorensen, 1993).

Findings from a review of longitudinal studies support that SES precedes obesity risk (Ball and Crawford, 2005). A systematic review determined that socioeconomic differences in BMI emerge by age five, with the majority of studies demonstrating an inverse association, in

particular for parental education (Shrewsbury and Wardle, 2008). Findings from the Millennium Cohort Study indicate that education of the primary carer is more strongly associated with obesity rates in in children age 5 than income (Brophy et al., 2009). In the Avon Longitudinal study of Parents and Children (ALSPAC) cohort, trajectory modelling of obesity on the basis of maternal education revealed differences emerging at age 4, which then widened with age (Howe, 2012). Parental income and education will determine food and lifestyle choices, access to health care services and housing conditions, all of which can affect child health.

These shifts in obesity prevalence in the lower socioeconomic groups in developed countries (McLaren, 2007b, Shrewsbury and Wardle, 2008), and more recently in low-mid income countries (Popkin et al., 2012) is termed the nutrition transition. This occurred with the accessibility of relatively cheap, low nutrient, high energy dense foods, leading to increased caloric intake (Popkin, 2001). Generally, socioeconomic disadvantage is associated with consumption of poorer diets which do not meet dietary recommendations: lower fruit and vegetable intake and higher intake of sugar and fat (Hanson and Chen, 2007). Socioeconomic disadvantage also affects infant feeding, and is associated with decreased duration or initiation of breastfeeding (Thulier and Mercer, 2009, Wijndaele et al., 2009a). Features of the built environment including neighbourhood safety, access to local facilities and food outlets, and social facilitation of unhealthy behaviours are important factors (Papas et al., 2007, Lovasi et al., 2009). Although parental SES often predicts obesity within the individual, intergenerational upward social mobility can have a positive impact (Cavaco et al., 2014), therefore SES represents an important target for interventions.

Overall, these findings demonstrate that obesogenic environments play an important role. An issue that remains unresolved is to disentangle the relative importance of environment factors and SES on later life obesity (Parsons et al., 1999).

1.2.4 The Developmental Origins of Health and Disease Hypothesis

1.2.4.1 *Background of the theory*

Hypotheses attributing early life factors in the development of subsequent disease later in life are relatively contemporary, as is the field of life course epidemiology. The seminal work by Barker and colleagues introduced the foetal origins of adult disease hypothesis (Barker, 1994). They noted a correlation between low birthweight and rates of coronary heart

disease in adulthood, leading to the theory that foetal undernutrition alters physiology and metabolism, leading to increased risk of heart disease in later life (Barker et al., 1989). The Barker hypothesis has since expanded into the Developmental Origins of Health and Disease (DOHaD) hypothesis, which has a broader scope. The hypothesis proposes that exposures in early life predispose an individual to many diseases in later life, including obesity (Samuelsson et al., 2008, Schellong et al., 2012), CVD (Barker et al., 1989) and T2DM (Whincup et al., 2008).

However, as early as 1962, James Neel noted that Human history was defined by periods of feast or famine, and those who had had fat reserves would have better survival and fertility (the 'Thrifty gene' hypothesis)(Neel, 1962). Neel's adaptive hypothesis has since been counteracted by the non-adaptive 'Drifty gene' hypothesis: that excess adiposity was not a survival advantage, but that genetic drift in genes encoding the upper body weight limit occurred when the risk of predation diminished (Speakman, 2008). The thrifty genotype has been hailed as 'too simplistic' (Reddon et al., 2018), in that everyone would be predisposed to obesity and therefore any high-risk SNPs would have appeared in the last 900,000 years (Speakman and Westerterp, 2013), which appears not to be the case (Wang and Speakman, 2016).

Commonality across these hypotheses is that susceptibility to disease is determined in developmentally critical periods, originating from prenatal and early life experience (Hales and Barker, 1992, Barker, 1995). These theories have shaped the discipline that is life course epidemiology as it stands today. Life course epidemiology examines how different aspects of biological, social and physiological disease risk factors impact at different stages over the life span, with potentially independent, cumulative or synergic effects (Ben-Shlomo et al., 2016). A life course approach can examine the effects of exposures at one time point on prospective outcomes, useful in the study of obesity due to its multi-dimensional aetiology.

Initially, this accumulation of risk was believed to have specific 'critical' or sensitive periods, operating in early life (i.e. foetal origins). However, longitudinal cohorts, which have the advantage of monitoring change over time, have also identified factors impacting at different time points. For example, the prenatal period, early infancy, the period of adiposity rebound and adolescence have been suggested to be critical time windows for early life programming of adult disease (Dietz, 1994). However, the term 'programming' has been criticised, as it implies disease is predetermined, and 'conditioning' has since been suggested

as a term to encompass how an individual is primed to respond to environmental challenges in a certain way, but it is not set in stone (Hanson and Gluckman, 2014). The DOHaD model has three key avenues; i.) Immediate homeostatic response; ii.) Predictive adaptive response; and iii.) Natural selection over many years (Ben-Shlomo et al., 2016). When there is mismatch between *in-utero* and postnatal environments, disease can occur. The predictive adaptive response, a phenotype with advantages in a predicted environment, could be harmful when there is a mismatch between a nutrient deprived intrauterine environment and a postnatal obesogenic environment.

1.2.4.2 *Evidence from human studies on mismatched environments*

One of the key examples of early life nutrient restriction in humans is the Dutch hunger winter of 1944-45, a result of the German occupation of the Netherlands during the World War II (WWII). It provides a rare opportunity to study the long-term effects of *in utero* nutrient restriction (<1000kcal/day) in humans previously well nourished. Pregnant women exposed to famine in early and mid-gestation, followed later by an adequate diet, had offspring who were more at risk of increased adiposity and impaired glucose tolerance as adults (Ravelli et al., 1998, Ravelli et al., 1999, Stein et al., 2007). Furthermore, the offspring of prenatally undernourished fathers (but not mothers) were more obese than offspring of parents who had not been undernourished, independent of paternal BMI, and thereby demonstrating potential transgenerational effects (Veenendaal et al., 2013).

In the Chinese famine of 1959-61, sex differences were also observed in that females born during the famine had a higher prevalence of OWOB but not males (Yang et al., 2008). It is interesting that these historical famines demonstrate sex differences in obesity outcomes.

Additionally, evidence of seasonal nutritional variation in The Gambia also supports the theory that alterations in maternal nutrient intake can affect birthweight (Rayco-Solon et al., 2005) and early mortality (Moore, 2016). The mechanisms have been examined using various animal models, and findings support that nutrient restriction leads to metabolic dysfunction in the offspring, both dependent (McKay et al., 2014) and independent of subsequent maternal food intake (Bispham et al., 2003). These findings highlight the detrimental effects of a mismatch between *in utero* insufficiency and postnatal nutrient intake.

1.3 Evidence for early life risk factors associated with obesity

With the earlier age of onset of childhood obesity, as well as the evidence supporting the DOHaD hypothesis, there is increasing interest in the early life factors contributing to obesity development.

Potential early life risk factors for obesity have been examined in multiple cohorts. A comprehensive systematic review of 30 prospective studies identified risk factors for childhood overweight to be; high birthweight, maternal smoking during pregnancy, early rapid weight gain, maternal pre-pregnancy overweight, whilst breastfeeding was protective (Weng et al., 2012). Another review determined maternal BMI, childhood growth and family SES to be probable risk factors (Brisbois et al., 2012). Similar themes emerge from these reviews surrounding maternal lifestyle influences, early life feeding and growth and socioeconomic disadvantage (Table 1.3).

A large review of risk factors for childhood obesity in the first 1,000 days found consistent associations for maternal pre-pregnancy BMI, prenatal tobacco exposure, maternal excess gestational weight gain, high infant birthweight, and accelerated infant weight gain (Woo Baidal et al., 2016). There was also some support for gestational diabetes, low SES, low maternal–infant bonding, and in infants; child care attendance (as a proxy for infection), antibiotic exposure, disturbed sleep, early introduction of solid food intake (Woo Baidal et al., 2016). Birthweight is often used as a proxy measure of factors affecting growth and development (Table 1.3).

Table 1.3 Summary table of the pre- and post-natal early life risk factors of OB, and proxy measures utilised in birth cohort studies lacking prenatal data

Category	Risk factor	Proxy measure
Pre-natal		
	Smoking	LBW/SGA
	Stress	LBW/SGA
Maternal factors	Maternal obesity	HBW/LGA
	Diabetes	HBW/LGA
	Maternal diet	HBW/LGA
	Maternal age	
Paternal factors	Nutrition & sperm quality	
	Smoking	
Post-natal		
	Infant feeding	
	Early life growth	
	Early life stress	Adverse childhood experiences
	Childhood infection	Antibiotic exposure

	Sleep issues	Sleep disturbance, night-time waking
	Socioeconomic status	Maternal education Parental socioeconomic status

LBW, low birthweight; HBW, high birthweight; SGA, small for gestational age; LGA, large for gestational age.

1.3.1.1 *Birthweight*

There is empirical evidence of a U-shaped curve between birthweight and adult obesity (Curhan et al., 1996), implying risk for both low and high birthweight infants.

1.3.1.1.1 Low birthweight (LBW)

Intrauterine growth restriction (IUGR) is classified as a birthweight below the 10th percentile, adjusted for gestational age. It can occur as a result of pre-eclampsia, placental insufficiency or maternal undernutrition (Villar et al., 2006, Salam et al., 2014), which can result from nutrient restriction, maternal smoking and stress. The foetus prioritises brain development at the expense of other body systems such as the renal or endocrine (Desai et al., 1996). In the case of the latter, it can lead to impaired Beta cell development and hence altered insulin secretion which predisposes the infant to glucose intolerance and diabetes (Portha et al., 2011). Pre-eclampsia is also more likely in mothers with a high BMI, pre-gestational diabetes, chronic hypertension, and in assisted reproductive technology pregnancies (Bartsch et al., 2016).

A potential mechanism as to how foetal undernutrition results in increased offspring adiposity could be through altered appetite regulation. In a rat model of undernourished mothers, food intake was significantly elevated at an early postnatal age in offspring, and continued to increase with age (Vickers et al., 2000). This hyperphagia when combined with a high fat diet resulted in increased offspring weight, indicating that both intrauterine and environmental factors are important. A rat model of IUGR demonstrated decreased plasma leptin and increased ghrelin, with a postnatal period characterised by excess food intake, catch-up growth and metabolic syndrome with comorbidities including obesity (Desai et al., 2005). Appetite stimulatory factors in the growth-restricted animals were at levels comparable to those of the fasting controls, demonstrating the potential for *in utero* programming of orexigenic hormones.

Other exposures that could suppress *in utero* growth (i.e. maternal smoking, stress and nutrient restriction) are examined in subsequent sections.

1.3.1.1.2 High birthweight (HBW)

The associated risk factors for large for gestational age (LGA) infants include pre-pregnancy obesity, excessive maternal weight gain, maternal or gestational diabetes mellitus (GDM), increasing parity and prolonged gestation (Jolly et al., 2003). LGA is defined as birthweight above the 90th percentile. LGA is related to childhood obesity, T2DM, CVD, and altered appetite and energy regulation in the offspring (Heerwagen et al., 2010, Dabelea and Crume, 2011, Frias and Grove, 2012).

Findings from the Southampton Women's Survey found that parity was associated with increased birthweight (Harvey et al., 2007), although this effect could be confounded by maternal obesity or SES. However, Hinkle et al. (2014) determined that the effects of parity were independent of factors related to maternal BMI or weight gain. The effects of parity appear to persist and are evident in adult offspring (Reynolds et al., 2010).

1.3.1.2 *Maternal factors*

1.3.1.2.1 Maternal body composition

In a large meta-analysis including 39 cohorts (265,270 births), both high pre-pregnancy BMI and gestational weight gain (GWG) were associated with risk of pregnancy complications, including gestational hypertensive disorders, GDM, and LGA (Santos et al., 2019). Obese mothers who had excessive GWG had the highest risk of pregnancy complications, and strikingly, 24% of pregnancy complications were attributed to maternal OWOB. Maternal body composition has also been associated with child overweight in many longitudinal birth cohorts (Baker et al., 2004, Harvey et al., 2007, Wright et al., 2010a, Bammann et al., 2014, Fairley et al., 2015b), yet Mendelian randomization analysis did not support a strong causal intrauterine effect of higher maternal BMI on offspring adiposity (Richmond et al., 2017).

However, the effects may be indirect via birthweight, as evidence suggests the relationship between maternal BMI and GDM on birthweight is casual (Tyrrell et al., 2016).

In England around half of women of childbearing age are OWOB (Craig et al., 2014). This has accompanying short-term risks of increased birth complications, increases the likelihood of GDM, and is associated with lower breastfeeding rates (Leddy et al., 2008). However, a randomised controlled trial (RCT) of antenatal dietary and lifestyle interventions in OWOB pregnant women (LIMIT study), was not successful in reducing the risk of adverse maternal

pregnancy and birth outcomes (Dodd et al., 2014), and further research is required into the mechanisms driving these associations.

1.3.1.2.2 Diabetes

Similar to the effects of maternal obesity, GDM is associated with HBW and increased risk of obesity and T2DM later in life (Bartsch et al., 2016). There is the increased risk of obesity in offspring of mothers who were diabetic before conception and those with GDM (Pettitt et al., 1983, Pettitt et al., 1988, Silverman et al., 1998). The risk is not likely due to genetic transmission, as siblings born when the mother was not diabetic do not have the same risk, implying that the disease state alters offspring phenotype (Dabelea et al., 2000).

An Australian trial for pregnancy treatment of GDM reduced macrosomia and pregnancy complications (Crowther et al., 2005), however there was no difference in offspring body composition in childhood (age 4-5) (Gillman et al., 2010). Other trials have also not found promising results, although long-term follow-up is needed (Guillemette et al., 2017).

1.3.1.2.3 Maternal age

Advanced maternal age, commonly considered as over 35 years of age, has been associated with GDM, gestational hypertension, preeclampsia, small for gestational age infants, spontaneous late preterm delivery, and caesarean section (Kahveci et al., 2018). Maternal age at birth is increasing in England (Office for National Statistics, 2017), therefore there could be important clinical implications.

Increasing maternal age could be synonymous with increases in parity, or to age-related metabolic changes relating to glucose regulation (Chandler-Laney et al., 2013). When controlling for all covariates, only maternal age below 25 and above 35 was associated with adverse offspring health outcomes (Myrskylä and Fenelon, 2012). Evidence for parity is mixed and warrants further investigation.

There are numerous biological and social theories surrounding this phenomenon of a U-shaped relationship between maternal age and offspring health outcomes. Older mothers are subject to physiological disturbances which occur with age, and a decline in reproductive functionality. There is increased risk of adverse outcomes in pregnancy, such as LBW and pre-term birth (Goisis et al., 2017, Sohn, 2017, Fuchs et al., 2018). In terms of metabolic health, it has been shown that offspring of both younger and older mothers had higher adult fasting glucose concentrations (Fall et al., 2015).

However, it is not necessarily the age of the mother that is the underlying cause, but the individual life circumstances, such as health behaviours in older mothers (Goisis et al., 2017), or age-related increases in maternal BMI (Sutcliffe et al., 2012). A study in Swedish youths (aged 19 in 2009), found that having an older mother was associated with lower self-rated health, including lower likelihood of regular exercise, and increased likelihood of obesity in adolescence (Barclay and Myrskylä, 2016).On the other hand, older maternal age has been shown to have many beneficial effects on child (up to age 5) health and development (Sutcliffe et al., 2012). Furthermore, sibling analysis has shown that although children born to older mothers are born smaller, as adults they are taller, more intelligent and are less likely to smoke (Carslake et al., 2017).

Although less recognised, paternal effects can also impact. Advanced paternal age has also been associated with adverse offspring health effects such as risk of stillbirth, some cancers, autism, and neurodevelopmental disorders (Malaspina et al., 2015, Nybo Andersen and Urhoj, 2017). Maternal and paternal age are likely to be similar, therefore there will be a high risk of collinearity and residual confounding, leading to heterogeneity in findings.

1.3.1.2.4 Maternal smoking

It has long been known that smoking during pregnancy has adverse outcomes on foetal development, in particular yielding LBW infants who are at increased risk of poor health outcomes, in particular when exposed during the third trimester (Kleinman and Madans, 1985, Lieberman et al., 1994). Maternal smoking during pregnancy is associated with a dose-dependent decrease in birthweight (Newnham, 1991), and is independently associated with later catch-up growth and childhood overweight (Harvey et al., 2007, Fairley et al., 2015a).

Systematic reviews support that maternal prenatal smoking is consistently associated with increased odds of childhood OWOB (Rayfield and Plugge, 2017). In a UK longitudinal birth cohort, maternal smoking during pregnancy was associated with a higher offspring BMI-z score and increased risk of child overweight as early as 3 years old (Fairley et al., 2015a). Whilst a Swedish cohort found maternal overweight and maternal smoking to be the greatest determinants of offspring overweight in young men (Koupil and Toivanen, 2007). There is uncertainty regarding a causal link between maternal smoking and offspring BMI in humans, as paternal smoking also demonstrates similar outcomes, which suggests factors operating within the shared environment (Howe et al., 2012). There is evidence for a causal

influence of maternal smoking during pregnancy on birthweight as the link was stronger for exposure for mothers than for fathers (Davey-Smith, 2008), which may suggest smoking has indirect effects on offspring body composition.

Under stringent experimental conditions, prenatal exposure to nicotine in rats resulted no differences in birthweight, but significant increases in weight were observed after 10 weeks of age (Gao et al., 2005). Although causal mechanisms were not explored (Gao et al., 2005), the increased adiposity could be a result the polycyclic aromatic hydrocarbons in tobacco smoke, as exposure to high doses in pregnant rats dysregulates lipogenesis leading to weight gain (Ortiz et al., 2013). However, doses examined in the rats were much higher than assumed exposure in human pregnancy.

Second-hand smoke could also have detrimental effects, as rodents and pups exposed during lactation also exhibit altered growth and metabolic complications (Santos-Silva et al., 2013). Rat pups had a lower body weight and length at weaning, but a 50% increase in fat mass and a 10% increase in food intake and a blood profile indicating metabolic disorder in adulthood (Santos-Silva et al., 2013). This is important regarding information given to new mothers, as even mothers who abstain from smoking during pregnancy but who restart postpartum may still be putting their children at risk.

Smoking prevalence during pregnancy has been found to negatively correlate with the socioeconomic variables such as maternal education (Cnattingius, 2004). Findings from the National Child Development Study, suggest that low social class and smoking during pregnancy influence the development of "high risk" adults, classified as those of a low birthweight and with a high BMI at age 33 (Power et al., 2003). The growth trajectories of the high-risk group showed that they gained weight steadily through life with a linear BMI trajectory. This suggests that *in utero* factors are of precedence, as factors restricting foetal growth were found to be associated with the high-risk phenotype, which could support theories on detrimental catch-up growth or metabolic programming in LBW infants.

1.3.1.2.5 Maternal diet

Dietary intake of nutrients has been linked to subsequent obesity in numerous studies, however due to difficulty in measuring this exposure in humans, the bulk of the evidence is from animal research. There is evidence to suggest that maternal exposure to a high fat diet (Venu et al., 2004, Bayol et al., 2007, Sun et al., 2012), polyunsaturated fatty acids

(Muhlhausler et al., 2010), sugar (Frazier et al., 2008), a low protein diet (Desai and Ross, 2011), vitamin restriction (Venu et al., 2004) and vitamin D deficiency (Morales et al., 2015) can impact on offspring adiposity. Furthermore, the timing of the nutritional exposure is complex, with differential effects observed based on stage of pregnancy or lactation, and therefore further discussion of specific nutrients is beyond the scope of this review.

1.3.1.3 Infant feeding

Infant feeding is another important factor that can affect early childhood growth and subsequent risk of obesity (Ong et al., 2002c, Oken et al., 2008). Exclusive breastfeeding is recommended for the first 6 months of life by the WHO (World Health Organization, 2011b), however UK breastfeeding rates fall short (McAndrew et al., 2012).

Human milk has the perfect combination of nutrients, hormones and immune factors required for infant development, and the beneficial effects of breastfeeding are well-established (Ip et al., 2007) (Victora et al., 2016). However, studies have demonstrated mixed results concerning offspring obesity (Koletzko et al., 2009, Beyerlein and von Kries, 2011, Marseglia et al., 2015). Systematic reviews and meta-analysis have found small, protective effects of breastfeeding on obesity in later life, although effect sizes decrease after adjustment for some confounding factors (parental obesity, maternal smoking and social class)(Arenz et al., 2004, Owen et al., 2005b, Oken et al., 2008).

A dose-response effect has also been observed between duration of breastfeeding and decreased risk of childhood obesity in a recent meta-analysis, particularly for durations over 7 months (Yan et al., 2014). Similarly breastfeeding was deemed protective, either long-term (over 2 years) or exclusive for 6 months, however only long-term breastfeeding remained significant in the multivariate analysis (Rathnayake et al., 2013). Long-duration breastfeeding exerted a protective effect with an age-dependent decrease in body fat in males born to overweight mothers (Buyken et al., 2008).

Findings should be interpreted with caution, as although most epidemiological studies demonstrate a protective effect (Armstrong et al., 2002, Harder et al., 2005), the effect is small and could signify publication bias (Dewey, 2003, Owen et al., 2005a) or residual confounding (Brion et al., 2011). The Promotion of Breastfeeding Intervention Trial (PROBIT) in Belarus, an intervention which successfully increased duration of exclusive breastfeeding,

found no differences in childhood adiposity or metabolic markers, therefore not supporting a protective role for breastfeeding on body composition (Martin et al., 2017).

Breast milk composition is determined by maternal diet and diabetes status and therefore in some circumstances breastfeeding could have a detrimental effect. Breastmilk composition can differ in obese and diabetic mothers with higher levels of fat, glucose and insulin (Young et al., 2012). This may explain why breastfed infants of diabetic mothers had a greater risk of overweight in childhood, with a dose-dependent relationship of breast milk ingested, independent of total milk intake and post-natal diet (Plagemann et al., 2002). In a mouse model of T2DM, cross-fostering pups from diabetic mothers to non-diabetic reduced body weight, perhaps due to the high lipid content of obese dam's breast milk (Reifsnyder et al., 2000), supporting an influence of both breastmilk and intrauterine exposure. A recent study also found that maternal obesity was associated with changes in the in human milk metabolites, and infant body weight (Venditti et al., 2019). Similar to diabetic mothers, obese women have higher levels of certain hormones and growth factors such as insulin, leptin, TNF-α, and IL-6 (Fields and Demerath, 2012, Andreas et al., 2014), but the role these factors has on infant growth is yet to be determined.

Many studies have found that formula-fed infants have different body composition trajectories, which could be a result of the differing milk composition compared to breast milk. As formula contains 1.5-2 times more protein than breastmilk this could accelerate growth velocity (the IGF-1 theory) and hence adiposity (Koletzko et al., 2005). A Cochrane review also reported that in LBW infants, high protein formula accelerates weight gain (Fenton et al., 2014). Furthermore, interventions using lower protein formula have demonstrated a positive effect on childhood body composition (reviewed by Redsell et al. (2016)).

Discrepancies in findings could also be attributed to the populations under study, for example due to genetic differences or differences in social attitudes towards breastfeeding, where there is not support for a causal relationship (Brion et al., 2011). SES is associated with likelihood of breastfeeding in the UK, with those in the more deprived groups less likely to breastfeed (Oakley et al., 2013) and evidence that the protective effect on BMI could be due to social patterning (Brion et al., 2011).

1.3.1.4 Early life growth

Greater infant weight gain has been consistently associated with increased risk of subsequent obesity in meta-analyses (Ong and Loos, 2006a, Druet et al., 2012, Zheng et al., 2018).

A deprived intrauterine environment is associated with rapid postnatal weight gain, leading to increased central adiposity at various life stages (Khandelwal et al., 2014). Although LBW infants are more likely to have catch-up growth (Ong et al., 2000), it is not necessarily detrimental and is essential for neurodevelopment, particularly in preterm or SGA infants (Belfort et al., 2011). This advocates for a balance between adequate early life growth and a need to differentiate 'unhealthy' catch-up growth.

Rapid growth occurs in all infants in the first year of life and is followed by a decline and a plateau, until growth commences an upward trajectory again at the onset of puberty. Adiposity rebound begins at the lowest BMI value, and if occurs early has been shown to be predictive of later adiposity (Rolland-Cachera et al., 2006). Findings from the ALSPAC cohort suggest various facets of early life growth are important risk factors for childhood obesity at 7 years, including birthweight, very early BMI or adiposity rebound, catch-up growth, weight at 8 months and weight gain in the 1st year, (Reilly et al., 2005b).

Infant weight gain was found to have moderate predictive ability for childhood obesity in a meta-analyses using 10 cohorts studies (area under receiving operating curve of 77%) (Druet et al., 2012). A 1 standard deviation unit in weight in the first year was associated with twice the odds of childhood obesity, and 23% increased odds of adult obesity. Recent systematic reviews have also supported that rapid weight gain (RWG) during infancy is associated later adiposity outcomes spanning from childhood to adulthood, but with higher odds in childhood (Zheng et al., 2018).

Regarding the timing of the exposure, using a life course approach, weight gain as early as the first week of life was a predictor of adult overweight status (for each 100-g increase OR 1.28, 95% CI 1.08 -1.52) suggesting that this very early period of life could have a lifelong impact (Stettler et al., 2005). RWG from birth to 1 year tends to be associated with higher odds of childhood adiposity than the timespan between birth to 2 years (Zheng et al., 2018). Infant fat mass at one year of age was significantly predicted by maternal age at delivery

(Chandler-Laney et al., 2013). Other factors that can affect early life growth include birthweight (section 1.3.1.1) and infant feeding (section 1.3.1.3).

1.3.1.5 Early life stress and childhood adversity

Early life stress is another proposed exposure in foetal programming of obesity. Stress is purported to contribute to obesity through hormonal regulation of appetite and eating behaviour (Torres and Nowson, 2007). A 2010 review by Entringer *et al*,. summarises multiple mechanisms that could mediate the effects of prenatal stress, with targets including metabolic, endocrine, or inflammatory systems (Entringer et al., 2010).

The stress response can be measured via corticotropin-releasing hormone (CRH) levels, which when released stimulate the hypothalamic–pituitary–adrenal (HPA) axis. When maternal CRH levels were measured in the late 2nd trimester of pregnancy there was an negative association with offspring BMI z-score at age 3, but an increase in central adiposity (Gillman et al., 2006). Studies using biochemical measures of stress are yet to yield results of a long-term follow-up, however animal studies have provided more evidence. In dams exposed to stress during the third week of gestation, as measured via corticosterone levels, offspring had higher birthweights, greater postnatal body weights, and evidence of impaired glucose tolerance (Tamashiro et al., 2009). Additionally, in mother-offspring bonnet monkeys, a 4 month period of imposed variable foraging demand in early life, which acts as a stressful exposure without food restriction, resulted in higher BMI and abdominal circumference in the offspring (Kaufman et al., 2007). There are parallels with this study to the current situation in the UK with increased reliance on food banks (Loopstra and Lalor, 2017), and uncertain food availability could be an important early life stressor.

In humans, stressful early life events are often defined as adverse childhood experience (ACEs), and there is a growing body of evidence that suggests these can impact on multiple health outcomes. The most frequent stressors examined are those which are direct, such as maltreatment, abuse or neglect, or indirect acting at the household or environment level, such as parental violence, parental separation or criminal behaviour (World Health Organization, 2011a). Other less common ACEs examined are extrafamilial (such as bullying, natural disasters)(Finkelhor et al., 2013, Mersky et al., 2017). For example, gestational exposure to extreme weather such as an ice storm (Dancause et al., 2012) or flooding (Dancause et al., 2015), have been shown to predict childhood BMI and adiposity.

A systematic review, summarising results from eight studies, found that the direct associations between ACEs and OWOB ranged 13–71% increased odds (Hughes et al., 2017). Adolescents who reported an ACE were more likely to have a higher BMI, and there were incremental increases in risk of overweight, obesity and severe obesity with increasing numbers of ACEs (Davis et al., 2019). Whereas a large UK study found 4+ ACEs were associated with 3x increased likelihood of morbid obesity, but not obesity (Bellis et al., 2013). Considering family member death as a stressful exposure, antenatal bereavement has also been associated with later life overweight (Li et al., 2010, Hohwü et al., 2014). These results may suggest that more severe or more frequent exposure may be related to more severe outcomes in terms of body composition.

In a longitudinal study examining childhood neglect and adolescent obesity, the potential mediators were impulsivity, depression, and compulsive eating behaviour (Shin and Miller, 2012). Findings from the ALSPAC cohort, found that ACEs (birth-16 years) were associated with lower risk of educational attainment and drug use and smoking, independent of SES (Houtepen et al., 2019). This suggests that ACEs may influence the likelihood of risky or compulsive behaviours, or more generally are associated with socioeconomic outcomes such as lower education, unemployment and poverty (Hughes et al., 2017).

There is considerable heterogeneity in the literature regarding the definition and measurement of ACEs, and many studies utilise questionnaire data or retrospective data. Many studies examine early life the stressors across childhood (<18 years)(Hughes et al., 2017), therefore further work examining very early life (as a critical period) stressful events on later obesity could yield insight.

1.3.1.6 Factors affecting colonisation of the gut microbiota

There is growing interest in the role of the gut microbiota in disease. Maternal pre-pregnancy BMI, antibiotics usage in the first 6 months and breastfeeding are factors that have been associated with establishment of the gut microbiota (Ajslev et al., 2011). Caesarean section is also an independent risk factor for childhood overweight, which may perhaps be mediated by bacterial colonisation of the gut microbiota (Tun et al., 2018), and the effects of caesarean section have been shown to be independent of maternal antibiotic usage (Mueller et al., 2015a).

Generally, observational studies suggest a link between antibiotic use and weight gain (Million et al., 2013). As microbial populations are largely established in early life this could link to disruptions in gut microbiota leading to altered physiology (Jernberg et al., 2010, Robinson and Young, 2010, Martin and Sela, 2013). This could impact very early on in development, as maternal antibiotic use in the 2nd or 3rd trimester has also been associated with higher childhood adiposity (Mueller et al., 2015b). However postnatal antibiotic use (1st year) has been associated with increased likelihood of overweight or obese in childhood (age 9) (Azad et al., 2014), and a review found antibiotic exposure (first 6-12 months) or childcare attendance (as a proxy), to be a risk factor for later child overweight in a small number of studies (Woo Baidal et al., 2016), suggesting effects may not just be due to *in utero* exposure.

Postnatal timing of the exposure may also be important, as a recent systematic review and meta-analysis found that exposure to antibiotics in the first 6 months postnatal was associated with OWOB (Rasmussen et al., 2018). Exposure between 6-24 months was not associated, which may suggest the first 6 months reflect a critical period. Some studies only find associations with >1 episode of antibiotics (Rasmussen et al., 2018), and evidence of a dose-response effect with recurrent courses of antibiotics (Shao et al., 2017). In line with this, antibiotic use throughout childhood was also associated with childhood weight gain (Schwartz et al., 2016), and small associations for pre and post-natal infections with obesity in early adulthood were uncovered in a large Danish cohort (Cocoros et al., 2013), suggesting an effect beyond early life on outcomes in later life.

Furthermore, the type of antibiotics may also have an impact, as in ALSPAC post-natal exposure to broad-spectrum antibiotics during the first 6 months of life was associated with greater impact on body mass in infancy (Trasande et al., 2013, Bailey et al., 2014).

There is increasing interest in obesity as an infectious disease of viral origin, termed 'Infectobesity'(Dhurandhar, 2011), likening the rapid spread of obesity to that of an infectious disease (Atkinson, 2007). For example, SMAM-1 is an avian adenovirus that acts on adipocytes and has been associated with human obesity (Dhurandhar et al., 1997). Ad-36 is another example of a adenovirus associated with human obesity (Esposito et al., 2012), with some evidence of a causative role, as twin studies discordant for infection have shown the infected twin to be heavier (Atkinson et al., 2005). Obese adults also take longer to shed influenza virus, and therefore there is the possibility that obesity may play a role in viral

transmission (Maier et al., 2018). There is however the risk of reverse causality, for example Heras et al. (2019) found that short-term consumption of a high-fat diet could increase susceptibility to listeria infection in mice.

These findings are interesting, but currently there is limited evidence for a causal relationship between infection and obesity.

1.3.1.7 Sleep

The first few months of life are important for development of healthy sleep patterns and circadian rhythm (Parmelee et al., 1964, Glotzbach et al., 1994). Short sleep duration in infancy has been associated with OWOB (Bell and Zimmerman, 2010) and fat mass in childhood (Reilly et al., 2005b, Bell and Zimmerman, 2010, Diethelm et al., 2011). Associations were not evident for older children or for day-time sleeping (Bell and Zimmerman, 2010). A systematic review examining risk factors for childhood obesity operating in the first 1,000 days found some evidence for curtailed infant sleep, which included quality and timing as well as duration (Woo Baidal et al., 2016). Most studies show weak-moderate associations for short sleep duration and childhood body composition and some evidence of a dose-response (Chen et al., 2008, Taveras et al., 2008). Obese children (6-7 years old) slept half an hour less, but this could perhaps reflect reverse causality (Heppe et al., 2012).

Many of the studies have been cross-sectional and findings from longitudinal cohorts are less encouraging. There was no observed association between sleep duration and BMI in the Longitudinal study of Australian Children (aged 0-7 years)(Hiscock et al., 2011), in the GenerationR study in pre-school children (Heppe et al., 2012), or in the Born in Bradford cohort (age 3)(Fairley et al., 2015a).

Differing results may be due to the definition of sleep problems, which is a great source of heterogeneity. The studies outlined in Table 1.4 are not intended as an exhaustive review of the literature but outline some of the differences in definitions of sleep disturbance or sleep problems, and the associations with measures of body composition. The definitions most frequently involve sleep duration, or alternatively, multiple occurrences of night-time waking plus another factor. For example definitions have also included: parental report of disturbance (Lozoff et al., 1985, Zuckerman et al., 1987), co-sleeping (Richman, 1981), extended duration of a waking event (Richman, 1981, Zuckerman et al., 1987), or frequent

waking and difficultly falling asleep. However there is no commonly accepted definition (Gaylor et al., 2001), and there is still a need for a single definition of infant sleep problems (Alamian et al., 2016).

Many children (20-30%) experience sleep problems in the early years (Sadeh et al., 2011). Defining sleeping issues is further complicated by the subjectivity in parental perception of sleep problems, and individual variation in infant sleeping patterns by developmental stage (Thiedke, 2001, Sadeh et al., 2011). ActiGraph measures, which use a wearable device that records body movement during sleep, can be translated to more reliable sleep–wake measures (Sadeh et al., 1991). In a cross-sectional study using ActiGraph measures, there was a positive correlation between short night time sleep at 6 months and greater weight-for-length at age 6 months (Tikotzky et al., 2010b). However, a longitudinal study found no association between ActiGraph infant sleep and adiposity later in infancy (36 months) (Klingenberg et al., 2013).

Currently studies on sleep have demonstrated mixed results, which may be due to differences in the definition, measurement, timing, severity of sleep problems, and age at outcome. There is a need for longitudinal studies with more objective measures of early sleep duration or disturbance.

Table 1.4 Sleep assessment definitions used in studies examining both sleep and childhood body composition

Author (year)	Criteria for sleep disturbance	Associations with childhood body composition
Zuckerman et al. (1987)	Either waking 3+ per night, an event lasting 1 hour, or parental report of severe disturbance to the mother's sleep	Increased risk of overweight in 11-12 year olds (Alamian et al., 2016)
Richman (1981)	Waking 5+ nights per week in addition to; co-sleeping with parents, waking >3 times per night, or duration of a waking event over 20 minutes	Increased risk of overweight in 11-12 year olds (Alamian et al., 2016)
Lozoff et al. (1985)	3+ waking occurrences per week and parental report of disturbance	No associations with overweight in 11-12 year olds (Alamian et al., 2016)
Taveras et al. (2014)	A score of 'curtailed sleep' (which included quality, timing and duration), from ages 6 months to 7 years	Sleep curtailment from infancy to school age was associated with higher odds of obesity in mid-childhood
Heppe et al. (2012)	Infant sleep duration, at age 2 years was dichotomised into "<11.5 h/night" and "≥11.5 h/night," in accordance to the mean sleep duration stated by the American Academy of Paediatrics	No association with preschool overweight (around age 4 years)
Reilly et al. (2005b)	Questionnaire data on sleep duration at age 3 years	Increased risk of obesity at age 7 years
Tikotzky et al. (2010a)	Various parameters from Actigraph sleep analysis	Sleep duration was negatively correlated with weight-to-length ratio measures at age 6 months
Chen et al. (2008)	Meta-analysis which used various measures of sleep duration from questionnaire data	Increased risk of OWOB (throughout childhood). There were reduced odds for each hour increase in sleep, and evidence of a linear dose-response in younger children (<10 years).

1.3.1.8 Cumulative risk factors

Whilst the exposures explored here could act in an independent manner, it is also plausible that they could interact, or have a cumulative influence on obesity (graphically represented in the Directed Acyclic Graph (DAG) in Appendix A).

The effect of multiple early life risk factors on childhood obesity has been investigated in a handful of prospective birth cohorts. In the Southampton Women's Survey, the presence of five risk factors (maternal obesity, excess GWG, smoking during pregnancy, low maternal vitamin D status), was associated with a 3.99 relative risk of OWOB at age 4 years, which increased to 4.65 at 6 years (Robinson et al., 2015).

Similarly, Gillman et al. (2008a) examined four combined modifiable early life risk factors (maternal smoking during pregnancy, excess GWG, short breastfeeding duration, and infant sleep duration), which were associated increased the risk of overweight (age 3), with sustained increased risk later in childhood (Gillman and Ludwig, 2013).

These studies demonstrate that clusters of socially patterned risk factors show evidence of accumulation of risk over time, and therefore could be important targets for interventions.

1.4 Summary of early life risk factors and adiposity

The increases in obesity prevalence in children are of great concern due to the tracking of weight from childhood to adulthood and will raise new issues for future generations. There are acknowledged problems with primarily using BMI to determine obesity, and its use in children due to differences in growth patterns and onset of puberty.

The majority of the literature on the developmental origins of obesity has focused on maternal exposures, yet risk factors for obesity operate at different stages across the life course (Parsons et al., 1999, Gillman et al., 2008a, Gillman et al., 2008b). Maternal risk factors during pregnancy, such as smoking, excessive GWG or maternal obesity, are important risk factors that can affect the likelihood of the offspring being overweight in childhood (Robinson et al., 2015). Birthweight, often used as a proxy for an adverse intrauterine environment, is well-studied factor that has demonstrated predisposition to both childhood and adult obesity (Ravelli et al., 1999, Carolan-Olah et al., 2015).

There is evidence and plausible hypothesis for several early life risk factors (summarised in *Table 1.3*). However there is uncertainty regarding a causal effects on BMI for exposures: maternal smoking, infant feeding, maternal age, sleep and caesarean birth, which may be due to confounding. Evidence from animal studies seems to support causal mechanisms for stress and infection. However, in epidemiological studies, there is more support for early life factors to indirectly impact on offspring BMI through birthweight and early life growth. Therefore, aside from birthweight, which has been studied in detail, the evidence for other early life modifiable risk factors is less robust, and factors have not been studied with respect to adult obesity (Monasta et al., 2010).

It is also important to note that as well as being independent risk factors, these factors are also likely to interact and influence one another (Appendix A). For example, early life

factors may be strongly influenced by SES, and perhaps more so in recent times due to widening social inequalities in childhood OWOB and changes in social patterning which may lead to clustering of exposures (Knai et al., 2012). Considering the multifactorial aspects of obesity, the interactions between pre-natal events, and post-natal behavioural and environmental factors could increase susceptibility, and ideally should be investigated collectively using longitudinal studies.

Table 1.5 A summary of the risk factors investigated, supporting hypotheses, potential confounders, and evidence (from studies listed in section 1.3).

Risk factor	Hypotheses	Potential confounder(s)	Reference(s)
Birthweight	Causal effect due to programming of body composition (fat and lean mass). Intrauterine growth restriction can lead to 'catch-up growth'.	Parity, gestational age, maternal BMI, maternal smoking, maternal diabetes	(Boney et al., 2005) (Parsons et al., 1999)
Rapid weight gain	A mismatch of intrauterine and post-natal conditions encourages rapid growth in early life.	Parity, gestational age Birthweight, infant feeding	(Ong and Loos, 2006b)
Breastfeeding (vs formula and duration)	Breastfeeding effects colonisation of gut microbiota and self-regulation of appetite. The nutritional composition of breastmilk may be beneficial or detrimental based on maternal characteristics.	SES, maternal BMI	(Armstrong and Reilly, 2002) (Yan et al., 2014)
Early weaning	Nutritional programming of metabolic systems and child growth. Could be dependent on diet quality and may have more of an impact in formula fed infants.	SES, breastfeeding	(Thompson, 2012) (Agostoni et al., 2008)
Parity	First-born children have lower birthweights and experience catch-up growth, compared to second-born.	Maternal age, maternal BMI	(Ong et al., 2002b)
Maternal age	The mechanisms are unclear. Biological mechanisms would suggest both young and advanced maternal age are associated with negative birth outcomes. However, the role may not be causal as maternal age is closely linked with maternal health and SES.	Maternal BMI, parity, SES	(Myrskyla and Fenelon, 2012) (Savage et al., 2013)
Adversity	Stress leading to altered HPA via glucocorticoid pathways, leading to fat deposition.	SES	(Tamayo et al., 2010) (Anda et al., 2006)
Sleep issues	Shorter sleep duration may affect appetite regulation and stress-related pathways	SES	(Bell and Zimmerman, 2010) (Anda et al., 2006)
Infection	Antibiotics can lead to disruption of gut microbiota, or early life infection could alter growth trajectories.	Breastfeeding	(Mueller et al., 2015a)
Caesarean	Could affect the establishment of microbiome	Breastfeeding, maternal BMI, birthweight (birth complications)	(Li et al., 2012)
SES	Socioeconomic inequalities affect many dimensions; the effects are multifactorial encompassing income, education, environment, and diet quality. Potential bidirectional relationship and transgenerational influences.	Maternal BMI	(Shrewsbury and Wardle, 2012) (Gibbs and Forste, 2014) (Howe et al., 2011)

1.5 Epigenetics as a mechanism linking early life factors and later disease

For the DOHAD theory to operate, there must be an underlying mechanism in which an exposure affects and then 'marks' an individual, with a lasting impact on health outcomes years later. The underlying mechanisms are hypothesised to be related to epigenetics. Epigenetic patterns are established early on in development and events *in utero* could possibly program the foetus to anticipate a specific environment postpartum, thereby acting as cellular memory. An insult at a critical period of development could lead to lasting and perhaps permanent effects (Ben-Shlomo and Kuh, 2002). It is plausible that epigenetic mechanisms could act as a mediator between early life exposures and obesity outcomes (such as for birthweight, section 1.5.3.1).

1.5.1 Introduction to epigenetics

Epigenetic modifications are heritable alterations to the genome, which do not change the underlying genetic code but can affect gene activity and expression. Epigenetic modifications can be grouped into broad categories of; chromatin remodelling and histone modifications, DNAm, and small non-coding RNA mechanisms. Within the cell nucleus, genetic material is tightly packaged into chromosomes made up of chromatin. Chromatin, the condensed form of DNA, is wrapped around structural histone proteins known as nucleosomes, forming the characteristic beads on a string appearance. Thereby, modifications which alter chromatin structure and hence packaging (and accessibility) of DNA can influence gene expression.

DNAm is the most well-characterised and well-studied epigenetic mark. DNAm is a cellular regulatory mechanism that involves the covalent addition of a methyl group (CH_3) on cytosine residues adjacent to guanine on DNA, and in mammals is catalysed by DNA methyltransferase enzymes (DNMTs)(Suzuki and Bird, 2008). The ten eleven translocation (TET) family of enzymes oxidise 5-methylcytosines and facilitate reversal of DNAm (Tahiliani et al., 2009, Ito et al., 2010). Methylation plays an essential role in early development and cell differentiation, but also in mediating gene expression thereby determining cell function (Sardina et al., 2018). Clusters of methylated cytosines tend to occur in promoter regions and are referred to as CpG islands (Suzuki and Bird, 2008). Generally, promoter methylation is associated with low or no transcription or gene silencing (Suzuki and Bird, 2008), whilst intragenic methylation is positively associated with gene expression (Jones, 1999). Gene body methylation was originally thought to be a mechanism for silencing repetitive elements (Yoder et al., 1997). However, whole-genome studies have indicated that due to exons being

more highly methylated than introns, with the occurrence of along exon-intron boundaries, it could play a role in alternative splicing (Laurent et al., 2010). DNAm may function in the formation of open chromatin structure and maintenance of enhancer elements (Wiench et al., 2011).

Epigenetic modifications occur during developmental processes and are particularly important in early development and cell specialisation (Hochberg et al., 2011). These times when the epigenome (the collective term for all the chemical modifications to DNA within the genome) is sensitive to change are referred to as critical or sensitive periods. During development, the epigenome is 'reset', and some epigenetic marks are re-established. There is widespread demethylation, followed by specific re-methylation, which must occur for cells to have specialised functions. Specific inherited epigenetics marks are generally at imprinted genes. However, modifications can also occur in response to environmental factors such as diet and lifestyle (Jaenisch and Bird, 2003). Patterns of methylation are now being recognised are more dynamic than previously thought, and there is growing evidence that DNAm can be influenced by environmental factors (section 1.5.3).

DNAm is a good candidate for the biological embedding of early life risk factors for obesity as; it has the capacity to modulate gene expression, can be influenced by environmental factors such as diet and lifestyle (Jaenisch and Bird, 2003), and has demonstrated associations with adiposity.

1.5.2 Epigenetic mechanisms and obesity

The Agouti mice models demonstrated the importance of maternal diet on offspring phenotype (Wolff et al., 1998). In mice, when the Agouti gene is unmethylated, the phenotype is yellow coat colour and a predisposition to obesity, hyperinsulinemia, cancer, and reduced lifespan. Whereas when the agouti gene is methylated (i.e. normal state), mice have a brown coat, with low risk of obesity and are healthier than their yellow counterparts. Other than the methylated allele, the mice are genetically identical. When pregnant female yellow mice are fed a methyl-donor rich diet, the pups had a brown coat and a healthy phenotype (Wolff et al., 1998) thereby demonstrating the importance of methyl-donor availability. Animal studies have demonstrated the potential for environmental and dietary influences on health and disease, and suggest a mediating role for DNAm linking early

development and postnatal body composition (Drake et al., 2005, Burdge et al., 2007, Godfrey et al., 2011).

There have been a number of human studies which have identified methylation patterns associated with later life adiposity ((for reviews see (Reynolds et al., 2013b, Rohde et al., 2018) and (van Dijk et al., 2015)).

Replication has been problematic in many candidate gene studies. There was a robust, large-scale BMI EWAS that utilised data from across multiple cohorts, and successfully validated 187 loci in independent cohorts (Wahl et al., 2016). However, regarding the direction of the effects, the majority of these loci were deemed a consequence rather than a cause of BMI. Others have also found that changes in BMI may occur prior to methylation changes (Richmond et al., 2016, Wahl et al., 2017), for example weight loss after bariatric surgery influences DNAm (reviewed by (Izquierdo and Crujeiras, 2019)). Overall, these studies provide evidence for epigenetic changes in relation to obesity or weight change.

There may also be a link between being biologically older and childhood body composition, as positive epigenetic age acceleration (measured using Horvath's epigenetic clock) at birth was associated with developmental characteristics longitudinally across childhood, including increased fat mass (Simpkin et al., 2017). DNAm has important roles in expression of imprinted genes, which are expressed from the parent of origin and which may play a role in the transgenerational development of obesity. Insulin-like growth factor 2 (IGF2), a key human growth factor, is an example of an imprinted gene differentially methylated in those periconceptionally exposed to famine (Heijmans et al., 2008a).

Overall, these studies establish a link between epigenetics and obesity. From candidate studies there have been many promising epigenetic biomarkers, with accumulating evidence linking DNAm changes and metabolic health outcomes. So far, for many of these genes the functional consequences of methylation changes remain uncertain.

1.5.3 Evidence for DNA methylation markers of early life exposures

There is growing interest in the use of malleable epigenetic markers as potential molecular mediators (i.e. intermediary), as biomarkers to identify those most at risk of subsequent disease, or as potential targets in disease treatment as a means of personalised medicine. There have been DNAm changes associated with some of the early life risk factors for subsequent obesity. More specifically, these DNAm changes have been linked with foetal

growth, birthweight, metabolism, and linked to maternal factors such as smoking and nutrition (Reynolds et al., 2013b). This section will focus on validated markers or markers from meta-analyses of human studies focusing on DNAm changes at individual CpG loci.

1.5.3.1 Birthweight

The Pregnancy and Childhood Epigenetics (PACE) Consortium, combines multiple cohorts with methylation data from 450K or EPIC arrays on newborns and children. It includes many large prospective studies such as ALSPAC and the Generation R study. Therefore, PACE can be used to conduct meta-analyses due to large sample size with much greater statistical power. In a large meta-analysis of 24 birth cohorts from the PACE consortium, birthweight was associated with differential methylation at 914 loci in neonatal blood (Küpers et al., 2019). Some of the loci demonstrated overlap with loci previously identified as associated with maternal smoking (*n*=55) (Joubert et al., 2016), and a few loci also with maternal BMI (*n*=3)(Sharp et al., 2017). There were a handful of CpG loci that were differentially methylated across the life course (in childhood, adolescence and adulthood), which mapped to genes; *KCNC4, GLI2, HOXC4, ZNF274, MIR548F5.*

The biological effects of glucocorticoids, steroid hormones hat play a role in foetal growth and development, are relayed by glucocorticoid receptors (GR). It is established that glucocorticoid exposure *in utero* is associated with low birthweight (Seckl, 2004). It has been found that increased methylation in the GR promoter, which leads to reduced expression of GR thereby reducing levels of glucocorticoid signalling, is associated with higher birthweights (Filiberto et al., 2011). In mothers, increased methylation (leading to reduced transcription) of the enzyme which inactivates glucocorticoid (11β-hydroxysteroid dehydrogenase type 2) (Alikhani-Koopaei et al., 2004), leads to higher circulating glucocorticoids, and has also been associated with low birthweight (Marsit et al., 2012).

Birthweight is often used as proxy for nutrient restriction. Those exposed to famine in the peri-conceptional period exhibited hypomethylation of IGF2 differentially methylated region compared to unexposed siblings (Heijmans et al., 2008b). Exposure late in gestation was not associated with changes in *IGF2* methylation, suggesting that very early development is a critical period in the establishment of environmentally patterned epigenetic marks. Mediation analysis also suggests that DNAm may be a mediator between early life famine

and subsequent increased BMI, interesting near genes with metabolic roles including energy metabolism (*PIM3*), glycolysis (*PFKFB3*), and adipogenesis (*METTL8*) (Tobi et al., 2018).

1.5.3.2 Maternal factors

1.5.3.2.1 Maternal BMI

In a large meta-analysis, maternal BMI was robustly associated with cord blood methylation at eight CpG loci within the PACE consortium (Sharp et al., 2017). Higher maternal BMI was associated with lower methylation at two CpG loci located in the *VIPR2* gene, a gene which encodes vasoactive intestinal peptide receptor 2 (VIPR2). The VIP pathway has been strongly associated with fat mass, and therefore may be important for obesity development (Liu et al., 2010).

In ALSPAC, increased GWG was associated with increased methylation at several CpG sites in offspring cord blood DNA (Morales et al., 2014). In addition, offspring DNAm differences have been observed in those born to both overweight and underweight ALSPAC mothers (Sharp et al., 2015a). In a small study, there was evidence of altered methylation in differentially methylation regions (DMRs) of imprinted genes in cord blood in association with parental obesity (Soubry et al., 2013). Thus far there has been inconsistency in the associations found in various cohorts for maternal BMI, and many of the associations between maternal BMI and offspring methylation at birth have not been replicated (Sharp et al., 2017). Although there is some evidence that the detrimental effects of maternal obesity are causally related to epigenetic alterations (Godfrey et al., 2017).

1.5.3.2.2 Maternal smoking

DNAm measures are particularly useful for exposures such as smoking when self-report data may be prone to bias (Dietz et al., 2010) and cotinine (metabolite of nicotine) measures are not available. When data were combined across studies (PACE consortium), there were many significant CpGs associated with maternal smoking, robust to different adjustment and analytical methods (Joubert et al., 2016). Associations were stronger for sustained smoking rather than smoking at any time during pregnancy (Joubert et al., 2016). *PRDM8* had the most CpGs associated on the array, and belongs to the SET domain family of histone methyltransferases (Fog et al., 2012), more specifically, this gene targets H3K9 of histones to repress transcription (Eom et al., 2009). Gene expression was analysed further in genes that had differential methylation, it was found that there was agreement for 6 genes.

PASK (PAS domain containing serine/threonine kinase), was one of the genes, and is a 'nutrient sensor', involved in the regulation of glucose and lipid homeostasis, and therefore may play a role in metabolic syndrome (DeMille and Grose, 2013, Zhang et al., 2015). Mice lacking *PASK* who are fed a high fat diet do not develop obesity (Zhang et al., 2015), therefore *PASK* deficiency can protect against diet-induced obesity.

Longitudinal analyses carried out in the ALSPAC cohort found that some CpGs that were altered in cord blood demonstrated reversible methylation, whilst others had persistent methylation (*AHRR, MYO1G, CYP1A1 and CNTNAP2*) across childhood and adolescence (Richmond et al., 2014). Recent work has been in the development of a methylation score, which could be used to predict previous smoking exposure. The score uses methylation across 15 CpG sites in 11 genes in order to predict prenatal smoke exposure (Richmond et al., 2018). This useful approach could be utilised to develop predictive models for other early life exposures.

1.5.3.2.3 Maternal age

Offspring of older mothers are at risk of adverse birth outcomes, which can impact via a number of mechanisms. As established by Horvath, age is associated with DNAm changes (Horvath, 2013). Furthermore, gene expression in human oocytes changes with age, which could affect biological function (Grondahl et al., 2010).

An EWAS (450K) in newborns from the Norway Facial Clefts Study identified hypomethylation within the KLHL35 gene in offspring born to older mothers (Markunas et al., 2016). The findings were replicated in the MoBa cohort and also in women in age 40-60 years, suggesting that methylation differences persisted across the life course.

Another smaller study examined methylation using the 27K array in a cohort of 168 newborns, and found a correlation between maternal age and methylation at genes related to neurological regulation, embryo development, and glucose regulation and metabolism (Adkins et al., 2011).

Increasing maternal age was associated with epigenetic differences in adult daughters, including an inverse association with methylation in the promoter region of *LHX8*, a gene related to female fertility (Moore et al., 2019). Furthermore *LHX8* expression is a marker for the activity of brown adipose tissue (Jespersen et al., 2013), with increased activity related to lower BMI (Cypess et al., 2009). Researchers also identified a set of genes associated with

maternal age that are important regulators of the development of various cancers and neurodevelopmental disorders.

Although there is a biological *in utero* connection between mothers and offspring, it is likely that maternal and paternal age will be highly correlated for most. Epigenetic changes that occur within sperm DNA have also been associated with offspring disease risk, in particular related to neuropsychiatric disorders (Jenkins et al., 2014).

1.5.3.3 Infant feeding

A systematic review examining breastfeeding and offspring DNAm found evidence of a negative association with promoter methylation in the *LEP* (Leptin), *CDKN2A* (tumour suppressor) and *SLC2A4* (glucose transporter) genes, in a handful of studies (5 in humans)(Hartwig et al., 2017). Breastfeeding was positively associated with *NPY* (encodes an orexigenic neuropeptide). Duration of breastfeeding is negatively associated with *LEP* methylation and with decreased infant growth (Obermann-Borst et al., 2013, Pauwels et al., 2019), therefore leptin may be a mediator of the developmental programming effects (Vickers and Sloboda, 2012).

Differential CpG methylation (buccal) in the *RXRA* gene was associated with duration of breastfeeding and with infant growth (Pauwels et al., 2019), a gene which has been previously associated with childhood fat mass (Godfrey et al., 2011).

Mischke and Plosch (2013), hypothesised that the effects of breast milk on DNA may be mediated by the microbiome (Mischke and Plosch, 2013).

These results overall do provide some evidence of a relationship between breastfeeding, infant growth and DNAm in some metabolic and appetite-related genes.

1.5.3.4 Early growth

There has not been an EWAS so far that has specifically examined RWG and childhood methylation, however some early life methylation changes have been observed. RWG is more likely in LBW infants; being SGA and weight gain in the first 3 months has been associated with lower *IGF2* DMR DNAm (Bouwland-Both et al., 2013). Specific DMRs of imprinted genes in umbilical cord blood and infant body composition were investigated in the Newborn Epigenetics Study (NEST) (Gonzalez-Nahm et al., 2018). Sex differences were identified, whereby lower weight-for-length z score at 1 year was associated with higher

methylation within mesoderm-specific transcript *MEST* in females, whilst in males higher methylation was observed in the paternally expressed gene 10 (*PEG10*) and *IGF2*.

Postnatal growth was associated with differential methylation in the *TACSTD2* gene using gene-specific pyrosequencing (Groom et al., 2012). Both gene expression and DNAm within the gene were also associated with childhood fat mass, however subsequent causal analyses demonstrated that the association was likely due to reverse causation or confounding.

1.5.3.5 Stress

In the first few years of life the immune system is sensitive to environmental stimuli (Danese and J Lewis, 2017). The pathways through which stress can impact on health could be through disturbance of inflammatory processes or via the hypothalamic–pituitary–adrenal axis (HPA). As a key pathway, many studies have focused on the components of the HPA axis and the glucocorticoid system, and how epigenetic alterations may disrupt of normal function in these pathways.

Early emotional experiences could lead to epigenetic alterations that impact on offspring obesity. This was observed in rat pups; offspring groomed by their mother had reduced anxiety in adulthood and associated epigenetic changes in the GR (Weaver et al., 2004a). Whereas pups raised by less nurturing mothers had hypermethylation within the GR receptor and increased stress response (Weaver et al., 2004b). In mice, early life stress (separation of pups from dams) was associated with DNAm changes and accompanying upregulated expression of the pituitary *POMC* gene, which mediates the adrenocortical response to stress (Wu et al., 2014b). Although human studies are limited, lower parental warmth has been associated with childhood overweight and therefore DNAm could mediate this association (Fairley et al., 2015a). Furthermore, there is some evidence for a role of GR methylation and stress in humans, determined using hippocampal tissue from the Quebec Suicide Brain Bank (Labonté et al., 2012). In the suicide completers who experienced childhood trauma, there was differential methylation in the promoter region of *NR3C1* (encodes glucocorticoid receptor) when compared to controls (Labonté et al., 2012).

The effects of natural disasters and a periods of hardship during pregnancy demonstrate that stress has the capacity to alter offspring methylation levels (Meaney and Szyf, 2005). Child abuse and exposure to partner violence can also alter offspring methylation (McGowan et

al., 2009, Radtke et al., 2011). In adolescents, adverse childhood experiences were associated with DNAm in genes also related to risk of obesity (*PCK2, CxCl10, BCAT1, HID1, PRDM16, MADD, PXDN, GALE*)(Kaufman et al., 2018). Findings for *PCK2* were most robust and were replicated across 2 cohorts, suggesting that *PCK2* methylation may be a mediator for intrafamilial childhood adversity (e.g., physical abuse, witnessing domestic violence) and BMI.

In a study on internationally adopted adolescents from deprived backgrounds, appropriately matched with non-adopted controls, differences in methylation patterns were investigated in an EWAS (Esposito et al., 2016). There were marked differences in cell type composition in adopted youths and differential methylation at 30 CpG loci, which mapped to enriched gene clusters related to neural and developmental functions. The adopted children were selected to model significant childhood adversity, although it is unclear specifically which forms of adversity they experienced but could include abuse, neglect, malnutrition or infection each with varying severity (Esposito et al., 2016).

A systematic review examining DNAm and childhood trauma summarised that numerous studies have identified differential methylation in the genes: *SLC6A4, BDNF, OXTR* and *FKBP5*, and most robustly in the GR *NR3C1* gene (Nöthling et al., 2019). For a comprehensive review on trauma-induced changes in DNAm see (Vinkers et al., 2015).

There have now been several epigenome-wide association studies for childhood adversity and DNAm, however as noted by (Houtepen et al., 2018a) many of these studies utilise a candidate gene approach or use susceptible populations (Houtepen et al., 2018a). Currently there are few studies that have examined very early life (rather than childhood) stress and DNAm in humans.

1.5.3.6 Infection

Epigenetic mechanisms are critical for normal development of the immune system, and the early postnatal period (0-12 months) reflects a time when the epigenome is amenable to environmental exposures affecting innate and adaptive immune responses (Martino et al., 2014).

It is plausible that pathogens could affect epigenetic processes, in particular viruses that reproduce undiscovered in host cells (reviewed by (Paschos and Allday, 2010)). There is evidence of DNAm changes in responses to Epstein-Barr virus infection (Birdwell *et al*.,

2014), human rhinovirus (McErlean et al., 2014), and HIV infection (Horvath and Levine, 2015, Nelson et al., 2017).

There is limited evidence on bacterial infections and DNAm in humans. Bacterial infections have been associated with rapid and active demethylation, rarely occurring at promoter regions, but localising to distal enhancer elements that regulate activation of key immune transcription factors (Pacis *et al*., 2015). Mycobacterium tuberculosis has been shown to alter the epigenome, but the mechanism is not fully understood (Kathirvel and Mahadevan, 2016). There is also evidence that some pathogens can disrupt histone modifications and reprogram host gene expression (reviewed by (Strunk et al., 2013)).

DNAm changes and early life exposure to antibiotics have not been examined. Maternal antibiotic use during pregnancy was examined in 397 pregnant women in DMRs of imprinted genes using bisulfite pyrosequencing (Vidal et al., 2013) in the NEST cohort. Differences were observed for *IGF2, H19, MEG3, PEG3* and *PLAG1,* with the latter also associated with birthweight. This could suggest that the association between prenatal exposure to antibiotics and birthweight may be mediated by changes in regulatory regions of some imprinted genes. DNAm at imprinted regions has also been investigated with respect to intrauterine infections. In cord blood, DNAm was increased at the *PLAGL1* DMR in pre-term infants with chorioamnionitis (a bacterial infection that occurs before or during labour), and for infants with funisitis (inflammation of the connective tissue of the umbilical cord), in the NEST cohort (Liu et al., 2013b).

1.5.3.7 Sleep

Overall there have been few studies examining sleep deprivation and DNAm, and non in children. In mice, sleep deprivation has a broad impact on DNAm in the cerebral cortex, in gene pathways involved in neuritogenesis and synaptic plasticity, and can increase the expression of DNA methyltransferases (Massart et al., 2014). In an EWAS in adults, even one night of sleep deprivation altered DNAm, particularly in the Notch and Wnt signalling pathways, which are important developmental pathways often dysregulated in cancer (Nilsson et al., 2016).

1.5.3.8 Transgenerational influences

Data from animal studies has demonstrated that the impact of a high fat diet has lasting effects for three successive generations in mice (Sarker et al., 2018). First generation

offspring of those exposed to prenatal famine have a higher BMI (particularly women) (Roseboom et al., 2000) and grandchildren exhibiting increased neonatal adiposity (Painter et al., 2008). It is possible that these effects are transmitted through transgenerational epigenetic inheritance but human data are sparse.

1.5.3.9 SES

There are a handful of studies examining the effects of early life SES on DNAm in peripheral blood (Demetriou et al., 2015). In young children (17 months, *n*=120), low maternal education was associated with a 1.3% increase (p=0.043) in *INSIGF* methylation; the overlapping region of *IGF2* and INS (insulin) (Obermann-Borst et al., 2012).

However, most studies examining SES have measured outcomes in adults. A genome-wide methylation analysis examined childhood SES and methylation in middle-age adult males (42-45 years) from the 1958 British Birth Cohort Study (Borghol et al., 2012). The study examined methylation of around 22,000 gene promoters using methylated DNA immunoprecipitation (MeDIP). Childhood SES was associated with differential methylation at 1252 sites, with enrichment in genes related to extra- and intracellular signalling and metabolism. This was in contrast with adult SES for which there were fewer (*n*=545) significant associations.

A genome-wide approach using the 27k array found 3 loci that had e small (less than a 5%) changes in methylation when comparing low and high early life SES (by occupation) in adults (mean age 33) (Lam et al., 2012), and no significant CpG loci were found for adult SES.

In a large sample (*n*=857) of healthy adults from the EPIC-Italy cohort, low life course SES was associated with lower methylation in pro-inflammatory genes (Stringhini et al., 2015). Lower life course SES has also been associated with epigenetic age acceleration (the difference between DNAm age and chronological age), using data from 3 large prospective cohorts (Fiorito et al., 2017). Furthermore, consistent results were found by Austin *et al*, with the observation that low SES in early life predicts age acceleration, whilst no associations found for later life SES (Austin et al., 2018). Both studies did not find attenuation of the association with upward social mobility. In additional to changes at individual CpG loci, global changes in methylation have been associated with SES (Subramanyam et al., 2013, Tehranifar et al., 2013).

These studies suggest that DNAm may provide a link between early life, or life course SES and disease risk. However, heterogeneity in study designs including the measure of SES and the measurement of DNAm make comparisons difficult.

1.6 Integrating the social determinants of health and epigenetic mechanisms with regards to obesity

There have been a number of theories to explain the rise in obesity (see section 1.2). A failure of previous frameworks is that they often fail to take into account the complexity of the interaction of early life factors, lifestyle, socioeconomic status and gene-environment interactions. There are a number of lifecourse models that more generally discuss lifecourse factors with regards to health, but few have attempted to model lifecourse factors specifically with regards to obesity.

There has been a life-course framework for obesity prevention proposed by Pérez-Escamilla and Kac (2013). Their approach uses a social ecological model which has layers from the individual, microsystem (home), mesosystem (neighbourhood), exosytem (larger environment), and macrosystem (social and health policies), and focuses on the consequences of maternal–child obesity. Hawkins *et al*, apply a similar approach, utilising the same factors but term these 'above water' levels, with the addition of 'below water' levels as the biological factors (Hawkins et al., 2018). Additionally, whilst many studies have examined many individual, independent factors (section 1.3), few studies have incorporated empirical data into a lifecourse framework model.

It is well established that social inequalities and lifestyle factors are significant contributors to health and are included in most theoretical models regarding the determinants of health. However, as epigenetics is a relatively new area of research, where it fits within models of the determinants of health has not yet been considered. Marmot and Wilkinson (2004) acknowledge many important factors impacting in early life that can affect health and wellbeing. Dahlgren and Whitehead's determinants of health model acknowledges hereditary factors and individual lifestyle factors (Dahlgren and Whitehead, 1992), however epigenetic factors would fall between the two. Their model also encompasses the socioeconomic, cultural, and environmental conditions, which collectively can have a great impact on health, and are factors which may have accompanying epigenetic changes.

In the Grossman model, health is treated as stock which depreciates over time without adequate investment in health (Grossman, 1972). For instance, as education increases, an individual is more efficient at taking care of their own health and therefore they increase investment in health. The Grossman model assumes individuals are born with a given level of health stock, however in terms of the DOHaD hypothesis, the level of health stock would potentially be determined by maternal and transgenerational influences, and early life exposures. Therefore, an individual with detrimental early life experiences embarks on life with less health stock and therefore lower health. These individuals would need greater investment in their health to simply maintain an equivalent level of health to that of an individual born with neutral health stock without negative exposures in early life. Risk could further be increased when combined with an obesogenic environment and underlying genetic susceptibility. Therefore a new conceptual model is required in order to incorporate these factors that can impact *in utero* via maternal exposures, on the individual level across the life course, or that have transgenerational influences, and ultimately influence 'health stock'.

In light of this, I propose a new model of the social determinants of health, which acknowledges the role of epigenetics and the interaction with behavioural-lifestyle, and socioeconomic factors (*Figure 1.1*). In this model each of the risk factors presented (circles) has the capacity to impact on DNAm, acting at the individual level, through maternal influences or transgenerationally. By applying this framework, the integration of the various sources of data, such as epigenetic, genetic and epidemiological data has the capacity to improve prediction models.

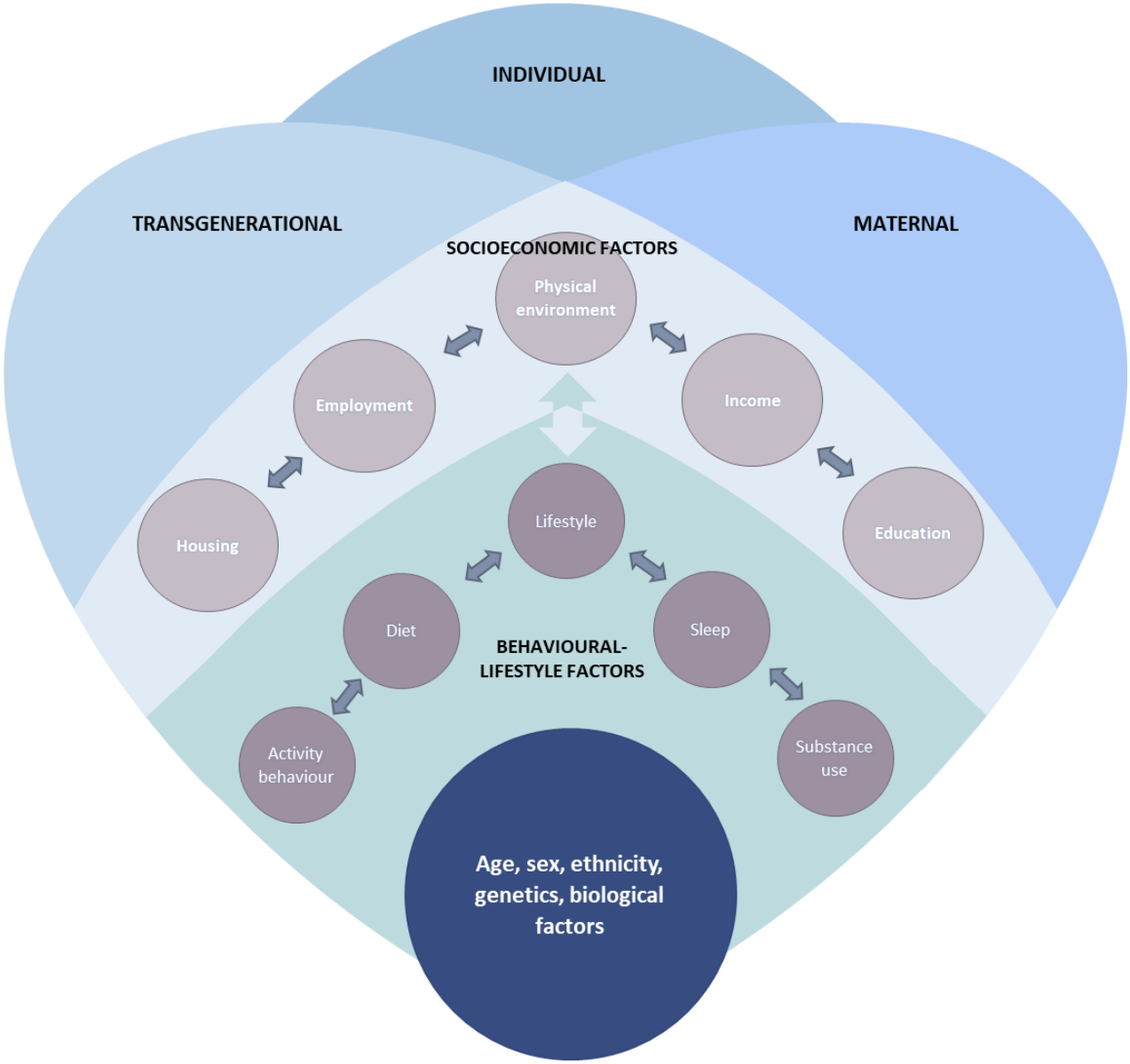


Figure 1.1 *Conceptual framework of factors influencing obesity risk.*

The framework encompasses the complex interactions of socio-economic, behavioural and lifestyle factors that contribute to the development of obesity. It is an adaptation of Dalgren and Whitehead's determinants of health model (1991), which incorporates the DOHaD hypothesis and the literature on early life risk factors of obesity.

In this model, the many determinants of obesity do not act in isolation, but are layered, interlinked (indicated by the arrows) and influence other factors (circles). The socio-economic factors influence the behavioural-lifestyle factors and vice versa. Each of these factors would have the capacity to impact on the individual level but also from maternal factors and transgenerationally. For example, diet on the individual level would encompass eating patterns, nutrient intake and early life feeding. On the maternal level, this refers to maternal diet during pregnancy and weaning. On the transgenerational level, this would refer to the individual's grandmother's diet. Within this hypothesis, each of the circled factors in theory could have associated epigenetic changes. The demographic factors (age, sex, ethnicity, genetics and biological factors including metabolism, puberty status etc), are positioned on the outer edge indicating that they are non-modifiable but acknowledging their effect on weight-related outcomes.

1.7 Summary

In summary, themes emerging from epidemiological studies indicate that nutrition and early feeding may be important, perhaps due to the relationship between to intrauterine conditions, birthweight, and catch-up growth. Aside from nutrition, maternal factors are of great importance as they represent a sustained *in utero* exposure. Very early postnatal life also represents a critical period, with the potential risk factors of infant feeding, adversity, sleep, and socioeconomic disadvantage. Additional research is required in order to determine the relative contribution of early life, lifestyle, and environmental factors to later life OWOB.

In studies examining early life risk factors, there are a lack of long-term follow-up data from childhood to adulthood, and a limitation of current research is that much is cross-sectional in nature or relies on retrospective data collection, which could be addressed by utilising longitudinal data sets. The timing of the exposure can also lead to differing effects in the offspring; therefore, further research is required in order to pinpoint critical periods in an infant's life. At present it is unclear if these early life factors of childhood obesity are risk factors for adult obesity in populations with low levels of childhood obesity, or if the risk is due to the tracking of obesity in childhood into adulthood (Freedman et al., 2001). Future studies ideally would examine multiple indices of adiposity to determine if exposures are related to overall body size, body fatness or the positioning of fat, each of which can carry different risks.

It is not only plausible, but also likely that epigenetic mechanisms could be a mediator between early life exposures and obesity outcomes. However, many of the earlier epigenetic studies used basic techniques and did not state whether they have adjusted for confounders (Obermann-Borst et al., 2012) or corrected for multiple testing in their analysis. Findings so far are suggestive and further replication studies are required using longitudinal cohorts. Investigation of the epigenetic changes associated with exposures may elucidate causal pathways and underlying mechanisms, and hence identify therapeutic targets. Furthermore, epigenetic markers could improve prediction models or highlight those 'high-risk' individuals who would benefit from intervention or monitoring, thereby guiding personalised interventions to improve public health.

1.8 Hypothesis and aims

This study will investigate the hypothesis that DNAm may be a mediating mechanism between early life exposures and subsequent obesity.

The aims of this thesis are:

1.) To investigate the impact of early life exposures and SES on childhood and adult adiposity using multiple indicators
2.) To investigate regional temporal changes on obesity, allowing scrutiny of the impact of modern environmental and socioeconomic factors
3.) i.) To identify gene-specific methylation differences in relation to those early life exposures found to influence obesity in children and adults

    ii.) To investigate the methylation differences in relation to the early life exposure(s) found to influence obesity in an adult population.

This study will address the limitations and unanswered questions summarised in section 1.7. Aim 1 examines the latency of effects; in particular, which early life exposures are associated with adult obesity, as it unknown whether some exposures that are associated with childhood obesity are also associated with adult obesity. Aim 1 also addresses whether associations for exposures differ across outcome measures of adiposity.

Aim 2 addresses the unanswered questions around cohort-timing and environmental effects, i.e. if exposures are associated with adult obesity in those without an early life obesogenic environment.

The final aim attempts to uncover novel changes in methylation in early life exposures that have demonstrated consistency from aims 1 and 2. This can be split into two parts; firstly to identify DNA methylation changes in children and adolescents, and secondly to quantify these methylation changes in adults, thereby examining the persistence of the effects.

1.9 Study design

Figure 1.2 outlines the study design and how data from each cohort was employed to address the aims.

Firstly, a literature search was carried out to explore the early life exposures implicated in obesity (section 1.3), to in order to determine exposures of interest. This study primarily

utilised epidemiological data from two North East birth cohorts; the Newcastle Thousand Families Study (NTFS) and the Gateshead Millennium Study (GMS). The objectives for aim 1 were to examine early life exposures with respect to adiposity outcomes in NTFS adults, and GMS children.

Using two cohorts from the same geographical area allows examination of these relationships over time and observation of the role of environment. To address aim 2, using both NTFS and GMS outcomes in childhood, the associations between early life factors, SES, and obesity were also investigated.

To fulfil aim 3, exposures of interest were investigated using DNAm data from the Avon Longitudinal Study of Parents and children (ALSPAC) cohort (Alspac Study Team, 2001), to examine exposures in relation to methylation and obesity. The basis of these findings guided methylation targets to examine in the NTFS samples.

The findings of this analysis could further understanding of the biological, social and economic factors which lead to health disparities over the life course.

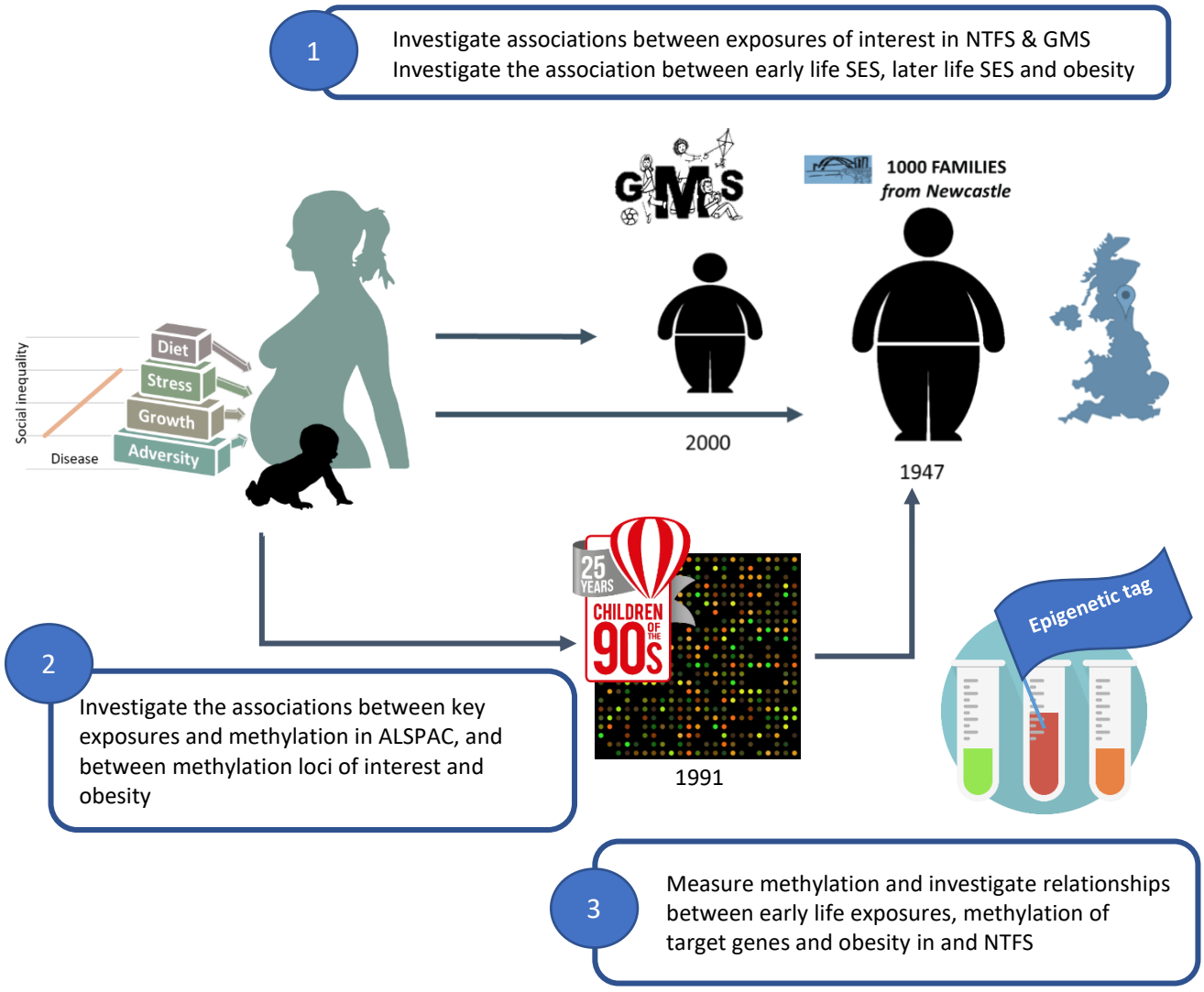


Figure 1.2 Overview of the study design, cohorts and aims
The years refer to the cohort year of birth. NTFS, Newcastle Thousand Families study, GMs, Gateshead millennium study; ALSPAC, Avon Longitudinal Study of Parents and children (AKA children of the 90s); SES, socio-economic status.

Chapter 2. Data and methods

The analysis undertaken in this thesis can be broadly arranged into two sections. First was the epidemiological analysis to determine the focal early life exposures associated with childhood and adult body composition in two North East cohorts (GMS and NTFS). The second part was the epigenetic analysis, whereby the focal early life exposures were investigated in a cohort with epigenetic data (ALSPAC). This also included a simplified version of the epidemiological analysis using the key early life exposures to substantiate the associations, and then investigated these exposures and DNAm in childhood and adolescence. Finally, to determine if DNAm patterns persist, the significant CpG loci identified in ALSPAC children were investigated in NTFS adults. This chapter details each of the cohorts and the methods for all analyses.

2.1 Datasets used

2.1.1 The Newcastle Thousand Families study (NTFS)

The Newcastle Thousand Families study (NTFS) is a birth cohort based in the North East of England. The study began in 1947, two years after the end of WWII. The study's original intent was to investigate the high infant mortality rate, which at the time was mostly attributed to acute infections. This conclusion was made prior to WWII by Sir James Spence, a paediatrician who undertook a review of the causes of death in children under 5. After the war, the NTFS was set up to investigate risk factors for infection, with the original intent to run the study for 1 year. However, the study continued and is now in its 8th decade.

Data were primarily prospectively recorded for the cohort. There are data available on a variety of early life factors, along with physical outcome measures in early and late adulthood, which makes it possible examine body composition in later life in relation to exposures in early life.

The cohort originally included nearly all (*n*=1142) babies born in Newcastle upon Tyne between May and June 1947 (Pearce et al., 2009). Data have been collected at numerous time points over the participant's life course (up until age 60 to date). The cohort have been followed extensively by utilising general practitioners, health visitors and schools throughout childhood (until they were 15 years old). Data collection was facilitated through the

placement of red spots on the study member's medical records to identify them, and hence they were referred to as 'red spots'. There were further intermittent follow-ups during adulthood (ages 18, 22, 32) with data collected on employment, anthropometrics, psychology, and crime.

At age 50 (49-51), the cohort were traced and took part in a physical assessment (*n*=412) and questionnaire (*n*=574). The clinical assessment covered a range of health outcomes through physical assessment including cardiovascular, metabolic, and musculoskeletal measures. Biological samples (serum and urine) were also obtained. Another similar clinical assessment took place at age 60. Comprehensive details on key findings and data collected can be found in the cohort profile (Pearce et al., 2009).

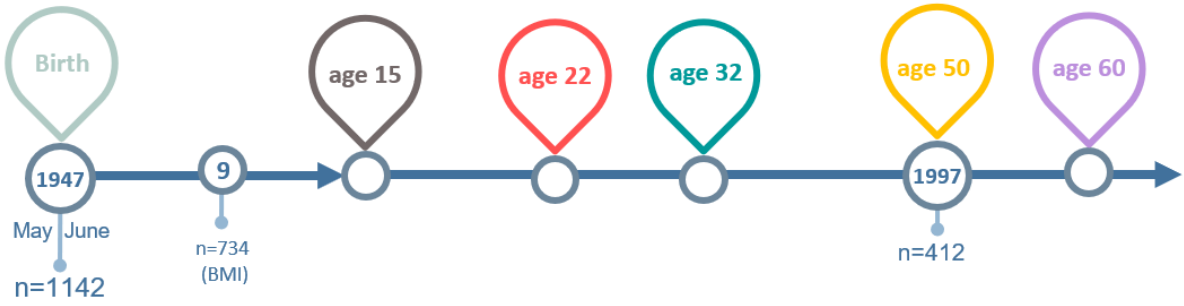


Figure 2.1 NTFS data collection to date.

There were multiple data collection sweeps throughout childhood to age 15, and then intermittent revisits at various time points during participant's lives, until the most recent age 60 follow-up.

2.1.1.1 Measurement of body composition outcomes

Measures of height and weight were taken throughout childhood at ages 3, 5, 9, 13, 14 and 15 years. At the age 50 and 60 clinical assessment, anthropometric measures (height, weight, waist and hip circumferences and bioelectrical impedance) were taken, which were performed in the morning after an overnight fast of at least 10 hours. An average of three measurements of bioelectrical impedance (Holtain) were used to estimate percentage body fat (BF%), using standard regression equations.

2.1.1.2 Measurement of exposure variables and covariates

In NTFS, maternal age, gestational age, and birthweight were recorded at the time of delivery and taken from hospital records. All other factors were recorded by health visitors.

Duration breastfed was the difference between the date of first and last time the participant had breast milk (in days), recorded by health visitors. This was analysed as a continuous and categorical variable. The duration of exclusive breastfeeding was the difference in days between the date breastfeeding commenced and the date of introduction of bottle-feeding with an alternative milk (dried, cow's or mixed feed). The age at weaning was the age at the introduction of solid foods. If date ranges were noted rather than a single date, the lowest date was taken. If no date for first breast milk was noted (but there was a date of cessation), date of birth was used.

Infections were reported throughout childhood by doctors, health visitors, or hospital referrals. Infections in the first year of life were used in this study. Only infection data with low risk of ascertainment bias were used. Bacterial infections included whooping cough, Tuberculosis (TB), scarlet fever, and pneumonia. Viral infections included measles, mumps, rubella, meningitis, bronchitis, and chicken pox. Infection in first year was defined as a dichotomous variable if the study member experienced any of the aforementioned infections ("had", "did not have"). Viral and bacterial infections were grouped initially, however as coefficients had opposite effects, viral and bacterial were then separated as variables.

Socioeconomic indicators in early life included; father's occupational social class, and household deprivation. To increase group sizes, social class was re-categorised into least advantaged (class IV and V), mid (class III skilled and unskilled) and most advantaged (class I and II).

Household deprivation was determined through indicators of inadequate housing conditions as assessed by Newcastle's public health department around time of birth. These were the presence of (0-4 factors) including; overcrowding, lack of hot water, shared toilet, and dampness or poor repair. As low numbers were observed in some categories, this variable was recoded to a binary variable representing either no housing issues (0) or evidence of housing issues (>=1).

2.1.1.2.1 Later life and lifestyle factors

Later life socioeconomic variables were determined from the age 50 questionnaire and included occupational social class, highest education level achieved and total household income (after tax). Although these are often used interchangeably in studies examining

health inequalities, they measure different components of SES and could have different underlying aetiologies in obesity development.

Social mobility was determined using the change in occupational social class using the three groups (least, mid and most advantaged) from birth to age 50. The highest level of achieved education was categorised (including equivalent qualifications) as none, O-level, A-level, and degree level and above. Previously, leaving school at an early age (≤17 years) has been associated with obesity (Wardle et al., 2002). Therefore, education was recoded to a binary variable indicating achieved education past secondary school level (greater than GCSE level, the main qualification undertaken by adults in the UK), indicating undertaking of higher education beyond compulsory school age (age 16).

Household income was a choice between 17 categories on the questionnaire. Household income was equivalised to account for household size. The Organisation for Economic Cooperation and Development (OECD) equivalence scales are used by the Statistical Office of the European Union and by the UK government to adjust household income according to household composition, acknowledging that resources required are not directly proportional to household size. To determine equivalised household income, the median value of the questionnaire income categories was divided by the square root of household size and then log2 transformed for normality. The square root scale is a method utilised in recent OECD publications, and denotes a household of four people requires twice as many resources than a house composed of one person. This method was used as it does not rely on knowing the ages of the household members (data were not available).

Data on lifestyle factors were taken from the self-report questionnaire at age 50. Study members were deemed as current smokers if they were current but not ex-smokers. Pack years (the number of packs of cigarettes smoked per day by the number of years the person has smoked) and categorical (current, ex, never) were also analysed. Physical activity level was derived from questionnaire responses based on frequency of various activities (based on the Medical Research Council's Physical Activity Questionnaire (Kuh and Cooper, 1992)) and was categorised to 'inactive', 'light activity', 'moderate activity' or 'heavy activity'. Whether a study member was single, married, widowed, divorced was also taken from the questionnaire and transformed to currently married/unmarried.

2.1.1.2.1.1 Energy intake

Dietary assessment was conducted at age 50 (using the EPIC food frequency questionnaire (FFQ)) and the FFQ EPIC Tool for Analysis (FETA) was used to calculate nutrient and food group data (Mulligan et al., 2014). These data were processed prior to use (for details see (Mann, 2017)). In accordance with the FETA guide, if there were 10+ missing answers nutrient data was defined as invalid.

The energy intake data were investigated prior to use, however due to some discrepancies these data were deemed invalid and were not included in the analysis (Appendix A).

2.1.2 The Gateshead millennium study (GMS)

The Gateshead millennium study (GMS), is a prospective birth cohort that recruited 1029 infants born to mothers resident in Gateshead during pre-specified weeks from June 1999 to May 2000. Gateshead is located on the southern bank of the River Tyne opposite the city of Newcastle upon Tyne in the UK. The original aims of the study were to investigate infant feeding behaviour and growth; however, it has since expanded to encompass multiple aspects of child health, including nutrition, physical activity, and well-being.

The cohort have been followed up intermittently throughout their early life and childhood with 15 phases of data collection so far. This has involved detailed questionnaires on growth, feeding, behaviour, illness and social factors (Parkinson et al., 2011), typically completed by the mother throughout study member's early life. After birth and in the days following, data were collected in the hospital or at home from mothers, midwives, and health visitors. Questionnaires in early life were completed by parents at around ages 6 weeks, 3 months, 4 months, 8 months, 12 months, 13 months, and 30 months. Schools also facilitated data collection throughout childhood and adolescence. Detailed information on the waves of data collection can be found in the cohort profile (Parkinson et al., 2011).

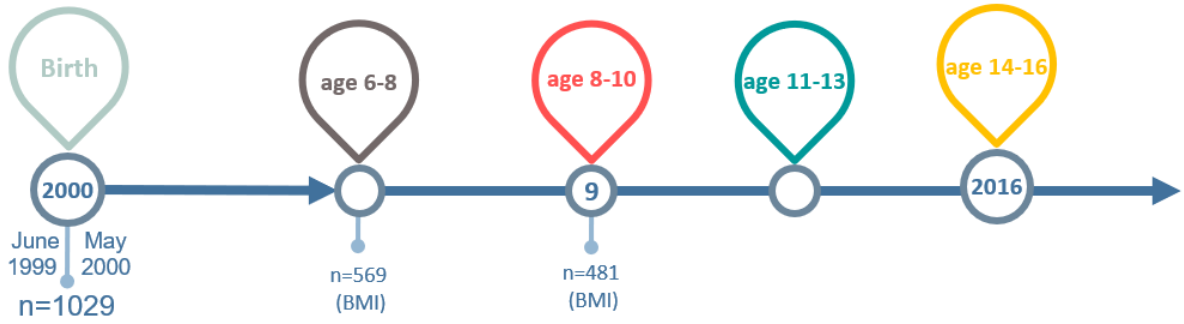


Figure 2.2 Data collection phases in GMS to date

2.1.2.1 *Measurement of outcome variables*

Measures of height and weight were taken throughout childhood. Height was determined using a Leicester portable height measure (Chasmors, London, UK), measured to the nearest 0.1 cm. Weight (kg) and bio-impedance (BIA) were measured using a Tanita TBF300MA in light clothing. Values for total body water, hydration, and lean mass were age and sex-specific (Wright et al., 2008b).

2.1.2.2 *Measurement of exposure variables and covariates*

Birthweight, gestation, number of previous children, mode of delivery, maternal age and postcode (for determination of Townsend deprivation score (Townsend et al., 1988)) were requested at recruitment. Birth order was analysed as a binary (first-born) variable. Maternal age was analysed as both a continuous measure and as categories of <25 years, 25-34 years and over 35 years. Pre-term was determined as a gestation length less than 38 weeks, normal was 38 to 41 and post-term was greater than 41 weeks. Weight was measured in a clinic at the 13-month health check.

All other variables were collected via questionnaires administered at regular intervals including adversity (4 month), sleeping (8 month) SES (birth and age 6-8) and infant feeding (recurrent questionnaires).

Parents were asked about feeding shortly after birth and in the 6-week, 3, 4, 8, and 12 month questionnaires. Parents were asked about mode of milk feeding, the cessation of breastfeeding and initiation of complimentary feeding. Duration breastfed (non-exclusive) in GMS was categorised by the study team as 'never', '<6 weeks', '>6 weeks' and '>4 months'. Breastfeeding was also categorised according to the time period of exclusivity (only breast milk), as '>4 months', '>8 months', '>12 months' or 'not exclusive (i.e. <4 months)'. Due to

the small group sizes in categories over 4 months upwards (average 2.8% prevalence) these were combined, and exclusive breastfeeding was analysed as 'not exclusive' and 'exclusive for 4+ months'. Formula fed was those who were never breastfed. Introduction of solid foods was in weeks.

Adversity in GMS was derived from the questionnaires at age 4 months that asked, 'Have you experienced any of the following in the last 12 months?' which therefore included up to 8 months prenatal exposure.

Infection data available for the first year were taken at the age 4 and 12 month questionnaires. Infections included any mention of infections, receipt of antibiotics, measles, mumps, influenza, rubella, chicken pox, whooping cough, bronchitis, gastroenteritis. These questionnaires also asked whether the child had been admitted to hospital and the reason for admittance. Responses that mentioned infection or a known infectious illness were determined to be an infection. These included terms "VIRUS" "VIRAL" "INFECTION" "RSV" "MENIN" "BRONC" "CHICKEN" "CROUP" "GASTRO" "PNEU" "MENEN" and "SEPTACEMIA". Data were then screened individually to check for any further spelling errors. Colds, influenza and ear infections were not included due to their high incidence in infants, with acute and minor effects.

The literature examining early life sleep and childhood overweight for the most part only examines sleep duration (Reilly et al., 2005a, Patel and Hu, 2008, Taveras et al., 2008), however sleep duration was not measured in GMS. Alternatively, there are various definitions of sleep problems in the literature (Table 1.4).

There were questions on sleep in the 8-month GMS questionnaire. Whilst there were no questions on sleep duration, parents were asked about disturbed sleep and issues falling asleep, which could affect sleep duration. The two questions were on a scale of 1 to 7 (ranging from always to never): 'how often did the child show fussing falling asleep'; 'how often did the child have disturbed sleep'. Parents often reported sleep issues, and frequent night-time waking and issues settling are common in infants (Armstrong et al., 1994). Therefore, to summarise the extremes and to capture those with the worst sleep problems, a dichotomous (two component) variable for sleep issues was determined as those who always (in the top 1 of 7) reported issues for both falling asleep and for disturbed sleep.

SES was assessed using an area-based indicator (Townsend score) and maternal education at birth. Townsend deprivation score was transformed into quintiles with 1 being the least advantaged and 5 the most advantaged. Maternal education was assessed at time of birth as the highest qualification attained. Deprivation was measured through ownership (both car and home ownership/mortgage) and employment (wage earner), at birth and in childhood. In childhood, parental occupation data were available at age 7-8 and 8-10. Upward mobility 0-8 was determined as the change from being deprived at birth (no ownership and no wage earner), to not deprived in childhood (ownership and wage earner).

Physical activity was measured over 7 days using Actigraph GT1M accelerometers (Actigraph LLC, Pensacola, Florida, USA) when participants were around 7 years of age. The Actigraph is an established, practical measurement of both activity (physical and sedentary) with high reliability and validity (de Vries et al., 2006, Penpraze et al., 2006). This was continuously worn on the right hip, attached with an elastic belt, but was permitted to be removed for water-based activities. Parents also completed a time log stating when the accelerometer was worn, however this was found to overestimate child physical activity (Basterfield et al., 2008). Data were processed manually prior to use as described previously (Basterfield et al., 2008, Basterfield et al., 2012a). Accelerometry measurement over 7 days is regarded as the optimum amount of time for measuring habitual physical activity (Ward et al., 2005) and shows good reliability in children (Penpraze et al., 2006). However in this cohort, 3 days of wear for over 6 hours/day was shown to produce acceptable reliability (Basterfield et al., 2011), therefore this was the criteria used to minimise loss due to missing data. Established, validated cut-off points were used to convert accelerometry data to levels of moderate–vigorous intensity physical activity (MVPA)(Puyau et al., 2002, Basterfield et al., 2008). Physical activity was analysed as the mean daily percentage of time spent in MVPA (%MVPA).

Seasonal differences in physical activity must also be considered in UK populations (Atkin et al., 2016). Generally, GMS children have been shown to be less active in the winter (Pearce et al., 2012a). Seasons were categorised as ‘spring’ (March to May, reference category, 23%), ‘summer’ (June-August, 21%), ‘autumn’ (September -November, 32%) and ‘winter’ (December-February, 23%).

2.1.3 The Avon Longitudinal Study of Parents and Children (ALSPAC)

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a birth cohort based in the former county of Avon, in the Bristol area (South West England). The original aim of ALSPAC was to further understand the ways in which physical and social environments interact, with a focus on the influence of genetic factors on health, behaviour and development of children over the life course (Alspac Study Team, 2001). The continuing aim has extended the study into a transgenerational resource for life course epidemiology, with enrolment of the original participant's children, siblings, and grandparents (Boyd et al., 2013).

The study recruited 14,541 pregnant women between 1990-92 with an expected delivery date between 1st April 1991 and 31st December 1992 (Boyd et al., 2013). There were 14,062 live births and 13,988 children alive at age 1 year. Through additional recruitment phases the sample has increased to 15,247.

Data collection has been via self-completion questionnaires and data linkage to external records. There has also been direct measurement of study members, through clinical assessments and biological samples (Boyd et al., 2013). At the clinical assessments around ages 7 and 17 years, blood samples were taken and DNA extracted. There were 68 data collection assessments between ages 4 weeks to age 18 years. The vast data collected have covered multiple dimensions of health, including social, genetic, physical, cognitive, environmental, and developmental factors. The ALSPAC data has been comprehensively coded and inspected by the ALSPAC study team. Complete details on recruitment and follow-up can be found in the cohort profile (Boyd *et al*., 2013).

2.1.3.1 *Measurement of outcome variables*

Weight and height were measured at annual clinics and at multiple time points throughout childhood (ages 9+, 10+, 11+, 13+, 15+). Anthropometric measures at (approximately) age 7 and 17 were used as outcomes in these analyses (when there were epigenetic data available). At age 7, height was measured to the nearest millimetre without shoes or socks using a Holtain stadiometer (Holtain Ltd, Crymych, Pembs, UK), whilst weight was measured using Tanita THF 300GS body fat analyser and weighing scales (Tanita UK Ltd, Yewsley, Middlesex, UK). At age 17, height was measured with a Harpenden stadiometer to the nearest mm, and weight using the Tanita Body Fat Analyser (Model TBF 401A) to the nearest 50g.

2.1.3.2 *Measurement of exposure variables*

Maternal age at delivery was originally categorised in ALSPAC as; <16 years, individual years between the ages 16-43, or >43 years. As the actual ages for those coded as <16 years and >43 years were unknown, this was categorised as done previously to "Less than 25" (young maternal age), "25-34" and "35+" (advanced maternal age) and was not analysed as a continuous variable.

Using medical records, gestational age was determined using the recorded date of the last menstrual period and date of delivery. Birthweight was taken from obstetric records. At 12 months, infants were weighed using the Seca 724 (or Seca 835 for children who could only be weighed with a parent).

Parity was determined from the 18-week questionnaire, and was defined as the number of previous pregnancies resulting in either a livebirth or a stillbirth. First-born was binary variable ('yes' if number of previous pregnancies was 0).

Occupational social class was determined as the highest category of parental social class using the 1991 British Office of Population and Census Statistics (OPCS) classification. Parental occupational social class was determined by the ALSPAC study team from questionnaire answers prior to use. Occupational social class was coded using the OPCS job codes to create six categories (I, II, III non-manual, III manual, IV and V). There was an additional category for armed forces, however there were no data on specific occupations or rank for those in the armed forces. Due to this, and also to harmonise with the occupational social class groupings in the other cohorts, armed forces were recoded to missing (total excluded, *n*=31, 0.2%).

There were no data on bacterial infection in ALSPAC; therefore, antibiotic exposure was used as a proxy. In the 6-month questionnaire, parents were asked whether the child had received antibiotics in the last 6 months (ages 0-6 months). This was coded to binary variables if the response was one or more episodes.

In ALSPAC, questions on adversity featured on both pre and post-natal questionnaires, therefore in order to harmonise with findings in GMS, three distinct adversity variables (prenatal, postnatal, and pre and post-natal) were investigated in ALSPAC. As the question on the GMS questionnaire specifically referred to child abuse of the study member, this will obviously only be relevant for the post-natal period. Therefore, in ALSPAC, any reference to

child abuse was only included in the post-natal period (8 month questionnaire), to reflect the study member's own exposure (rather than the mother's). In ALSPAC this referred to if the mother or partner were 'emotionally or physically cruel to children,' which may reflect events occurring within the household (other children), but that may not be directed towards the study member. Neither NTFS nor GMS had detailed information on the dimension(s) of the abuse. Questionnaire responses for adversity were only extracted up to age 8 months, as the next questionnaire was administered at age 21 months and would include responses out of the specified exposure period (parental to 12 months postnatal).

Maternal smoking was self-reported on the questionnaires at 18 and 32 weeks gestation and 8 weeks postnatal, and asked if cigarettes were smoked during pregnancy. Maternal smoking during pregnancy was coded as 'yes' if the mother answered yes to smoking on any of the three questionnaires. If there were any missing data at any of the time points, this was coded to missing (as it cannot be assumed that the mother did not smoke at this time point), i.e. mothers had to say they did not smoke at all time points to be counted as non-smokers during pregnancy. However, if the answer was yes to smoking at any of the time points, this was counted as a 'yes'.

2.1.4 The Accessible Resource for integrated Epigenomic Studies (ARIES)

The Accessible Resource for integrated Epigenomic Studies (ARIES) is a subset of the ALSPAC cohort. The ARIES project aims to find links between exposure, phenotype, genotype, and methylation data. Genome-wide DNAm analysis was done for a subset of 1,018 mother and child pairs at three time points in children and two in mothers, as well as profiling various tissues for references. ARIES selection was based on availability of DNA samples at the three time points (birth, childhood (age 7.5) and adolescence (age 17.1)).

In the offspring, DNAm was quantified from cord blood at birth, and peripheral blood in childhood (mean age 7.5 years) and adolescence (mean age 17.1 years). This was done using the Illumina Infinium HumanMethylation450 BeadChip. Pre-processing, quality control and estimation of cell type proportions were carried out by the ALSPAC team prior to researcher use (Relton et al., 2015).

The sub-sample was mostly representative of the main study population; however, ARIES mothers were slightly older, less likely to have a manual occupation and were less likely to smoke during pregnancy (Relton et al., 2015).

2.2 Definition and measurement of outcomes, exposures and covariates

2.2.1 Outcomes: Body composition measurements

Previous studies have found differing results when assessing BMI (continuously) to categorical obesity with regards to maternal exposures and offspring adiposity (Sharp et al., 2015b). Therefore, it is valuable to examine health risks associated with a higher BMI using both BMI on a continuous scale and indicators for overweight and obesity. As BMI is a proxy, alternative measures were analysed in order to determine if risk factors vary by outcome measure. To encompass each aspect of defining obesity this study will use BMI, BMI categories, a measure of central obesity (WhtR or WHR), and a measure of body fat (%BF or FMI). Comparing all three measures could uncover more about risk factors-specific mechanisms. These methods are discussed in detail in sections 1.1.4 and 1.1.5.

Outcomes were BMI, BF% and WHR in adults, and BMIz, WHtR and FMI in children. Body mass index (BMI) was calculated as weight (kg) divided by height (m) squared. In children, BMI is confounded by age-related physiological variation, and so needs to be assessed with respect to a growth reference. The UK90 reference was used as it is the most appropriate reference, recommended for use in British children (Wright et al., 2002). The UK90 growth reference was determined from a sample of 32,222 measurements taken from 12 surveys between 1978 and 1994. It was sampled from a UK population with ages ranging from 0-23 years, and is used for population monitoring and published figures using Health Survey for England and National Child Measurement Programme data. Childhood BMI was transformed to z-scores (adjusted for age and sex) using the UK90 growth reference with the Zanthro program in STATA (Vidmar et al., 2004). The program calculates standard deviation z-scores using Cole's LMS method. L, M, and S represent the skew, mean, and coefficient of variation of the measurement as it changes with the x variable (age)(Cole, 1990).

In adults, BMI was classified into weight categories according to the World Health Organisation guidelines (World Health Organization, 2000) (*Table 1.1*). In children, weight categories were grouped as 'normal weight' or 'overweight/obese' using BMIz. In line with previous analyses in this cohort (Parkinson et al., 2017), clinical cut-offs were used to determine weight categories, whereby healthy weight is between the 2nd and 91st centiles, overweight greater than 91st centile and obese greater than the 98th centile (SACN, 2012).

Clinical cut-offs are recommended by NICE in a clinical setting for individual children (Dinsdale et al., 2011).

Measures of waist obesity and body fat were also considered as important adiposity outcomes in NTFS and GMS (section 1.1.4). In adults, waist-hip ratio (WHR) was calculated as waist circumference divided by hip circumference, measured by research nurses at the clinical examination. A WHR >0.9 in men and >0.85 in women corresponds to increased risk of metabolic complications (World Health Organization, 2011c). Waist-to-height ratio (WHtR) was used in children and is calculated as the waist circumference (cm) divided by height (m). This study utilised the acknowledged cut-off of WHtR >0.5 to signify waist obesity (Ashwell et al., 2012). Body fat was determined by bioelectrical impedance (BIA) in both cohorts. In adults, BF% was the proportion of fat mass relative to total body weight. In children, fat mass index (FMI), a relative measure of body fat, was calculated as fat mass (kg) divided by height (m) squared.

2.2.2 Definition of early life

Firstly, the definition of early life must be addressed. There are many definitions of early life with emphasis on the first 1000 days, which covers conception to around 24 months. The first 1000 days concept appears to have originated from the importance of ensuring adequate nutrition in early life, most frequently in the developing world. The idea appears to have been taken on by DOHaD researchers, with the first 1,000 days of an infant's life having a great impact on the child's growth and development, including cognitive development and immune function. However, this arbitrary figure relates to nutrition-related factors for cognitive development, and therefore this definition may be less important for non-nutritional factors.

A medical definition would determine infancy as the first year of life. For the purpose of this study, and due to the loose definitions of critical and sensitive periods, exposures will focus on conception to the first year of life. However, most of the data collected in the cohorts was postnatal.

2.2.3 Defining early life exposures

This study hypothesised that a number of early life factors would be associated with subsequent OWOB, based on the literature review presented in Chapter 1. Table 2.2

summarises the definition of the early life factors and covariates, which were harmonised across cohorts for consistency. Exposures were selected based on the availability of data from the NTFS and GMS.

Birthweight was analysed as both a continuous (z-score) and categorical variable. Originally birthweight was recoded into low (<2.5kg), normal (2.5kg - 4.5kg) and high birth (>4.5kg) weight categories. However, due to the low frequency of low birthweight and high birthweight (in NTFS, Table 3.5), these categories were not analysed further. Instead, birthweight was converted to weight-for-gestational age z-scores (method outlined in section 2.1.1.1) and was used to determine categories of small (SGA) and less than the 10th or greater than the 90th percentile respectively.

Similarly, weight at 12 months was transformed into a z-score (see section 2.2.1 for details). RWG was analysed as a dichotomised variable, defined as greater than +0.67 change in weight for age z-score from birth to 12 months. This change in is equivalent to crossing one major growth centile band on a standard child growth chart between the two time points(Ong and Loos, 2006b).

In addition, in the cohort comparison (chapter 5) due to the large difference in birthweight z-scores between the cohorts, weight gain conditional on birthweight was also considered. Conditional weight gain, or thrive index (TI), accounts for normal catch-up growth from low birthweight, as a linear measure of weight gain adjusted for regression to the mean (Wright et al., 1994). TI was calculated as weight for age z-score at around 12 months, minus the birthweight z-score multiplied by the regression coefficient (r, between the two weight measures) for the cohort (i.e. TI birth to 12 months = $\text{z-score}_{12m} - r \times \text{z-score}_{birth}$). Similar to RWG, rapid thrive (RT) was defined as TI >0.67.

Maternal age was categorised into 3 groups; less than 25, 25-34 and over 35. Advanced maternal age, defined as over 35 years, is associated with adverse pregnancy outcomes such as LBW, still birth and labour complications, preterm delivery and chromosomal defects (Jacobsson et al., 2004, Cleary-Goldman et al., 2005).

The categories for duration breastfed determined in GMS were ‘never’, ‘<6 weeks’, ‘6 weeks – 4 months’ and ‘>4 months’. Equivalent categories formed using NTFS data to match when comparing the cohorts.

Adverse events in the first year of life were based around potentially stressful exposures of social, monetary and care issues. These events included parental separation, police involvement, child abuse, debt and illness or death of parent or sibling. Adversity was a dichotomous variable defined as experiencing any of the following; parental separation, police involvement, abuse, debt or death of a family member, in the first year of life.

To increase group sizes, occupational social class was recoded to most advantaged (I, II), mid (III) and least advantaged (IV, V) (Table 2.1). Household deprivation coding was cohort specific and are explained in sections 2.1.1.2 (NTFS) and 2.1.2.2 (GMS).

Table 2.1 Occupational social class recoding

Occupational social class groupings	**Occupational social class recoded**
I Professional occupations **II Managerial and technical occupations**	Most advantaged
III(N) Skilled occupations: non-manual **III(M) Skilled occupations: manual**	Mid advantaged
IV Partly skilled occupations **V Unskilled occupations**	Least advantaged

Table 2.2 Definitions of the early life exposures across the cohorts

Exposure	**Class**	**Definition**	**NTFS**	**GMS**	**ALSPAC**
Maternal age	Continuous	individual years	Harmonised		Analysed only as a categorical variable
	Categorical	<25, 25-34, 35+ years	Harmonised		
Birthweight (Bwt)	Continuous	Bwt (kg) was normalised to the UK90 growth standard and transformed into z-scores, standardised for sex and gestational age	Harmonised		
SGA, LGA	Categorical	<10th percentile z-score small for gestational age (SGA), and >90th percentile for large for gestational age (LGA) infants	Harmonised		

Caesarean	Categorical	Mode of delivery was caesarean	Not measured	Vaginal or caesarean birth	Not analysed in this cohort
First-born	Categorical	Birth order was analysed as a binary (first-born) variable	Data not available^	Harmonised	
SES (birth)	Categorical	Varied by cohort	Father's occupational social class (SOC90)	Townsend score (quartiles)	Highest category of parental social class (OPCS) classification
Household deprivation	Categorical	Specific to the time-period	No housing issues, or evidence of housing issues (overcrowding, lack of hot water, shared toilet, dampness or poor repair)	Deprived (no ownership) or not deprived indicated by presence of ownership (both car and home ownership/mortgage) and employment (wage earner)	Not analysed in this cohort*
Rapid weight gain (RWG)	Categorical	>0.67 SD change (birth to ~12 months) in weight for age z-score	Harmonised		
Rapid thrive (RT)	Categorical	TI >+0.67 SD	Harmonised		
Breastfeeding	Continuous and categorical	The duration of any kind of breast feeding	Harmonised (in the cohort comparison)		Not analysed in this cohort*
Exclusive breastfeeding	Continuous and categorical	The duration of only breastmilk	Harmonised		Not analysed in this cohort*
Weaning	Continuous and categorical	The age when solid foods were introduced into the diet	Harmonised		Not analysed in this cohort*
Adversity	Categorical	Experiencing any of the following; parental separation, police involvement, abuse, debt, death or illness in the family	0-12 months	From 8 months pre-natal to 4 months post-natal	Three separate variables of: Prenatal, postnatal, and both pre and postnatal

*Not analysed in this cohort – only exposures that were important exposures in NTFS and GMS were analysed further in ALSPAC

^ Electronic data were not available

2.2.4 Definition of a confounder

A confounder is a variable associated with the exposure of interest and the outcome, which when unaccounted for can lead to biased regression estimates. Kleinbaum *et al*,. define a

confounder as that whose presence or absence from the full model changes the coefficient of the primary explanatory variable by 10% or more (Kleinbaum et al., 1982). This definition was kept in mind when adjusting models for potential confounders, and as confounders were incorporated into the models their effects on the coefficients of other variables was noted to assess their impact. Key confounders were however identified from the literature and are listed in Table 1.5.

2.3 Methods for the epidemiological analysis

2.3.1 General analytical strategy

A systematic approach was taken to determine the early life exposures associated with subsequent body composition in each of the cohorts. The primary focus of this work was on the North East cohorts; NTFS and GMS, and the workflow for the statistical analysis of these cohorts is presented in Figure 2.3. Additional analyses are detailed in section 2.3.7.

In order to compare results across cohorts the data were harmonised as coherently as possible (see Table 2.2). Twins/non-singleton births (in the original cohorts: GMS, *n*=36; NTFS, *n*=28; ALSPAC, *n*=264) were excluded to satisfy assumptions of independence within the regression models, and due to differential intrauterine environments (affecting foetal growth) compared to singleton pregnancies. All childhood weight-related variables were transformed to z-scores using the same growth reference, and underweight participants (z-score <-2 or BMI<18.5kg/m^2) were excluded, as underweight shows inverse associations with all-cause mortality (Prospective Studies Collaboration, 2009), and therefore should not be combined with normal weight (the reference group).

The ALSPAC cohort was utilised primarily for the epigenetic data. However, prior to the epigenetic analysis, the associations between exposures and body composition were examined to determine the strength and magnitude of the associations. The analysis in the ALSPAC cohort was a streamlined analysis utilising only BMIz and OWOB, in order to justify the focal exposures for the epigenetic analysis. Although some exposures have been analysed with regards to childhood body composition in ALSPAC (Reilly et al., 2005a) some were not comparable in their definitions or adjusted for the same confounders, therefore it was necessary to run the models for comparability.

2.3.2 Descriptive characteristics, sample representativeness and sex differences

Summary statistics are provided for each exposure and outcome variable for all cohorts. All data were checked for the presence of extreme/implausible values, and any invalid measures were checked for coding errors (cross-referenced with original data where possible), recoded or removed.

Early life exposures were analysed longitudinally with respect to body composition in later life. Attrition is common in longitudinal cohorts and therefore body composition measures were not available for the entire original sample at the later time point(s).

Missing data can reduce the statistical power, can lead to biased estimates and can reduce the representativeness of the samples. There are three types of missing data: missing completely at random (MCAR), missing at random (MAR), and missing not at random (Kang, 2013). Ideally a study will be well-planned in order to be representative of a population at baseline and with systematic data collection at subsequent follow ups to minimise data loss. However, attrition is inevitable in long-term studies. The most common approach to dealing with missing data is to exclude those with data missing and performing a complete case analysis. It is advised that researchers seek to understand the reasons for the missing data (Kang, 2013). In large sample sizes and if the assumption of MCAR is satisfied, then list wise deletion can produce unbiased estimates. However, if the sample size is small and MCAR is not satisfied then results may be biased. A method of dealing with missing data is imputation, which uses the other variables available in the dataset to estimate the missing variable, which has the benefit of retaining the initial cohort sample size. However, imputation assumes MAR and many imputation procedures assume data are normally distributed, therefore when these conditions are not met this could lead to incorrect coding of variables. In this study, as many variables were not normally distributed and were either binary or categorical variables, a complete case analyses was done.

The impact of attrition and whether there were differences between those with and without body composition measures (the sub-sample) was assessed. This was done using t-tests, Mann-Whitney tests or chi-square tests for parametric, non-parametric or categorical exposures respectively. The reasons for the missing data was partly due to retention of the more advantaged groups, and therefore this was addressed using weighting in the cohort

comparison (section 2.3.7.3). Sex differences (if anticipated), were also assessed using these methods in the sub-samples (those with body composition measures).

2.3.3 Correlations between exposures and body composition measurements

To determine whether BMI is an adequate proxy measurement, the strength of correlations between weight outcome measures were assessed with Pearson correlation. Correlations between exposure variables were also assessed. The correlation coefficient (r), can range from +1 to -1, with r>0 indicating a positive association, and r<0 indicating a negative association. Results are presented as scatter graphs with Pearson correlation coefficients (r).

2.3.4 Socioeconomic differences in infant feeding

In the case of infant feeding, which is often socially patterned in the UK, SES is an important confounder. Therefore, differences in infant feeding by socioeconomic groups (birth) were assessed. In NTFS, the data were continuous and therefore this was done using either analysis of variance test (parametric data), or a Kruskal-Wallis equality-of-populations rank test (non-parametric). In GMS, the breastfeeding data were categorical; therefore, a Chi^2 test was used.

2.3.5 Examining the associations between early life factors and subsequent body composition

Linear regression modes were used to examine associations between exposures and continuous outcome variables (i.e. BMI/BMIz). Linear regression, in its simplest form (bivariate), can be used to establish the strength of the relationship between the dependent variable (y, outcome) and independent variable (x, exposure), which can be viewed as the equation:

$$y = \alpha + \beta x + \varepsilon$$

Where α is the constant or y-axis intercept, and ε is the residual model error. β is the coefficient of x or the slope of the regression line, and x is the independent variable.

Linear regression uses the principle of least squares, whereby it aims to minimise the sum of all squared deviations of the observed data points from the best fitting regression line, and is sometimes referred to as ordinary least squares (OLS) regression.

An extension of bivariate regression is multiple regression, which includes multiple independent variables. In this model, the β coefficient is adjusted for the effect of any additional covariates included in the model.

The assumptions of linear regression, and how these were addressed are in Box 2.1 and Table 2.3 respectively.

Although linearity is assumed, performing a simple linear regression using a binary independent variable is not invalid, it is equivalent to performing a two-sample t-test (although addressing slightly different objectives) however with the capacity to adjust for covariates.

Variable selection is an important step when estimating a linear model that explains the data in the simplest way. When there are many covariates, stepwise regression is helpful method for identifying key predictors. The goal is to produce a parsimonious and accurate model as it excludes variables that do not explain the variation in the outcome, and aims to retain independent variables that best predict the outcome. This approach is useful for explanatory model building; however, it may exclude important explanatory variables and known confounders, particularly in very large datasets where other methods may be more appropriate (see (Smith, 2018)). Another approach could be to leave all variables in the model (full model), however this may lead to many independent variables. This may introduce issues with collinearity or insufficient sample sizes to accurately estimate a model with many degrees of freedom. Alternatively, existing knowledge could guide variable selection, which would thereby only include the key exposures. A disadvantage to this approach is that relies on previous knowledge of the relationships, which might not apply in different populations.

Therefore, a combined approach was taken here, utilising stepwise regression to identify the key exposures explaining variation in the outcome, but then models were adjusted for key confounders using a theory-driven approach (*Table 1.5*).

Key assumptions of linear regression

- A linear relationship between the exposure and outcome
- Homogeneity of residuals (homoscedasticity, residuals should have constant variance)
- Independence of errors (residuals should not be correlated with y or x)

Other important considerations

- The residuals are normally distributed (necessary for hypothesis tests to be valid rather than estimation of coefficients)
- Any potential issues with outliers or influential data points (high leverage points, a measure of those which deviate from the mean)
- There should be little collinearity between predictor variables
- Model specification, the model should be appropriately specified with relevant variables and no important missing (omitted) variables
- Predictor variables should be accurately measured

Box 2.1 Key assumptions and considerations of linear regression

Table 2.3 Assumptions of linear regression and how these were investigated in STATA

Issue/assumption	How the assumption was checked
Linear relationship	Using a scatter plot of the x and y variable
Homoscedasticity	Using a scatter plot of the standardised residuals against the predicted values
Independence of errors	This should not be an issue utilising cross-section data with independent study members (twins excluded)
Normality of residuals	Can be examined using a kernel density plot of the residuals with the normal density line overlaid, a standardised normal probability (p-p) plot, or quantiles of a variable quantiles plot (q-q)
Influential data	Examined using plots that shows the leverage by the residual squared
Multi-collinearity	Examined the variance inflation factor (variance inflation factors below 10 accepted)
Model specification	The STATA linktest command performs a model specification test based on the Goodness of link test (Pregibon, 1980)
	The regression specification error test (RESET) for omitted variables implemented in STATA ovtest (Ramsey, 1969)

Firstly, to determine the relationship between exposures and outcomes of interest, separate (bivariate) models were examined for each exposure and outcome. Next, adjusted regression models were constructed. Stepwise forward regression models were constructed using STATA stepwise modelling, with a p value of <0.1 for inclusion, informed by bivariate analyses. Covariates were added manually to models, with model fit informed by the Bayesian information criterion (BIC), whereby the model with the lowest BIC is preferred (Schwarz, 1978). The BIC is a model selection criterion. Adding more variables to a model can result in overfitting; the BIC applies a penalty accounting for the number of parameters to identify the best model.

Logistic regression was performed to examine the relationships between each exposure and binary outcome variables (i.e. obesity, or OWOB). Coefficients or odds ratios (OR) with corresponding 95% confidence intervals and level of significant (p) are presented. Good

model fit for the multivariable logistic models was determined by Hosmer and Lemeshow's test, no evidence of collinearity and no observations that deviate in an influential manner.

In addition, to determine the impact of SES on the relationship between early life factors and body composition, bivariate models were adjusted for SES. Models were sequentially adjusted for SES at birth, SES at age 9, and at both time points, to examine their relative effects on regression coefficients. Similarly, the impact of lifestyle factors (at time of body composition measurement) was examined in GMS children by adjusting regression models for physical activity, and in NTFS adults by adjusting the path model (outlined in section 2.3.6) for lifestyle factors (smoking and physical activity).

2.3.6 Examining the pathways between early life factors and BMI

When examining the relationship between X (exposure) and Y (outcome), all multivariate approaches (with more than two variables) involve the questions of moderation and mediation. Moderation asks if a third variable interacts with X, whereas mediation asks whether it intervenes on the X-Y relationship. Approaches involve statistical methods dealing with correlation and partial correlation. For example, these include: multiple regression, general linear models, linear mixed models, path analysis, or structural equation modelling for continuous data (Garson, 2017).

Path analysis, which is a form of structural equation modelling, is one way to examine the relationships between variables, and as a means of determining which assumptions best fit the data at hand (Garson, 2017). Additional information is acquired using path analysis than through addition of variables and product terms (interactions) to regression models (Garson, 2017).

Alternatively, in the econometrics literature, instrumental variable methods (Sobel, 2008), such as two-stage least-square (2SLS) regression models are often used (Cameron and Trivedi, 2010). In short, in a regression of x on y, when there is endogeneity, x may be correlated with the error term of y which thereby incorporates the effect of unmeasured variables (i.e. confounders) (Angrist and Imbens, 1995). The 2SLS models address this by using an instrumental variable that predicts X but is uncorrelated with the y error term, thereby resolving the issue of endogeneity (Antonakis et al., 2014). A key benefit of this method is that it doesn't require all confounders to be measured and incorporated.

However, this approach has several assumptions and requires good instrumental variables, which makes it difficult to apply appropriately (Podsakoff et al., 2011).

If causal inference is the aim, there are numerous other methods that can be employed (Antonakis et al., 2014). However, if the aim is to model the relationships between variables and depict the relationships graphically, then path analysis is well-equipped. Therefore, a path analysis approach was used to investigate the life course impact of early life risk factors (0-1 year), SES and lifestyle on subsequent BMI. Path models were used to uncover the associations between variables, to disentangle the relative influence of each risk factor, and the indirect pathways to BMI.

The initial baseline path model included the variables that demonstrated significant associations in the multivariable linear regression model. Confounders and exposures not included in the adjusted model were sequentially included, the model estimated and assessed for model fit. The model was grown adding paths between variables with a priori hypotheses (*Table 1.5*, hypothesised relationships presented in the DAG, Appendix A) and modification indices (suggested paths to improve model fit, as a measure of change in the likelihood ratio chi-square), then non-significant paths were removed until a good model fit was achieved.

Good model fit was determined to be a non-significant χ^2, root mean square error of approximation (RMSEA)<0.05 and non-significant PCLOSE, and also comparative fit index (CFI) and goodness-of-fit index (GFI) both >0.95. All direct paths with p<0.05 were modelled and standardised β coefficients are presented. Confidence intervals were determined using bootstrapping (50,000 iterations for two rounds of thinning). Standardised β coefficients are presented, which represent partial regression coefficients between connected variables, controlling for all prior variables (Garson, 2008).

The indirect effect is the product of each component path, the direct effects are straightforward relationships (not going through any other variable) and the total effects are the sum of direct and indirect effects.

A p value <0.05 was used to denote significance throughout. All statistical analyses were done in STATA 14 and updated to version 15 in 2017 (StataCorp, College Station, TX) and path diagrams were constructed using SPSS Amos (SPSS Inc, Chicago, IL).

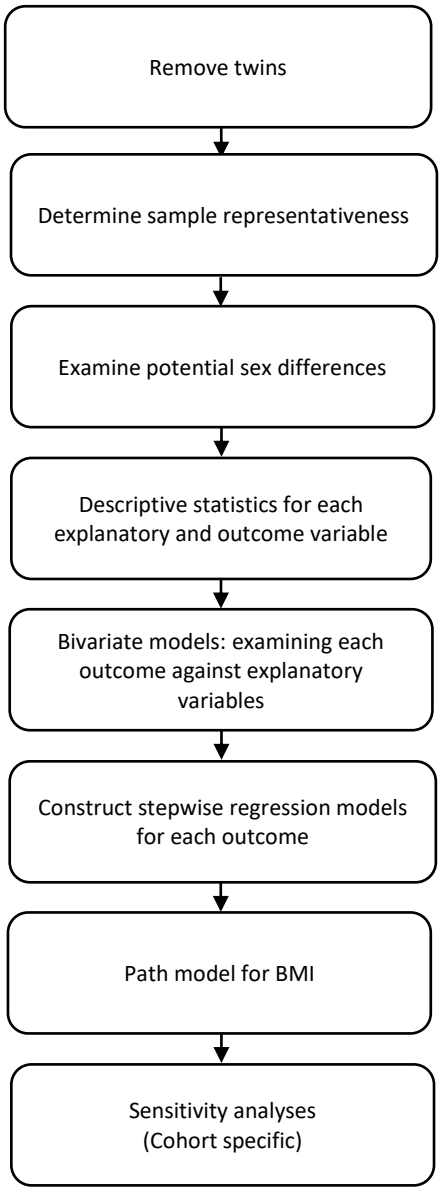


Figure 2.3 Workflow of the statistical analysis in the NTFS and GMS cohorts

2.3.7 Additional chapter-specific sensitivity analyses

2.3.7.1 *Sensitivity analysis by ethnicity (chapter 4)*

Whilst it is acknowledged that OWOB disproportionately affect minority groups, there was no evidence that early life risk factors varied by ethnicity in the Born in Bradford cohort (Fairley et al., 2015a), a cohort with a larger proportion of ethnic minorities. Due to the small proportion of non-Caucasian children, the sample was not stratified by ethnicity. However, results were checked for robustness by religion due to a notable proportion of ultra-

orthodox Jewish (UOJ) families in GMS. Sensitivity analysis was carried out to assess whether differences in religion may affect estimates in GMS children, which could represent socioeconomic differences or differences in feeding practices (as determined previously in the GMS cohort (Wright et al., 2010b)).

2.3.7.2 *Robust regression (chapter 4)*

There were some outliers in FMI model, therefore a robust regression model was utilised. In comparison to OLS, robust regression produces a model which is less affected by outliers or influential observations (Rousseeuw and Leroy, 2005). It uses iteratively reweighted least squares whereby each point is assigned a weight and then coefficients are estimated using OLS. With further iterations, weights are reassigned, with points further from the model predictions having smaller weights. The coefficients are then recomputed using the weights. Estimation continues until convergence is achieved.

The workflow for the statistical analysis is specified in section 2.3.5 (Figure 2.3).

2.3.7.3 *Inverse probability weighting (chapter 5)*

Additional analyses were carried out in chapter 5 (cohort comparison), including testing for interactions within models and inverse probability weighting (IPW). In the regression models, interactions between explanatory variables and SES, were tested within the model using likelihood ratio tests.

As cohorts were not representative of the original sample, inverse probability weighting (IPW) was carried out as sensitivity analysis, using variables that were significantly different between the cohorts (Appendix D, Table X). IPW applies weighting to the sample to account for the imbalanced representation of exposures. However repeating analyses with weighting minimally altered results, therefore unweighted results are presented.

2.3.7.4 *Investigation of early life growth and BMI trajectories (chapter 5)*

Utilising the longitudinal data available, rapid thrive was investigated further with regards to BMIz over the life course. Height and weight data were available at various points in childhood and adolescence for GMS (ages 0, 1, 3, 6-8, 8-10, 14-16) and NTFS (ages 0, 1, 9, 13), which were transformed into z-scores as described previous (section 2.2.1). BMIz trajectories were plotted by to examine the average growth patterns based on early life growth (rapid thrive).

2.4 Methods for the epigenetic analysis

2.4.1 DNA methylation arrays and considerations

The Illumina Infinium® HumanMethylation450K BeadChip assay (450K array) is a popular, cost-efficient method for large-scale profiling of DNAm. The array determines genome-wide methylation status of over 485,000 CpG sites. Essentially, it utilises bisulfite treated DNA to determine the relative proportions of methylated and unmethylated fragments (Pidsley et al., 2013).

There are multiple platforms for measuring CpG methylation across the genome, with perhaps the most popular being the 450K array (>450,000 CpG sites), or the newer EPIC platform (>850,000 CpG sites). There is also the capacity to build a custom array, however this limits comparability between studies.

The 450K array covers most known genes, with many probes focused in promoter regions, however the array also covers non-CpG island loci including in the; gene body, 3' untranslated region (UTR) and intergenic sequences. Whilst the array has good coverage of known genes and previously identified methylated sites, it spans fewer than 2% of CpG sites across the genome.

Prior to the DNAm analysis, DNA is bisulfite modified, which converts unmethylated cytosine residues to uracil (and subsequently replaced by thymine), whilst methylated cytosines are resistant to bisulfite conversion. The array determines methylation levels by quantifying the proportion of cytosine and thymine bases. Estimated cytosine methylation levels are expressed as beta values (β), which range from 0 (0% or no cytosine methylation) to 1 (100% or complete cytosine methylation), at each CpG site.

$$\beta = \frac{\text{Methylated signal}}{\text{Methylated signal + unmethylated signal +100}}$$

Figure 2.4 The calculation of beta values in methylation arrays.
Beta values are calculated as the proportion of the signal due to methylated signal over the total signal. To ensure the denominator is not zero, a constant (100) is added. Beta values range from 0-1, with higher values indicating higher methylation and vice versa.

With this feature in mind, it is important to remember that, as a relative proportion across all cells, a 50% methylation level could represent half of all cells fully (100%) methylated and half of all cells completely unmethylated (0%), or alternatively that 50% the of alleles are methylated (allele-specific methylation), or various combinations of these two factors. Bisulfite sequencing techniques can provide more information about methylation patterns (Song et al., 2013), with the caveats being that the region sequenced is relatively small, and any allele-specific driven methylation patterns can be extrapolated only from the region sequenced (Fang et al., 2012, Kuleshov et al., 2014). Therefore, if genotype data are not available, the influence of other genomic loci nearby or elsewhere cannot be discerned (for example the influence of single nucleotide polymorphism (SNPs)).

The 450K array uses two types of probes that quantify methylation in different ways. Type I use a similar design to that exploited in the older 27K beadchip array, which use a single colour but with two different probes to capture methylated and un-methylated. Whilst type II probes use one probe but two different colours for methylated (red) and un-methylated (green). The design of the array creates some issues in analysis, for example; the distribution of methylation values differ by probe; type II probes show greater variability and are often less reproducible (Dedeurwaerder et al., 2011). Furthermore, some probes exhibit cross-reactivity or contain polymorphisms (Chen et al., 2013). Difference in colour channel performance (related to chip lot and scanner) can affect calculation of beta values. Normalisation methods can help to combat these issues.

At present, the 450K array is the most widely used platform, however with the advent of the larger EPIC array it is likely that studies in the future will begin to transition to the newer platform.

The key limitations of the 450K array (cell types, SNPs, batch effects) can be quelled using appropriate statistical techniques, as described further in the following sub-sections.

2.4.1.1 Adjusting for cell type heterogeneity

Determining the proportion of cell types is an important factor in epigenetic analysis. Ideally, DNAm would be measured in the tissue relating to the phenotype of interest, however obtaining such tissues is often not feasible and considered invasive in human population studies. DNAm is most frequently measured in blood, as it is an easily accessible tissue,

therefore acting as a surrogate medium. Whole blood is composed of erythrocytes (red blood cells), leukocytes (white blood cells) and platelets (Figure 2.5). Leukocytes comprise of granulocytes (neutrophils, eosinophils, basophils) and agranulocytes (monocytes, lymphocytes), whilst subtypes of lymphocytes include B cells, T cells and natural killer (NK) cells. These cells are all nucleated and therefore contain genomic DNA.

However, this presents a challenge when analysing because DNAm is cell-type specific (Illingworth et al., 2008, Mill and Heijmans, 2013), and therefore variations in cell proportions will give different methylation profiles. Factors influencing DNAm marks and cell-type proportions can be related to; age, phenotype or disease (Mill and Heijmans, 2013). Therefore, cell composition, being both associated with the methylation level and in some cases also with the exposure/outcome of interest, can confound the relationship between CpG methylation and the exposure/outcome.

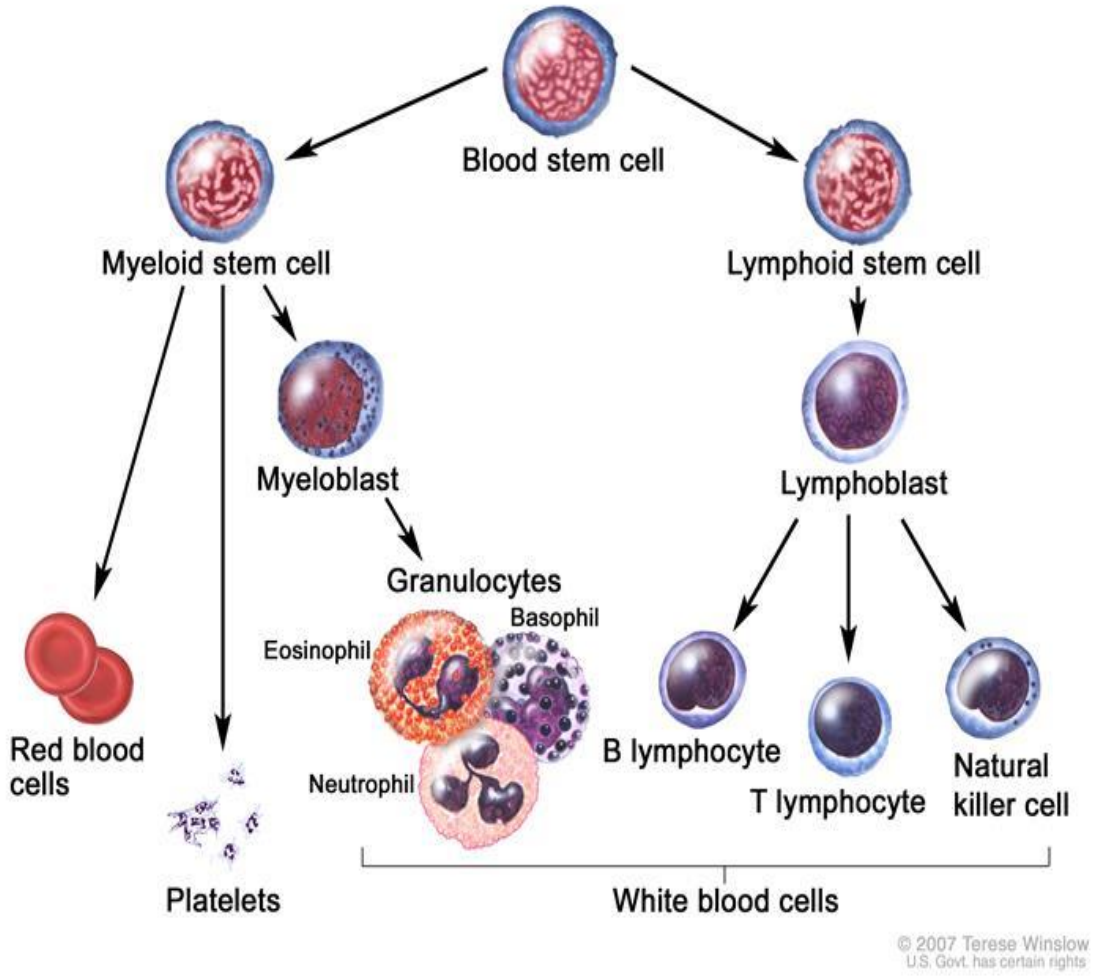


Image from NIH National Cancer Institute
Image: NCI Dictionary of cancer terms, "Blood cell development", National Cancer Institute, Available at https://www.cancer.gov/publications/dictionaries/cancer-terms/def/white-blood-cell. Accessed January 2019.

Figure 2.5 Lineage of blood cell development.
Stem cells differentiate to form different white blood cells, platelets or red blood cells.

Discrepancies arising as a result of not adjusting for cell type composition were first noted by Liu *et al*,. in an epigenome-wide association study (EWAS) comparing rheumatoid arthritis cases and controls (Liu et al., 2013a). They found that there were many false positive associations due to the ratio of granulocytes to lymphocytes, but after correction these associations were no longer significant. Re-analysis of previous studies found that much of the observed variability in DNAm could be explained by cell composition (Jaffe and Irizarry, 2014a), highlighting the importance of adjusting for cell composition when using whole blood (Adalsteinsson et al., 2012). In an ideal situation, individuals would have their blood cellular subtypes sorted and profiled, to determine the proportion of each cell type.

However, this often is not often carried out, therefore computation methods and algorithms have been developed for estimation.

Cell proportions were calculated in ALSPAC using the Houseman reference-based algorithm (Houseman *et al*., 2012). The Houseman algorithm is a popular method to estimate cell proportions and has been independently validated (Houseman et al., 2012, Accomando et al., 2014). Other reference-based methods exist (Newman et al., 2015), however the Houseman reference-based method is often preferred (Kaushal et al., 2017). Alternatively, methods not utilising cell references use unsupervised deconvolution methods instead, such as surrogate variable analysis, which shows robust sensitivity and specificity (Leek and Storey, 2007, Houseman et al., 2016).

Phenotypic variation in cell-type composition could confound analyses. On the other hand, it could also represent an important physiological change in response to an exposure or disease, which may be related to the phenotype of interest. When searching for biomarkers (related to an exposure) that are associated with a disease outcome, to regress out variation from cell counts could potential disregard important loci. Obesity is an acknowledged chronic, inflammatory condition. It has been associated with inflammatory indicators including C-reactive protein (Visser et al., 1999a) and white blood cell counts (Bastard et al., 2006, Farhangi et al., 2013). A degree of inflammation is a component of the obesity phenotype, therefore to find novel biomarkers associated with this phenotype, biomarkers were investigated in models with and without adjustment for cell composition and findings were compared.

2.4.1.2 **Methods of adjusting for confounding**

There are many potential sources of confounding in DNAm studies. Microarrays are run in separate batches due to scale, therefore major sources of variation are row, slide or chip effects, processing date or operator (Leek et al., 2010). Thoughtful experimental design and randomisation can help mitigate these effects. Direct adjustment will only account for the influence of known confounding variables (such as batch, age, or sex). The epigenome is complex and influenced by the environment in a manner that is still not fully understood, and even in the best-designed studies, there will be many unmeasured factors, which could have an impact on DNAm and lead to spurious findings.

Methods that account for unmeasured confounders are useful in longitudinal studies on human populations where it would be impossible to; measure or to know all potential confounding factors, or if they were known, to attempt to integrate them all into a statistical model. Many methods have been proposed for dealing with the heterogeneity, either by removing non-specific variation (not just from cell types) from unmeasured confounders (such as those utilising surrogate variables) (Gagnon-Bartsch and Speed, 2012), or methods that were designed specifically for methylation data, such as those developed by Houseman *et al,* (Houseman et al., 2012, Houseman et al., 2014).

The aim of surrogate variable analysis (SVA) is to remove the unwanted variation, whilst retaining differences due to primary variable of interest; thereby identifying consistent differences between groups and removing latent variation. These methods are also capable of dealing with variation from batch, slide or chip , which cannot always be adequately corrected for (Teschendorff et al., 2009).

SVA finds sources of variation from the methylation data itself, and models these as singular vectors (surrogate variables) derived from singular value decomposition, which will be linearly uncorrelated. The surrogate variables are then included as covariates in the regression model (Leek and Storey, 2007). Independent SVA (ISVA) is a modified version of SVA whereby the surrogate variables are deemed independent. In support of ISVA, known confounding factors such as age and batch are obviously statistically independent variables and are linearly uncorrelated, and therefore it would be appropriate to model these as independent variables. ISVA was shown to perform best at capturing a known specific biological signature when compared to other adjustment methods (Teschendorff et al., 2011). However that may not hold true for all datasets (Teschendorff et al., 2011), and the best method will be dependent on the tissue type under investigation, and whether the analysis is reference-based or reference-free (Teschendorff and Zheng, 2017). A thorough study compared each of the common methods (Houseman's reference-based method, RefFreeEWAS, SVA, ISVA, EWASher and RUV) using extensive simulations. There was no method that performed perfectly for all parameters measured, however the authors concluded that SVA was the most robust (and 'safest') method (McGregor et al., 2016). In summary, there is support for both SVA and ISVA as high-performing adjustment methods, therefore both were utilised in these analyses.

2.4.1.3 **Exclusion criteria**

The 450K array relies on the specific hybridisation of genomic DNA to probes on the chip, and therefore alterations in the DNA such as SNPs, repetitive sequences or insertions/deletions can interfere with accurate measurement of DNAm (Chen et al., 2013). Known SNPs can be cross-referenced and removed prior to or post-analysis.

It is also recommended to exclude identified non-specific probes to reduce the risk of false discovery, as these may not hybridise specifically. SNPs and non-specific probes (Chen et al., 2013) were removed in epigenetic analyses.

2.4.2 Measurement of DNA methylation in ARIES

DNAm was quantified for the ARIES sub-sample using the 450K array (Illumina Inc., CA) according to the standard protocol. Cord and peripheral blood samples were collected using standard procedures and DNA was extracted. DNA was bisulfite-converted using the Zymo EZ DNA Methylation kit (Zymo, Irvine, CA) prior to hybridising DNA to the BeadChip.

Sample handling, measurement of DNAm and quality control were carried out by the ALSPAC team at the University of Bristol. A semi-random approach was used to distribute samples across slides to ensure each time point was well represented, and to minimise potential confounding by batch effects. Samples failing quality control thresholds (average probe P-value ≥0.01) were repeated.

ALSPAC methylation data were pre-processed with background correction and subset quantile normalisation in R (version 3.0.1), using the Touleimat and Tost pipeline (Touleimat and Tost, 2012, Relton et al., 2015). The data are processed prior to researcher use, primarily applying the following adjustments. Slide effects were regressed out on the raw betas before normalisation. Functional normalisation implemented in the R package meffil was used to normalise the data, which is a between-array normalization method and an extension to quantile normalisation (Fortin et al., 2014). This removes unwanted technical variation by regressing out the variation explained by control probes. Normalisation was applied separately by gender for the sex chromosomes, and there was also a background correction and a dye-bias correction applied. Many ARIES samples have been successfully genotyped, and any samples that failed genotype quality control due to sample swaps, gender mismatches, or relatedness issues between mothers and offspring were excluded. Gender mismatches (n=411) for individuals were assessed by comparing genotype probes to SNP-chip

data, or by assessing sex-chromosome methylation. Sample type, whether cells were from white cells from buffy coat or peripheral blood lymphocytes from whole blood, did not explain much variation in the child or adolescent samples. Any duplicates were removed (age 7, *n*=10; age 17, *n*=10) with retention of the sample with the highest number of detected probes.

The proportions of CD8 T cells, CD4 T cells, NK cells, B cells, monocytes and granulocytes in the samples were estimated using *the estimate CellCounts function* (using the Houseman *et al*. method (Houseman et al., 2012)), in the minfi Bioconductor package in R by the ALSPAC team. Cell counts were corrected for in the linear adjusted models (see *Figure 2.6*). Sensitivity analysis was run with models without correction for cell counts to see if correction influenced results.

DNAm levels are presented as 'beta' values (β-value), which represents the proportion of cells methylated at each individual locus, ranging from 0-1 (0-100% methylated).

2.4.3 Statistical analysis

There were three different analyses undertaken to examine the relationship between DNAm data (outcome) and early life exposures, including analysis of differentially methylation positions, differentially methylated regions, and differentially methylated positions in a subset of candidate loci.

The significantly differentially methylated loci were analysed further with respect to phenotype; to determine if the loci were related to subsequent obesity; and via annotation of the gene region to highlight functional characteristics.

2.4.3.1 Differentially methylation positions

An epigenome wide association study (EWAS) examines genome-wide epigenetic marks (DNA methylation) in order to determine differential sources of variation for a given exposure or phenotype. For example, this could be to distinguish between cases or controls, or between exposed and unexposed. It is also possible to conduct an EWAS on a continuous variable, thereby assessing the linear relationship between the exposure/phenotype and epigenetic marks. Using methylation data from the ARIES subset of the ALSPAC cohort, methylation at individual CpG sites was investigated for each exposure in individual EWAS's. All EWAS and bioinformatics analyses were done in Rstudio version 3.3.2.

For the statistical analysis using the ALSPAC and ARIES data, the associations between exposures of interest and DNAm at 7 and 17 years, at more than 450,000 CpG sites were examined (for varying numbers of participants depending on exposure data availability (see Table 6.9)).

The EWAS utilised a linear model, with DNAm as the outcome and the exposure of interest as the independent variable, as the baseline model. Age and cell type composition are important considerations in epigenetic analysis (although many of the DNAm changes that have been associated with age are actually due to age-related changes in cell composition) (Jaffe and Irizarry, 2014b). Therefore, models were adjusted for age, sex and cell proportions. Four models were run in total; unadjusted, adjusted for covariates, surrogate variable adjusted and independent surrogate variable adjusted (Figure 2.6). Cell counts included; B cells, CD4T cells, CD8 T cells, granulocytes, monocytes and NK cells. Analyses for each variable were done for complete cases. As sensitivity analysis, all models were also run without cell counts included.

Model 1 (no adjustment)

$$\text{DNAm} = \text{exposure}$$

Model 2 (adjusted for covariates)

$$\text{DNAm} = \text{exposure} + \text{age} + \text{sex} + \text{cell counts}$$

Model 3 (SVA)

$$\text{DNAm} = \text{exposure} + \text{age} + \text{sex} + \text{cell counts} + \text{surrogate variables}$$

Model 4 (ISVA)

$$\text{DNAm} = \text{exposure} + \text{age} + \text{sex} + \text{cell counts} + \text{independent surrogate variables}$$

Figure 2.6 EWAS adjustment models run using the Meffil R package

All analyses were done in R, with the package Meffil (Min et al., 2017). The package employs SVA or ISVA methods which have been shown to successfully account for unmodelled or unknown confounding factors (such as batch) (Leek and Storey, 2007, Teschendorff et al., 2011). Meffil simultaneously computes unadjusted, adjusted, SVA and ISVA models, thereby allowing results to be compared (Min et al., 2017) (Figure 2.6). Whilst all models are computed by the software, only SVA and ISVA models were considered, as the 'none' and 'all' models will not capture the residual variation associated with technical (batch) effects.

In order to minimise the influence of outliers in methylation data, beta values were winsorised at the level of 5% (95th percentile cut-off). Winsorising adjusts extreme values so that they are transformed to match the next closest value.

When computing multiple tests, there is the increased potential to find a significant result through chance, which would lead to a false positive. Correction for multiple testing was applied using a false discovery rate (FDR) threshold of $p<0.05$. The Benjamini-Hochberg method (FDR) determines the threshold for the expected proportion of false positives (type I errors) (Benjamini and Hochberg, 1995). In contrast to the more conservative Bonferroni correction (specified p value (i.e. p=0.05) divided by number of comparisons), which treats all p values as equal, the FDR assesses P values based on ranking. Bonferroni correction can lead to false negatives and runs the risk of discarding significant observations, whilst FDR adjustment aims to have the smallest number of false signals appearing as significant, but with potentially more type I errors. In summary, the Bonferroni cannot control for type II and the FDR cannot for type I, therefore, both the FDR and Bonferroni p values were assessed.

Results of EWAS models were analysed using Q-Q plots. These show how well the specified distribution fit the results, and if the quantiles of theoretical and actual distributions agree, the points will lie on the line y=x. Departures from linearity indicate issues with the data. The Q-Q plots were visually inspected to determine the best fitting model. If the same loci appeared in different models, and the Q-Q plots for that variable were equivalent, then results from the SVA models were prioritised as it is more robust (McGregor et al., 2016).

2.4.3.2 Differentially methylated regions

Instead of looking for changes at individual CpG loci, another means of identifying phenotypic differences in methylation is to analyse regions. DMRs, which are stretches of CpG loci, may have more of a functional effect on gene expression than individual CpG loci (Jones and Baylin, 2002). Additionally, if changes in DNAm are small but persistent across a region, there is more statistical power to detect them collectively as DMRs (Robinson et al., 2014).

DMRcate is an R package which uses Kernel smoothing (non-parametric estimation of the underlying curve/structure) for the estimation of DMRs, and allows adjustment for covariates (Peters et al., 2015). A study using simulated data found that DMRcate outperformed the other methods studied in terms of precision, but with a slightly lower sensitivity for change in betas (Martorell-Marugan et al., 2018). Its predictive performance was better than Bumphunter and Probe Lasso (two common methods used in DMR analysis),

and similar to that of comb-p, which often performs best but is implemented in python programming language (Peters et al., 2015). As the intention was to implement analyses using R in order to integrate the analysis with other Bioconductor tools, DMRs were analysed using the Bioconductor R package DMRcate (Peters et al., 2016).

DMRcate utilises the limma R package to apply Bayesian linear model methods (Peters et al., 2016). Similar to the models that were ran analysing single CpG loci, DNAm was the outcome and the exposure was the independent variable. However, when using DMRcate it is preferable to use M values (log logit transformed Beta values, M=log(beta/1−beta)), from a statistical standpoint to deal with any homoscedasticity (Du et al., 2010). Firstly, the t-statistic for the linear model is computed, and then kernel smoothing is applied with a Gaussian kernel bandwidth for smoothed-function estimation (lambda), scaled by a scaling factor C (for bandwidth). DMRcate applies correction for multiple testing using the Benjamini-Hochberg method to determine the significant CpGs. DMRs are computed using the specified lambda for CpGs within that distance from one another, and the p value for the DMR is calculated using Stouffer's method (Stouffer et al., 1949). For optimal prediction of DMRs using 450K array data, the recommended settings are (lambda=1000 and C=2); as the Gaussian kernel is calculated as lambda/C = sigma, this translates to one standard deviation of Gaussian kernel equal to 500 base pairs. Power in DMRcate increases when lambda is smaller or C is larger (Odom et al., 2018).

DMRcate was executed using the recommended settings on the M-values, for each exposure and time point, and the models were run with and without adjustment for cell counts. Whilst ISVA and SVA have many benefits, unlike other methods such as ComBat (which adjusts for batch effects using an empirical Bayes framework (Johnson et al., 2007)), they do not directly adjust methylation data, meaning there are no adjusted data for downstream analyses. Therefore, in order to keep analysis in line with the previous models, the surrogate variables (that were calculated using meffil in the EWAS models) were included as covariates in the DMR models. The surrogate variables from the SVA model were used as this is deemed the 'safest' model (McGregor et al., 2016). Probes ≤2 nucleotides distance to a SNP with a minor allele frequency > 0.05, or probes in the list of cross-reactive probes, were filtered out (Chen et al., 2013, Pidsley et al., 2016). In this array, these steps filtered out approximately 15,518 of 453,723 (~3.3%) loci. In short, the analysis involved running model 3 in Figure 2.6, both with cell counts and without.

2.4.3.3 BMI-associated candidate loci approach

The previous methods have focused on determining new loci primarily concentrated on the exposures. In order to narrow down loci to those suspected to be related to adiposity, a candidate gene approach was taken using CpG loci found to be associated with BMI. Therefore, instead of the 485,000 CpG loci (on the 450K), this BMI-associated, smaller subset of CpG loci were instead analysed as outcomes (n=187). This aims to increase the likelihood of finding a biomarker that is associated with adiposity, as well as the exposure. The candidate loci were selected from a robust, large-scale EWAS that utilised data from across multiple cohorts (Wahl et al., 2016). Epigenetic loci were investigated with respect to BMI in cohorts of European and Indian-Asian descent (Wahl et al., 2016), which were then validated in other population-based studies, and some loci via other methods (e.g. pyrosequencing, *n*=4). After validation, 187 genetic loci were identified as associated with BMI in adults (Appendix E).

These loci were chosen as this was a powerful study that included multiple analytical components to produce robust findings. The results were: replicated in separate samples, robust to method of analysis, independent of cell heterogeneity and correlated with multiple tissue levels, and demonstrated some clinical significance. The study had a large sample size (EWAS *n*=5,387, replication *n*=4,874) with a diverse population. DNAm was primarily investigated in blood; as this is useful for clinical purposes, but there were also moderate to strong correlations between tissues (i.e. blood and adipose tissue). The candidate loci also mapped to genes with roles such as in lipid metabolism, inflammation, and metabolic, cardiovascular, respiratory and neoplastic diseases, suggesting functional roles in disease.

The early life exposures studied here have been associated with subsequent changes in BMI. Therefore, it is plausible to hypothesise that if DNAm is a mediating mechanism between exposure and outcome, these early life exposures may influence methylation of those loci which are associated with BMI.

Using the 187 loci as candidate loci, EWAS were run for individual exposures, at both time points using the Meffil R program (Min et al., 2017). This involved running the same models outlined in Figure 2.6, but in this case DNAm refers to the 187 individual loci.

2.4.3.4 Phenotypic differences in methylation

Methylation at specific, significant loci was examined graphically at the two time points (age 7 and 17) by phenotype. Within the same individual, the change in methylation over time (from age 7 to 17), was also examined with respect to the exposure using the Student's t-test.

To assess whether methylation varied by body composition, differentially methylated loci were also investigated with respect to both the exposure and outcome of interest (OWOB). A statistical method to test for differences between two or more groups is using an analysis of variance (ANOVA) test. A one-way ANOVA is appropriate when there is one independent variable. The first assumption of the ANOVA is that residuals are normally distributed. This can be determined by inspecting the residual plots. Homogeneity of variance is another important assumption of the one-way ANOVA. The Bartlett's test for equal variances can provide information as to whether this assumption holds true (the null hypothesis is that all groups are equal), however this test is based on the assumption that the samples are normally distributed. When there is violation of normality, a more robust test for homogeneity of variances is Levene's test, which is less influenced by departures from normality. Similar to Bartlett's test the null hypothesis is that the variance of all groups is equal. Another assumption of the ANOVA is that groups are independent (i.e. study members will belong to one group only).

The null hypothesis for an ANOVA is sometimes referred to as the omnibus null hypothesis, as when it is rejected, the ANOVA test statistical indicates that group means are not equal, i.e. there is a difference between groups, however it does not indicate which group is different. Post-hoc analysis, can be used to identify where the differences lie, and can also take into account correction for multiple comparisons (e.g. Bonferroni or Benjamini-Hochberg), and adjusting the accepted p value threshold. A caveat being, that when the estimates are conservative, adjusting for multiple testing increases the likelihood of type II errors (false negatives).

For data that are not normally distributed, there is a non-parametric version of this test (Kruskal-Wallis, KW). Similarly, the KW test assumes that the groups have the same distributions, but it does not make the same strict assumptions as the ANOVA and is instead

rank-based. However, in substituting data to ranks, information is lost, which makes it a less powerful test than a one-way ANOVA.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

Figure 2.7 The ANOVA null hypothesis

Visual inspection (box plots) of the data aided interpretation, and histograms of residuals were used to determine if the data fit the normality assumption. Bartlett's test (Snedecor and Cochran, 1983) was used to test if groups had equal variances, or Levene's test in the case of non-normality. If data fit the assumptions of normality and variance, a one-way ANOVA was used with Bonferroni adjustment for multiple testing, or in the case of non-normality, the KW test was used. For the KW test, Dunn's multiple comparisons test using rank sums (Dunn, 1964) was applied using the user-written STATA program (Dinno, 2015). The package applies correction for multiple comparisons, and results were Bonferroni adjusted by multiplying the p-values in each pairwise test by the total number of pairwise tests. These analyses were done in STATA version 15.1 (StataCorp, College Station, Texas, USA) using standard commands aside from those specified.

2.4.3.5 Annotation of significant CpG loci

The human reference genome (GRCh37/hg19 assembly) accessible on the UCSC Genome Browser was used for conceptualisation of the gene region (Kent et al., 2002). The genome browser is accessible online and allows interactive visualisation of sequence data and gene regions from many species, including all vertebrates and some invertebrates. The human GRCh37/hg19 assembly was used, which is compatible and consistent with bioinformatics tools for 450k analysis. Significant CpG loci were mapped to the genomic location (including nearest gene and associated CpG island) using the Illumina ilmn12.hg19 annotation (Hansen, 2015). To examine if there was consistency in methylation within the CpG island, correlations between significant CpG loci (with the other CpG loci in the island) were investigated, including separately by exposure.

If the CpG loci identified from the EWAS were located in proximity to a gene or within an island, any other CpG sites on the 450K which mapped to the gene and/or island were also examined (individually as outcomes) with respect to the exposure using linear regression

(adjusted for age, sex and white blood cells). Similarly, significant CpG loci were investigated at the alternative time point (i.e. DNAm in adolescence was also examined at the specified loci if in childhood DNAm changes were significant).

2.4.3.6 Examining associations for rapid thrive

Due to the similarities between RT and RWG, and in an attempt to distinguish effects which may be partly due to birthweight, loci that were significant for RWG were analysed with respect to RT. This was done using a linear model, adjusted for age, sex, and the SVAs (from the RWG EWAS model), with separate models with or without cell counts.

2.5 Methods for Lab analysis

Loci of interest identified from EWAS's in the ALSPAC cohort at age 7 (chapter 5) were investigated further in NTFS adults. There were no pre-existing DNA methylation array data for the NTFS samples, which is costly to obtain. Therefore, isolation and amplification of the loci of interest was achieved using DNA primers. Firstly, the surrounding gene regions were identified and checked for suitability. To differentiate the methylated positions, the samples underwent bisulfite modification, and the genetic material was amplified. Finally, the samples were sequenced to determine the proportion of methylated residues, and data were analysed statistically to examine the relationships between DNAm and the phenotype.

2.5.1 Identification of epigenetic loci

There were significant (P_{FDR}<0.05) associations (in adjusted models) identified for RWG at two CpG loci. These loci were cg01379158 (NT5M) and cg11531579 (no associated gene) and were both associated with a 1% increase in methylation (p=0.02) in those who had RWG. Furthermore, at both CpG loci the highest methylation was in those with RWG and had OWOB. On these criteria, these loci were deemed good candidates for further investigation in line with selection criteria (Box 2.2).

Gene loci selection criteria

- A significant change in methylation at a CpG site associated with an early life exposure (False Discovery Rate (FDR) adjusted p≤ 0.05)
- Significant changes in the adjusted (surrogate variable analysis or independent surrogate variable analysis adjusted) models (p<0.05)

Box 2.2 Gene loci selection criteria

2.5.2 *In silico* bisulfite conversion and primer design

Firstly, the region of interest was identified using the UCSC genome browser (Hinrichs et al., 2018). The CpG site was entered into the search box and the chromosomal location confirmed. A region of 1000 base pairs (bp), i.e. 500 bp both upstream and downstream of the target locus was used for primer design. The region also encompassed neighbouring CpG sites that do not feature on the 450K array.

The region was inspected for SNPs and repeat elements, which could interfere with gene amplification and sequencing. This was done using the drop-down menus to display repeat elements and to display common SNPs (SNPs with >= 1% minor allele frequency (MAF) that map only once to the assembly), flagged SNPs (SNPs < 1% MAF, or unknown meaning they could be >1%) and multiple SNPs (SNPs that map to >1 place on the assembly).

The sequence of the 1000bp region was annotated further in Microsoft Word, this included denoting CpG sites, SNPs and primer placement (Box 2.3 and Figure 2.8). The primer sequence is based on the bisulfite-converted sequence. Methylated cytosines (methylated CpG sites) are resistant to bisulfite modification (overviewed in section 2.5.4). Therefore, after bisulfite conversion, all non-methylated cytosines are converted to thymine (represents the non-CpG sites and non-methylated CpG sites), and the only cytosine residues that remain are methylated cytosines (Figure 2.8). Therefore in order to design primers, the sequence must be transformed to the bisulfite sequence (replacing all potentially non-methylated cytosine (C) with Thymine (T)) (i.e. the cytosine residues not adjacent to Guanine), using the method outlined in Box 2.3.

1. The genomic sequence was extracted from the UCSC
2. Using find and replace (CTRL+H), all spaces were removed by finding '^p' and leaving replace clear
3. Next, all CpG sites were found using the criteria: find 'CG', replace with 'XG'
4. Next all non-CpG site C residues were replaced with T using: Find 'C', replace with 'T'
5. Finally, the Cytosines adjacent to Guanine were converted back to C's using: Find 'X', replace with 'C'

Box 2.3 in silico bisulfite conversion

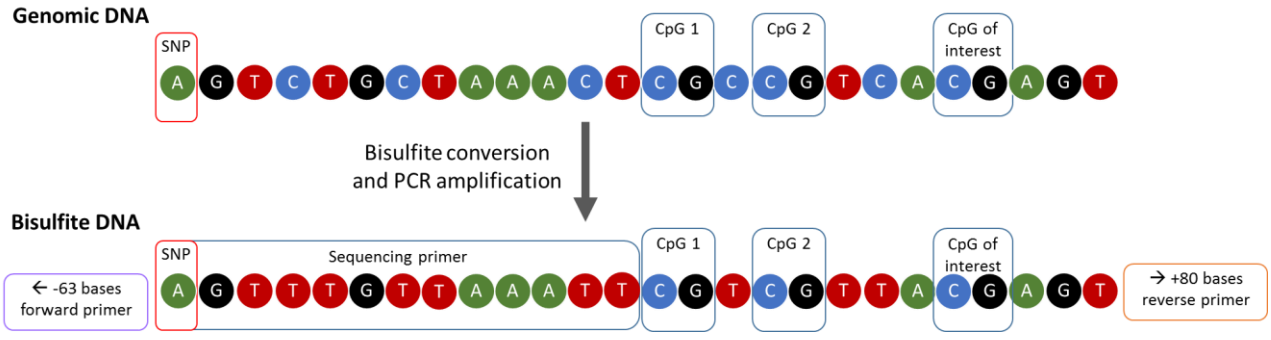


Figure 2.8 Bisulfite conversion and primer placement around the region of interest.
After bisulfite treatment of the genomic DNA strand and amplification, all non-methylated cytosines have been converted to thymine. The sequencing primer is a few bases upstream of the CpG of interest. The forward primer is 63 bases upstream of the sequencing primer, and the reverse primer is 80 bases after the last base. The region contains 3 CpG sites and a potential SNP.

Primers were designed using Methprimer (Li and Dahiya, 2002), an online tool for designing

bisulfite-conversion-based Methylation primers. Primers for bisulfite sequencing should be non-specific regarding methylation status, i.e. the primers will bind whether the CpG is methylated or unmethylated (Li and Dahiya, 2002). Therefore, when designing primers, it is important to avoid CpG sites.

The original genomic DNA sequence was entered as the input sequence, with the target specified as 500 base pairs (bp) along and 2 bp (CG) in length (e.g. 500,2) where the target CpG was located. In the initial primer design (Table 1), all parameters were set to default, including an optimum product size of 200 bases, a primer melting temperature (T_m) of 55^{o}C (range 50-60 ^{o}C) and primer size of 25bp (range 20-30bp). An ideal product size is 150-200bp, as larger fragments can be more difficult to amplify due to DNA fragmentation during bisulfite modification (Patterson et al., 2011). Similar annealing temperatures for the forward and reverse primer were also considered when choosing primer sets to minimise amplification bias (Shen et al., 2007). If no primers could be identified, then the 'optimum criteria' were 'relaxed' (Table 2.4).

Primer sets from the Methprimer output were reviewed in turn with reference to the annotated bisulfite sequence. If the forward or reverse primer set contained any SNPs or repeat elements, then another primer set was selected and inspected. If no primers were identified that did not contain SNPs, then the SNP allele frequency was inspected.

Table 2.4 The optimum and relaxed criteria utilised for primer design

	Optimum criteria	**Relaxed criteria**
Product size (bp)	200	100-300
Melting temperature (^{o}C)	55	50-70
Primer length (bp)	25	16-25
SNPs	Does not contain SNPS	Can contain low frequency SNPs

Oligo nucleotides for the primers were obtained from IDT (Integrated DNA Technologies (IDT), IA, USA). The reverse primers were labelled with biotin, which allows for sepharose beads to bind when preparing the samples for pyrosequencing.

2.5.3 Sample storage and quality

DNA was extracted from peripheral blood samples previously by Pearce et al. (2012c) using the Nucleon BACC2 kit (Tepnel Life Sciences, UK). DNA samples were stored at -80 °C in individual wells of four 96-well plates, in either Tris-EDTA (TE) buffer or water, with an adhesive plate seal. The sample quality, condition, freeze-thaw cycles, and previous treatment prior to retrieving samples was unknown. Therefore, the quality of all samples was assessed using a NanoDrop 2000 (Thermo Fisher Scientific, MA, USA) prior to use. As the samples had been stored for a long time, the seals had become brittle and did not appear airtight, meaning that for some samples there was minimal material left. In an attempt to salvage any remaining DNA, 10µL of warmed (37°C) nuclease free water was added to wells, the solution was agitated using pipetting, and samples were left for ~1 hour before removal. Although biological material can degrade over time, DNA is fairly robust.

2.5.4 Bisulfite modification of genomic DNA

First demonstrated by Frommer et al. (1992) and CIark et al. (1994), bisulfite modification allows selective conversion of non-methylated cytosines in genomic DNA, for analysis of DNAm. It is considered the gold standard method for DNAm studies requiring resolution of single nucleotides (Clark et al., 2006). Bisulfite treatment deaminates (removal of the amine group) unmethylated cytosines converting to uracil in single stranded DNA, whilst 5-methylcytosine (5-mC) remains unchanged (as the methyl group prevents conversion). Therefore, following gene amplification, the uracil residues will be amplified as thymine, and the 5-mC as cytosine, allowing differentiation between methylated and unmethylated bases upon sequencing (Clark et al., 2006).

Genomic DNA was bisulfite converted using the EZ DNA Methylation Gold™ kit (Zymo Research, Cambridge Bioscience, UK) according to the manufacturers protocol. Using this kit, bisulfite conversion is quick, as DNA is denatured and bisulfite converted in a single step in individual reaction tubes. Using spin columns, the samples are desulphonated and cleaned up ready for downstream use. The kit states a conversion efficiency (of non-methylated C residues are converted to U) of >99%, and DNA recovery of 75% (Manufacturers handbook, Zymo EZ DNA Methylation Gold™ kit, Zymo Research). Generally, this kit performs well in terms of conversion (Holmes et al., 2014), and is one of the preferred options when high DNA recovery is required (Kint et al., 2018). Samples that were below the threshold of 95%

for bisulfite conversion efficiency (determined using the PyroMark software, see) were repeated.

The manufacturer's protocol recommends converting DNA in the range of 500 pg - 2µg. The samples ranged in concentration from 0-100 ng/µL. Therefore, all samples were diluted with water to contain either 50ng (or less in the low DNA samples, for which all material was used) or 100ng of starting genomic DNA (gDNA) in a volume of 20µL.

To these samples, 130µL of prepared CT conversion reagent (containing sodium metabisulfite) was added and tubes placed in a Bio-Rad thermocycler S1000 (Hercules, CA, USA) using the following conditions: 98°C for 10 minutes, 64°C for 2.5 hours and a 4°C hold. To bind the sample to the column, the samples were transferred into columns with the addition of 600µL of M-Binding Buffer and centrifuged at full speed, and the flow through was discarded. Next, samples were washed with 100µL of M-Wash Buffer, and desulphonated using 200µL of Desulphonation Buffer for 20 minutes at room temperature, and then spun. The samples were then washed twice using 200 µL of M-Wash Buffer. Finally, the samples were eluted into a clean microcentrifuge tube using 12 µL of M-Elution buffer.

2.5.5 PCR optimisation

Polymerase chain reaction (PCR) is a molecular biology method used amplify stretches of DNA. PCR optimisation is required to determine the most appropriate annealing temperature (T_a) for the primer set that gives the best product. An optimised reaction was deemed as that which produced a clear, single band when visualised on the gel, with no secondary products (potential primer dimers). Optimisation was carried out using samples that were 50ng or 100ng (prior to bisulfite conversion) to ensure that the PCR would work at low DNA concentrations. PCR master mix was prepared in a PCR hood that had been treated with ultra-violet light (minimum 20 minutes) prior to use to minimise the risk of contamination. PCR master mix was prepared containing GoTaq® Hot Start Green Master Mix (Promega), forward and reverse primers, and nuclease free water in the volumes specified in Table 2.5. The volume of bisulfite DNA (bsDNA) was dependent on the sample concentration: for samples containing 50ng of gDNA 2µL of bsDNA was used, whilst 1µLof bsDNA was used for 100ng reactions (plus 9 and 8µL of water respectively). All assay validation steps used HEK293T (Human embryonic kidney) cell line DNA.

Table 2.5 PCR master mix composition, volume and reagent supplier

Reagent	Amount (µL)	Supplier
GoTaq® Hot Start Green Master Mix	12	Promega (WI, USA)
Forward primer (10pmol/µL) =10µM	1	IDT (IA, USA)
Reverse primer (10pmol/µL) 10µM	1	IDT (IA, USA)
Nuclease free water	8 or 9	Qiagen (Hilden, Germany)
bsDNA (100ng, 50ng)	1 or 2	Cell line DNA or NTFS DNA
Total	23	

A no template control was included in all PCR assays, which substituted bsDNA for water, to detect the presence of any DNA contamination. Methylated controls were also included (0% and 100% methylated), ensuring primers could bind irrespective of methylation level. Sources of methylated and unmethylated bisulfite converted DNA were commercial (unmethylated EpiTect control genomic DNA, Qiagen, Hilden, Germany).

All PCR reactions were carried out using the Bio-Rad thermocycler S1000 (Hercules, CA, USA). The estimated annealing temperature (Ta) was determined using the most respected theoretical primer Ta calculation by Rychlik et al. (1990) (Box 2.4). The average Ta (of the forward and reverse primers) was used as a mid-point to base the temperature range for gradient PCR.

A temperature gradient PCR was carried out for assays using human cell-line bisulfite converted DNA (HEK 293T cell line) in order to determine the optimal T_a. For gradient PCR, the 8 rows of the thermocycler each have a different temperature setting for the annealing stage. The thermocycler determines the temperature settings based on the range inputted. The PCR reactions were then run with the reaction conditions outlined in Box 2.5.

- Ta=0.3*Tm PRIMER + 0.7*Tm PRODUCT - 14.9
- Forward = 0.3*49.5 + 0.7*68.4 -14.9 = 14.85+32.98= 47.83
- Forward(edit)=0.3*49.4 + 0.7*68.4 -14.9 = 14.82+32.98= 47.80
- Reverse=0.3*52.3 + 0.7*68.4 -14.9 = 15.69+32.98= 48.67

Box 2.4 An example annealing temperature (Ta) calculation for the forward and reverse primers for cg11531579
Calculations are according to the empirical formula by Rychlik et al,. (Rychlik et al., 1990)
Ta, annealing temperature; Tm, melting temperature.

1. 1 cycle of 95 °C for 10 minutes
2. 50 cycles of:

   95 °C for 40 seconds,

   Ta °C for 40 seconds,

   72 °C for 40 seconds;

3. 1 cycle of 72 °C for 5 mins
4. 4°C ∞

Box 2.5 PCR reaction conditions
Ta, annealing temperature.

2.5.6 Gel electrophoresis

PCR products were visualised on agarose gels. During the validation stage, this gave an indication of the amount of DNA (strength of the band), the PCR product size, and whether there was any contamination or secondary products.

Agarose gels contained 1.5% agarose powder dissolved in 1X Sodium boric acid (SB) buffer, microwaved until dissolved. GelRed (Cambridge BioSciences, Cambridge, UK), a fluorescent nucleic acid dye, was added (2µL per 100ml gel) for staining, which fluoresces when exposed to ultra-violet light allowing visualisation. Fragments less than 100bp were likely a result of primer dimers.

For the gels, 2µL of PCR product from each reaction was run alongside a 100bp DNA ladder (New England Biolabs, Ipswich, MA, USA), as a guide to fragment size. Gels were run at 90 Volts for 30 minutes and were visualised using the Odyssey Fc viewer (Li-cor Biosciences Ltd, Lincoln, NE, USA).

2.5.7 Pyrosequencing as a targeted approach for quantifying DNA methylation

Measuring methylation: Cytosines can be either methylated or unmethylated and are therefore binary states. Methylation reflects an average across a whole sample.

Pyrosequencing is a high-throughput CpG methylation analysis platform, which detects level of methylation at multiple, individual CpG sites. It is an accurate, reproducible method and considered the "gold standard" in DNAm analysis (Kurdyukov and Bullock, 2016).

It involves the real-time, sequence-based detection and quantification of DNAm. Using sequencing by synthesis, the sequential incorporation of nucleotides complementary to the template DNA leads to detection of nucleotide in the form of a light signal (Tost and Gut, 2007)(Figure 2.9).

Prior to pyrosequencing, template bisulfite DNA must be amplified using a biotin-labelled primer. After denaturation the biotin-labelled single strand is isolated and hybridised to the pyrosequencing primer.

The hybridised PCR product is incubated with the required enzymes (DNA polymerase, adenosine triphosphate (ATP) sulfurylase, luciferase, and apyrase) and substrates (adenosine 5' phosphosulfate (APS) and luciferin). One of the four deoxyribonucleotides (dATP, dCTP, dGTP and dTTP) is added to the reaction according to the dispensation order (determined by the pyrosequencing computer software). When complementary nucleotides are introduced, catalysed by DNA polymerase, pyrophosphate is released. The pyrophosphate along with APS, is enzymatically converted by ATP sulfurase to ATP. When

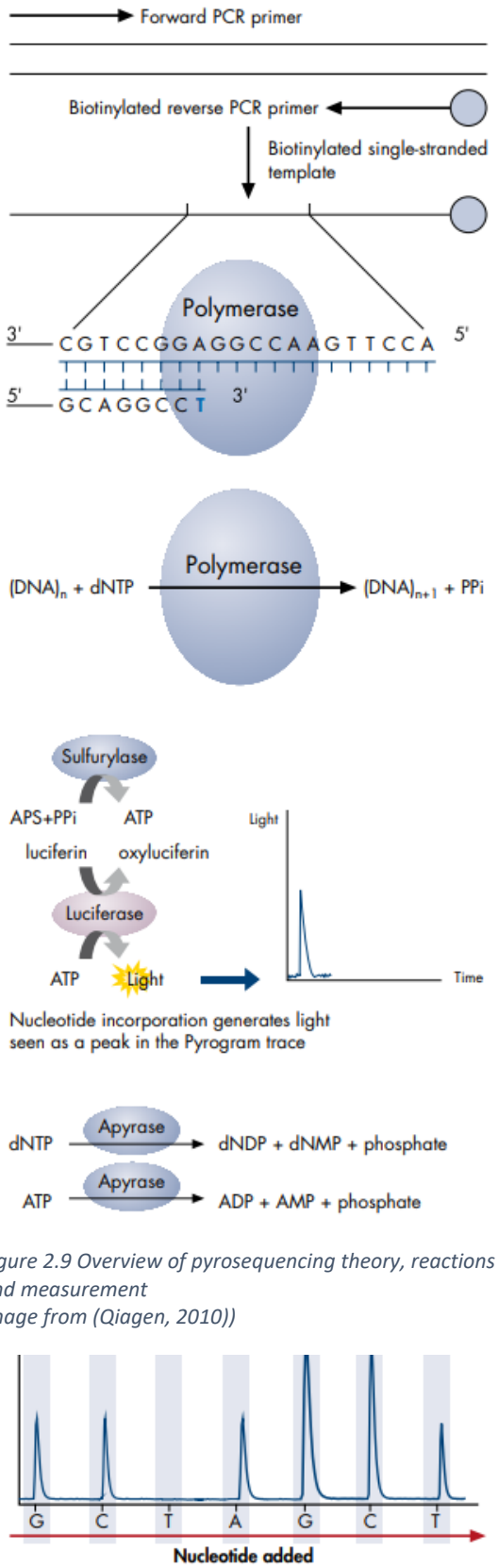


Figure 2.9 Overview of pyrosequencing theory, reactions and measurement
Image from (Qiagen, 2010))

ATP is present, Luciferase converts luciferin to oxyluciferin, which releases visible light in proportion to the number of nucleotides incorporated. This light signal is detected by sensors in the pyrosequencer, and is emitted as a peak, the height of which is proportional to the number of nucleotides in the raw data output (pyrogram). Once unincorporated nucleotides have been degraded by Apyrase, the next nucleotide is added and continues until the DNA strand is elongated. The methylation level is quantified as the proportion of C to T as indicated by the peaks and presented as percentage methylation (Figure 2.10).

2.5.8 Pyrosequencing assay design

Pyrosequencing assays require an additional 'sequencing primer'. This primer ideally would be ~10-15bp and start just before the CpG of interest, or a few bases upstream (<15bp). The sequence of the sequencing primer needs to avoid CpG sites to ensure binding whether the sequence is methylated or unmethylated. The sequencing becomes less reliable further away from the sequencing primer, hence it is beneficial to have the sequencing primer immediately before the CpG of interest. The sequencing entry (entered for the pyrosequencing assay) was the sequence (<100bp) that immediately follows the sequencing primer and includes the CpG of interest plus any additional CpGs in the region of interest (in this case this included 2 additional CpGs) (Table 2.6). CpG 2 and 3 in the analysis sequence are not featured on the 450K array. The sequencing entry also contained a low frequency SNP (rs190517174, A: 99.641% (4990 / 5008); G: 0.359% (18 / 5008)). This was entered into the PyroMark ID software as 'R' to denote either A or G.

Table 2.6 Dispensation order for the cg11531579 assay.

CpG	Analysis sequence	Dispensation order
cg11531579	RGTTTGTTAAATTC/TGTC/TGTTAC/TGAGT	TAGTGCTGATCAGTCGTGATCGAG

C/T indicates the CpG sites in the sequence. Dispensation order (the sequence in which nucleotides are dispensed) is determined by the PyroMark software to maximise efficiency.

2.5.9 Pyrosequencing protocol

Both the binding buffer and annealing buffer solutions were prepared according to the protocol. The Binding Buffer Solution was prepared to a total of 70µL per sample (beads, binding buffer and Milli-Q H_2O) as per the measurements in Table 2.7. This was added along with 10µL of PCR product to a 96 well plate. The plate was sealed and mixed (1400rpm) for 10 minutes. To a PyroMark Q96 Plate Low, 40µL of Annealing Buffer Solution was added,

comprising of 38.4µL Annealing Buffer (20 mM Tris-Acetate - 2 mM $MgAc_2$, Qiagen, Hilden, Germany) and 1.6µL Sequencing Primer (10pM)(Table 2.7).

Table 2.7 Preparation of the binding buffer and annealing buffer solutions

Reagent	Per well (µL)	Source
Binding buffer solution		
1x Binding Buffer	40	Qiagen (Hilden, Germany)
Streptavidin-sepharose beads	2	VWR International (Leicestershire, UK)
Milli-Q Water	28	
Total per well	**70**	
Annealing buffer solution		
Annealing buffer (20 mM Tris-Acetate – 2 mM $MgAc_2$)	38.4	Qiagen (Hilden, Germany)
Sequencing primer (10pM)	1.6	IDT (IA, USA)
Total per well	**40**	

After 10 minutes of mixing, the 96 well plate was transferred to the PyroMark® Q96 ID Vacuum Workstation (Qiagen, Hilden, Germany). The sepharose beads in the binding buffer solution bind to the biotin labelled primer (in the samples). The sepharose-bound samples are then isolated and immobilised using the vacuum block tool. The bound-samples then undergo denaturation (to single stranded DNA) and washing in the wells of the vacuum workstation. Firstly, the samples were rinsed with 70% ethanol. Next, the samples were denatured in the denaturing buffer (0.2 M NaOH), and then washed in the wash buffer well (10X, pH 7.6 (10 mM Tris-Acetate)). Finally, the samples were released into the PyroMark Q96 Plate containing annealing buffer solution. The plate was sealed, the samples heated for 2 minutes at 80^{o}C, and then left to cool to room temperature for 10 minutes to allow annealing of the sequencing primer to the single stranded biotin-labelled PCR product. The reagent cartridge was loaded with nucleotides, enzyme and substrate according to the volumes calculated by the PyroMark software.

Samples were run in duplicate from the same PCR reaction (section 2.5.10). Replicates that were not within 5% methylation of one another were repeated. Methylated controls (0% and 100%) were included in each plate and should be comparable between plates. To indicate the presence of contamination, negative controls from PCR (no template control) and pyrosequencing (binding buffer solution and no PCR product) were included on all plates. Pyrosequencing reactions were carried out using the PyroMark Q96 ID system (Qiagen, Hilden, Germany), in 96-well PyroMark Q96 plates (2 batches) on the same day.

The PyroMark software output also includes a bisulfite modification quality check (pass indicates conversion efficiency >95%), to check that all non-methylated cytosines have been converted to thymine. Furthermore, the software also includes a peak height quality check for all CpG sites. For any samples that did not pass the quality check, the pyrograms were visually inspected and if there were no apparent issues in the bases specified or the peak heights, were corrected to 'passed' status. Some samples had very low peak heights and noisy traces and therefore failed (*n*=11). This was likely due to low DNA concentrations as all the samples which failed due to low peak heights were those that had <50ng of starting DNA, therefore these were repeated with 3µL of bsDNA. An example of a pyrogram output from the PyroMark software is shown in Figure 2.10.

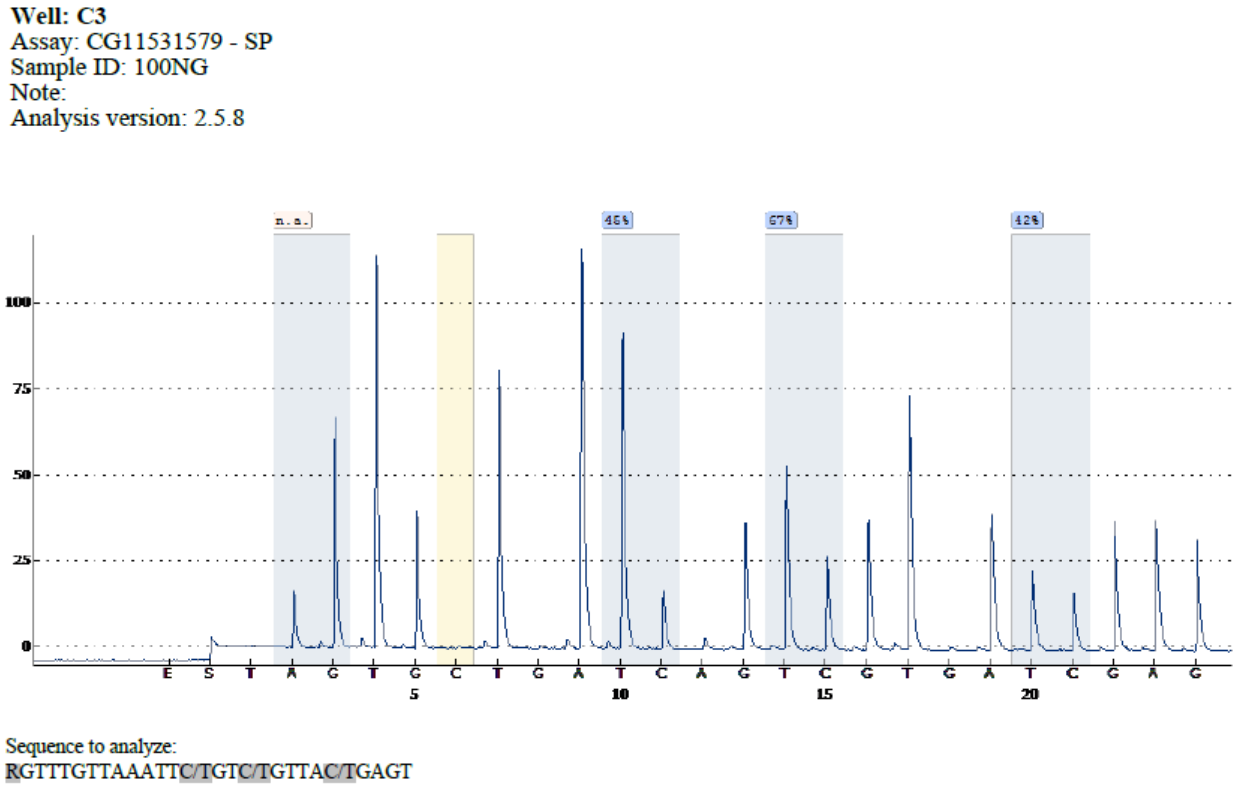


Figure 2.10 Example pyrogram of cg11531579 assay.
Peak height is presented on Y axis and dispensation order on the X-axis. The blue shaded areas show the CpG sites, with corresponding % methylation above. The yellow shaded indicates the bisulfite modification quality check. The peaks indicate that light was emitted when that nucleotide was added. CpG 1 is 45% methylated, CpG 2 is 67% methylated and CpG is 42% methylated. The n.a. corresponds to the potential SNP which was not a feature of this analysis.

2.5.10 Pyrosequencing validation

Pyrosequencing assays were validated across a known range of DNAm concentrations to ensure that DNAm is quantifiable across a range of values. Validation was done using DNA

methylated control DNA (Epitect, Qiagen, Hilden, Germany) between 0% and 100% methylation (Table 2.8). For each CpG, the expected vs observed methylation was plotted on a scatter graph, and the linear equation of the line and the correlation between values were determined.

Table 2.8 Composition of the standards for 10ul between 0-100 % methylation DNA using control DNA

% Methylation of DNA	Methylated DNA %	Unmethylated DNA %	Methylated DNA (µL)	Unmethylated DNA (µL)
0	0	100	0	10
5	5	95	0.5	9.5
10	10	90	1	9
25	25	75	2.5	7.5
50	50	50	5	5
75	75	25	7.5	2.5
90	90	10	9	1
95	95	5	9.5	0.5
100	100	0	10	0

2.5.11 How to define outliers in DNA methylation analysis

There were some influential outliers in DNAm and therefore steps were taken to address these. The influence of outliers can be large when the sample size is relatively small, which can affect the mean and increase variability. In statistical models, many of which rely on mean differences, this can influence results and model validity. Statistical inference will rely on a standard deviation which measures normal spread of data, but extreme values at either side of the distribution increases the standard deviation and decreases the likelihood of finding a statistically significant difference, which increases likelihood of type II error (Cousineau and Chartier, 2010). There are various ways to define outliers and handle outliers; however there is no unanimous consensus.

One route would be to determine if the values obtained are consistent with findings from other published studies. There has not been another study which has analysed DNAm at this locus using pyrosequencing, and it is acknowledged that alternative methods for quantifying DNAm (such as arrays) may differ in the values obtained. However, blood methylation levels

in adults have been investigated in a handful of cohorts with freely accessible 450K array data. CpG methylation at this locus has been measured in similar populations (European decent) in cohorts in Australia and in the Netherlands. In the Brisbane Systems Genetics Study (BSGS), DNAm was measured in 614 individuals from 117 families of European descent in Brisbane, Australia (Powell et al., 2012). Families consisted of adolescent monozygotic (*n* = 67 pairs) and dizygotic (*n* = 111 pairs) twins, their siblings (*n* = 119), and their parents (*n* = 139). The children were on average 14 years old (range 9-23), adults were 47 years old (range 33-75) (Figure 2.12), and beta values at this CpG were between 0.05-0.12 (Figure 2.11). The authors state that any measurement greater than five interquartile ranges (determined/decided by comparing estimates with and without outliers) from its nearest quartile was set to missing in order to avoid the influence of outliers (Powell et al., 2012). Therefore, the range of methylation values at this CpG could have differed prior to adjustment. In the BSGS, methylation at this CpG (cg11531579) positively correlated with age (Pearson correlation =0.34, p<0.0001) (Figure 2.12).

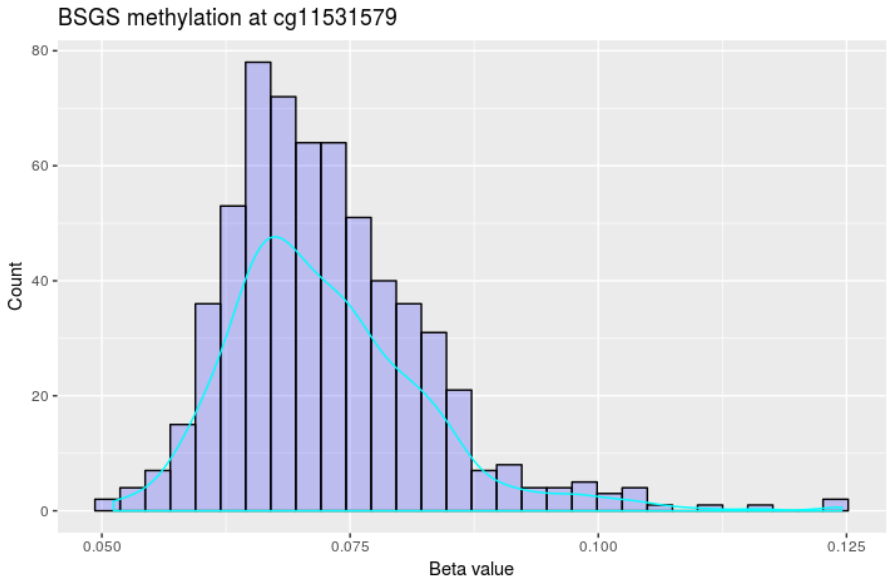


Figure 2.11. Histogram of methylation (beta values) at cg11531579 in the Brisbane Systems Genetics Study (BSGS).
N=614. Line represent Kernel density estimate. Methylation ranged from 0.05-0.12 (median 0.07).

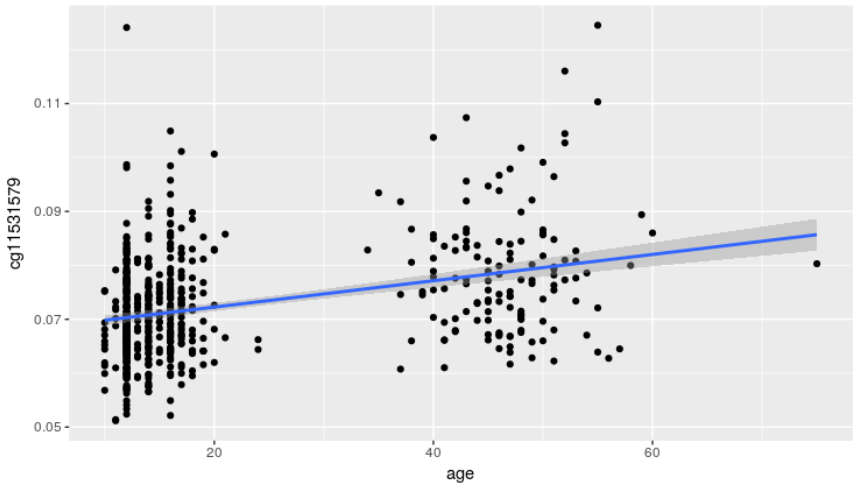


Figure 2.12 Plot of methylation (beta values) at cg11531579 by age in the Brisbane Systems Genetics Study (BSGS). Line represents the linear fit with shaded 95% confidence intervals. Pearson correlation =0.34, p<0.0001. Age ranged from 10-75 years.

Another study examined methylation in elderly individuals. In the B-Vitamins for the PRevention Of Osteoporotic Fractures (B-PROOF) study, a randomised controlled trial which involved B12 and folic acid supplementation, methylation levels were quantified in 87 individuals aged 65-75 years from the Netherlands using the 450K array (Kok et al., 2015).

In the B-PROOF study, the range of methylation values was greater than those in the BSGS cohort, with beta values ranging from 0.05 to 0.3 (Figure 2.13). The data presented from the B-PROOF study are the filtered data, normalised using the Subset-quantile Within Array Normalization (SWAN) procedure (Maksimovic et al., 2012) (available in the R package minfi (Aryee et al., 2014)).

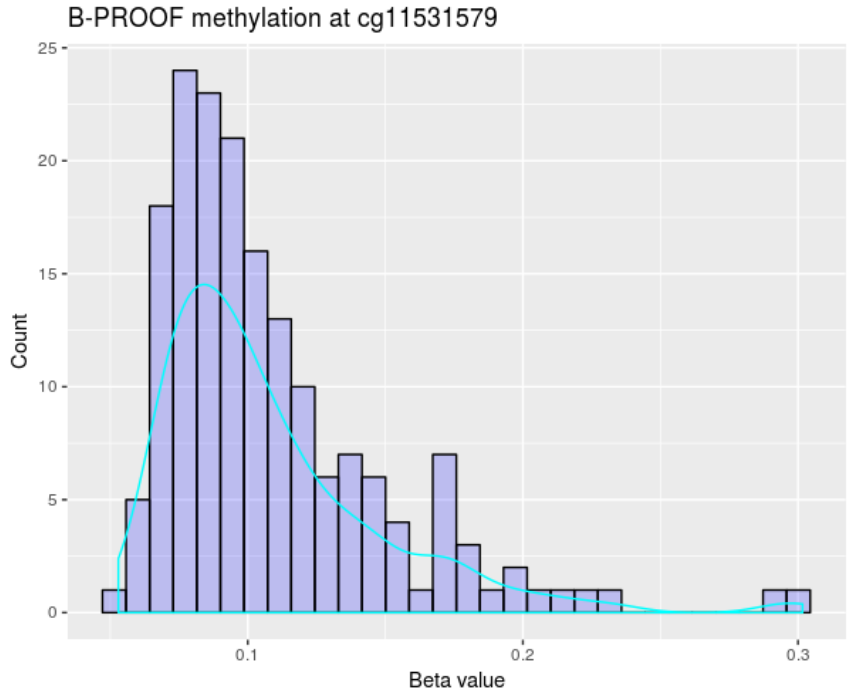


Figure 2.13 Methylation (beta values) at cg11531579 in the B-PROOF study.
Betas were normalised betas using the SWAN normalisation procedure. Line represent Kernel density estimate. Methylation values ranged from 0.05-0.30, median methylation was 0.10. Baseline age ranged from 65-75. Methylation did not vary with age within this age range.

Another option for defining outliers would be to use the statistical formula that defines outliers as greater than 1.5 x IQR. Using this formula on NTFS data would exclude data points with methylation greater than 11.2% (*n*=16), which is in line with methylation levels from the Brisbane study (max beta value was 0.12, aka 12%). However, this would exclude many values that could be valid, as seen in the B-PROOF study, whereby the upper observed methylation levels were 32% (beta value 0.32). The upper values in these two large studies are very dissimilar (12% - 32%) (Table 2.9) and choosing one of these values as part of the criteria would give notably different results.

In NTFS, DNAm was centred around 3%, with a minority of samples that had very high methylation (Figure 2.14). Therefore, using a percentile cut-off (90th percentile, >13.4%) would exclude those with very high methylation, leaving observations that are more representative of the sample generally.

Different cohorts adopt different methodological processes for dealing with outliers in their data, and different normalisation pipelines. This makes comparing the statistical outliers with values obtained in previous studies problematic, as the data presented are often processed data (such as with B-PROOF and BSGS), with limited information provided on the processing steps applied.

However, as there is consistency in the values specified using the statistical formula, 90th percentile and the Brisbane study, there is some agreement for an upper limit of 12% (Table 2.9). Therefore, the definition of 'high methylation' utilised here was methylation >12%, which is reasonable considering the characteristics of the data. Methylation increases at this locus with age (Figure 2.12), which could explain some differences observed, and the Brisbane study included ages more similar to this cohort, whereas the B-PROOF study included older participants.

Sensitivity analysis was done excluding the outliers. Outliers were defined as those with >12% methylation, in line with findings from previous studies (Table 2.9). Statistical analyses were repeated with outliers excluded to see how this influenced results.

It is worth noting that other factors such as disease or lifestyle can influence methylation patterns (Robertson, 2005, Anderson et al., 2012). In NTFS, no variables (sex, smoking, exercise, SES, RWG, birthweight, alcohol intake) predicted whether an individual had outlier methylation. Sample sizes were too small to investigate whether high methylation levels were associated with disease (i.e. cancer or infection).

Table 2.9 Threshold methylation values for the different approaches to defining outliers for cg11531579

Method	Study	Threshold methylation value
Statistical method (>1.5*IQR)	NTFS	11.2
90th percentile	NTFS	13.4
Comparison with other data sets	Brisbane	12
	B-PROOF	32
	ALSPAC	12

Beta values are presented as % methylation for comparison. IQR, inter-quartile range.

2.5.12 Statistical analysis

Pyrosequencing results were analysed using STATA version 15 (STATA Corp., Texas, USA), and methylation values are reported as percentages (0-100%). Average (mean) DNAm was calculated at each individual CpG locus from the duplicate pyrosequencing reactions. From this, overall mean methylation (across the region spanning 3 CpG loci), was calculated as the average of the 3 values at the 3 loci. The correlation between CpG loci was assessed using Pearson's correlation, and were also examined stratified by RWG.

There were outliers in DNAm values (very high methylation), which positively skewed the data, therefore the median was the preferred summary statistic. Outliers were detected across all CpG loci, and those which had high methylation at 1 CpG site tended to also have high methylation at the other loci. The impact of the outliers was addressed by using the statistical methods for non-parametric, positively skewed data. Furthermore, sensitivity analysis was carried out to assess the impact of the outliers.

Summary statistics are presented for DNAm by RWG and by body composition (OWOB, OB). Phenotypic (RWG, OWOB, OB) differences in DNAm were compared using Wilcoxon signed-rank test(s). DNAm between sub-phenotype groups (no RWG/RWG and OWOB/healthy weight) were compared using the non-parametric Kruskal-Wallis test. This test is less sensitive to outliers, or non-normal distributions, and therefore is a more appropriate test to use than the ANOVA. A post-hoc Dunn's test was used to adjust for multiple comparisons. The association between DNAm and phenotype was assessed further in regression models (detailed methods in section 2.5.12.1), with adjustments for sex. Batch (i.e. pyrosequencing plate) was initially included as a factor variable in regression models, however inclusion did not affect estimates (<10% change in coefficients) and therefore batch was not considered a confounder in these analyses. Stratified analysis was carried out to examine methylation in only those who had RWG in childhood (the exposed group), with regards to adult OWOB.

Analyses of the epigenetic datasets (to determine outliers) was done using R version 3.5.2, and reports methylation (beta values, ranging from 0-1).

2.5.12.1 Statistical models

Both the relationship between the early life exposure (RWG) and later life DNAm, as well as the relationship between DNAm and outcome (body composition) were of interest, and were addressed using an adaptation of the 'meet in the middle' approach. This strategy has been proposed as a means to identify intermediary biomarkers related to both exposure and disease outcome (Vineis and Perera, 2007), and has been previously applied in the metabolomics literature (Chadeau-Hyam et al., 2011, Assi et al., 2015). In short, this approach evaluates the following relationship: exposure → intermediate biomarkers of exposure ← disease. In this approach, omics data from prospective cohorts is used to identify molecules that represent intermediate markers of early effect, which are used to link exposures with disease endpoints (Vineis and Perera, 2007). The approach is flexible in

that it can utilise data from a prospective study, or can crossover between exposure biomarkers and disease risk markers from case-control studies. If searching for causal associations, the causal nature of an association is reinforced if it is found in all three steps, however causality is not justifiable without formal mediation analyses.

Statistical models were utilised to examine the associations between the exposure(RWG)→DNAm, and the DNAm→outcome (body composition) associations to gauge clues about DNAm as a possible intermediary. These models had to consider the characteristics of the data and the model assumptions, therefore a variety of models were utilised (Table 2.10).

The relationship between body composition outcomes (OWOB, OB) and DNAm (exposure) were assessed using logistic regression. As sensitivity analysis, stratified analysis examined associations in the exposed (those who had RWG) population (Appendix F, Table XXIV) however associations did not differ.

There were indications from the scatter plots that the relationship between BMI and DNAm was non-linear (outlined in section 2.5.12.1.1). Therefore, fractional polynomial models were used in order to determine the appropriate power transformation for these data. Median and Tobit regression models were used to examine the relationship between DNAm and RWG, to account for the positive skew in the dependent variable (DNAm).

Table 2.10 Summary of the statistical models, exposures, outcomes and age at measurement

Statistical model	**Reasoning**	**Exposure (age)**	**Outcome (age)**	**Sensitivity** [b]
Linear regression	1. To model a perceived linear relationship 2. Non-linear relationship was analysed with fractional polynomial terms	DNAm [a] (age 50)	BMI (ages 50 and 60)	✓
Logistic regression	Binary outcome variable	DNAm [a] (age 50)	OWOB (ages 50 and 60) OB (ages 50 and 60)	✓ ✓
Median regression	Skewed outcome variable	RWG (0-12 months)	DNAm [a] (age 50)	✓
Tobit regression	Skewed outcome variable with many zero's	RWG (0-12 months)	DNAm[a] (age 50)	✓

[a] DNA methylation (DNAm) refers to separate models for the 3 individual CpG loci, and for average methylation.
[b] Sensitivity analysis was done for CpG3 only (as it was the CpG of interest).

2.5.12.1.1 Fractional polynomial terms

Fractional polynomial (FP) models can be used to compare fit in regression models using non-linear functions (Royston and Altman, 1994). The combinations of powers (−2, −1, −0.5, 0, 0.5, 1, 2, 3) are each fitted to the model until the best fitting model (lowest deviance) is achieved. The FP degree of the polynomial (the largest exponent) is termed m.

Linear regression using the derived fractional polynomial terms was utilised with BMI as the outcome and DNAm as the predictor.

2.5.12.1.2 Median regression model

Standard linear regression uses the mean of the dependent variable (Y) to make inferences about the data, whereas quantile regression makes predictions for a given quantile of Y, such as the 50th (median). Median regression is a semi-parametric form of regression analysis which is more robust to outliers than standard linear regression (Koenker and Bassett Jr, 1978). It does not require the same assumptions as linear regression regarding a parametric distribution of the residuals or constant variance. Median regression was utilised to examine the relationship between methylation (age 50) with the exposure, RWG, adjusted for confounders (sex, birthweight). Robust estimates of standard error are reported, to account for heteroscedasticity (demonstrated using residual plots). These results were compared with those from the Tobit models.

2.5.12.1.3 Tobit model

A Tobit model was also used to estimate the linear relationship between methylation and RWG. The Tobit model is a censored regression model, which estimates linear relationships between variables when there is censoring in the outcome variable from either above or below (Tobin, 1958). Censored distributions are a mixture of both discrete and continuous distributions. Values above or below a threshold are censored (i.e. unknown above the upper or below lower limits), such that all values take on that specified value, even if the true value may be higher or lower. For example, the Tobit model will regard observations below the lower limit as lying somewhere between the limit and zero.

Censoring from below would apply in this case to accommodate the many methylation values that equal 0 (Humphreys, 2013). The presence of many zeros in the dependent

variable causes issues when using an ordinary least squares model, as the data are positively skewed. However, it is assumed that these are true zeros in this case. Here the outcome (methylation) has both a lower and upper limit of detection (as identified in the assay validation), meaning methylation is frequently reported as 0% because the sensitivity of this assay (the smallest value of methylation that was observed at 0% methylation in the validation) was ~3% (Figure 2.14). In this case, a Tobit model would be appropriate, with lower censoring at 3%, which would mean that any observation less than 3% is not known exactly but then takes on that value. Upper censoring was also applied with the upper detection limit from the assay validation.

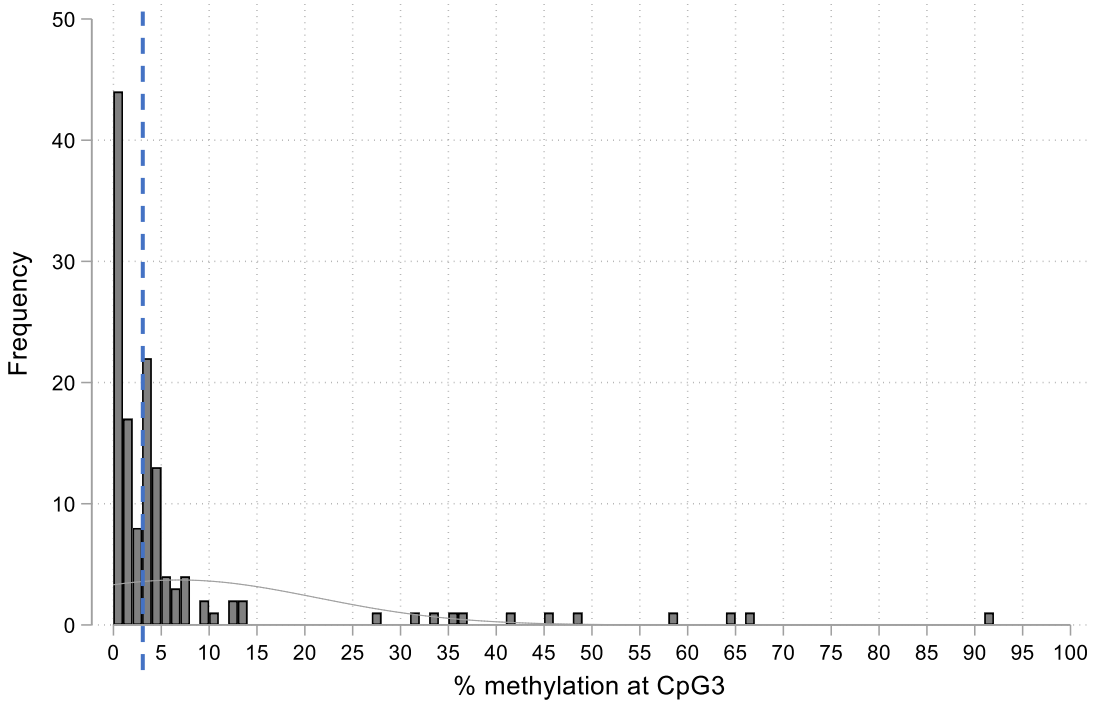


Figure 2.14 Observed distribution of methylation values at CpG3 (cg11531579).
If the lower limit of the assay is 3% (dashed line), then *n*=80 of these samples fall above the limit. Therefore, the distributions for samples with over 3% methylation is known, but only know the number of observations below 3% is known. Median regression and Tobit regression were utilised to address this issue.

2.5.13 Potential effects of single nucleotide polymorphisms

A potential explanation for the high methylation observed in some samples, is that SNPs either upstream or downstream could also influence methylation (Gibbs et al., 2010, Chen et al., 2013). In order to investigate if methylation patterns were being influenced by SNP effects, the region nearby the CpG of interest was sequenced. Samples were selected with high methylation (>12% methylation). There were 7 samples with high methylation, which

had DNA remaining. These were randomly matched (sex, RWG, BMI (age 50)) with other 'low methylation' (<5%) samples using the STATA program ccmatch (Cook, 2015).

Genomic DNA primers were designed using the region 800 bp either side of the CpG of interest, forming a product size of 865 bp. PCR reactions were optimised as described previously (section 2.5.5) using a temperature gradient at both high (100ng) and low (50ng) amounts of DNA. Once optimised, PCR was carried out on the NTFS DNA samples, which for many samples used the remaining DNA. An estimated 20ng of gDNA (where available) was used for each PCR reaction in a total reaction volume of 24uL. Some samples failed (likely due to very low DNA concentrations), with no visible bands when visualised using gel electrophoresis, leaving 3 pairs (*n*=6) samples for sequencing.

PCR clean up and sequencing (of the left and right strands) was carried out by the University of Sheffield Core Genomic Facility. After sequencing, the left and right reads need to be aligned, however all the left reads failed sequencing, and therefore the reverse complement of the right sequence was used for analysis.

The sequence traces were visually inspected for quality. Sequences were aligned to the reference sequence using genescreen, a desktop program for alignment of multiple DNA sequences that highlights SNPs (Carr et al., 2011). The SNPs were compared between matched pairs. Linkage disequilibrium, when genetic variants are inherited in a non-random manner, was examined using the web-based tool: LDlink (Machiela and Chanock, 2015). Between two genetic variants, D prime (D') indicates the allelic segregation, and R squared (R^2) measures the correlation of alleles, with values ranging from 0 (weak correlation) to 1 (strong correlation).

Chapter 3. Exploring the relationship between early and later life exposures and obesity in middle-age

3.1 Introduction

This chapter addresses the first aim of the thesis and examines if early life factors and SES are associated with body composition in NTFS adults (aged 49-51 years), in order to determine which (if any) early life factors have a lasting effect into middle age, irrespective of various lifestyle and socioeconomic influences. The investigation of early life factors for obesity risk is a relatively recent concept, hence early life factors have been investigated in few pre-obesogenic cohorts so far. Additionally, the majority of the literature concentrates on childhood obesity, and there is limited data as to whether these are risk factors in adults, due to a lack of long-term follow-up data in many studies (Parsons et al., 1999). Data from NTFS, which commenced in 1947, provides the opportunity to investigate early life factors for in a pre-obesogenic, post-war cohort of middle-age adults to address these shortcomings.

These are important questions to ask as they might direct whether interventions or resources should focus on addressing the early life exposures, or on modifying environment and lifestyle, in order to reduce the burden of obesity.

3.2 Aims

The analysis in this chapter utilises data from a pre-obesogenic environment cohort to i.) Determine if early life exposures were associated with adult body composition in the NTFS cohort; ii.) Examine if risk factors vary for different outcomes i.e. using proxy measures or direct measures of body composition and distribution; and iii.) Examine the relative contribution of early life factors, SES and lifestyle to adult BMI, as well as the relationships between these factors.

3.3 Participants and Methods

3.3.1 Exposure and outcome data

The NTFS dataset is described in detail in chapter 2.1.1. A large proportion of the cohort were traced at age 50 (*n*=866, 88%), with 113 lost to follow-up, which included those who died (after age 1), moved abroad or were untraceable.

Early life exposures used in this analysis included birthweight (continuous and categorical), occupational social class at birth, maternal age (continuous and categorical), infant feeding (breastfeeding, exclusive breastfeeding and weaning age), bacterial infections, adverse events, and housing problems. Sex and gestational age were additional covariates.

Data on lifestyle factors (smoking, physical activity) and socioeconomic position were derived from the self-completed questionnaire, and nutrient intake from the food frequency questionnaire, at age 50. Multiple dimensions of body composition were analysed as outcomes including; BMI and obesity (OB), body fat percentage (BF%) and waist-hip ratio (WHR). The coding of the early and late life factors is in section 2.1.1.2.

3.3.2 Statistical methods

Descriptive statistics are presented for each demographic, explanatory and outcome variable. A potential issue with longitudinal data is attrition (loss of participants over time), which reduces sample size and can leave a biased remaining sample. Apart from sex (fewer males), the cohort followed up at age 50 were shown to be representative of the original cohort for the early life factors examined in a previous study (Lamont et al., 2000). This was reassessed for the study members who had BMI measures at age 50, with respect to the early life variables of interest in this study (methods in section 2.3.2).

Relationships between BMI and obesity (BMI>30kg/m^2) and each of the early life and later life factors were examined using linear (BMI, WHR, %BF) and logistic (obese) regression models, and path analysis. For details on statistical methods see 2.3.5.

3.4 Results

3.4.1 Sample representativeness

Questionnaires were returned by 574 members and 412 attended a full clinical assessment between ages 49-51. There were some differences between the original cohort and those who attended the clinic at age 50 (Table 3.1). There were a greater proportion of women ($p<0.01$), and fewer from the lower occupational social class group. Therefore, sex and SES at birth were included as covariates in adjusted models.

Table 3.1 Differences in early life variables between NTFS study members present at birth and at age 50

Variable		Mean/Median/(%)		P value
		Did not attend	Attended	
Birthweight (z-score)		-0.2	-0.1	0.57
Maternal age (years)		28.3	28.7	0.28
Housing problems (%)		58	52	0.06
Duration Breastfed (days)		132.9	119.6	0.16
Female (%)		44.6	56.5	<0.01
SES (birth, %)	Least advantaged	33	27	0.06
	Mid	54	62	
	Most advantaged	13	11	

P-values calculated from t-tests, chi-square tests or Mann-Whitney tests as appropriate

3.4.2 Sex differences

There were no *a priori* reasons to anticipate sex differences in early life exposures, however, socioeconomic differences between men and women have been observed when examining obesity (Wardle et al., 2002). Therefore, a Wald test was performed for each outcome for the socioeconomic variables (occupational SES at birth and middle age, education and income) to determine if there were significant differences between men and women. These results indicated that the models for WHR and BF% (but not BMI) should be stratified by sex. Furthermore, as there are sex-specific cut-offs for WHR (Table 1.1) and BF% (Table 3.2) for men and women, it is appropriate to stratify. There were also biological differences in support of stratifying by sex for these outcomes, whereby most men had an 'at-risk' WHR (82% men, only 21% women), and most men had obese levels of body fat (86%, compared to just 56% of women).

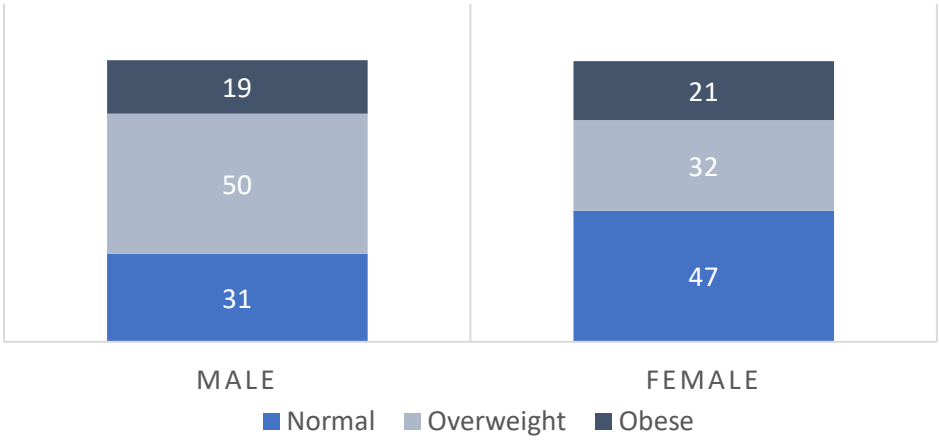
Table 3.2 Percentage BF cut-offs by sex for adults (aged 40-59) using the Gallagher classification

Age (years)	Category	Sex	
		Men	Women
40–59	Underweight	<11.0%	<23.0%
	Healthy	11.0%–22.9%	23.0%–34.9%
	Overweight	23.0%–28.9%	35.0%–40.9%
	Obese	≥29.0%	≥41.0%

3.4.3 Descriptive characteristics

3.4.3.1 Outcome measures

At age 50, the cohort had the same proportion of healthy weight (40%) and overweight (40%), with the remainder obese (20%) (Figure 3.1). These figures are in line with UK 1997 averages (age 45-54) of 43.6% overweight and 22.1% obese (Health and Social Care Information Centre, 2014). There were significant differences in BMI categories between sexes; with 50% of NTFS men classed as overweight compared to 32% of women.



	Healthy weight (n)	(%)	Overweight (n)	(%)	Obese (n)	(%)	Total	P
Males	54	31.2	86	49.7	33	19.1	173	**0.001**
Females	103	46.2	73	32.7	47	21.1	223	

Figure 3.1 Proportion (%) of NTFS study members in each weight category at age 50 stratified by sex. Number of observations (n). Chi-square p value shown for differences in distributions between sexes

BMI values for males followed a normal distribution, whilst the female values were skewed by some values greater than 40Kg/m^2 increasing the range of values (Appendix B, Figure VII). According to the Gallagher classification for age 40-59, the mean BF% for both men (36.4%)

and women (41.7%) (*Figure 3.1*) was in the category for greater risk of suffering from obesity related health conditions (Gallagher et al., 2000). As to be expected, women had a higher mean BF%, however males had a higher BMI and WHR (Table 3.3). Mean WHR in females was within healthy limits, whilst the average value for men was in the 'at-risk' category.

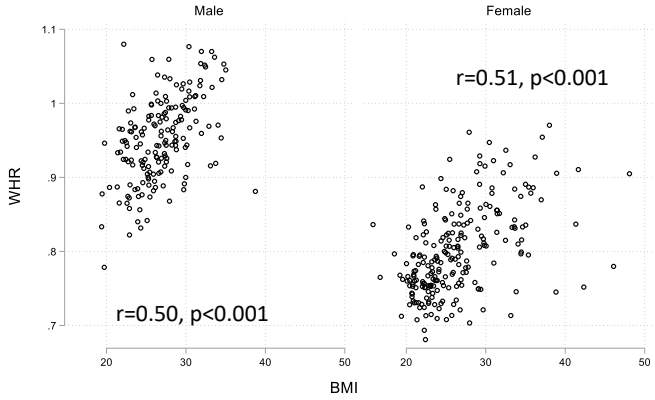
Table 3.3 NTFS continuous outcome variables at age 50 stratified by sex

Outcome			**Male**					**Female**			
variable	n	Mean	SD	Min	Max	n	Mean	SD	Min	Max	P
BMI	173	26.92	3.60	19.39	38.74	225	26.48	5.28	15.81	48.09	**0.02**
Body fat (%)	172	36.43	7.16	15.40	54.40	223	41.74	9.12	14.40	61.80	**<0.001**
WHR	172	0.95	0.06	0.78	1.08	225	0.80	0.06	0.68	0.97	**<0.001**

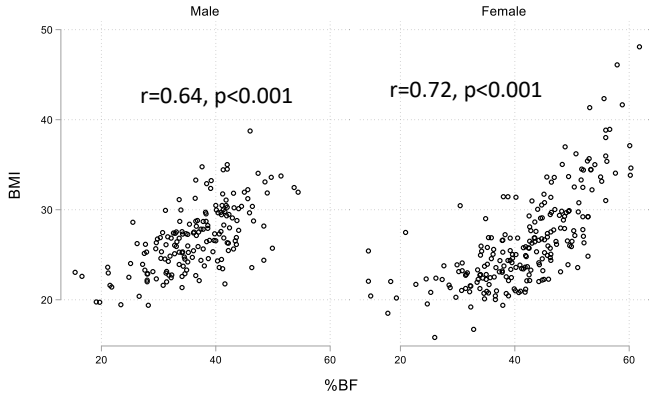
When testing for differences in distributions, parametric t-tests (WHR) were conducted or Mann-Whitney tests (BMI, BF%) were used for skewed distributions. Number of observations (n), standard deviation (SD), minimum (min) and maximum (max) values and P values for differences between sexes.

3.4.3.2 Correlations between adiposity outcomes

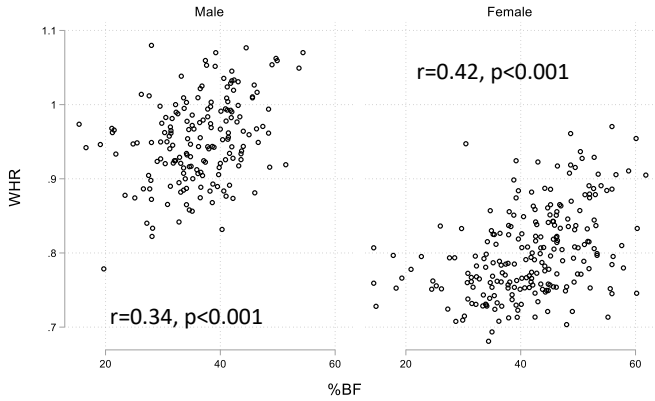
Correlations between measures of adiposity were stronger for females than males (Figure 3.2). The strongest correlation for both sexes was between BMI and body fat, suggesting that BMI is a good predictor of body fat in this population. There were also moderate positive correlations between WHR and body fat (Figure 3.2). There was a weak correlation between BF% and WHR, more so for males (r=0.34) than females (r=0.42).



A. WHR and BMI



B. BMI and BF%



C. WHR and BF%

Figure 3.2 Correlations between NTFS outcome measures stratified by sex.
A.) BMI z-score and body fat percentage (BF%).
B.) waist-hip ratio (WHR) and BMI z-score.
C.) WHR and body fat percentage.
r, Pearson correlation coefficient. Explanatory variables

3.4.3.3 *Exposures*

Descriptive data are presented for all explanatory variables Table 3.4 (continuous or ordinal) and Table 3.5 (categorical variables). Aside from females having a higher birthweight z-score (Table 3.4), there were no other significant differences between males and females for the early life variables (Table 3.5).

Mean birthweight z-score for this population was lower (-0.18) than the growth reference, which likely reflects the earlier date of birth (Cole et al., 1995). The average maternal age was 28 years old, with the majority of mothers aged 25-34.

Most infants were in the normal birthweight category (Table 3.5). There were significant differences in birthweight z-score between males and females with significantly more females born LGA (12.5% compared to 4%). Around 42% of study members experienced a period of rapid growth from birth to twelve months.

Around a third (30%) of study members who came to clinic at age 50 were breastfed for over 6 months, whilst 19% were never breastfed (Table 3.5). Infant feeding was not related to SES (Appendix B, Figure VIII). Over half of the cohort experienced some form of housing problem, a third experienced an adverse event before they were three years old and 18% had a bacterial infection within the first year (Table 3.5).

Table 3.4 Early life differences in continuous/ordinal variables between NTFS males and females

Variable	All				Male		Female		
	n	Mean (SD)	Min	Max	n	Mean (SD)	n	Mean (SD)	P
Birthweight (kg)	398	3.39 (0.49)	1.93	4.88	173	3.41 (0.47)	223	3.38 (0.51)	0.57
Birthweight (z-score)	398	-0.12 (1.04)	-3.17	3.02	173	-0.28 (0.93)	223	0.01 (1.11)	**0.01**
Gestational age (weeks)	395	39.89 (0.98)	33	44	171	39.94 (0.81)	222	39.86 (1.10)	0.39
Maternal age (years)	398	28.72 (5.89)	17	45	173	28.44 (5.70)	223	28.87 (6.02)	0.47
Breastfed (days)	377	122.94 (119.76)	0	443	168	121.88 (118.68)	207	124.15 (121.21)	0.71
Exclusively breastfed (days)	387	69.41 (62.60)	0	291	170	65.70 (61.81)	215	72.28 (63.21)	0.46
Introduction of solids (days)	379	158.76 (42.45)	15	345	168	157.17 (43.95)	209	160.03 (41.43)	0.52
No. infections in first year	393	1.09 (1.13)	0	7	172	1.21 (1.20)	219	1.01 (1.07)	0.11

Number of observations (n), standard deviation (SD), minimum (min), maximum (max) values, and P values (p) for t-tests or Mann-Whitney tests for differences between sexes presented in columns.

Table 3.5 Descriptive statistics of categorical early life variables stratified by sex, for the NTFS age 50 sub-sample

Categorical variables	Male (%)	Female (%)	Total (%)	p
Centile categories of birthweight (*n*=395)				
SGA (*n*=44)	11.7	10.7	11.1	**0.014**
Normal (*n*=316)	84.2	76.8	80	
LGA (*n*=35)	4.1	12.5	8.9	
Weight categories of birthweight (*n*=398)				
LBW (*n*=10)	1.7	3.1	2.5	0.598
Normal (*n*=382)	97.1	95.1	96	
HBW (*n*=6)	1.2	1.8	1.5	
RWG (*n*=163)				
No (*n*=95)	59.7	57.3	58.3	0.759
Yes (*n*=68)	40.3	42.7	41.7	
Maternal age at birth (*n*=398)				
Less than 25 (*n*=109)	27.2	27.6	27.4	0.99
25-34 (*n*=217)	54.9	54.2	54.5	
35+ (*n*=72)	17.9	18.2	18.1	
Housing problems (*n*=396)				
None (*n*=192)	48	48.9	48.5	0.85
Housing issues (*n*=204)	52	51.1	51.5	
Breastfeeding (*n*=219)				
None (*n*=42)	16.7	20.9	19.2	0.73
<4 weeks (*n*=29)	15.6	11.6	13.2	
4 wks - 6 month (*n*=83)	36.7	38.8	37.9	
6 months+ (*n*=65)	31.1	28.7	29.7	
Occupational social class at birth (*n*=390)				
Least advantaged (*n*=107)	29.8	26.7	27.4	0.29
Mid (*n*=240)	56.5	65.3	61.5	
Most advantaged (*n*=43)	13.7	9.0	11.0	
Bacterial infection in first year (*n*=301)				
No (*n*=248)	78.9	85.0	82.4	0.17
Yes (*n*=53)	21.1	15.0	17.6	
Viral infection in first year (*n*=307)				
No (*n*=218)	42.2	57.8	71.0	0.80
Yes (*n*=89)	43.8	56.2	29.0	
Any early life adverse event (first 12 months) (*n*=352)				
No (*n*= 303)	50.0	50.0	74.7	0.50
Yes (*n*= 49)	44.9	55.1	25.3	

Row sample sizes (n), column percentages (%) shown and Chi-square p-value (p) presented

In terms of later life characteristics (Table 3.6), fewer women were educated, and 40% of women had no qualifications, compared to 27% of men. The majority of the cohort undertook light activity. At age 50, 27% of included NTFS participants smoked, corresponding to 28% of women and 25% of men (Table 3.6). This is inconsistent with national (PHE) figures which show higher proportions of male smokers (ONS, 2014), although differences were not significant (Table 3.6). Over three quarters of the cohort were married.

Table 3.6 Descriptive statistics of explanatory later life categorical variables related to growth stratified by sex, for the NTS age 50 sub-sample.

Later life variables	Male (%)	Female (%)	Total (%)	p
Occupational social class at 50 (*n*=375)				
Least advantaged (*n*=52)	11.6	15.6	13.9	0.295
Mid (*n*=132)	39	32.2	35.2	
Most advantaged (*n*=191)	49.4	52.1	50.9	
Education (*n*=382)				
None (*n*=131)	26.8	39.9	34.3	0.004
O-level (*n*=131)	32.3	35.8	34.3	
A level (*n*=71)	25	13.8	18.6	
Degree (*n*=49)	15.9	10.6	12.8	
Educated past secondary school (*n*=382)				
No (*n*=262)	59.1	75.7	68.6	0.001
Yes (*n*=120)	40.9	24.3	31.4	
Physical Activity (*n*=387)				
Inactive (*n*=43)	7.7	13.7	11.1	0.156
Light Activity (*n*=191)	54.8	45.2	49.4	
Moderate (*n*=90)	21.4	24.7	23.3	
Heavy Activity (*n*=63)	16.1	16.4	16.3	
Current smoker at age 50 (*n*=395)				
No (*n*=288)	74.7	71.6	72.9	0.485
Yes (*n*=107)	25.3	28.4	27.1	
Current marital status (*n*=393)				
Not married (*n*=79)	17.8	21.9	20.1	0.313
Married (*n*=314)	82.2	78.1	79.9	

Row sample sizes (n), *column percentages* (%) *shown* and *Chi-square p-value* (p) presented

3.4.4 Relationships between early life exposures and later life BMI and obesity

In order to determine which early and later life factors might be predictive of middle-age BMI and obesity, each of these factors were examined individually (Table 3.7). In bivariate (unadjusted) analyses, study members were over twice as likely to be obese if they had a bacterial infection in the first year of life. However, there were no significant associations for viral infections. Those who smoked and participated in heavy physical activity were also significantly less likely to be obese. Current smokers had a significantly lower BMI. However, there were no significant differences in BMI or obesity likelihood between ex-smokers and non-smokers, or for number of pack years (Appendix B, *Table VI*). Those who did heavy physical activity had a significantly lower BMI and were 90% less likely to be obese (

Table 3.8).

A lower BMI was associated with having an older mother (aged over 35), whilst being educated and married was associated with a higher BMI. Socioeconomic advantage was associated with lower BMI. There were no associations for birthweight, adversity, overcrowding, housing problems, or income with either BMI or obesity.

Table 3.7 Bivariate (unadjusted) associations between exposures of interest and outcomes obesity (OB) (logistic regression) and BMI (linear regression) in NTFS study members (age 50)

		BMI			OB		
		Coef	CI	p	OR	CI	p
Sex	Female	-0.40	[-1.29,0.49]	0.38	1.12	[0.68,1.84]	0.66
Birth	Birthweight (z-score)	-0.08	[-0.49,0.33]	0.71	1.10	[0.87,1.40]	0.42
	Gestation (weeks)	0.25	[-0.15,0.66]	0.22	1.18	[0.90,1.55]	0.23
Maternal age	Continuous	-0.07	[-0.14,0.00]	0.06	0.99	[0.94,1.03]	0.50
	<25	0.45	[-0.59,1.48]	0.40	1.37	[0.79,2.38]	0.26
	25-34	Ref		.	Ref		.
	35+	-1.46	[-2.63,-0.29]	**0.01**	0.75	[0.36,1.55]	0.44
Infant feeding	Never breastfed	Ref		.	Ref		.
	<4 weeks	0.55	[-1.52,2.63]	0.60	2.29	[0.70,7.49]	0.17
	4 wk – 6 months	0.26	[-1.46,1.97]	0.77	1.53	[0.53,4.40]	0.43
	6 Months +	0.68	[-0.94,2.31]	0.41	1.52	[0.55,4.20]	0.42
	Breastfed (days)	0.00	[-0.00,0.01]	0.14	1.00	Ref	0.37
	Exclusive (days)	0.01	[-0.00,0.02]	0.15	1.00	[1.00,1.01]	0.29
	Weaning age (days)	-0.01	[-0.02,0.01]	0.35	1.00	[0.99,1.00]	0.37
Early life	Rapid weight gain	0.76	[-0.44,1.97]	0.22	1.22	[0.51,2.91]	0.66
	Any infection	1.22	[0.33,2.10]	**0.01**	2.03	[1.21,3.39]	**0.01**
	Number of infections	0.42	[0.02,0.82]	**0.04**	1.36	[1.11,1.67]	**<0.01**
	Bacterial infection	1.88	[0.53,3.23]	**0.01**	2.60	[1.34,5.02]	**0.01**
	Viral infection	0.29	[-0.82,1.39]	0.61	1.51	[0.83,2.73]	0.18
SES (childhood)	Least advantaged	Ref		.	Ref		.
	Mid	-0.92	[-1.93,0.08]	0.07	0.55	[0.32,0.95]	**0.03**
	Most advantaged	-1.40	[-2.98,0.18]	0.08	0.35	[0.13,0.99]	**0.05**
	Housing score	-0.02	[-0.43,0.38]	0.91	1.12	[0.89,1.39]	0.33
	Overcrowding	0.33	[-0.66,1.31]	0.52	1.65	[0.98,2.78]	0.06
Adversity	Any adverse event	0.16	[-2.02,2.33]	0.89	1.93	[0.66,5.61]	0.23
SES (later life)	Social class	Ref		.	Ref		.
	Mid	-0.63	[-2.04,0.79]	0.39	0.53	[0.26,1.08]	0.08
	Most advantaged	-0.59	[-1.95,0.76]	0.39	0.40	[0.20,0.80]	**0.01**
	Educated	0.51	[-0.46,1.48]	0.30	0.93	[0.53,1.62]	0.79
	No qualifications	Ref		.	Ref		.
	GCSE/O-level	-1.11	[-2.20,-0.02]	0.05	0.58	[0.31,1.08]	0.09
	A level	0.29	[-1.01,1.58]	0.67	0.86	[0.43,1.74]	0.68
	Degree	-0.50	[-1.96,0.95]	0.50	0.54	[0.22,1.32]	0.18
	Income	-0.31	[-0.75,0.13]	0.17	0.79	[0.62,1.01]	0.06
	Married	1.01	[-0.11,2.12]	0.08	1.65	[0.82,3.29]	0.16
Lifestyle	Inactive	Ref		.	Ref		.
	Light activity	-1.25	[-2.75,0.24]	0.10	0.73	[0.35,1.52]	0.40
	Moderate activity	-1.99	[-3.64,-0.35]	**0.02**	0.43	[0.18,1.01]	**0.05**
	Heavy activity	-2.90	[-4.65,-1.15]	**<0.001**	0.16	[0.05,0.52]	**<0.001**
	Smoker	-1.32	[-2.31,-0.34]	**0.01**	0.53	[0.28,0.99]	**0.05**

Coefficients (coef) and odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p); Ref, reference group. Bold indicates significant at p<0.05.

3.4.4.1 *Socioeconomic factors*

Those in the most advantaged socioeconomic group compared to the least at birth had a lower BMI in the unadjusted analysis (coef=-0.43, p=0.06)(Appendix B, Table III). Being educated to O-level was associated with a lower BMI compared to those with no qualifications (coefficient=-1.11, p<0.05), however, there was no trend across increasing educational attainment (Appendix B, Figure IX).

As this cohort experienced great upward social mobility (previously reported by (Forrest et al., 2011)), this was investigated further regarding BMI. Social mobility was not associated with BMI, but those consistently in the mid advantaged occupational social class group had 2.6 kg/m2 lower BMI (p=0.01) on average (Appendix, Table III). However, increasing social advantage was associated with lower odds of obesity across multiple levels (Appendix, Table III). Odds were lowest for those always in the most advantaged groups compared to always in the least (OR=0.09, p=0.005).

There were no associations for other SES variables including experiencing adversity, housing issues or overcrowding (a composite of housing issues) in the first year, and any adiposity-related outcome measures.

3.4.5 Multivariable regression models for BMI and obesity

The adjusted model for BMI shows that the significant early life exposures were older maternal age and bacterial infection in the first year, adjusted for sex, early and later life SES, breastfeeding and lifestyle (smoking and physical activity) (

Table 3.8). Physical activity level (PAL), smoking and social class (age 50) were also independent significant predictors of BMI. These factors explained 17% of the variation in BMI.

Adjusting for social class at birth attenuated the associations for sex, older maternal age, illness, and duration breastfed (

Table 3.8). The association between BMI and being married was attenuated by SES and was not significant after adjusting for smoking.

Table 3.8 Multivariable linear regression model for early and later life factors and BMI in NTFS study members age 50 (*n*=262).

BMI (age 50)		Coef	CI	p
Female		-0.8	[-1.88,0.28]	0.147
Maternal age				
	<25	0.43	[-0.81,1.67]	0.498
	25-34	Ref		.
	35+	-1.52	[-2.96,-0.08]	**0.038**
Social class (birth)		Ref		.
	Mid	-1	[-2.28,0.28]	0.125
	Most advantaged	-0.78	[-2.88,1.32]	0.465
Bacterial infection		2.12	[0.70,3.53]	**0.003**
Duration breastfed (weeks)		0.03	[-0.00,0.06]	0.079
PAL	Inactive	Ref		.
	Light activity	-2.30	[-4.17,-0.42]	**0.017**
	Moderate activity	-3.09	[-5.14,-1.03]	**0.003**
	Heavy activity	-3.89	[-6.18,-1.60]	**0.001**
Smoker		-2.31	[-3.57,-1.06]	**<0.001**
Social class (age 50)		Ref		.
	Mid	-2.02	[-3.74,-0.29]	**0.022**
	Most advantaged	-1.58	[-3.27,0.11]	0.067

Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p); Ref, reference group. Bold indicates significant at p<0.05. All covariates are presented.

In multivariable analysis; infections, smoking and heavy PAL remained significant predictors of obesity (Table 3.9). SES at birth was no longer significant, and instead later life SES was a significant predictor of obesity.

Table 3.9 Multivariable logistic regression model for early and later life factors and obesity in NTFS study members age 50 (*n*=275)

Obese (age 50)		OR	CI	p
Female		1.22	[0.62,2.40]	0.57
Social class (birth)		Ref		.
	Mid	0.55	[0.26,1.18]	0.13
	Most advantaged	0.52	[0.14,1.98]	0.34
Bacterial infection		2.75	[1.30,5.79]	**0.01**
PAL	Inactive	Ref		.
	Mild	0.62	[0.24,1.57]	0.31
	Moderate	0.41	[0.13,1.24]	0.11
	High	0.06	[0.01,0.53]	**0.01**
Smoker		0.35	[0.15,0.81]	**0.01**
Social class (age 50)		Ref		.
	Mid	0.29	[0.11,0.75]	**0.01**
	Most advantaged	0.36	[0.14,0.89]	**0.03**

Odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Ref, reference group. Bold indicates significant at p<0.05. All covariates are presented.

In summary, bacterial infection was the only early life risk factor associated with both increased BMI and obesity in adjusted models. Being more advantaged in later life, and the lifestyle factors; smoking and heavy physical activity were associated with reduced odds of obesity and a lower BMI.

3.4.6 Relationships between early life exposures and alternative measures of adiposity

The relationships between risk factors and BF% and WHR were also considered. These were investigated separately by sex.

3.4.6.1 Body fat percentage analysis

There were differences in associations between risk factors and BF% in males and females (Appendix B,Table IV). In bivariate models for men, social class at birth, income and smoking were all significantly associated with decreased body fat, whilst high social class at 50, being educated and being married were associated with higher body fat (Appendix B, Table IV).

In females, there were significant, positive associations for duration breastfed and bacterial infection in the bivariate models, whilst moderate and heavy physical activity was associated with decreased BF% (Table 3.10).

In the multivariable model for males, social class at birth and smoking were significant predictors of lower BF% (Table 3.10). Father's social class (birth) remained a significant predictor after adjustment for the study members own social class in middle age. The coefficients and strengths of the associations increased for social class at 50 and smoking after controlling for social class at birth. Income was no longer significant after controlling for occupational social class (age 50). Whilst higher social class at birth was associated with decreases in body fat, social class at time of measurement was associated with increased body fat (p=0.06), although the effects of SES at birth were more significant (p<0.001).

In the multivariable model for females, bacterial infections were associated with increased body fat and heavy physical activity with decreased body fat in women (Table 3.10). These relationships were not influenced by social class and there was strong support for the model without SES at birth.

Table 3.10 Multivariable linear regression model for associations between early and later life factors and BF% in NTFS males and females (age 50).

BF% Males	**Coef**	**CI**	**p**	**BF% Females**	**Coef**	**CI**	**p**
SES (birth)	Ref		.	Bacterial infection	4.33	[0.36,8.30]	**0.03**
Mid	-3.74	[-6.17,-1.31]	**<0.001**	Breastfed (weeks)	0.07	[-0.02,0.15]	0.13
Most advantaged	-5.88	[-9.39,-2.37]	**<0.001**	PAL			
SES (age 50)	Ref		.	Inactive	Ref		.
Mid	1.49	[-2.25,5.22]	0.43	Light activity	-3.2	[-7.71,1.30]	0.16
Most advantaged	3.60	[-0.21,7.41]	0.06	Moderate activity	-3.49	[-8.46,1.47]	0.17
Smoker	-4.02	[-6.54,-1.49]	**<0.001**	Heavy activity	-6.31	[-11.82,-0.79]	**0.03**
n	161			n	155		
Adjusted R^2	0.12			Adjusted R^2	0.06		

Reference category for SES was least advantaged. Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p); Ref, reference group. Bold indicates significant at p<0.05.

3.4.6.2 Waist-to-hip ratio analysis

Associations between risk factors and WHR in males and females are presented in Table V (Appendix B). There were differences in predictors of WHR by sex, with relatively few

significant risk factors for women. Similar to BF%, many socioeconomic variables were associated with WHR in males and some were also in females (Appendix B, Table V).

Increasing social advantage at birth was associated with a lower WHR in males (b coefficient=-0.05, p=0.002). Birthweight was less significant after controlling for SES (age 50). Later life socioeconomic variables including high occupational social class at 50, income and university education were also associated with a decreased WHR. However, these factors were no longer significant after adjusting for social class at birth. Similar to results for BF%, father's social class (birth) remained a significant predictor after adjustment for the study members own SES. In the adjusted model, higher social class at birth was the only significant predictor of decreased WHR in men and explained around 7% of the variation (adjusted R2=0.07).

There were no early life and few later life exposures associated with WHR in women. In the bivariate analysis, there was a significant association for a decreased WHR for those in the mid social class group at 50 compared to low, and for household income (Table V). After adjusting for smoking, household income was not significant. There was a small, positive association between smoking and WHR. In the final model, the only significant predictor of WHR in women was mid-social class at age 50, however this explained little of the variation in WHR (adjusted R^2=0.01) (Table 3.11).

Table 3.11 WHR multivariable regression models for NTFS males (*n*=162) and females (*n*=210)

WHR Males	**Coef**	**CI**	**p**	**WHR Females**	**Coef**	**CI**	**p**
Social class (birth)	Ref		.	Social class (birth)	0	Ref	.
Mid	-0.02	[-0.04,0.00]	0.053	Mid	-0.01	[-0.03,0.01]	0.36
Most advantaged	-0.05	[-0.07,-0.02]	**0.002**	Most advantaged	-0.02	[-0.05,0.02]	0.30
Social class (age 50)	Ref		.	Social class (age 50)	Ref		.
Mid	-0.01	[-0.04,0.02]	0.342	Mid	-0.03	[-0.05,-0.00]	**0.04**
Most advantaged	-0.01	[-0.04,0.02]	0.374	Most advantaged	-0.01	[-0.03,0.01]	0.45
Birthweight (z-score)	-0.01	[-0.02,0.00]	0.081				

Reference category for SES was least advantaged. Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p); Ref, reference group. Bold indicates significant at p<0.05. All covariates are presented.

3.4.6.3 Summary of the associations across different outcome measures

For BMI and obesity associations in adjusted models were similar and were significant for bacterial infection, later life SES, physical activity and smoking, with an additional association for older maternal age for BMI (Table 3.12).

There were differences in the significant associations for BF% and WHR between men and women (Table 3.12). There was similarity in that later life occupational social class was associated with WHR in both sexes. Early life SES was associated with both WHR and BF% in males.

Associations that were consistent across 3 or more outcome measures were: bacterial infection, later life SES, physical activity and smoking.

Table 3.12 Summary of results across adiposity outcomes from adjusted regression models in NTFS study members

			Outcome measures			
			BMI	Obesity	Body fat %	WHR
Exposures	Individual	Sex	-	-		
	Maternal age	Older mother	✓	-	-	-
	Birth	Birthweight	-	-	-	✓(M)
		Occupational social class	-	-	✓(M)	✓(M)
		Breastfeeding	-	-	✓(F)	-
		Housing	-	-	-	-
	Childhood	Bacterial infection	✓	✓	✓(F)	-
		Adversity	-	-	-	-
		Rapid weight gain	-	-	-	-
	Adulthood	Education	-	-	-	-
		Occupational social class	✓	✓	✓(M)	✓(M, F)
		Income	-	-	-	-
		Married	-	-	-	-
		Physical activity	✓	✓	✓(F)	-
		Smoking	✓	✓	✓(M)	-

Tick denotes significant association in bivariate regression. M; males, F; females.

3.4.7 Pathways between early and later life factors and BMI

The path models (Figure 3.3) demonstrate the relationships between each of the risk factors and BMI, as well as the relationships between the factors themselves.

In the path model, breastfeeding was associated with increased BMI. The model also shows that a longer duration breastfeeding was associated with reduced likelihood of bacterial infections. Those who were educated and more advantaged were less likely to be smokers. Occupational social class in middle age had a similar direct effect on BMI to father's occupational social class at birth.

Infection in the first year had the largest positive direct and total effect. The socioeconomic variables, education and occupational social class at 50, had direct effects on BMI until the lifestyle variables were considered.

Lifestyle factors explained around half of the variation in BMI; when included in the path model the variation explained increases almost two-fold due to the indirect effects. Lifestyle factors (including PAL and smoking) were endogenous and predicted by socioeconomic variables. The effects of later life social class were mediated through lifestyle variables once they were included in the model, rather than being a direct predictor. The model explained 13% of the variation in BMI, a lower value than the multivariable regression model, which may be due the use of binary variables.

The path models indicate that early life and socioeconomic factors explain relatively small differences in BMI (7%), but similar in proportion to the lifestyle factors.

In multivariable models, education was not directly associated with outcomes, although it predicted other variables in the path model (see Figure 3.3).

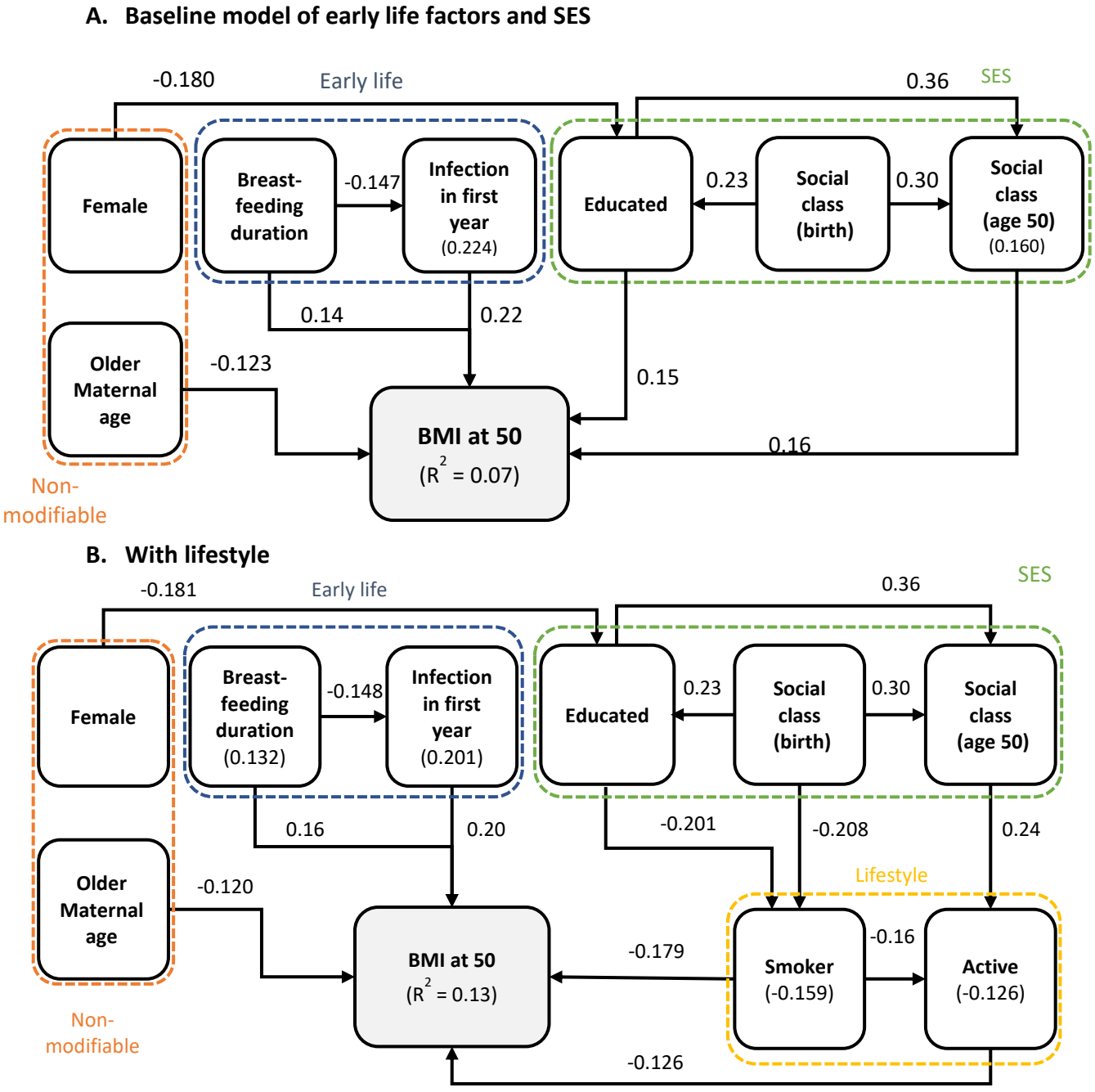


Figure 3.3 Path models of the relationship between early and later life variables, and BMI (age 50).
Models are presented without (A) or with (B) adjustments for lifestyle. Arrows show the direction of the effect. Standardised coefficients are shown. All direct effects are represented by solid lines and are significant at p<0.05. Total effects are presented in brackets with associated p values.

3.5 Discussion

In summary, exposures that were consistently associated across different outcome measures were; bacterial infection, occupational social class (at birth and middle age), smoking and physical activity. There was some evidence for maternal age, birthweight and breastfeeding, although less consistent.

Although prevalence of obesity was similar between the sexes there were more men who were overweight, therefore men on average had a higher BMI compared to females. This is consistent with findings for this age group by PHE (Health and Social Care Information Centre, 2014). Research suggests this could be due to the fact that men are less likely to recognise themselves as overweight or be dissatisfied with their weight (Sullivan and Brown, 2013, Tsai et al., 2016). Alternatively, the value or awareness of the social ideals related to thinness may explain the lower rates of overweight in women (McLaren and Kuh, 2004, McLaren, 2007a).

BMI correlated moderately with body fat and WHR, however different exposures were associated with different adiposity outcomes. BMI had the most significant explanatory variables and similar associations were observed when examining obesity. There were comparatively fewer exposures associated with measures of WHR and BF%, however SES emerged a consistent factor. There was a socioeconomic impact on adiposity that spanned across outcomes and varied by sex. SES at birth was not significantly associated with BMI or obesity in middle age, but was an important predictor of other later life factors (such as smoking, education and mid-life SES). Early life SES was an independent predictor of WHR in males, whereas SES in later life was a predictor of WHR in females. Similar to WHR, SES at birth also predicted body fat in males, as did SES in middle age.

Those who had a bacterial infection in the first year of life were more likely to be obese and have a higher BMI or higher body fat (women only) at age 50. This association was independent of SES, amongst other potential confounding factors. Taking into account the era, infection in the first year could encompass some aspects of SES that occupational social class does not, such as adverse living conditions that increase susceptibility to infection.

There was a weak association between birthweight and WHR in males. The standardised birthweight is similar to that of the National survey of Health and Development (NSHD), a larger UK birth cohort that commenced in 1946, thereby allowing comparisons to be made.

Similar findings from the NSHD support a small inverse association between birthweight and WC (Kuh et al., 2002) and higher lean mass (Bann et al., 2014). However, there was no associations for birthweight categories and adiposity outcomes. In NTFS, only 3% of study members were born LBW, which is dissimilar to current data from the office of national statistics which shows rates of low birthweight at ~7% in 2015 (ONS, 2015). Similarly, whilst figures for HBW were only 2% in NTFS, UK averages were around 11% in 2015. Therefore, perhaps the relatively low proportion of high and low birthweight could explain the lack of association. Furthermore, children born preterm or very low birthweight could possibly have lower odds of survival therefore reducing the proportion of children born LBW.

In contrast to other studies, older maternal age was associated with a lower adult BMI and breastfeeding with a slightly higher BMI, which may be a peculiarity of the historic cohort (discussed further in section 5.5).

Similar to other studies (Reiner et al., 2013), there was an inverse relationship between physical activity and obesity. Current smokers in NTFS had a lower BMI, which is in agreement with the literature (Eisen et al., 1993), however some studies find that heavy smoking is associated with increased weight (Chiolero et al., 2008). Additionally, NTFS men who were current or ex-smokers had larger WHR. This has also been noted in other studies (Canoy et al., 2005, Chiolero et al., 2008), and although similar findings are often observed for women, a study of older Dutch women also observed similar sex differences (Visser et al., 1999b). NTFS women had relatively healthy WHR compared to the men, which may explain these differences.

Thus far, there are indications that some factors in the peri- and post-natal period could affect obesity development. There were obvious differences in males and females in relation to body composition and also by exposure, which could suggest that sex-specific strategies are required for tackling overweight and obesity. The important exposures with respect to adult body composition were SES at birth, older maternal age, and bacterial infection in the first year, the latter of which requires further investigation. The early life exposures and lifestyle factors explained relatively a low percent of the variation in adult body composition, suggesting that other factors are contributing.

Chapter 4. The influence of early life exposures on childhood body composition

4.1 Introduction

This chapter will address the thesis aim of investigating the impact of early life exposures and SES on multiple indicators of childhood adiposity. The analysis will use data from the Gateshead Millennium study (GMS), a birth cohort from Gateshead that commenced in the year 2000. This cohort contains a wealth of early life data, and anthropometric data collected throughout childhood. There have already been a great deal of studies which examine multiple risk factors for childhood obesity from multiple cohorts (Fairley et al., 2015a)(reviewed in section 1.3). As of yet, the early life factors associated with childhood OWOB in children in North East England have not been determined. The North East is the region with the highest proportion of children starting school with obesity in the UK (NHS Digital, 2017), and therefore it is important to understand the factors driving this early onset of obesity, and potentially if these factors are different to other parts of the UK.

The previous chapter examined risk factors across different measures of adiposity (BMI, overweight and obesity, central obesity and body fat) in NTFS adults. Therefore, in keeping with previous analyses, this chapter will examine if early life risk factors are associated across similar components of body composition in GMS children.

In the previous chapter, it was determined that lifestyle factors were important intermediate factors between SES and body composition. Therefore, in line with the previous analysis, this chapter will also examine whether lifestyle (physical activity) can modify the effect of an exposure that occurs during a critical developmental period. Lifestyle may act in a synergistic, antagonistic, clustered or independent manner (Jacob et al., 2015).

4.2 Aims

The aims of this chapter were to i.) Determine if early life exposures were associated with childhood overweight or obesity in the GMS cohort; ii.) Examine if risk factors vary for outcomes i.e. using proxy measures or direct measures of body composition and distribution; and iii.) Examine the relationship between early life risk factors and childhood body composition considering the impact of lifestyle (physical activity).

4.3 Participants and methods

This analysis uses early life data from birth to 13 months, and outcome data from the age 6-8 follow-up. The current cohort consists of all traceable study members who did not withdraw. At the follow-up between ages 6-8, anthropometric measures, physical activity, and food intake were measured.

4.3.1 Anthropometric variables

BMIz, waist-to-height ratio (WHtR) and fat mass index (FMI) at age 6-8 were analysed as outcomes, calculation of these is in section 2.2.

4.3.2 Early life exposure data

There were data on several early life risk factors in GMS including; birthweight, rapid weight gain (RWG), first-born, maternal age, adversity, infection and SES, and the covariates sex, gestation and physical activity (moderate–vigorous intensity physical activity (MVPA)). The definitions and measurement of these factors is outlined in section 2.2.

Maternal education was chosen as the main SES indicator at birth as it represents economic resources and social characteristics related to knowledge and health literacy (Galobardes et al., 2006), and has been shown to have the strongest influence on pregnancy outcomes (Mortensen et al., 2008) (Parker et al., 1994).

4.3.3 Statistical analysis

SES, sex, birthweight categories, maternal age, first-born, adverse events, sleep issues, infection, RWG, and breastfeeding were categorical variables. All other variables were analysed as continuous or binary variables.

It was not anticipated that there would be differences between genders in the relationships between early life risk factors and obesity in children (no strong correlations between sex and early life factors, Appendix B). However, to check this assumption, sex-exposure interactions were examined for each outcome, as outlined in section 2.3.2. There were no significant differences, and therefore the sample was not stratified by sex.

Additionally, the impact of childhood physical activity was investigated, by adding MVPA to the multivariable models (adjusted for season) and evaluating the effects on the coefficients.

4.4 Results

4.4.1 Sample representativeness

There were data for 619 study members included at the age 6 to 8 follow-up, representing 60% of the cohort. Study participants included in these analyses were comparable to the original cohort for most early life exposures (Table 4.1 and Table 4.2). There were differences in that the sample were more advantaged, in terms of Townsend score, higher achieved maternal education and less material deprivation (Table 4.2), leading to more even distribution across socioeconomic strata than the original sample (Parkinson et al., 2011). Those measured at age 6-8 had a significant longer gestational age, had slight older mothers, more experienced adversity, were breastfed for longer and were less likely to be formula fed (*Table 4.1* and *Table 4.2*). These factors were likely related to the socioeconomic differences in the 6-8 sample (Appendix C, Table VII).

Table 4.1 Differences in continuous and ordinal early life variables for GMS children with and without body composition data at age 6-8.

Variable	**Without data**					**With data**					**P value**
	n	Mean	SD	min	max	n	Mean	SD	min	max	
Birthweight (z-score)	424	-0.03	0.99	-2.74	4.52	569	-0.02	1.04	-3.48	3.51	0.95 [a]
Gestation length (weeks)	424	39.35	1.88	27	43	569	39.57	1.55	29	43	**0.040** [a]
Maternal age (years)	424	27.2	6.15	15.27	44.41	569	28.46	5.79	16.09	45.75	**<0.001** [a]
Age weaned (weeks)	269	14.49	3.42	3	33	479	14.59	3.2	5	42.5	0.67 [a]

P values derived from tests for differences in means (T test [a]), median values (Rank-sum [b]) n, sample size; SD, standard deviation; min, minimum; max, maximum. Bold indicates significant at p<0.05.

Table 4.2 Differences in categorical variables for GMS children with and without body composition data at age 6-8.

Early life variables	All		Without data		With data		P value
	n	%	n	col %	n	col %	
Sex	993		424		569		
Male	505	50.9	223	52.6	282	49.6	0.34
Female	488	49.1	201	47.4	287	50.4	
First-born	991		423		568		
No	528	53.3	227	53.7	301	53	
Yes	463	46.7	196	46.3	267	47	0.83
Gestation categories	993		424		569		
Pre-term	107	10.8	56	13.2	51	9	
Normal	823	82.9	338	79.7	485	85.2	0.061
Post-term	63	6.3	30	7.1	33	5.8	
Birthweight	993		424		569		
SGA	89	9	36	8.5	53	9.3	
Normal	810	81.6	351	82.8	459	80.7	0.69
LGA	94	9.4	37	8.7	57	10	
Caesarean	993		424		569		
No	844	85	365	86.1	479	84.2	0.41
Yes	149	15	59	13.9	90	15.8	
Maternal age at birth	993		424		569		
Less than 25	326	32.8	164	38.7	162	28.5	
25-34	547	55.1	215	50.7	332	58.3	**0.003**
35+	120	12.1	45	10.6	75	13.2	
Breastfeeding	953		405		548		
Never	468	49.1	232	57.3	236	43.1	**<0.001**
<6wk	237	24.9	101	24.9	136	24.8	
>6wk	89	9.3	26	6.4	63	11.5	
>4m	159	16.7	46	11.4	113	20.6	
Exclusively breastfed (>4 months)	249		262		511		
No	468	91.6	232	93.2	236	90.1	0.21
Yes	43	8.4	17	6.8	26	9.9	
Formula fed only	953		405		548		
No	485	50.9	173	42.7	312	56.9	
Yes	468	49.1	232	57.3	236	43.1	**<0.001**
Rapid weight gain	813		282		531		
No	567	69.7	205	72.7	362	68.2	0.18
Yes	246	30.3	77	27.3	169	31.8	
Adversity	934		398		536		
No	719	77	327	82.2	392	73.1	**0.001**
Yes	215	23	71	17.8	144	26.9	
Sleep issues (8 months)	644		201		443		
No	532	82.6	162	80.6	370	83.5	0.36
Yes	112	17.4	39	19.4	73	16.5	
Infection (0-12 month)	994		425		569		
No	895	90	394	92.7	501	88	**0.015**
Yes	99	10	31	7.3	68	12	

Socioeconomic variables	All		Without data		With data		P
	n	%	n	col %	n	col %	
Townsend quintile	987		424		563		
1 Least advantaged	188	19.0	84	19.8	104	18.5	
2	201	20.4	81	19.1	120	21.3	
3	221	22.4	94	22.2	127	22.6	**0.021**
4	223	22.6	113	26.7	110	19.5	
5 Most advantaged	154	15.6	52	12.3	102	18.1	
Maternal education (birth)	915		382		533		
None	142	15.5	78	20.4	64	12.0	
GCSE	534	58.4	227	59.4	307	57.6	**<0.001**
A level	111	12.1	41	10.7	70	13.1	
Degree	128	14.0	36	9.4	92	17.3	
Deprived at birth	991		422		569		
No	483	48.7	163	38.6	320	56.2	**<0.001**
Yes	508	51.3	259	61.4	249	43.8	
Occupational social class (childhood)	373		13		360		
Least advantaged	110	29.5	1	7.7	109	30.3	
Middle	133	35.7	6	46.2	127	35.3	0.21
Most advantaged	130	34.9	6	46.2	124	34.4	
Upward mobility (0-8 years)	994		425		569		
No	945	95.1	425	100	520	91.4	**<0.001**
Yes	49	4.9	0	0	49	8.6	

Row sample sizes and column percentages (col %) shown. P value represents the Chi-square test statistic. Bold indicates significant at $p<0.05$.

4.4.2 Exposures

Descriptive statistics for those measured at age 6-8, plus sex differences are presented in Table 4.3, Table 4.4 and Table 4.5. The mean age of the cohort was around 7.5 years. There was an even proportion of males and females. The majority had a normal birthweight, and there were 9.3% SGA and 10% LGA. In terms length of gestation, 85% of children were born normal term, 9% were pre-term and 5.8% post-term. In the sub-sample, 47% were first-born children, and mean maternal age was 29 years old. A large proportion of children were never breastfed (43.1%). Around a quarter of children experienced adversity in the first year and 16.5% of children had sleep issues in the first 8 months. There was a fairly even split between parental socioeconomic groups at age 6-8, with 34.4% in the most advantaged group, 35.3% mid and 30.3% in the least advantaged group.

Males were weaned slightly earlier than females, on average spent more time doing MVPA (Table 4.3), and were more likely to have an infection in the first year (Table 4.4). There were no other significant differences between males and females, however there were some differences in socioeconomic variables that were borderline significant (p<0.1) (Table 4.5).

Table 4.3 Summary statistics and differences in early life continuous variables stratified by sex for the GMS cohort (age 6-8)

Variable	**Male**					**Female**					
	n	mean	SD	min	max	n	mean	SD	min	max	p
Age (years)	282	7.44	0.46	6.42	8.42	287	7.46	0.45	6.42	8.58	0.53 [a]
Bwt (z-score)	282	-0.02	1.04	-2.74	3.45	287	-0.02	1.04	-3.48	3.51	0.98 [a]
Gestation (weeks)	282	39.53	1.69	29	42	287	39.61	1.4	34	43	0.41 [a]
Maternal age (years)	282	28.81	5.84	16.59	45.75	287	28.12	5.72	16.09	43.46	0.08 [a]
Weaned (weeks)	379	14.31	3.12	3	32	369	14.81	3.42	4	42.5	**0.04** [a]
MVPA (%)	241	4.71	2.61	0.58	15.03	238	4.08	2.19	0.34	13.88	**0.01** [b]

Number of observations (N), mean, standard deviation (SD), minimum (min) and maximum (max) values and P values (p) for t-tests or Wilcoxon rank-sum tests for differences presented. a T-test or b Ranksum tests for differences between males and females. Bold indicates significantly different at p<0.05.

Table 4.4 Summary statistics and differences in categorical early life variables stratified by sex, for the GMS cohort (age 6-8)

Binary variables	**All (n)**	**% Yes**	**Male (n)**	**% Yes**	**Female (n)**	**% Yes**	**p**
First-born	568	47	281	47	287	47	0.99
Caesarean	569	15.8	282	15.6	287	16	0.89
Exclusively breastfed (>4 months)	262	11.6	12	8.5	26	9.9	0.41
Formula fed	548	43.1	270	39.6	278	46.4	0.11
RWG (0-12 months)	531	31.8	263	31.6	268	32.1	0.90
Adversity (0-12 months)	536	26.9	267	29.6	269	24.2	0.15
Sleep issues (8 months)	443	16.5	222	17.1	221	15.8	0.72
Infection (0-12 months)	569	12	282	15.6	287	8.4	**0.008**
Categorical variables	All (n)	%	Male (n)	%	Female (n)	%	p
Gestation	569		282		287		0.086
Pre-term	51	9	31	11	20	7	
Normal	485	85.2	231	81.9	254	88.5	
Post-term	33	5.8	20	7.1	13	4.5	
Categories of birthweight	569		282		287		0.73
SGA	53	9.3	29	10.3	24	8.4	
Normal	459	80.7	225	79.8	234	81.5	
LGA	57	10	28	9.9	29	10.1	
Maternal age	569		282		287		0.40
Less than 25	162	28.5	81	28.7	81	28.2	
25-34	332	58.3	162	57.4	170	59.2	
35+	75	13.2	39	13.8	36	12.5	
Breastfeeding	548		270		278		0.41
Never	236	43.1	107	39.6	129	46.4	
<6wk	136	24.8	70	25.9	66	23.7	
>6wk	63	11.5	35	13	28	10.1	
>4m	113	20.6	58	21.5	55	19.8	
Season accelerometry	241		238		479		
Spring	56	23.2	56	23.5	112	23.4	
Summer	53	22	48	20.2	101	21.1	0.41
Autumn	83	34.4	71	29.8	154	32.2	
Winter	49	20.3	63	26.5	112	23.4	

Sample sizes (n) and column percentages (col %), and Chi-square test statistic presented (p). Bold indicates significantly different at p<0.05.

Table 4.5 6 Summary statistics and differences in socioeconomic categorical variables stratified by sex, for the GMS cohort (age 6-8)

	All (n)	%	Male (n)	%	Female (n)	%	p
Townsend score	278		285		563		0.079
1 Least advantaged	46	16.5	58	20.4	104	18.5	
2	54	19.4	66	23.2	120	21.3	
3	61	21.9	66	23.2	127	22.6	
4	54	19.4	56	19.6	110	19.5	
5 Most advantaged	63	22.7	39	13.7	102	18.1	
Maternal education	269		264		533		0.44
None	27	10	37	14	64	12	
GCSE	155	57.6	152	57.6	307	57.6	
A level	36	13.4	34	12.9	70	13.1	
Degree	51	19	41	15.5	92	17.3	
Childhood SES	177		183		360		0.19
Least advantaged	46	26	63	34.4	109	30.3	
Mid	64	36.2	63	34.4	127	35.3	
Most advantaged	67	37.9	57	31.1	124	34.4	
Upward mobility (0-8 years)	282		287		569		0.060
No	264	93.6	256	89.2	520	91.4	
Yes	18	6.4	31	10.8	49	8.6	

Sample sizes (n) and column percentages (col %), and Chi-square test statistic presented (p).

4.4.3 Infant feeding and SES

There were significant differences between Townsend quintiles and the duration of breastfeeding, in that the most advantaged breastfed for longer and were less likely to be formula fed (Appendix C, Table VII). There was no statistically significant differences in weaning age between the Townsend quintiles (Kruskal-Wallis p = 0.07).

4.4.4 Outcomes

According to BMIz, at age 6-8 there were 116 (21%) children who were OWOB, and of those 52 (9.3%) were obese (Table 4.7). Whilst for FMI, 5.2% of children had a fat mass index >=91st centile, and 1.1% had a FMI>=98th centile. Fewer were classified as obese using WHtR than using BMI (7%). There were no significant differences in outcomes between sexes. Mean BMIz (0.45) and FMI (4.06) were similar in males and females (Table 4.7).

Table 4.7 Descriptive characteristics of body composition outcomes stratified by sex in the GMS cohort (age 6-8)

Categorical outcomes	**All**		**Male**		**Female**		
	n	col %	n	col %	n	col %	P value
All	562		278		284		0.56
Healthy weight	446	79.4	218	78.4	228	80.3	
OWOB	116	20.6	60	21.6	56	19.7	
Waist OB	505		488		993		0.98
No	470	93.1	454	93	924	93.1	
Yes	35	6.9	34	7	69	6.9	
Continuous outcomes	**All**		**Male**		**Female**		
	n	mean (SD)	n	mean (SD)	n	mean (SD)	P value
BMIz	569	0.45 (1.11)	282	0.44 (1.15)	287	0.46 (1.08)	0.85
FMI	567	4.06 (1.92)	282	4.11 (1.84)	285	4.00 (2.00)	0.49

Proportion (%) of study members in each weight category with Chi^2 test statistic for differences between sexes. Mean and standard deviation (SD) for BMIz and FMI, with p value for differences between sexes (T-test).

There were strong correlations (r>0.7) between FMI with both BMIz and OWOB (Table 4.8). Whilst the weakest associations, although still moderate, were between FMI and BMIz and categorical OWOB (r<0.55). Waist OB showed the weakest correlations with the other measures, however again these were still modest (r>0.62). Overall, there was good correlation between the alternative measures of obesity.

Table 4.8 Correlations between the body composition measures in the GMS cohort (age 6-8)

	BMIz	**FMI**	**OWOB**	**Waist OB**
BMIz	1			
FMI	0.749	1		
OWOB	0.763	0.701	1	
Waist OB	0.622	0.664	0.667	1

Pearson correlation coefficients between outcome measures.
All correlations p<0.0001.

4.4.5 Relationship between early life exposures and childhood body composition

In bivariate analyses, a higher birthweight and adversity were both associated with a higher BMIz and FMI and increased likelihood of OWOB (Appendix C,

Table XI). RWG was associated with increased BMIz. MVPA was negatively associated with all outcomes.

The exposures consistently significant across outcomes in bivariate analyses were birthweight and adversity (

Figure 4.1). Weaning was the only feeding-related variable that demonstrated significant associations in the bivariate models, albeit only for BMIz. Rapid weight gain also demonstrated significant associations with BMIz.

The effects of these early life variables were investigated further in multivariable models adjusted for known confounders, including early and later life SES, and other early life exposures (Table 4.9).

Anticipated confounding variables (Table 1.5) of the relevant exposures were included in the models. Including breastfeeding duration in the models made little difference to the estimates, and there was strong support for the models without breastfeeding (indicated by BIC), therefore it was not included. Using categorical maternal age rather than continuous explained more of the variation in outcomes. The impact of physical activity on the early life variables was also investigated in separate adjusted models (Table 4.10).

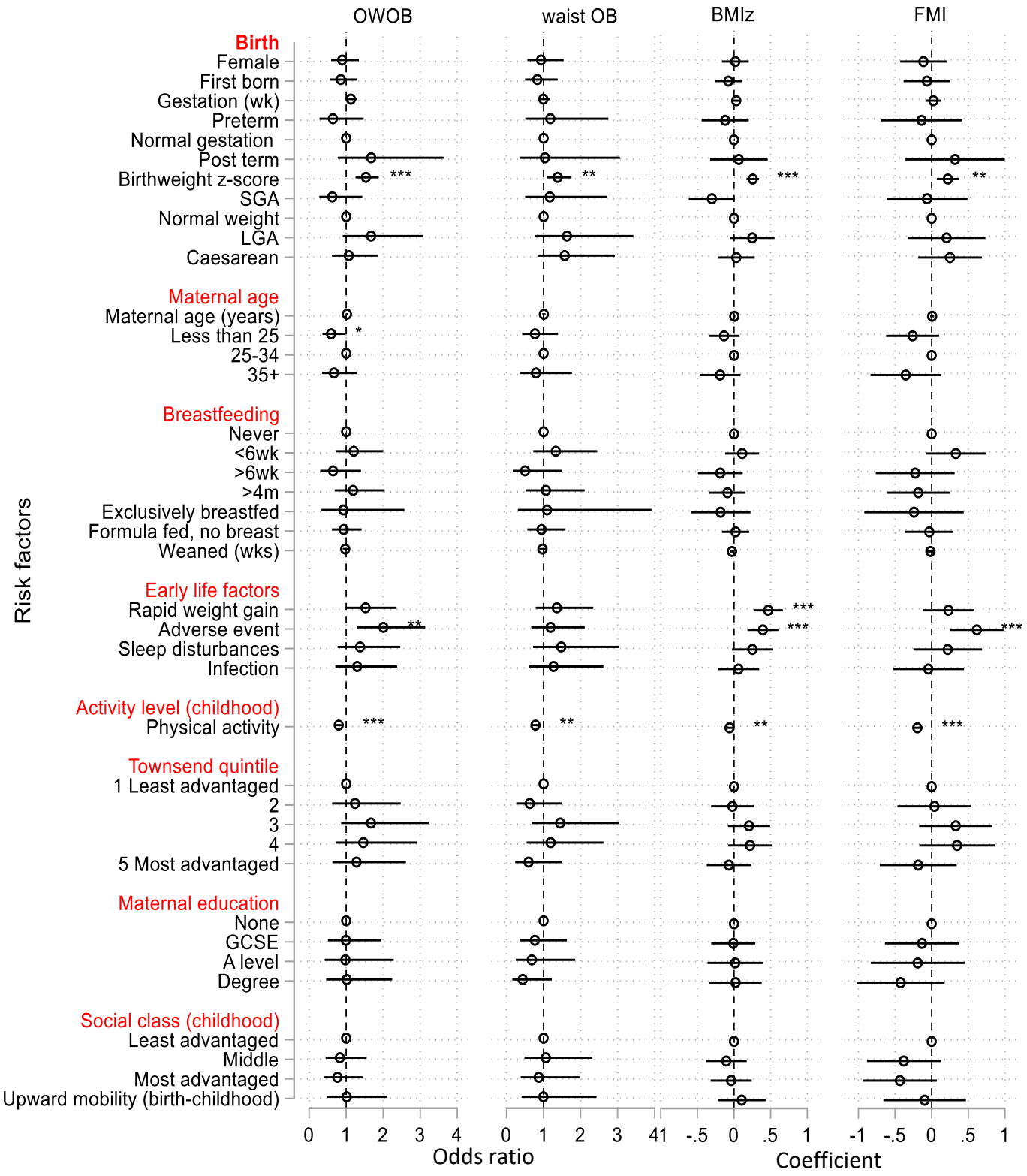


Figure 4.1 Forest plot of bivariate models for each outcome (x-axis) and exposure (y-axis) in GMS (age 6-8). Odds ratios presented for the outcomes OWOB and waist OB, and coefficients for BMIz and FMI, with respective 95% confidence intervals and the corresponding level of significance (indicates p<0.05, ** p<0.01, ***p<0.001). Physical activity is adjusted for season.*

Table 4.9 Multivariable regression models for early and childhood factors and all adiposity outcomes in GMS (age 6-8).

Variable	OWOB			BMIz			FMI			Waist OB		
	OR	CI	p	Coef	CI	p	Coef	CI	p	OR	CI	p
Female	0.85	[0.42,1.71]	0.65	-0.03	[-0.27,0.20]	0.78	-0.06	[-0.49,0.38]	0.80	1.07	[0.48,2.42]	0.87
Gestation length (wk)	1.30	[0.99,1.69]	0.055	0.04	[-0.04,0.11]	0.30	0.03	[-0.11,0.17]	0.66	0.92	[0.71,1.20]	0.54
First-born	1	[0.48,2.09]	0.99	-0.01	[-0.27,0.24]	0.94	0.01	[-0.46,0.48]	0.97	1.05	[0.45,2.42]	0.92
Maternal age												
Less than 25	0.27	[0.09,0.85]	**0.025**	-0.35	[-0.69,-0.00]	**0.047**	-0.65	[-1.28,-0.02]	**0.043**	0.29	[0.07,1.16]	0.080
25-34	Ref		.	Ref		.	Ref		.	Ref		.
35+	0.63	[0.22,1.81]	0.39	-0.09	[-0.43,0.25]	0.60	-0.45	[-1.07,0.17]	0.16	1.18	[0.39,3.59]	0.77
Adversity	2.59	[1.25,5.35]	**0.010**	0.31	[0.04,0.57]	**0.023**	0.64	[0.15,1.13]	**0.010**	1.06	[0.44,2.55]	0.90
Birthweight z-score	2.61	[1.73,3.94]	**<0.0001**	0.44	[0.31,0.57]	**<0.001**	0.43	[0.19,0.68]	**0.001**	1.69	[1.09,2.64]	**0.020**
Rapid weight gain	3.86	[1.69,8.82]	**0.001**	0.81	[0.52,1.10]	**<0.001**	0.71	[0.18,1.24]	**0.009**	2.07	[0.79,5.48]	0.14
Wean age	0.87	[0.75,1.03]	0.10	-0.06	[-0.11,-0.01]	**0.012**	-0.08	[-0.17,0.00]	0.061	0.82	[0.70,0.97]	**0.023**
Maternal education												
None	Ref		.	Ref		.	Ref		.	Ref		.
GCSE	1.47	[0.27,8.03]	0.66	0.4	[-0.11,0.92]	0.13	0.89	[-0.06,1.84]	0.066	2.61	[0.30,22.46]	0.38
A level	1.3	[0.20,8.44]	0.78	0.42	[-0.16,1.00]	0.16	0.75	[-0.33,1.82]	0.17	1.11	[0.10,12.96]	0.93
Degree	1.87	[0.30,11.82]	0.51	0.33	[-0.26,0.93]	0.27	0.39	[-0.70,1.49]	0.48	1.08	[0.10,11.82]	0.95
SES (childhood)												
Least advantaged	Ref		.	Ref		.	Ref		.	Ref		.
Mid	0.37	[0.14,0.97]	**0.042**	-0.23	[-0.54,0.09]	0.16	-0.65	[-1.24,-0.07]	**0.028**	0.54	[0.18,1.61]	0.27
Most advantaged	0.44	[0.17,1.20]	0.11	-0.09	[-0.43,0.26]	0.61	-0.53	[-1.16,0.11]	0.10	0.89	[0.30,2.63]	0.84
N	261			265			265			275		
pseudo R^2/adjusted R^2	0.207			0.219			0.104			0.111		

Coefficients (coef) or odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Ref indicates reference category for factor variables.
Bold indicates significant at p<0.05.

4.4.6 Associations across different outcome measures

There was consistency across measures in bivariate analyses (Appendix C,Table VIII). In bivariate analyses, birthweight z-score was positively associated with all outcome measures. MVPA was negatively associated with all outcomes. Whilst younger maternal age was associated with lower odds of OWOB, lower BMIz and FMI. Adversity was associated with increased odds of OWOB, increased BMIz and increased FMI (but not waist OB). The effects of adversity could be attributed to socioeconomic factors including debt and parental separation (Appendix C, Table IX).

However, there were no associations for any other early socioeconomic factors (Appendix C,Table VIII) for any of the weight outcomes. Although, being in the mid socioeconomic advantaged group (compared to the least advantaged) in childhood was associated with lower odds of OWOB and a lower BMIz. There was no association for breast or formula feeding with any of the outcome measures, however weaning age was negatively associated with BMIz and lower odds of waist OB. For BMIz only, there was a positive association ($p<0.1$) between sleep issues in the bivariate model, however this was no longer present after adjustment.

Overall, there was good agreement across OWOB, BMIz and FMI for the early life factors; younger maternal age, adversity, birthweight and rapid weight gain (Table 4.9). There were differences in associations for waist OB, whereby there were only significant associations for birthweight (increased odds) and weaning (decreased odds).

4.4.6.1 The influence of physical activity

MVPA was negatively associated with all outcomes, with the most significant association being for FMI (Table 4.10). For BMIz, adjusting for MVPA did not lead to great changes in most significant coefficients. However, it removed the significant association for younger maternal age. Adjusting for MVPA removed any significant associations for SES. After adjustment, the coefficients for birthweight decreased across all outcomes, and birthweight was no longer significant associated with waist OB (Table 4.10). The association between weaning and lower odds of waist OB was no longer significant after controlling for MVPA. After adjustment for MVPA, the odds ratio notably increased for RWG and OWOB (OR=3.8 in model without MVPA, compared to OR=4.8 in adjusted model).

Table 4.10 Multivariable regression models with and without adjustment for MVPA for outcomes in GMS (age 6-8).

		OWOB			OWOB adjusted for MVPA			BMIz			BMIz adjusted for MVPA		
Exposures		OR	CI	p	OR	CI	p	coef	CI	p	coef	CI	p
Maternal age	Less than 25	0.27	[0.09,0.85]	**0.025**	0.16	[0.04,0.66]	**0.011**	-0.35	[-0.69,-0.00]	**0.047**	-0.32	[-0.67,0.03]	0.07
	35+	0.63	[0.22,1.81]	0.39	0.45	[0.12,1.68]	0.24	-0.09	[-0.43,0.25]	0.60	-0.16	[-0.49,0.18]	0.35
Adversity		2.59	[1.25,5.35]	**0.01**	3.08	[1.32,7.18]	**0.009**	0.31	[0.04,0.57]	**0.023**	0.28	[0.02,0.54]	**0.034**
Birthweight z-score		2.61	[1.73,3.94]	**<0.001**	2.19	[1.38,3.49]	**0.001**	0.44	[0.31,0.57]	**<0.001**	0.36	[0.23,0.50]	**<0.001**
Wean age		0.87	[0.75,1.03]	0.1	0.90	[0.76,1.07]	0.23	-0.06	[-0.11,-0.01]	**0.012**	-0.06	[-0.11,-0.02]	**0.009**
Rapid weight gain		3.86	[1.69,8.82]	**0.001**	4.82	[1.78,13.09]	**0.002**	0.81	[0.52,1.10]	**<0.001**	0.82	[0.53,1.11]	**<0.001**
Maternal education	GCSE	1.47	[0.27,8.03]	0.66	1.21	[0.16,9.46]	0.85	0.4	[-0.11,0.92]	0.13	0.46	[-0.07,0.98]	0.09
	A level	1.3	[0.20,8.44]	0.78	1.04	[0.11,9.50]	0.97	0.42	[-0.16,1.00]	0.16	0.52	[-0.06,1.11]	0.08
	Degree	1.87	[0.30,11.82]	0.51	1.87	[0.22,15.53]	0.56	0.33	[-0.26,0.93]	0.27	0.44	[-0.16,1.04]	0.15
SES (childhood)	Mid	0.37	[0.14,0.97]	**0.042**	0.46	[0.15,1.39]	0.17	-0.23	[-0.54,0.09]	0.16	-0.11	[-0.42,0.21]	0.50
	Most advantaged	0.44	[0.17,1.20]	0.11	0.51	[0.16,1.66]	0.27	-0.09	[-0.43,0.26]	0.61	0.07	[-0.28,0.41]	0.71
MVPA		-			0.72	[0.58,0.90]	**0.003**	-			-0.07	[-0.12,-0.02]	**0.007**
N		261			240			0.219			0.247		
R^2/pseudo R^2		0.207			0.293			0.260			0.303		
		FMI			**FMI adjusted for MVPA**			**Waist OB**			**Waist OB adjusted for MVPA**		
Exposures		coef	CI	p	coef	CI	p	OR	CI	p	OR	CI	p
Maternal age	Less than 25	-0.65	[-1.28,-0.02]	**0.043**	-0.67	[-1.30,-0.04]	**0.037**	0.29	[0.07,1.16]	0.08	0.3	[0.06,1.61]	0.16
	35+	-0.45	[-1.07,0.17]	0.16	-0.49	[-1.10,0.11]	0.11	1.18	[0.39,3.59]	0.77	1.37	[0.37,5.13]	0.64
Adversity		0.64	[0.15,1.13]	**0.01**	0.63	[0.16,1.10]	**0.009**	1.06	[0.44,2.55]	0.90	1.04	[0.38,2.89]	0.93
Birthweight z-score		0.43	[0.19,0.68]	**0.001**	0.26	[0.02,0.50]	**0.033**	1.69	[1.09,2.64]	**0.020**	1.39	[0.85,2.27]	0.20
Wean age		-0.08	[-0.17,0.00]	0.061	-0.08	[-0.16,0.01]	0.077	0.82	[0.70,0.97]	**0.023**	0.86	[0.72,1.02]	0.084
Rapid weight gain		0.71	[0.18,1.24]	**0.009**	0.69	[0.17,1.21]	**0.009**	2.07	[0.79,5.48]	0.14	2.73	[0.89,8.35]	0.078
Maternal education	GCSE	0.89	[-0.06,1.84]	0.066	0.95	[0.00,1.90]	**0.049**	2.61	[0.30,22.46]	0.38	1.42	[0.15,13.18]	0.76
	A level	0.75	[-0.33,1.82]	0.17	0.94	[-0.11,2.00]	0.079	1.11	[0.10,12.96]	0.93	0.73	[0.06,9.68]	0.81
	Degree	0.39	[-0.70,1.49]	0.48	0.55	[-0.53,1.63]	0.32	1.08	[0.10,11.82]	0.95	0.79	[0.07,8.99]	0.85
SES (childhood)	Mid	-0.65	[-1.24,-0.07]	**0.028**	-0.54	[-1.11,0.03]	0.062	0.54	[0.18,1.61]	0.27	0.51	[0.14,1.82]	0.30
	Most advantaged	-0.53	[-1.16,0.11]	0.10	-0.44	[-1.07,0.18]	0.17	0.89	[0.30,2.63]	0.84	0.81	[0.23,2.95]	0.76
MVPA		-			-0.23	[-0.32,-0.13]	**<0.001**	-			0.70	[0.53,0.93]	**0.013**
N		265			243			275			246		
R^2/pseudo R^2		0.104			0.189			0.111			0.167		

Models were additionally adjusted for covariates; sex, gestation and first-born, and MVPA models also for season. Coefficients (coef) or odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Reference categories for factor variables were; least advantaged for SES; no maternal qualifications for maternal education; and maternal age 25-34. Bold indicates significant at p<0.05.

4.4.7 Sensitivity analysis for childhood FMI

The model for FMI did not fit regression diagnostic criteria. There was a slightly positive skew of the residuals and the normal probability plot demonstrated deviation from normality towards the tails (Appendix C, Figure XI). The non-normal residuals suggested that this model does not explain all trends in the dataset.

Model building was redone to find the best model for FMI (*Table 4.11*). Despite being confounding variables, there was support (BIC) for the model to not include the socioeconomic variables; maternal education, Townsend and occupation social class in childhood, or the feeding variables (Table 4.11).

The skewed residuals could be due to the high values of FMI observed leading to positive skew in the values. Taking the definition of an outlier as greater than the upper quartile plus 1.5 times the inter-quartile range (IQR) (>Q3 + 1.5*IQR), excluding outliers removed from the sample those with a FMI greater than 8.19 (n=16 in adjusted model). This appeared to correct the skewed residuals (Appendix C, Figure XI) which appeared more normally distributed, although statistically were not (Shapiro-Wilk $p<0.005$). Further investigation into which data points deviated from the mean (high leverage points using Cook's distance), found no obvious, identifiable pattern. Therefore, the variables (Table 4.11) are a better model fit for those with a FMI within the middle of distribution for FMI, but not for those with high FMI.

Table 4.11 Multivariable models for the FMI model and the model excluding the outliers in GMS (age 6-8).

	FMI			FMI without outliers		
Exposures	coef	CI	p	coef	CI	p
Female	-0.14	[-0.46,0.19]	0.42	-0.15	[-0.41,0.10]	0.24
Gestation (weeks)	0.04	[-0.06,0.15]	0.44	0	[-0.08,0.08]	0.94
First-born	0.02	[-0.33,0.37]	0.90	0	[-0.28,0.27]	0.99
Maternal age (years)						
Less than 25	-0.25	[-0.66,0.16]	0.23	-0.22	[-0.54,0.10]	0.18
25-34	Ref		.	Ref		.
35+	-0.38	[-0.86,0.10]	0.12	-0.59	[-0.97,-0.20]	**0.003**
Adversity	0.67	[0.30,1.04]	**<0.001**	0.51	[0.22,0.80]	**0.001**
Birthweight (z-score)	0.36	[0.19,0.53]	**<0.001**	0.16	[0.03,0.30]	**0.020**
RWG	0.56	[0.19,0.94]	**0.004**	0.30	[-0.00,0.59]	0.051
N	499			473		
Adjusted R^2	0.058			0.048		

Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Ref indicates reference category for factor variables. Bold indicates significant at p<0.05. All covariates are presented.

The study members with the greatest fat mass also had very high BMI and waist OB, therefore these data are probably valid measurements, and it would be incorrect to simply exclude those with high FMI. The models fit other regression diagnostic criteria (i.e. homoscedastic, no omitted variables, no collinearity) except for normality of residuals. This was due to skewed residuals, which were prone to non-normality at the tails (Appendix C, Figure XI). Therefore, to obtain reliable estimates without excluding data points, robust regression was used.

The robust regression included the same variables as the basic model (model 1). The coefficients for birthweight, RWG and adversity were smaller and less significant when using robust regression compared to linear regression (Table 4.12). In addition, the coefficient for older maternal age was significant in the robust regression model. Other coefficients remained similar. Therefore, linear regression may lead to an overestimation of the coefficients and significance (for birthweight, RWG and adversity).

Table 4.12 Comparison of the FMI adjusted linear regression model and the robust regression model in GMS (age 6-8).

FMI	**Model 1 - Linear**			**Model 1 - Robust**		
	coef	CI	p	coef	CI	p
Female	-0.14	[-0.46,0.19]	0.42	-0.23	[-0.50,0.04]	0.10
Gestation length (wk)	0.04	[-0.06,0.15]	0.44	-0.02	[-0.10,0.07]	0.69
First-born	0.02	[-0.33,0.37]	0.90	0.0	[-0.29,0.29]	0.98
Adversity	0.67	[0.30,1.04]	**<0.001**	0.50	[0.20,0.81]	**0.001**
Birthweight z-score	0.36	[0.19,0.53]	**<0.001**	0.17	[0.02,0.31]	**0.021**
Rapid weight gain	0.56	[0.19,0.94]	**0.004**	0.34	[0.02,0.65]	**0.037**
Maternal age Less than 25	-0.25	[-0.66,0.16]	0.23	-0.20	[-0.54,0.14]	0.25
25-34	Ref		.	Ref		.
35+	-0.38	[-0.86,0.10]	0.12	-0.60	[-1.00,-0.20]	**0.004**
N	499			499		
Adjusted R^2	0.058			0.045		

Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Ref indicates reference category for factor variables. Bold indicates significant at p<0.05. All covariates are presented.

4.4.8 Sensitivity analysis for demographic factors

A small proportion of the age 6-8 sample with BMIz measures were ultra-Orthodox Jewish (*n*=9, 1.6%). As there are established differences in feeding practices within this group, any models that included feeding variables (BMIz, OWOB, waist OB) were re-run excluding ultra-Orthodox Jewish (UOJ) study members. There were no significant differences in prevalence of OWOB or waist OB, or differences in the mean BMIz or FMI between this group and the remainder of the sample, however this may be due to the small sample size.

Despite the low proportion, repeating analyses without ultra-Orthodox Jewish study members did have an impact on the regression coefficients and odds ratios. For the OWOB model, the odds ratios for younger maternal age and mid childhood SES decreased after excluding UOJ (Table 4.13). The model indicated that for OWOB, coefficients were underestimated, and therefore for the majority of the cohort there were larger associations for adversity, RWG and birthweight with childhood adiposity. Results remained similar for BMIz and waist OB.

Although sample sizes were too small to investigate the UOJ group separately, these results could suggest that early life risk factors are not the same for this group or have a smaller magnitude.

Table 4.13 Multivariable regression models for UOJ sensitivity analysis for all outcomes in GMS (age 6-8).

		BMI			without UOJ		
Variable		**coef**	**CI**	**p**	**coef**	**CI**	**p**
Maternal age	Less than 25	-0.35	[-0.69,-0.00]	**0.047**	-0.40	[-0.76,-0.03]	**0.032**
	25-34	Ref		.	Ref		.
	35+	-0.09	[-0.43,0.25]	0.60	-0.07	[-0.42,0.28]	0.69
Adversity		0.31	[0.04,0.57]	**0.023**	0.31	[0.04,0.58]	**0.027**
Birthweight z-score		0.44	[0.31,0.57]	**<0.001**	0.45	[0.32,0.59]	**<0.001**
Rapid weight gain		0.81	[0.52,1.10]	**<0.001**	0.82	[0.51,1.12]	**<0.001**
Wean age		-0.06	[-0.11,-0.01]	**0.012**	-0.06	[-0.11,-0.00]	**0.036**
Maternal education		Ref		.	Ref		.
	GCSE	0.40	[-0.11,0.92]	0.13	0.48	[-0.07,1.03]	0.088
	A level	0.42	[-0.16,1.00]	0.16	0.49	[-0.12,1.11]	0.12
	Degree	0.33	[-0.26,0.93]	0.2	0.39	[-0.24,1.01]	0.23
SES (childhood)		Ref		.	Ref		.
	Mid	-0.23	[-0.54,0.09]	0.16	-0.24	[-0.57,0.09]	0.15
	Most advantaged	-0.09	[-0.43,0.26]	0.61	-0.09	[-0.45,0.27]	0.63
n		265			254		
R^2		0.219			0.206		

		OWOB			without UOJ		
Variable		**OR**	**CI**	**p**	**OR**	**CI**	**p**
Maternal age	Less than 25	0.27	[0.09,0.85]	**0.025**	0.25	[0.08,0.81]	**0.021**
	25-34	Ref		.	Ref		.
	35+	0.63	[0.22,1.81]	0.39	0.68	[0.23,1.98]	0.48
Adversity		2.59	[1.25,5.35]	**0.01**	2.87	[1.36,6.06]	**0.006**
Birthweight z-score		2.61	[1.73,3.94]	**<0.001**	2.67	[1.75,4.07]	**<0.001**
Rapid weight gain		3.86	[1.69,8.82]	**0.001**	4.32	[1.83,10.19]	**0.001**
Wean age		0.87	[0.75,1.03]	0.10	0.88	[0.74,1.04]	0.12
Maternal education		Ref		.	Ref		.
	GCSE	1.47	[0.27,8.03]	0.66	1.59	[0.28,8.93]	0.60
	A level	1.3	[0.20,8.44]	0.78	1.3	[0.20,8.63]	0.79
	Degree	1.87	[0.30,11.82]	0.51	1.9	[0.29,12.28]	0.50
SES (childhood)		Ref		.	Ref		.
	Mid	0.37	[0.14,0.97]	**0.042**	0.33	[0.13,0.89]	**0.028**
	Most advantaged	0.44	[0.17,1.20]	0.11	0.39	[0.14,1.06]	0.066
n		261			250		
Pseudo R^2		0.207			0.205		

		Waist OB			without UOJ		
Variable		**OR**	**CI**	**p**	**OR**	**CI**	**p**
Maternal age	Less than 25	0.29	[0.07,1.16]	0.080	0.32	[0.08,1.29]	0.11
	25-34	Ref		.	Ref		.
	35+	1.18	[0.39,3.59]	0.77	1.23	[0.40,3.76]	0.72
Adversity		1.06	[0.44,2.55]	0.90	1.02	[0.41,2.54]	0.96
Birthweight z-score		1.69	[1.09,2.64]	**0.020**	1.63	[1.04,2.55]	**0.033**
Rapid weight gain		2.07	[0.79,5.48]	0.14	1.9	[0.69,5.21]	0.21
Wean age		0.82	[0.70,0.97]	**0.023**	0.82	[0.69,0.97]	**0.021**
Maternal education		Ref		.	Ref		.
	GCSE	2.61	[0.30,22.46]	0.38	2.59	[0.30,22.61]	0.39
	A level	1.11	[0.10,12.96]	0.93	1.11	[0.09,12.97]	0.93
	Degree	1.08	[0.10,11.82]	0.95	1.09	[0.10,12.13]	0.94
SES (childhood)		Ref		.	Ref		.
	Mid	0.54	[0.18,1.61]	0.27	0.57	[0.19,1.75]	0.33
	Most advantaged	0.89	[0.30,2.63]	0.84	0.93	[0.31,2.81]	0.89
n		275			263		
Pseudo R^2		0.111			0.102		

Models included those with infant feeding variables, with and without ultra-orthodox Jewish (UOJ) study members for outcomes BMI, OWOB and waist OB. Models were additionally adjusted for sex, gestation and first born (for which all had non-significant p values) Coefficients (coef) or odds ratios (OR) are presented with confidence intervals (CI) and the corresponding level of significance (p). Ref indicates reference category for factor variables. No qualifications was the reference category for maternal education, and least advantaged for SES. . Bold indicates significant at $p<0.05$.

4.4.9 Path analysis

In the path model, birthweight, adversity, RWG and physical activity had direct paths to BMIz. All but MVPA had positive coefficients. The direct effects of birthweight and RWG were similar in magnitude, however RWG had the largest total effect on BMIz of all the exposures. Cumulatively, the exposures in the model explained 23% of the variation in BMIz.

Higher maternal education was associated with increased likelihood of being first-born and increased birthweight, independent of maternal age. Maternal education was not a direct predictor of BMIz, but had a small significant total effect, mediated through the early life exposures. Lower birthweights were observed for first-born children. Older maternal age was associated with lower likelihood of being first-born, as to be anticipated, however neither of these factors had significant total effects on BMIz.

There was an interesting relationship between birthweight, RWG and BMIz. Whilst a higher birthweight was associated with a higher BMIz (age 6-8), higher birthweights were associated with decreased likelihood of RWG. However, the total effect of birthweight on BMIz remained significant despite the attenuating effect via RWG, suggesting this remains an important factor. As well as direct and total effects, birthweight also had a significant indirect effect on BMIz.

Females spent less time doing MVPA compared to males. Removing MPVA and season from the model decreased the R^2 by around 1% and minimally altered coefficients (less than 10% change). Therefore, the effects of MPVA on predicting BMIz in this model are minimal. When building the path model, including weaning age minimally altered coefficients and did not increase $R^{2,}$ and resulted in a direct path with BMIz ($p<0.1$) but no paths to other variables. As it was not significant at $p<0.05$ it did not meet criteria for inclusion in the path model. Breastfeeding and Townsend score were not predictors of any variables in the model.

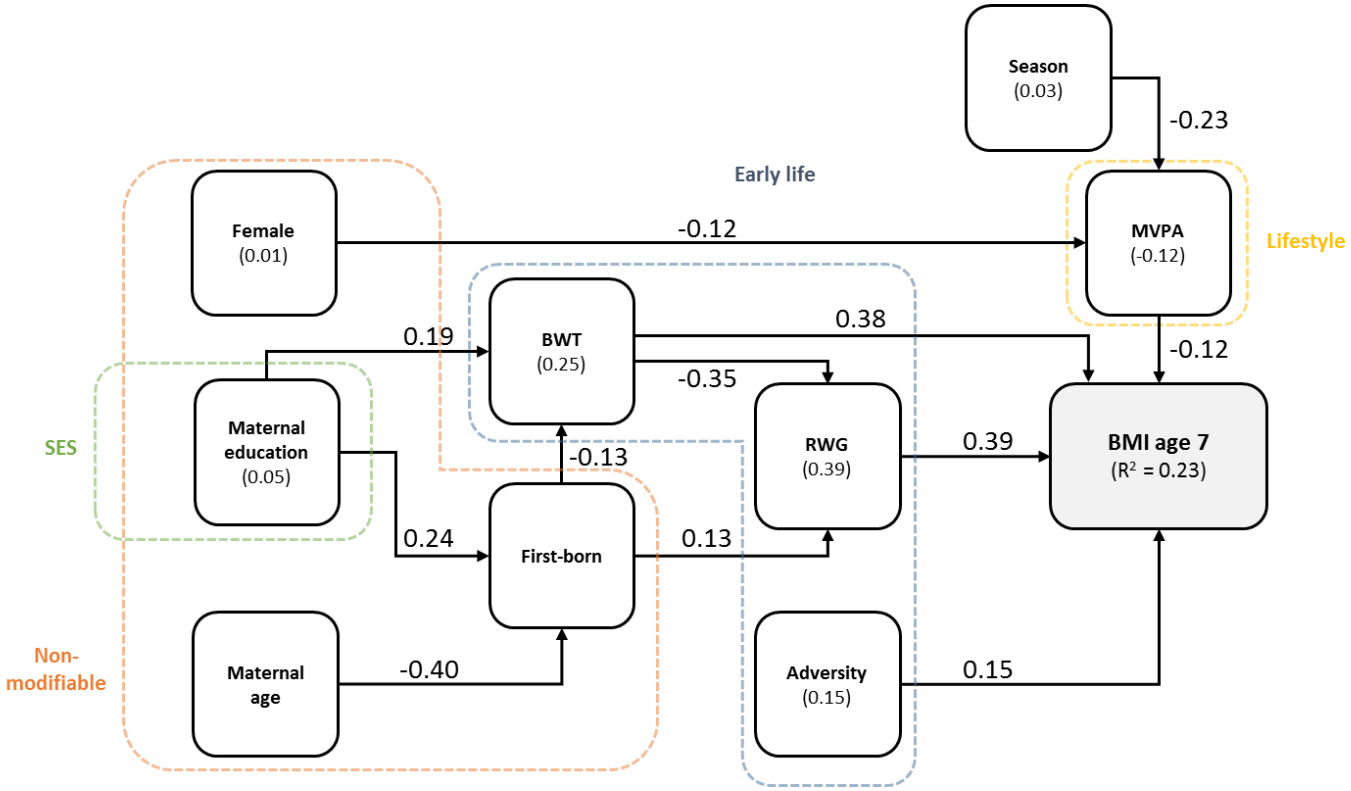


Figure 4.2 Path model showing the relationships between early life and childhood variables with BMIz at age 6-8.

Arrows show the direction of the effect. Standardised coefficients are shown. All direct effects are represented by solid lines and are significant at p<0.05. Total effects (if significant) are presenting in brackets with associated p values. There were significant indirect effects on BMIz for maternal education (0.05), female (0.01), season (0.03) and birthweight (0.14).

4.5 Discussion

4.5.1 Summary

Early life risk factors for childhood adiposity were examined for the outcomes BMIz, OWOB, waist OB and FMI in GMS children aged 6-8. For every outcome, there was an association for birthweight, and associations for young maternal age, adversity, and RWG across most outcomes (bar waist OB). These associations were evident even after adjusting for confounding factors including SES and physical activity. Age at weaning did not demonstrate consistent associations with body composition. Overall adversity, parental separation, money issues were significant predictors of adiposity even after adjustment for parental SES (Appendix C, Table IX). The low prevalence of specific adversity exposures (i.e. for death) may have affected the likelihood of finding associations.

Further exploration of the pathways to BMIz demonstrated that RWG was an intermediate factor between birthweight and BMIz. Maternal education was associated with birthweight,

and parity, which were also both associated with RWG. Adversity and MVPA independently predicted BMIz.

The associations noted for RWG were large, with 4 fold increased odds of OWOB and large increases in BMIz and FMI, with estimates increasing after adjusting for PAL. However, when excluding the FMI outliers and using robust regression, the estimates decreased suggesting that this relationship is stronger in children with higher fat mass.

Decreases in adiposity were evident across all outcomes for physical activity. Adjusting for activity removed any significant associations between SES and outcomes, and no early life exposures predicted waist OB after adjusting for activity. When MVPA was included in the multivariable models, the model explained a much larger proportion of the variation in the outcome variables, more so for OWOB and FMI (R^2 increased by 9%). Although in the path model, when activity was excluded it made little difference to estimates or overall variation explained. This may suggest that although activity is an important factor for obesity and fat mass, the early life and maternal factors and the pathways between them are the main influences on childhood BMIz.

There were no associations for caesarean birth, which is in contrast to systematic reviews and meta-analyses, which generally find a positive association (Darmasseelane et al., 2014, Kuhle et al., 2015). Findings from the Pelotas study from Brazil (Barros et al., 2017), and a matched sibling-pair design study (Rifas-Shiman et al., 2018) have attributed the association to unmeasured confounding. Differences might also arise when distinguishing between elective and emergency c-sections, for example considering underlying issues related to an emergency c-section such as maternal pre-pregnancy BMI or macrosomic infants. The underlying biological mechanism for caesarean birth is yet to be determined, but may be related to the microbiome (Masukume et al., 2018), which demonstrates great interindividual variation and could create heterogeneity in results (Kuhle and Woolcott, 2017).

There were null associations for sleep problems, which may be due to the definition of sleep problems used. The majority of studies analysing sleep problems use sleep duration, which was not measured in this cohort. There is also the possibility that higher BMI precedes sleep problems (Wang et al., 2019).

There were no associations for early life infection, however this was a rather crude measure extrapolated from questionnaire data. More accurate infection data could come from data linkage with NHS records.

4.5.2 Sensitivity analyses

Sensitivity analyses based on religion revealed some differences. Previous research in this group in GMS has found differing growth patterns, including weight faltering in infancy (Wright et al., 2010b). This was attributed to delayed introduction of solids and longer duration breastfeeding (Wright et al., 2010b). Excluding this demographic also meant the exposures; adversity, birthweight, and RWG corresponded to higher odds of OWOB and increased FMI and BMIz. This suggests that the effect sizes were underestimated, and perhaps implies that these risk factors may not be risk factors in different populations. Despite previous findings that ultra-orthodox infants were lighter at age 13 months, the prevalence of overweight or adiposity in childhood (age 6-8) was not different between ultra-orthodox children and the rest of the cohort. It was not possible to discern if early life risk factors differed between groups, and due to the small sample size results should be interpreted with caution.

4.5.3 Choice of outcome measure

There were strong correlations between all outcome measures, more so between BMIz, FMI and OWOB, whilst waist OB had slightly weaker correlations with the other outcomes (however all Pearson r>0.6). Studies have found that correlations between body fat and BMI in children can range vastly (from 0.22 to 0.9) and are stronger in those with higher fat mass (Wells, 2000, Freedman et al., 2005b).

Despite the strong correlations, there were some differences between risk factors and outcomes (Figure 4.3). In accordance with the work of Basterfield *et al* (2012), there were differences in early life risk factors based on choice of outcome measure. However, contrary to their work, in these analyses there was good consistency in associations for exposures for BMIz, FMI and OWOB. The risk factors for increased BMIz (maternal age, birthweight, adversity and RWG) also correspond to increased odds of OWOB, which is perhaps not surprising as BMIz and OWOB are based on the same measures.

Height is a confounding factor when determining childhood adiposity. The outcome measures accounted for height, however there still may be residual correlation. The use of

FMI in addition to primary outcomes allows greater sensitivity to detect early life exposures associated with adiposity. Although, in the study by Steiberger *et al*. (2005), the correlations between BMIz and body fat quantified by DXA were stronger among children with a higher percentage of body fat (r = 0.9) than among leaner children (r = 0.5) (Steinberger et al., 2005). This suggests that BMI is more reflective of body fat in children with higher fat mass.

Ideally studies would use direct measures of body fat. Whilst BIA has its limitations, many of the other sophisticated methods for determining body fat are less suitable for small children as they require subjects to lie still for an amount of time. In this cohort the risk factors for FMI were the same as those for OWOB, suggesting that BMI (for determining OWOB) may be an adequate proxy measure.

Overall these results indicate that the risk factors; birthweight, RWG, adversity and maternal age were consistently associated with adiposity in this cohort, with some support for early weaning and low SES. Physical activity was associated with reduced adiposity across all outcomes.

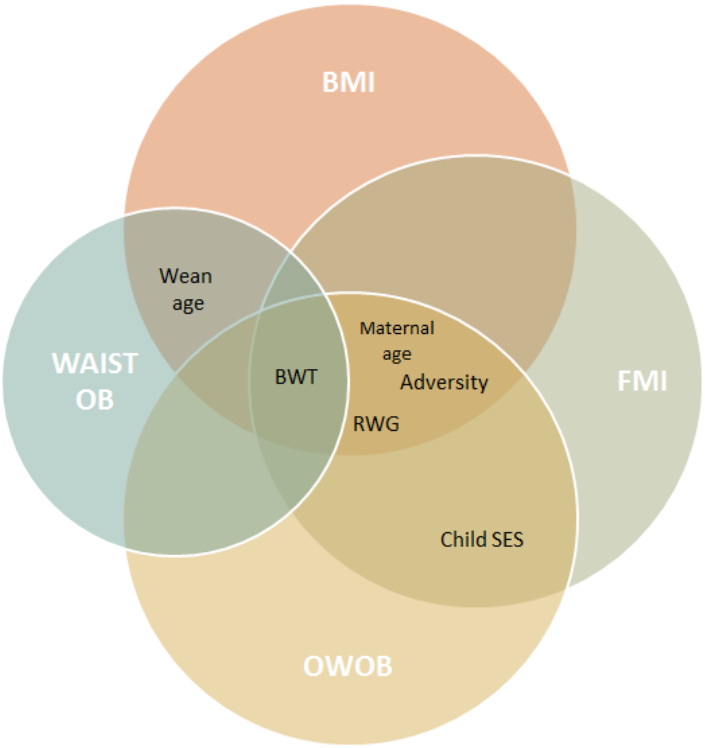


Figure 4.3 Venn diagram for significant associations (p<0.05) in multivariable models across outcomes.

Associations are from the multivariable models not adjusted for PAL and outliers not excluded for FMI. BWT, birthweight, Child SES, childhood socioeconomic status; RWG, rapid weight gain.

Chapter 5. The influence of early life factors and the environment on childhood obesity over time in two regional birth cohorts

5.1 Introduction

Increases in childhood obesity in the last 20 years have been somewhat attributed to social change and a changing environment. All of the environmental influences and conditions of life that encourage OWOB have been termed the obesogenic environment (Lake et al., 2011). Over this time, there has also been widening social inequalities in childhood overweight (NHS Digital, 2016, Bann et al., 2018), and changes in the social patterning of obesity (Knai et al., 2012).

Early life risk factors for childhood obesity have mostly been examined in modern cohorts. However, multiple early life risk factors have not previously been examined in children in a historic (i.e. pre-obesity epidemic) cohort. Therefore, it is unknown whether these early life risk factors are only risk factors in conjunction with an obesity-promoting environment or are distinct, biologically embedded risk factors.

This chapter addresses the second aim of the thesis and investigates the regional temporal changes on obesity, and the impact of modern environmental and socioeconomic factors. The analysis uses data from both the 1947 Newcastle Thousand Families Study (NTFS) and the 2000 Gateshead Millennium Study (GMS)), two cohorts from the same region- North East of England born over 50 years apart. Using cohorts from the same region controls for some baseline area-level differences. The cohorts had data collected on body composition at various ages throughout childhood, and both cohorts had measures of body height and weight at ages 9 and 13. The analysis in this chapter focused on body composition at age 9, due to a lower level of missing data, and to minimise the bias of puberty on results (which typically begins between around age 11 and 12 years for girls and boys respective (Marshall and Tanner, 1969, Marshall and Tanner, 1970)).

The previous chapters determined that the early risk factors for childhood body composition (in GMS) were birthweight, maternal age, adversity and rapid weight gain (RWG). Breastfeeding was associated with BMI in NTFS adults. SES had indirect effects in both adults and children. Therefore, these were the exposures investigated in this analysis. Whilst there

were other associations found (bacterial infection (NTFS adults), physical activity (NTFS adults and GMS children)), comparable data were not available for these exposures for both cohorts.

This chapter will investigate and compare the relationship between early life and societal factors on childhood body composition in the two cohorts. Using cohorts from the same area but at different time points provides a unique opportunity to investigate if, and how the escalating obesogenic environment may have transformed early life risk factors for obesity.

5.2 Aims

The first aim of this chapter was to determine if an increasingly obesogenic environment has altered the impact of early life factors and societal factors on BMI and the likelihood of being overweight/obese (OWOB) in childhood. The analysis will examine if risk factors have changed over time, in two regional birth cohorts that were not subject to the same confounding and environmental influences.

Secondly, this chapter aimed to assess the relative importance of each of the early life factors and SES on childhood BMI by understanding the pathways between them.

5.3 Participants and methods

Details on the two cohorts utilised in this chapter are outlined in sections 2.1.1 (NTFS) and 2.1.2 (GMS). Anthropometric measures were taken at around age 9 for participants from both cohorts and were used to calculate BMIz and OWOB (for details see methods section 2.2.1). There were several early life factors directly comparable across the two cohorts and measured at similar time points (Table 5.1). These included; SES, maternal age, birthweight, breastfeeding, adversity and RWG (in the first year). Due to the large difference in birthweight z-scores between the cohorts, weight gain conditional on birthweight was also considered (rapid thrive). Recoding of the variables was similar for both, except for SES and adversity (Table 5.1).

Table 5.1 Description of the earlexposures and any differences between the cohorts

Time point	Variable	Description	Type of variable	Cohort diff-erences?
Birth	Maternal age	Years and groups (<25, 25-34, 35+)	Continuous and categorical	No
	Birthweight	Birthweight z-score, using the British 1990 growth reference	Continuous	No
	Gestation	Weeks (continuous) and categorised as; pre-term, <38 wks; post-term >41 wks	Continuous and categorical	No
	Socioeconomic status (SES)	In NTFS Social class based on occupation In GMS Townsend score (quintiles)	Categorical, 5 categories with 1 being the most advantaged and 5 the least	Yes
First year	Rapid weight gain (RWG)	If experienced a 0.67 SD change in weight for age z-score	Dichotomous	No
	Rapid thrive (RT)	If experienced a 0.67 SD change in conditional weight gain ($z\text{-score}_{12m} - r \times z\text{-score}_{birth}$)	Dichotomous	No
	Breastfeeding	Predefined categories (GMS) of never, <4 weeks, 4 week-6 months and 6 months+ For NTFS, weeks were transformed into the same categories to match GMS	Categorical	No
	Adversity	Experiencing any of the following; parental separation, police involvement, abuse, debt, death or illness in the family; - In the first year in NTFS (0-12 months) - From 8 months pre-natal to 4 months post-natal in GMS	Dichotomous	Yes

SD, standard deviation.

Statistical analysis

Sex, SES, adverse events, RWG, RT and breastfeeding were analysed as categorical variables. Maternal age and birthweight were analysed as both categorical and continuous variables, and gestation as a continuous variable. Outcomes were BMIz and OWOB at age 9.

The statistical analysis is outlined in section 2.4.3. Briefly, this included testing cohort representativeness of sub-sample of participants at age 9 and examining baseline differences in the cohorts, using t-test, chi-squared or Wilcoxon rank-sum tests as appropriate. Variation between socioeconomic groups and body composition (OWOB, BMIz and height), were analysed using Chi-square tests or one-way ANOVA. Logistic (OWOB) and linear (BMIz) regression analyses were carried out as described in section 2.3.5. Associations between exposures and outcomes were estimated, with separate analyses for each cohort, and with adjustment for SES.

Although BMI assesses weight independent of height, it remains correlated with height in children (Metcalf et al., 2011b)(residual correlation). In GMS, there was a moderate correlation ($r=0.4$, $p<0.0001$) between height and BMIz (age 9), whereas there was no correlation in NTFS ($p=0.3$). Furthermore, as there was a notable height difference between the cohorts, height was included as a covariate in the multivariable models.

Odds ratios (OR) with corresponding 95% confidence intervals for OWOB and explanatory variables were estimated using logistic regression. Multivariable analysis was possible only in the GMS cohort due to few OWOB study members in NTFS.

5.4 Results

5.4.1 Sample differences in early life exposures

For the children in the original cohorts (NTFS *n*=1142, GMS *n*= 1029), measures of height and weight were available for 734 members of NTFS and 481 of GMS at age 9. Study members measured at age 9 differed significantly from the remainder of the original study members in both cohorts (Table 5.2). In GMS those with BMI measures (age 9) had a higher mean maternal age, were more advantaged at birth, had a longer duration of breastfeeding and a greater proportion had experienced an adverse event, compared to the original sample. The NTFS age 9 sample also had a higher mean maternal age, the sample was less advantaged and had an overall shorter duration of breastfeeding, and a slightly longer length of gestation.

Table 5.2 Descriptive statistics and sample representativeness of those with BMI measures at age 9 in the NTFS and GMS cohorts for all early life exposures and covariates

	NTFS					GMS				
Continuous	**Measured**		**Not measured**			**Measured**		**Not measured**		
	n	Mean (SD)	n	Mean (SD)	p	n	Mean (SD)	n	Mean (SD)	p
Birthweight (z-score)	251	-0.23 (1.11)	734	-0.11 (1.04)	0.14	506	-0.02 (1.00)	481	0.01 (1.03)	0.94
Maternal age (years)	244	26.99 (5.42)	734	28.97 (5.84)	**<0.001**	506	27.1 (6.08)	481	28.82 (5.72)	**<0.001**
Gestation (weeks)	246	39.67 (1.77)	727	39.87 (1.07)	**0.032**	506	39.42 (1.82)	481	39.55 (1.55)	0.22

	NTFS				GMS			
Categorical	**Total**	**Measured**	**Not measured**	**p**	**Total**	**Measured**	**Not measured**	**p**
	N	Col %	Col %		N	Col %	Col %	
Sex	1,097	363	734		988	507	481	
Male	561	54.5	49.5	0.11	501	51.7	49.7	0.53
Female	536	45.5	50.5		487	48.3	50.3	
Birthweight	973	246	727		987	506	481	
SGA	111	15	10.2		87	8.7	8.9	
Normal	784	78	81.4	0.10	806	82.4	80.9	0.77
LGA	78	6.9	8.4		94	8.9	10.2	
Gestation categories	983	250	733		987	506	481	
Pre-term	34	7.2	2.2		105	11.9	9.4	
Normal	907	87.2	94.0	**<0.001**	819	81.2	84.8	0.32
Post-term	42	5.6	3.8		63	6.9	5.8	
RWG (0-12 months)	354	17	337		808	354	454	
No	213	58.8	60.2	0.91	562	70.9	68.5	0.46
Yes	141	41.2	39.8		246	29.1	31.5	
RT (0-12 months)	354	17	337		808	354	454	
No	219	58.8	62	0.79	572	72.9	69.2	0.25
Yes	135	41.2	38		236	27.1	30.8	
Maternal age at birth	978	244	734		987	506	481	
Less than 25	272	36.5	24.9		322	39.1	25.8	
25-34	544	54.5	56	**<0.001**	545	50.6	60.1	**<0.001**
35+	162	9	19.1		120	10.3	14.1	
Breastfeeding categories	460	114	346		948	483	465	
None	65	8.8	15.9		465	56.1	41.7	
<6 weeks	73	21.1	14.2	**0.004**	237	25.1	24.9	**<0.001**
>6wk	143	21.9	34.1		89	6.8	12.0	
>4m	179	48.2	35.8		157	12.0	21.3	
Adversity	346	97	249		928	475	453	
No	298	89.7	84.7	0.23	715	80.8	73.1	**0.005**
Yes	48	10.3	15.3		213	19.2	26.9	
SES at birth	1021	310	711		981	506	475	
Least advantaged	158	11.3	17.3		186	20.2	17.7	
2nd to least	162	15.8	15.9		200	18.2	22.7	
Mid advantaged	577	52.6	58.2	**<0.001**	221	23.1	21.9	**0.001**
2nd to most	92	14.5	6.6		222	26.5	18.5	
Most advantaged	32	5.8	2.0		152	12.1	19.2	

Number of study members in each category (n) and corresponding column percentage (Col %) or mean and standard deviation (SD). P values shown for Chi-square test for significant differences between the those with data (BMI measured) and those without for categorical variables, and t-tests for continuous variables. SGA, Small for gestational age; LGA, Large-for-gestational age; RWG, rapid weight gain; RT, rapid thrive; SES, socioeconomic status.

5.4.2 Cohort differences in outcome measures

GMS children on average had a significantly ($p<0.001$) higher BMIz (+0.52 z-score) and were taller (+8.4cm) than NTFS children (Table 5.3). Of the 734 study members, there were few children with overweight (5%) or obesity (3%) in the NTFS in 1956, with the majority (93%) having a healthy weight. Of the 481 in GMS sample, 33% were OWOB, of these 11% were obese (Table 5.3).

5.4.3 Descriptive characteristics of socioeconomic groups and weight outcomes

In both cohorts, there were no significant differences in the prevalence of children with OWOB by socioeconomic group (at each time point), and no trend for BMIz (ANOVA $p>0.05$) (Table 5.3). Although there were no significant differences, in both cohorts the 2nd to most advantaged group (at birth) had the highest mean BMIz (age 9).

There were socioeconomic differences in height in NTFS, which were significant (after Bonferroni correction) at birth between the least advantaged and both the mid (+3.0cm, $p<0.001$), and the most advantaged (+5.9cm, $p=0.04$)(Table 5.3). There were no differences for childhood SES. Whereas in GMS, height was associated with childhood SES with significant differences between the least and most advantaged (+2.1, $p=0.056$). However, there were no significant height and SES interactions on BMIz within the cohorts.

When examining SES groups in childhood, in GMS the least advantaged group (at age 9) had a greater proportion of children with OWOB (32%), although this was not significant ($p=0.07$) (Table 5.3). In contrast, in NTFS the most advantaged group (at age 9) had the highest proportion of OWOB (8.5%); and this was comprised exclusively of children with obesity.

Table 5.3 Body composition measures (age 9) by SES categories at birth and age 9 in the GMS and NTFS cohorts

	NTFS					GMS				
	n	Healthy (%)	OW-OB (%)	Mean BMIz (SD)	Mean height (SD)	n	Healthy (%)	OW-OB (%)	Mean BMIz (SD)	Mean height (SD)
All	734	93.1	6.9	0.08 (-0.88)	127.4 (7.40)	481	76.5	23.5	0.60 (1.06)	135.8 (6.34)
SES at birth										
Least advantaged	123	91.9	8.1	0.07 (-0.90)	125.2 (7.31)	84	76.2	23.8	0.63 (1.04)	135.7 (5.82)
2nd to least advantaged	113	98.2	1.8	-0.01 (-0.82)	126.6 (7.26)	108	78.7	21.3	0.58 (1.07)	136.5 (6.03)
Mid	414	92	8	0.11 (-0.89)	128.2 (7.31)	104	76	24	0.61 (1.1)	135.9 (6.63)
2nd to most advantaged	47	91.5	8.5	0.18 (-0.98)	127.7 (7.59)	88	70.5	29.5	0.72 (1.05)	135.0 (6.68)
Most advantaged	14	92.9	7.1	-0.03 (-0.87)	131.1 (6.66)	91	80.2	19.8	0.5 (1.02)	135.4 (6.53)
Total	711					475				
P value		0.22 [a]		0.66 [b]	**0.001** [b]		0.59 [a]		0.74 [b]	0.52 [b]
SES at 9										
Least advantaged	250	94	6	0.04 (-0.92)	127.4 (6.78)	87	67.8	32.2	0.63 (1.13)	134.9 (5.88)
Mid	379	92.9	7.1	0.08 (-0.85)	127.9 (7.31)	110	80.9	19.1	0.53 (1.01)	136.0 (6.34)
Most advantaged	47	91.5	8.5	0.17 (-0.85)	128.6 (6.73)	105	79	21	0.70 (0.99)	137.0 (6.32)
Total	676					302				
P value		0.77 [a]		0.63 [b]	0.50 [b]		0.07 [a]		0.49 [b]	0.06 [b]

Category totals (N); corresponding row percentages (row %); Standard deviation (SD); Socioeconomic status at birth (SES) was fathers occupational social class in NTFS or Townsend quintile in GMS. Bold indicates p<0.05.
a Chi-square test p value presented for differences between socioeconomic group and weight categories
b ANOVA p value for differences between socioeconomic groups

5.4.4 Cohort differences in early life exposures

Descriptive statistics for all variables are shown in Table 5.4, with tests for significant differences between cohorts. Mean birthweight was higher in GMS children and length of gestation was marginally longer in NTFS (Table 5.4). There were no differences in the distribution of birthweight categories (SGA, LGA) between the cohorts. In GMS there were more children who were never breastfed and there was overall shorter duration of breastfeeding: just under half of GMS children were never breastfed (49%), whereas breastfeeding for over 4 months was more commonplace in NTFS (39%). Rapid growth (RWG and RT) was more common in NTFS than in GMS. In GMS there were fewer older mothers (35+) and a greater proportion of younger mothers (<25 years).

There were few in the most advantaged socioeconomic groups in NTFS at both birth and age 9 (Table 5.4). The composition of socioeconomic groups was notably different between the cohorts at age 9 (occupational social class): only 6.8% of the NTFS cohort were in the highest occupational group, compared to 34.9% in GMS ($p<0.0001$) (Table 5.4).

Table 5.4 Descriptive statistics for early life exposures and covariates and baseline differences between the cohorts

	Cohort				
	NTFS		GMS		
Continuous variables	n	Mean (SD)	n	Mean (SD)	P value
Birthweight (z-score)	1,002	-0.15 (1.06)	993	-0.02 (1.02)	0.009
Maternal age (years)	995	28.48 (5.80)	993	27.92 (5.97)	0.035
Gestation (weeks)	990	39.82 (1.28)	993	39.48 (1.70)	<0.0001
Categorical variables	n	Col %	n	Col %	P value
Sex	1,114		994		
Male	570	51.2	506	50.9	0.905
Female	544	48.8	488	49.1	
Gestation categories	990		993		
Pre-term	34	3.4	107	10.78	
Normal	914	92.3	823	82.88	**<0.0001**
Post-term	42	4.2	63	6.34	
Categories of birthweight	990		993		
SGA	115	11.6	89	9.0	
Normal	796	80.4	810	81.6	0.10
LGA	79	8.0	94	9.5	
RWG	360		813		
No	218	60.6	567	69.6	**0.002**
Yes	142	39.4	246	30.4	
RT	360		813		
No	227	63.1	577	71.0	**0.007**
Yes	133	36.9	236	29.0	
Maternal age (years)	995		911		
Less than 25	276	27.7	326	32.8	
25-34	554	55.7	547	55.1	**0.004**
35+	165	16.6	120	12.1	
Breastfeeding categories	469		993		
Never	68	14.5	468	49.1	
<6 weeks	75	16.0	237	24.9	**<0.0001**
>6 weeks	143	30.5	89	9.3	
>4 months	183	39.0	159	16.7	
Adversity	352		934		
No	303	86.1	719	77.0	**<0.0001**
Yes	49	13.9	215	23.0	
SES at birth	1,036		987		
Least advantaged	158	15.3	188	19.1	
2nd to least advantaged	165	15.9	201	20.4	**<0.0001**
Mid	589	56.9	221	22.4	
2nd to most advantaged	92	8.9	223	22.6	
Most advantaged	32	3.1	154	15.6	
SES at age 9	718		373		
Least advantaged	265	36.9	110	29.5	
Mid	404	56.3	133	35.7	**<0.0001**
Most advantaged	49	6.8	130	34.9	

Number of study members in each category (n) and corresponding column percentage (Col %) or mean and standard deviation (SD). P values shown for Chi-square test for significant differences between NTFS and GMS for categorical variables, and t-tests for continuous variables. Bold indicates p<0.05. SGA, Small for gestational age; LGA, Large-for-gestational age; RWG, rapid weight gain; RT, rapid thrive; SES, socioeconomic status.

5.4.5 Relationship between childhood BMI, early life risk factors and SES

Similar to descriptive analyses, there were no significant direct linear associations between early life or childhood SES and BMIz at age 9 in either cohort (Table 5.6, unadjusted model). In NTFS unadjusted regression models, females had a significantly lower BMIz. RWG and RT were associated with an increased BMIz, each to a similar degree. Higher maternal age was associated with a small decrease in BMIz. These associations remained after adjustment for SES at birth and at age 9. No other variables demonstrated significant ($P<0.05$) associations with BMIz in NTFS.

In the GMS cohort, birthweight and adversity were associated with an increased BMIz, with adjustment for SES at both time points attenuating these effects (Table 5.6). Both RWG and RT were also associated with increased BMIz, however the coefficient for RT was much larger (RWG $b=0.32$, $p<0.05$; RT $b=0.70$, $p<0.001$). After adjusting for SES, older maternal age was significantly associated with a lower BMIz in GMS.

There was no evidence for interactions between explanatory variables and SES at birth for pooled data from both cohorts (all p values >0.1). Interactions between SES at birth and explanatory variables were also investigated separately by cohort. In NTFS, there was a significant interaction between maternal age (continuous) and those in the 2nd to most advantaged group at birth compared to the least advantaged (-0.07 reduction in BMIz, $p=0.02$), which remained after adjustment for SES at age 9, however there was no overall trend (Table 5.5). This indicates that the slopes of the best-fitting regression lines between maternal age and childhood BMI are not parallel for every socioeconomic group. Group sizes were too small to investigate adversity, categories of birthweight and maternal age in NTFS. There were no significant interactions between SES and exposures in GMS.

Table 5.5 The significant exposure-socioeconomic status interaction(s) from the unadjusted and adjusted for SES (age 9) bivariate regression models in the NTFS cohort (age 9)

SES (at birth) & exposure interaction effects						
NTFS	Unadjusted			Adjusted for SES (age 9)		
	Coef	CI	p	Coef	CI	p
Maternal age						
*Least advantaged	Ref			Ref		
*2nd to least advantaged	-0.01	(-0.05,0.03)	0.60	-0.01	(-0.05,0.03)	0.48
*Mid	0.02	(-0.01,0.05)	0.15	0.02	(-0.01,0.05)	0.28
*2nd to most advantaged	-0.07	(-0.13,-0.01)	**0.020**	-0.06	(-0.12,-0.00)	**0.036**
*Most advantaged	0.08	(-0.05,0.21)	0.21	0.08	(-0.05,0.21)	0.22
n	711			657		

* p<0.05, ** p<0.01, *** p<0.001. Coef, coefficient; CI, 95% confidence interval; Ref, reference category. SES, socioeconomic status. Bold indicates significant at p<0.05.

Table 5.6 Linear regression associations for the early life exposures and BMIz at age 9 years with unadjusted, adjusted for SES (birth) and adjusted for SES (age 9) estimates presented for the NTFS and GMS cohorts

NTFS										
		Unadjusted			**Adjusted for SES (birth)**			**Adjusted for SES (age 9)**		
Exposure		Coef	CI	p	Coef	CI	p	Coef	CI	p
Female		-0.23	(-0.36,-0.11)	**<0.001**	-0.24	(-0.37,-0.11)	**<0.001**	-0.25	(-0.38,-0.11)	**<0.001**
Birthweight z-score		0.05	(-0.01,0.11)	0.092	0.05	(-0.01,0.11)	0.134	0.04	(-0.02,0.11)	0.20
Birthweight categories	SGA	-0.14	(-0.35,0.07)	0.20	-0.15	(-0.37,0.08)	0.20	-0.18	(-0.41,0.05)	0.13
	Normal	Ref		.	Ref		.	Ref		.
	LGA	0.09	(-0.15,0.32)	**0.46**	0.08	(-0.15,0.32)	0.49	0.04	(-0.21,0.29)	0.75
Maternal age (years)		-0.01	(-0.02,-0.00)	**0.035**	-0.01	(-0.02,-0.00)	**0.023**	-0.01	(-0.03,0.00)	**0.022**
Maternal age categories	< 25	0.09	(-0.06,0.24)	0.26	0.11	(-0.05,0.27)	0.19	0.15	(-0.02,0.31)	0.079
	25-34	Ref		.	Ref		.	Ref		.
	35+	-0.13	(-0.30,0.04)	0.14	-0.13	(-0.30,0.04)	0.13	-0.11	(-0.29,0.07)	0.22
Adverse events		-0.09	(-0.39,0.21)	0.55	-0.03	(-0.35,0.30)	0.88	0.07	(-0.29,0.43)	0.71
RWG		0.29	(0.10,0.48)	**0.002**	0.27	(0.08,0.46)	**0.006**	0.27	(0.08,0.47)	**0.005**
RT		0.23	(0.04,0.42)	**0.01**	0.21	(0.01,0.40)	**0.021**	0.20	(-0.00,0.40)	**0.031**
Breastfeeding categories	Never	Ref			Ref			Ref		
	<6 weeks	0.19	(-0.17,0.54)	0.30	0.18	(-0.16,0.52)	0.30	0.18	(-0.16,0.52)	0.30
	>6 weeks	0.12	(-0.17,0.42)	0.41	0.11	(-0.18,0.39)	0.47	0.11	(-0.18,0.39)	0.47
	>4 months	0.08	(-0.22,0.38)	0.60	0.08	(-0.20,0.36)	0.58	0.08	(-0.20,0.36)	0.58
SES at birth										
Least advantaged					Ref			Ref		
2nd to least advantaged					-0.07	[-0.30,0.16]	0.53	-0.01	[-0.25,0.23]	0.92
Mid					0.04	[-0.14,0.22]	0.65	0.05	[-0.16,0.26]	0.61
2nd to most advantaged					0.11	[-0.19,0.41]	0.48	-0.02	[-0.36,0.32]	0.90
Most advantaged					-0.07	[-0.58,0.44]	0.78	-0.28	[-0.90,0.34]	0.38
SES age 9										
Least advantaged								Ref		
Mid								0.02	[-0.14,0.18]	0.82
Most advantaged								0.25	[-0.11,0.61]	0.17
Height (age 9)		0	(-0.00,0.01)	0.29	0	(-0.01,0.01)	0.46	0.01	(-0.00,0.02)	0.12

GMS										
		Unadjusted			**Adjusted for SES (birth)**			**Adjusted for SES (age 9)**		
Exposure		Coef	CI	p	Coef	CI	p	Coef	CI	p
Female		-0.03	(-0.22,0.16)	0.78	-0.05	(-0.24,0.15)	0.64	-0.15	(-0.39,0.09)	0.23
Birthweight z-score		0.21	(0.12,0.30)	**<0.001**	0.23	(0.13,0.32)	**<0.001**	0.22	(0.10,0.33)	**<0.001**
Birthweight categories	SGA	-0.2	(-0.53,0.13)	0.23	-0.19	(-0.53,0.15)	0.28	-0.15	(-0.61,0.30)	0.51
	Normal	Ref		.	Ref		.	Ref		.
	LGA	0.22	(-0.09,0.53)	0.16	0.25	(-0.07,0.57)	0.12	0.25	(-0.14,0.64)	0.21
Maternal age (years)		0	(-0.02,0.01)	0.72	0	(-0.02,0.02)	0.88	-0.02	(-0.04,0.01)	0.18
Maternal age categories	< 25	-0.12	(-0.34,0.10)	0.30	-0.14	(-0.37,0.10)	0.25	-0.11	(-0.41,0.20)	0.50
	25-34	Ref		.	Ref		.	Ref		.
	35+	-0.28	(-0.56,-0.00)	**0.049**	-0.27	(-0.55,0.01)	0.055	-0.44	(-0.77,-0.10)	**0.011**
Adverse events		0.42	(0.20,0.64)	**<0.001**	0.43	(0.21,0.65)	**<0.001**	0.34	(0.07,0.61)	**0.014**
RWG		0.36	(0.16,0.57)	**0.001**	0.37	(0.16,0.58)	**0.001**	0.32	(0.05,0.59)	**0.019**
RT		0.63	(0.43,0.83)	**<0.001**	0.65	(0.44,0.86)	**<0.001**	0.70	(0.45,0.96)	**<0.001**
Breastfeeding categories	Never	Ref			Ref			Ref		
	<6 weeks	-0.05	(-0.30,0.19)	0.67	-0.04	(-0.29,0.21)	0.75	-0.04	(-0.29,0.21)	0.75
	>6 weeks	-0.28	(-0.59,0.04)	0.085	-0.24	(-0.57,0.10)	0.17	-0.24	(-0.57,0.10)	0.17
	>4 months	-0.16	(-0.42,0.09)	0.21	-0.17	(-0.45,0.10)	0.22	-0.17	(-0.45,0.10)	0.22
SES at birth										
Least advantaged					Ref			Ref		
2nd to least advantaged					-0.09	(-0.39,0.22)	0.58	-0.21	(-0.61,0.18)	0.29
Mid					-0.06	(-0.36,0.25)	0.72	-0.29	(-0.69,0.10)	0.15
2nd to most advantaged					0.07	(-0.25,0.38)	0.68	-0.21	(-0.61,0.19)	0.30
Most advantaged					-0.18	(-0.50,0.14)	0.27	-0.18	(-0.59,0.23)	0.38
SES age 9										
Least advantaged								Ref		
Mid								-0.09	(-0.40,0.22)	0.57
Most advantaged								0.05	(-0.27,0.37)	0.76
Height (age 9)		0.07	(0.05,0.08)	**<0.001**	0.07	(0.05,0.08)	**<0.001**	0.07	(0.05,0.08)	**<0.001**

The unadjusted model is the relationship between the exposure and BMIz, with models further adjusted for SES at birth, and then SES age 9.
Coef, coefficient; CI, 95% confidence interval; p, p value; Ref, reference category. SES, socioeconomic status; RWG, rapid weight gain; RT, rapid thrive.

The multivariable analyses were stratified by cohort, and RT (rather than RWG) was used as it controls for some of the differences in birthweight. In the stratified multivariable model for NTFS, females had a lower BMIz (-0.22, CI -0.41,-0.02) and those who experienced RT in the first year had a higher BMIz (0.20, CI 0.01,0.40) (Table 5.7). Those from the most advantaged group at birth had a significantly lower BMIz, however there was no overall trend by SES. There was no significant interaction between maternal age and SES after adjustment for the other covariates.

The significant predictors of BMIz in GMS were RT, birthweight and adversity, adjusted for covariates; maternal age, gestation, SES and first-born (Table 5.7). RT was associated with greater increases in BMIz in GMS. The early life factors explained a greater proportion of the variation in BMIz in GMS (GMS R^2=17%, NTFS R^2=3%, Table 5.7). There were no significant interactions between early life exposures and SES at birth in the adjusted model for either cohort.

Adjusting for height attenuated the effects (by 10-30%) for all significant associations in GMS. Whilst adjusting for height made little difference to estimates for early life variables in NTFS. The variance inflation factors were low (<2.5) for all variables, therefore this is less likely to be due to multi-collinearity. This may be because height is a mediator on the pathway for early life factors and BMIz in GMS (Figure 5.1)

Table 5.7 Multivariable fully adjusted linear regression models for BMIz (age 9) by cohort

	NTFS			GMS		
	coef	CI	p	coef	CI	p
Female	-0.22	(-0.41,-0.02)	**0.029**	-0.16	(-0.39,0.06)	0.16
RT	0.22	(0.01,0.43)	**0.038**	0.50	(0.23,0.76)	**<0.001**
Birthweight z-score	0.02	(-0.08,0.11)	0.72	0.17	(0.05,0.29)	**0.006**
SES at birth						
Least advantaged	Ref		.	Ref		.
2nd to least	0.05	(-0.31,0.41)	0.79	-0.24	(-0.62,0.14)	0.22
Mid advantaged	0.1	(-0.21,0.41)	0.54	-0.34	(-0.73,0.05)	0.084
2nd to most	0.04	(-0.45,0.53)	0.88	-0.19	(-0.58,0.19)	0.33
Most advantaged	-0.93	(-1.87,0.01)	0.052	-0.18	(-0.57,0.22)	0.39
Height (cm)	0	(-0.02,0.01)	0.59	0.04	(0.02,0.06)	**0.019**
Adversity				0.30	(0.05,0.55)	**<0.001**
Adjusted R^2	0.022			0.212		
n	313			269		

Models were additionally adjusted for height, maternal age, gestation and SES at age 9.
* p<0.05, ** p<0.01, *** p<0.001. Bold indicates p<0.05. Coef, coefficient; CI, 95% confidence interval; Ref, reference category.; n, number of observations. RT, rapid thrive; SES, socioeconomic status.

5.4.6 **Early life predictors of childhood overweight/obesity**

In addition to BMIz, early life exposures were examined as risk factors for childhood OWOB. In terms of socioeconomic factors, in NTFS there were lower odds of OWOB between the second to least advantaged group (compared to the least at birth) ($p<0.05$), but there was no overall trend (Table 5.8). Whilst in GMS, mid-socioeconomic advantage (age 9) corresponded to 50% lower odds of OWOB compared to the least advantaged.

Birthweight, adversity, RT and RWG were associated with higher odds of OWOB in GMS children in unadjusted models (Table 5.8). In NTFS, the only early life exposure significantly associated with OWOB was RT (unadjusted model, Table 5.8).

Table 5.8 Bivariate (unadjusted) logistic regression models for overweight/obese (age 9) by cohort

	NTFS			GMS		
	Unadjusted			Unadjusted		
Variable	OR	CI	p	OR	CI	p
SES at birth						
Least advantaged	Ref	.		Ref	.	
2nd to least	0.20	(0.04,0.95)	**0.043**	0.87	(0.44,1.71)	0.68
Mid advantaged	0.98	(0.47,2.05)	0.95	1.01	(0.52,1.99)	0.97
2nd to most	1.05	(0.31,3.53)	0.94	1.34	(0.68,2.65)	0.40
Most advantaged	0.87	(0.10,7.35)	0.90	0.79	(0.38,1.62)	0.52
SES at age 9						
Least advantaged	Ref	.		Ref	.	
Middle	1.2	(0.63,2.31)	0.58	0.50	(0.26,0.96)	**0.036**
Most advantaged	1.46	(0.46,4.60)	0.52	0.56	(0.29,1.07)	0.079
Female	1.02	(0.58,1.80)	0.95	0.96	(0.63,1.47)	0.86
Birthweight z-score	1.17	(0.90,1.53)	0.25	1.28	(1.04,1.57)	**0.020**
Birthweight categories						
SGA	0.75	(0.26,2.15)	0.59	0.75	(0.34,1.67)	0.48
Normal	Ref	.	.	Ref	.	.
LGA	1.17	(0.44,3.08)	0.75	1.31	(0.68,2.54)	0.43
Maternal age (years)	0.97	(0.92,1.02)	0.23	0.98	(0.95,1.02)	0.37
Maternal age categories						
<25	1.36	(0.72,2.60)	0.35	0.89	(0.54,1.45)	0.63
25-34	Ref		.	Ref		.
>35	0.86	(0.38,1.94)	0.72	0.5	(0.24,1.03)	0.06
Adversity	0.85	(0.18,3.91)	0.83	1.78	(1.12,2.85)	**0.015**
RWG	1.71	(0.77,3.75)	0.19	1.65	(1.04,2.60)	**0.033**
RT	2.24	(1.01,4.96)	**0.046**	2.38	(1.51,3.75)	**<0.001**
Breastfeeding categories						
None	Ref	.		Ref	.	
<6 weeks	2.36	(0.41,13.46)	0.19	0.93	(0.55,1.58)	0.80
6 weeks-4 months	2.72	(0.58,12.74)	0.12	0.47	(0.21,1.05)	0.067
>4 months	2.58	(0.55,12.05)	0.13	0.85	(0.48,1.49)	0.57

OR, Odds ratio; CI, 95% confidence interval; Ref, reference category. * p<0.05, ** p<0.01, *** p<0.001. Bold indicates p<0.05. SGA, Small for gestational age; LGA, Large-for-gestational age; RWG, rapid weight gain; RT, rapid thrive; SES, socioeconomic status.

OWOB was examined further in GMS as group sizes were too small to sufficiently estimate a multivariable model for NTFS. The associations of increased odds of OWOB with higher birthweight and RWG, and reduced odds with increasing SES remained significant after adjustment (*Table 5.9*). RWG had the greatest effect on OWOB at age 9. There were no significant interactions between exposures and SES, with OWOB in GMS. Adjusting for height increased the variation explained and removed significant associations for birthweight.

Table 5.9 Multivariable fully adjusted logistic regression models for OWOB (age 9) in GMS

	GMS (RWG)			GMS (RT)			RT, adjusted for height		
	OR	CI	p	OR	CI	p	OR	CI	p
Bwtz	2.12	(1.47,3.04)	**<0.001**	1.72	(1.25,2.38)	**0.001**	1.35	(0.95,1.94)	0.098
RWG	3.65	(1.73,7.70)	**0.001**						
RT				3.98	(2.06,7.68)	**<0.001**	2.12	(1.02,4.40)	**0.044**
Adversity	1.84	(0.96,3.53)	0.066	1.84	(0.95,3.56)	0.071	1.63	(0.81,3.27)	0.17
SES (age 9)									
Mid	0.37	(0.16,0.88)	**0.024**	0.36	(0.15,0.86)	**0.021**	0.35	(0.14,0.87)	**0.024**
Most advan-taged	0.41	(0.17,0.96)	**0.041**	0.39	(0.16,0.93)	**0.035**	0.33	(0.13,0.82)	**0.017**
Ht (age 9)							1.14	(1.07,1.22)	**<0.001**
n	269			269			269		
Pseudo R^2	0.107			0.126			0.186		
BIC'	36.8			31.3			19.7		

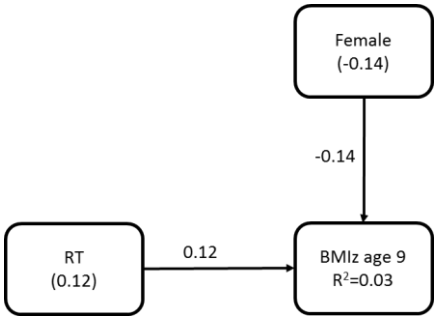
*OR, Odds ratio; CI, 95% confidence interval; Ref, reference category; N, number of observations.. * p<0.05, ** p<0.01, ***, p<0.001. Bold indicates p<0.05. Bwtz, birthweight z-score; RWG, rapid weight gain; RT, rapid thrive; SES, socioeconomic status; ht, height. Models are additionally adjusted for sex, SES at birth, gestation, maternal age. SES reference category is least advantaged. A difference of 5.5 in BIC' provided positive support for the model with rapid thrive rather than RWG.*

5.4.7 Path analysis of the predictors of childhood BMI

Figure 5.1 illustrates the adjusted path models for BMIz. In both cohorts, rapid thrive was directly associated with increased BMIz, but no variables were associated with RT. Sex was associated with decreased BMIz in NTFS, however, similar to the multivariable regression model (Table 5.7), there were no other exposures directly or indirectly associated with BMIz in NTFS.

In GMS, there were direct relationships between birthweight, RT, adversity and height with BMIz. All exposures had significant total effects (the sum of direct and indirect effects) on BMIz except SES at birth. Increasing SES was positively associated with birthweight z-score. RT and birthweight z-score demonstrated positive associations with height. Adversity, RT and birthweight z-score all had similar positive, direct effects on BMIz. Dissimilar to NTFS, the largest total effect came from height in GMS. The GMS model explained 24% of the variation in BMIz, whilst the NTFS model explained a very small proportion of the variation (3%). For the GMS path model there was strong support for the model without SES at age 9 and without breastfeeding, neither of which altered estimates, improved model fit, or increased overall variance explained.

A – NTFS



B - GMS

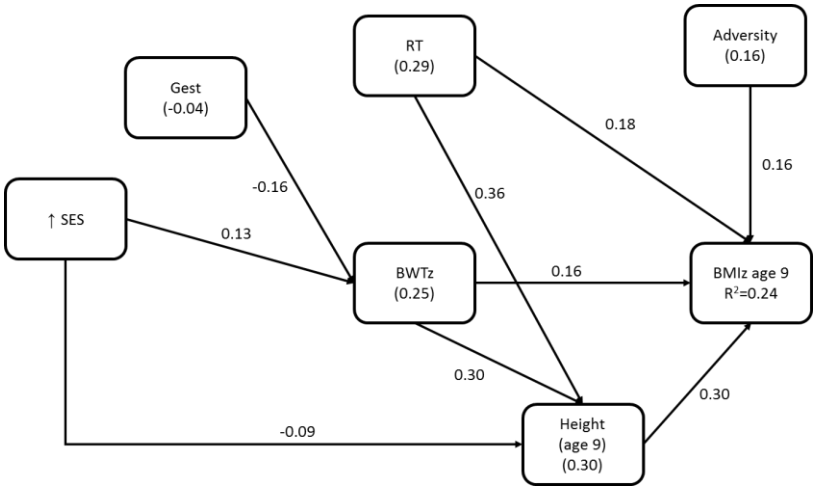


Figure 5.1 The path models for early life factors and BMIz (age 9) for NTFS (A) and GMS (B) cohorts.

Arrows show the direction of the effect. Standardised coefficients are presented, which represent partial regression coefficients between connected variables, controlling for all prior variables (Garson, 2008). All direct effects are represented by solid lines and are significant at p<0.05. Total effects (if significant) are presenting in brackets with associated p values. The NTFS model was additionally adjusted for height, birthweight, maternal age, gestation length, SES (birth) however these had no significant paths that improved model fit (informed by modification indices). The GMS model was additionally adjusted for sex and maternal age (no significant paths). There were significant (p<0.05) indirect paths to BMIz for gestation (-0.04), RT (0.11), birthweight z-score (0.09). Gest, gestation length; SES, socioeconomic status (birth); RT, rapid thrive.

5.4.8 Investigating rapid thrive

RT was the risk factor associated with increased odds of OWOB and higher BMIz in both cohorts. There were no variables that predicted RT (Figure 5.1) in the cohorts.

The effects of RT on mean BMI z-scores over time were examined further by cohort (Figure 5.2). By definition, those with RT show a sizeable increase in z-score from birth to 12 months. Within cohorts, those who had RT tended to have higher BMI z-scores throughout childhood than those who did not. However, in NTFS, at age 13 (the last available time point before adulthood) although the z-score was greater, there was no significant difference between those who had RT and those who did not. Similar to NTFS, in GMS those who had

RT had a large initial increase in z-score (0-12 months) (albeit to a lower average z-score), however subsequently they have a higher BMIz throughout childhood and an overweight BMIz (z-score>=1).

Those who did not have RT demonstrate a decrease in z-score after birth, which over time stabilises around the average (z-score of 0) in NTFS, but in GMS gradually increases.

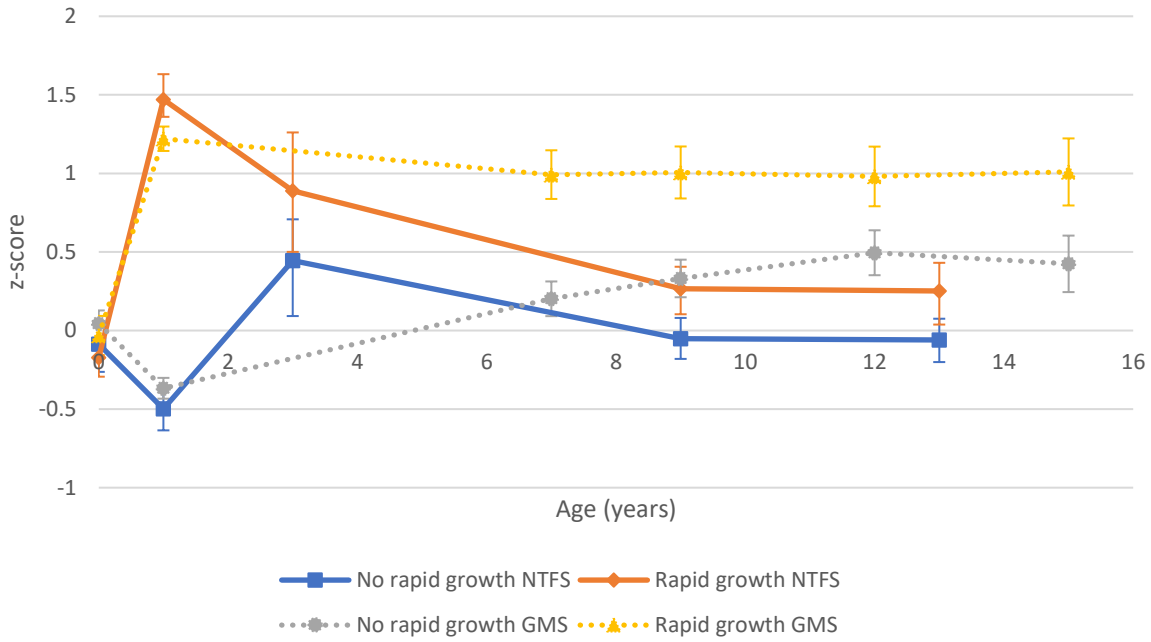


Figure 5.2 The change in average z-score over by cohort and RT
Error bars represent 95% confidence intervals. Average values were used for the interval ages for GMS.

5.5 Discussion

These data provided an opportunity to analyse how early life factors influencing BMI have changed over a long time period in one geographical area. There were more early life exposures significantly associated with childhood body composition and they explained more variation in BMIz in the modern (GMS) cohort.

Early life and childhood experiences of NTFS study members were vastly different to those of GMS children and warrant further discussion to provide context for the results. Food was heavily regulated due to state enforced rationing leading to reductions of sugar, meat and fats and individuals were entitled to their fair share (Zweiniger-Bargielowska, 2000). Rationing did not cease immediately after the war but continued until 1954. This reduced disparities of the previous decade between the social classes regarding energy and nutrient intakes, with calorie and protein intakes between working and middle classes stable between 1944-1956 (Zweiniger-Bargielowska, 2000). This period of time also saw an increase in activity levels due to longer working hours, more women in the workplace and

reduced vehicle usage as a result of petrol rationing. These factors will have impacted both the mother during pregnancy and also the offspring during the first few years of life.

A comparison between two of the British birth cohorts found those born in 1958 were not heavier than those born in 1946 until adolescence, when the weight trajectories for the later cohort had a faster rate of gain (Li et al., 2008). Further differences emerged by mid-adulthood in measures of waist and hip circumferences, with higher rates of obesity in the 1958 cohort. Compared to modern day, this cohort experienced the obesity epidemic at a later age (around mid-thirties) meaning a shorter exposure to an obesogenic environment. This is consistent with results here, whereby obesity did not emerge until later life (chapter 3). This coincides with a time of technological change, whereby households and industry were shifting toward labour-saving devices (Lakdawalla and Philipson, 2009), therefore lifestyles became more sedentary. This additional time also meant more women could take up employment, and with this came a change in what was deemed traditional home cooking and diet.

Post-war was a time of rising affluence in Britain, which was seen here in the social mobility of NTFS. The boom of the economy was associated with increased travel due to decreases in transport costs and the accompanying introduction of new, cheaper foods. The food environment shifted; with the focus was no longer solely on sustenance, and restaurant and fast food chains began to emerge. These factors could partly explain the latency in obesity onset.

Early life factors previously explained a relatively small proportion of the variation in BMI (2% in NTFS compared to 22% in GMS), suggesting that early life factors in combination with modern environments, have become more important over time in determining childhood body composition.

5.5.1 Summary of results

At baseline, the two cohorts differed in birthweight, maternal age, gestation length, RWG, duration breastfed, adversity, and SES. On average birthweight increased, there were fewer older mothers and more young mothers, RT was less likely, breastfeeding duration was shorter and more experienced adversity. Height has also increased vastly between the two cohorts, and in the modern cohort was associated with increased BMI and odds of OWOB (in

GMS). The social class differences in height seen in NTFS have diminished over time. However, in the modern cohort socioeconomic differences in OWOB have emerged.

There were large differences in birthweight between the cohorts, and birthweight was negatively associated with RWG (section 4.4.9), therefore RT was also analysed to further understand the effects of conditional weight gain. RT corrected for some of the effects of catch-up growth (from low birthweight), suggesting that RWG encompassed some of the indirect, and potentially developmental programming effects of birthweight. RT was associated with increased BMIz in both cohorts, however in GMS the effects were greater.

There were socioeconomic gradients in height in NTFS, in accordance with earlier observations by Wright and Parker (2004). These differences were much smaller in GMS. Indeed, socio-economic differences in height, which have decreased over, time (Bann et al 2018). Height modulated the effects of many exposures on BMIz (notably for NTFS), which may suggest that some of these associations were related to lean mass rather than fat mass. Other studies have suggested that the relationship between RWG and subsequent adiposity is merely a marker of fast growth and later height (Wright et al., 2012). This analysis instead suggests instead that the effects of rapid growth were previously unrelated to height (NTFS), but are now somewhat mediated through height (GMS).

Both higher birthweights and RT were associated with increased height in GMS, which may suggest they are important growth-related factors on the pathway to obesity, however it is difficult to untangle the complex relationships between height, BMI and SES. Adjusting for height could also introduce bias, if it lies on the causal pathway. There is a lack of consistency as to whether other body composition measures (such as height) should be further adjusted for (Tu et al., 2005).

Although anthropometric measures were utilised at age 9 to minimise the bias from puberty, there are reports of children entering puberty as young as age 6 years (Herman-Giddens et al., 1997), therefore some children may have been more developed. Early onset of puberty may be more likely in children with a higher BMI (Freedman et al., 2002, Kaplowitz, 2008, Li et al., 2017b), therefore it may not be appropriate to adjust for pubertal status if BMI is causally related to early puberty (Mumby et al., 2011).

As there were few OWOB study members in NTFS it was not possible to analyse OWOB in a multivariable model. However, this highlights that one of the main differences between the

cohorts is the relatively low levels of OWOB in the historical cohort. In spite of this, the bivariate analyses suggested that RT that was associated with OWOB in NTFS, which is in line with findings for BMIz.

Rapid infancy growth was the prominent factor in these analyses, due to the positive associations with BMIz, and also OWOB in the modern cohort, suggesting that despite the changing environment, RT has remained a consistent risk factor over time. Whereas adversity and birthweight may perhaps be risk factors exacerbated by modern-day obesogenic environments.

Chapter 6. The Avon Longitudinal study of parents and children (ALSPAC)

6.1 Introduction

The previous chapters determined the risk factors for childhood and adult obesity in the North East cohorts. The risk factors, which were consistently associated with adiposity across both cohorts, were birthweight, RWG, maternal age, and SES. It was necessary to examine RWG conditional on birthweight (rapid thrive), in the cohort comparison due to the large differences in birthweight, although findings were similar. Adversity was associated with multiple outcome measures in GMS. Recorded bacterial infection was associated with increased odds of obesity in adults in NTFS, however the GMS questionnaire data on infections was unreliable, therefore it remains uncertain whether the association was present in GMS children.

A strength of the North East region cohorts is that despite the time differences between them, there were similarities in the early life influences on childhood BMI (Chapter 5). Although they include comprehensive data, a limitation of the North East cohorts is that they are relatively small samples in comparison to the larger epidemiological cohorts. Replication of the findings in a larger cohort, from another part of the UK, would reinforce that these exposures are indeed early life risk factors for subsequent obesity. To address these shortcomings, this chapter used data from the ALSPAC cohort, a large, longitudinal birth cohort from the South West of England with vast exposure data (Boyd et al., 2013).

This chapter addresses the third aim of the thesis and focuses on the analysis of early life risk factors in the ALSPAC cohort to identify gene-specific methylation differences in relation to focal early life exposures. Firstly, the epidemiological associations between focal early life exposures and body composition outcomes in childhood and late adolescence were examined. For consistency in methods and exposures, the exposures studied in NTFS and GMS were harmonised with ALSPAC data. The ALSPAC cohort also has epigenetic data, thereby allowing early life exposures to be investigated with respect to changes in DNAm.

The second section of this chapter examines the hypothesis that epigenetics is a mechanism linking early life factors and later life obesity. Whilst DNAm biomarkers have been identified for some exposures, others have either; not been investigated, have been measured

differently, or not been validated (section 1.5.3). Less is known regarding whether these early life exposures leave lasting, measurable effects on DNAm in later life. Thereby these data provide the opportunity to determine the focal early life exposures, harmonised across cohorts, and examine these with respect to methylation in childhood and adolescence. This analysis could elucidate whether these risk factors are mediated through epigenetic programming and if the epigenetic marks are stable over time.

6.2 Aims

The first aim of this chapter was to establish whether early life risk factors for subsequent obesity in the North East cohorts are also risk factors in the ALSPAC cohort. To address this aim, early life exposures were investigated with respect to body composition in childhood and adolescence, using statistical regression methods.

The second aim addresses the overall thesis aim 3(i); to determine which of the focal early life exposures (from the analyses in NTFS and GMS cohorts) demonstrate a relationship with DNAm childhood or adolescence. This was addressed using epigenome-wide analyses at individual CpG loci for each exposure.

6.3 Participants and methods

6.3.1 Data selection and recoding

The ALSPAC data dictionary was used to identify and match the exposures with the variables from NTFS and GMS. For the early life variables, data from mother-focused questionnaires at 18 and 32 weeks prenatal, and 8 weeks postnatal were used, along with data from the child-focused questionnaires at 4 weeks, 15 months and age 17. Partner questionnaire data were used at age 8 for childhood SES. Clinic data from ages 7 and 17 were used for body composition measures.

BMIz and OWOB were utilised as outcomes in these analyses, in line with the outcomes that were comparable across all three cohorts and time points. These were coded as described in section 2.2.1.

There were several early life risk factors with demonstrated consistent associations in GMS and NTFS, and for which data were available in the ALSPAC cohort. These included; birthweight, RWG, RT, infant feeding, parity, maternal age, adversity, antibiotics exposure

(proxy for bacterial infection), and SES. The definitions and measurement of these factors is outlined in 2.2.3.

Social class in childhood were based on partner's self-reported data at age 8 using the 1991 British Office of Population and Census Statistics (OPCS) classification (see section 2.1.3 for details), as there were no maternal data at this time point.

6.3.1.1 ARIES recoding

ALSPAC collected blood at multiple time points. DNA was extracted and DNAm was measured for 1000 mother-child pairs for those who had DNA available at birth, age 7 and ages 15 (TF3) or 17. The age 7 sample was used to represent childhood methylation levels. The age 17 sample was used as this sample was more representative of adults, and study members are likely to have finished the majority of their childhood growth (whereas the TF3 sample included those as young as 14.6 years).

EWAS were run using the meffil R program (detailed in the next section), which allows binary or continuous variables as inputs in models, and therefore some categorical explanatory variables were recoded. The variables: RWG, antibiotic exposure (0-6 months), prenatal adversity and postnatal adversity remained coded as previously described (section 6.3.1).

Birthweight was associated with outcomes across the cohorts, however this exposure has been studied previously in the ALSPAC cohort and no significant loci (for childhood methylation) were identified (Simpkin et al., 2015). As birthweight was not examined as an independent exposure, both RWG and RT were examined which may somewhat differentiate birthweight effects. There was a lower risk of adiposity for mid and most advantaged compared to the least, therefore SES was recoded to a binary variable for "least advantaged" (increased risk of adiposity). Maternal age was examined as a separate binary variables of either young (<25 years), or advanced (>45 years) maternal age (compared to the reference category of 25-34 years).

6.3.2 Statistical analysis

The ALSPAC epidemiological data was analysed using the same statistical approach as in the cohort comparison (of GMS and NTFS) (see section 2.3.1). Briefly, this firstly involved examining bivariate relationships for all focal early life risk factors. The effects of SES on the coefficients were examined sequentially. Then, the same multivariable models were run as that for the cohort comparison (chapter 5) for means of comparison. However, this resulted

in models that did not meet regression criteria. Therefore, to find the best predictive model for the explanatory variables under investigation, regression models were re-estimated using stepwise regression. The BIC was used as a decisive factor for model selection, with lower model BIC preferred. The aim was to form a parsimonious model, i.e. a model that has the greatest explanatory ability, using the least number of variables necessary.

6.4 Results

6.4.1 ALSPAC sample characteristics

6.4.1.1 Outcomes

There were a total of 7,868 (50.5% female) participants at the age 7 follow-up, and 4,858 (43.8% female) at age 17 with body composition measurements (*Table 6.1*). Using BMIz to determine weight categories, 13% of study members had an OWOB BMIz at age 7, and 21% at age 17 (Table 6.1). Accordingly, mean BMIz was also higher at age 17.

Table 6.1 Descriptive characteristics of body composition outcomes for ALSPAC participants age 7 and 17

	Age 7		Age 17	
	n	Mean (SD)	n	Mean (SD)
BMIz	7,868	0.19 (1.01)	4,858	0.47 (1.13)
	n	%	n	%
Healthy weight	6,841	86.95	3,827	78.78
OWOB	1,027	13.05	1,031	21.22
Total	7,868		4,858	

Proportion (%) of study members in healthy weight or overweight/obese (OWOB) and mean BMIz and standard deviation (SD). n, sample size; %, column percentage.

6.4.1.2 Exposures

Descriptive statistics for those with data available at each follow-up are shown in

Table 6.2. Compared to the original cohort, at both ages 7 and 17; there were more females, more first-born children, fewer younger mothers and more who experienced adversity. There were also more SGA and fewer LGA, however overall mean birthweight z-score and gestational length was greater. The sample was also more advantaged at both time points. At age 17, there were fewer who had antibiotics compared to the original cohort. There were no differences for RWG at either time point or antibiotics at age 7.

Table 6.2 Summary characteristics of ALSPAC participants (all) and those with body composition measures (age 7 and 17)

	All		Age 7			Age 17		
Continuous variable	n	Mean (SD)	n	Mean (SD)	T-test p	n	Mean (SD)	T-test p
Birthweight (z-score)	13457	0.09 (1.0)	7336	0.15 (0.97)	**<0.001**	4486	0.15 (0.98)	**<0.001**
Gestation (weeks)	14178	38.39 (5.56)	7428	39.5 (1.8)	**<0.001**	4538	39.5 (1.75)	**<0.001**
Categorical variable	All		Age 7			Age 17		
	n	col %	n	col %	Chi^2 p	n	col %	Chi^2 p
Sex	14409		7868		**0.050**	4858		**<0.001**
Male	7387	51.3	3975	50.5		2127	43.8	
Female	7022	48.7	3893	49.5		2731	56.2	
First-born	11356		6953		**0.002**	4262		**<0.001**
No	6277	55.3	3763	54.1		2211	51.9	
Yes	5079	44.7	3190	45.9		2051	48.1	
Birthweight categories	13457		7336		**<0.001**			**0.004**
SGA	1467	10.9	852	11.6		529	11.8	
Normal	10913	81.1	6008	81.9		3637	81.1	
LGA	1077	8	476	6.5		320	7.1	
Gestation categories	14178		7428		**<0.001**			**<0.001**
Pre-term	2027	14.3	728	9.8		445	9.8	
Normal	11116	78.4	6147	82.8		3770	83.1	
Post-term	1035	7.3	553	7.4		323	7.1	
RWG	1184		853		0.76			0.67
No	822	69.4	590	69.2		375	68.8	
Yes	362	30.6	263	30.8		170	31.2	
RT	1,184		853		0.51	545		0.69
No	813	68.7	581	68.1		371	68.1	
Yes	371	31.3	272	31.9		174	31.9	
Maternal age at birth	13641		7428		**<0.001**	4538		**<0.001**
Less than 25	3309	24.3	1180	15.9		706	15.6	
25-34	8981	65.8	5341	71.9		3243	71.5	
35+	1351	9.9	907	12.2		589	13	
Adversity pre-natal	14997		7868		**<0.001**	4858		**<0.001**
No	10347	69	5183	65.9		3232	66.5	
Yes	4650	31	2685	34.1		1626	33.5	
Adversity post-natal	14997		7868		**<0.001**	4858		**<0.001**
No	11415	76.1	5628	71.5		3553	73.1	
Yes	3582	23.9	2240	28.5		1305	26.9	
Pre and post-natal	14997		7868		**<0.001**	4858		**0.001**
No	12960	86.4	6645	84.5		4132	85.1	
Yes	2037	13.6	1223	15.5		726	14.9	
Antibiotics (0-6 months)	11016		6923		0.48	4243		**0.031**
No	7460	67.7	4705	68		2925	68.9	
Yes	3556	32.3	2218	32		1318	31.1	
SES at birth	10540		6678		**<0.001**	4115		**<0.001**
Least Advantaged	589	5.6	273	4.1		163	4	
Mid	4058	38.5	2324	34.8		1323	32.2	
Most advantaged	5893	55.9	4081	61.1		2629	63.9	
SES in childhood	3553		3038		**0.007**	2069		**<0.001**
Least Advantaged	413	11.6	333	11		215	10.4	
Mid	1079	30.4	920	30.3		595	28.8	
Most advantaged	2061	58	1785	58.8		1259	60.9	

Differences between the original sample (all), and either the age 7 and 17 samples, assessed using T-tests for continuous variables, and Chi-square (Chi^2) tests for differences between the whole sample and those measured at either age 7 or 17. n, sample size; col %, column percentage; SD, standard deviation. Bold indicates significant at p<0.05.

6.4.2 Associations between early life factors and adiposity outcomes

In bivariate analyses, birthweight z-score and LGA were associated with increased BMIz and increased odds of OWOB at both time points (Appendix E,

Table XI). SGA was negatively associated, and gestation length positively associated with BMIz at age 7, but there were no associations for OWOB or outcomes at age 17. RWG was positively associated with BMIz at age 7 and 17, and OWOB at age 7. Younger maternal age (<25 years) was associated with higher BMIz (0.24, p<0.001) and 50% increased odds of OWOB at age 17. Whilst older maternal age (35+) was associated with a lower BMIz at age 17.

Interestingly, pre-natal and pre and post-natal adversity (collectively) were associated with BMIz at age 7 and 17, and OWOB at age 17, but not OWOB at age 7 (Appendix E,

Table XI). At age 7, the coefficients were small, almost doubling by age 17. The effect sizes were greater for pre- and post-natal adversity combined. There were no associations for post-natal adversity alone.

SES at birth and in childhood were not directly associated with BMIz (age 7), but the most advantaged group (both time points) was associated with lower BMIz and OWOB at age 17, with similar effect sizes. In addition, the association for childhood SES and OWOB was also seen at age 7.

Table 6.3 Summary of ALSPAC bivariate associations for early life exposures and outcomes (age 7 and 17)

Early life exposures		Outcome measures			
		BMI		OWOB	
		Age 7	Age 17	Age 7	Age 17
Maternal	Age	-	✓ (young & old)	-	✓ (young)
Birth	Birthweight	✓	✓	✓	✓
	Occupational social class	-	✓	-	✓
Early life	Antibiotics (0- 6 months)	-	-	✓	✓
	Adversity Prenatal	✓	✓	-	✓
	Postnatal	-	-	-	-
	Pre and postnatal	✓	✓	-	✓
	Rapid weight gain	✓	✓	✓	-
	Rapid thrive	✓	✓	✓	-
Childhood	Occupational social class	-	✓	✓	✓

Tick represents a significant association between exposure and specified outcome measure in the (unadjusted) bivariate model

6.4.2.1 *Multivariable cohort-comparative models*

In order to compare models with those in GMS and NTFS cohort comparison, RT was used in the multivariable models (for models with RWG see Appendix E). In the multivariable analyses, the combined pre and post-natal adversity variable was used as it had the largest coefficient and was most significant in the bivariate models. As a sensitivity test, additional models were estimated with the other adversity variables as controls, however they were not significantly associated and did not improve model fit. Similar to analyses in the cohort comparison, the impact of childhood SES on the early life factors was investigated further.

At age 7 the associations for RT decreased but remained after adjustment for SES at both time points. Associations for birthweight were similar but were not significant for OWOB (age 7) after adjustment for childhood SES. After adjustment, RT was the only early life factor associated with both outcomes in childhood.

The combined multivariable models including early life factors and SES explained more the variation in adiposity outcomes in childhood. All age 7, BMI models fit regression criteria (regression diagnostic plots in Appendix E). However, the model for BMIz (age 7) had slightly positively skewed residuals, therefore ordinal regression (OWOB age 7 multivariable model) results which do not assume normality of residuals may be more valid.

All models for OWOB (age 17) had non-significant p values and very low pseudo R^2 value which indicates that the early life factors included were poor at predicting OWOB in adolescence.

There was also very strong support for all models without adjustment for SES in childhood (informed by BIC). There was also a large decrease in sample size, which may be also result in poor model specification. Therefore, instead parsimonious models were estimated (section 6.4.2.2).

Table 6.4 Multivariable linear and logistic regression for BMIz and OWOB at ages 7 and 17, adjusted for SES

Age 7										
BMIz		Unadjusted			Adjusted for SES (birth)			Adjusted for SES (childhood)		
		Coef	CI	p	Coef	CI	p	Coef	CI	p
Female		-0.06	[-0.19,0.07]	0.37	-0.06	[-0.20,0.07]	0.34	-0.05	[-0.25,0.14]	0.61
Gestation		0.01	[-0.03,0.06]	0.56	0.01	[-0.03,0.06]	0.6	0.01	[-0.05,0.08]	0.69
Birthweight z-score		0.19	[0.12,0.26]	**<0.001**	0.18	[0.11,0.26]	**<0.001**	0.16	[0.06,0.27]	**0.003**
Maternal age	Less than 25	-0.04	[-0.25,0.16]	0.68	0.06	[-0.17,0.29]	0.60	0.35	[-0.07,0.77]	0.10
	35+	-0.13	[-0.32,0.06]	0.17	-0.13	[-0.33,0.07]	0.20	0.02	[-0.26,0.29]	0.90
First-born		0.07	[-0.07,0.21]	0.31	0.05	[-0.09,0.19]	0.48	0.12	[-0.09,0.34]	0.26
RT		0.77	[0.63,0.91]	**<0.001**	0.77	[0.62,0.91]	**<0.001**	0.67	[0.46,0.87]	**<0.001**
Adversity		0.02	[-0.15,0.19]	0.83	-0.03	[-0.22,0.15]	0.72	0.04	[-0.25,0.32]	0.80
SES (birth)	Mid				0.27	[-0.08,0.61]	0.13	-0.17	[-0.81,0.47]	0.60
	Most advantaged				0.28	[-0.06,0.62]	0.11	-0.06	[-0.70,0.57]	0.85
SES childhood	Mid							0.13	[-0.24,0.51]	0.49
	Most advantaged							-0.01	[-0.38,0.35]	0.94
n		825			772			382		
Adjusted R^2		0.158			0.154			0.112		

OWOB		Unadjusted			Adjusted for SES (birth)			Adjusted for SES (childhood)		
		OR	CI	p	OR	CI	p	OR	CI	p
Female		0.66	[0.42,1.02]	0.063	0.66	[0.42,1.04]	0.076	0.77	[0.38,1.54]	0.46
Gestation		1.01	[0.87,1.16]	0.92	1.02	[0.88,1.18]	0.82	1.01	[0.80,1.26]	0.95
Birthweight z-score		1.41	[1.11,1.80]	**0.005**	1.31	[1.02,1.68]	**0.035**	0.97	[0.66,1.41]	0.86
Maternal age	Less than 25	0.85	[0.43,1.69]	0.65	1.14	[0.55,2.36]	0.72	1.77	[0.53,5.83]	0.35
	35+	0.62	[0.29,1.31]	0.21	0.56	[0.26,1.24]	0.15	0.29	[0.07,1.28]	0.10
First-born		1.15	[0.73,1.83]	0.55	1.09	[0.68,1.76]	0.71	1.11	[0.53,2.36]	0.78
RT		3.93	[2.54,6.08]	**<0.001**	3.77	[2.41,5.89]	**<0.001**	3.01	[1.54,5.88]	**0.001**
Adversity		0.99	[0.55,1.77]	0.97	0.93	[0.49,1.74]	0.81	1.61	[0.67,3.88]	0.29
SES (birth)	Mid				1.33	[0.36,4.89]	0.66	0.38	[0.06,2.58]	0.32
	Most advantaged				1.67	[0.46,6.07]	0.44	0.64	[0.10,4.32]	0.65
SES childhood	Mid							1.08	[0.33,3.58]	0.90
	Most advantaged							0.6	[0.19,1.93]	0.40
n		825			772			382		
Pseudo R^2		0.091			0.086			0.085		

Age 17										
BMIz		Unadjusted			Adjusted for SES (birth)			Adjusted for SES (childhood)		
		Coef	CI	p	Coef	CI	p	Coef	CI	p
Female		-0.08	[-0.27,0.12]	0.43	-0.1	[-0.30,0.10]	0.33	-0.08	[-0.35,0.20]	0.59
Gestation		0	[-0.06,0.07]	0.91	0	[-0.07,0.07]	0.98	0.01	[-0.09,0.10]	0.91
Birthweight z-score		0.06	[-0.05,0.17]	0.30	0.05	[-0.07,0.16]	0.42	0.09	[-0.07,0.24]	0.27
Maternal age	Less than 25	0.2	[-0.12,0.51]	0.23	0.27	[-0.09,0.63]	0.14	0.45	[-0.24,1.15]	0.20
	35+	-0.11	[-0.39,0.18]	0.47	-0.08	[-0.38,0.22]	0.59	0.04	[-0.35,0.44]	0.83
First-born		0.04	[-0.17,0.25]	0.72	0.01	[-0.21,0.24]	0.90	0.11	[-0.19,0.41]	0.48
RT		0.21	[0.00,0.43]	**0.05**	0.23	[0.00,0.45]	**0.045**	0.18	[-0.11,0.47]	0.22
Adversity - Pre and post-natal		0.31	[0.03,0.59]	**0.03**	0.26	[-0.04,0.56]	0.088	0.19	[-0.22,0.61]	0.36
SES (birth)	Mid				0.42	[-0.14,0.98]	0.14	-0.15	[-1.13,0.83]	0.77
	Most advantaged				0.29	[-0.26,0.84]	0.29	-0.46	[-1.43,0.52]	0.36
SES (childhood)	Mid							-0.22	[-0.79,0.35]	0.44
	Most advantaged							-0.1	[-0.64,0.44]	0.71
n		527			498			270		
Adjusted R^2		0.011			0.009			0.001		

OWOB		Unadjusted			Adjusted for SES (birth)			Adjusted for SES (childhood)		
		OR	CI	p	OR	CI	p	OR	CI	p
Female		0.98	[0.63,1.52]	0.93	0.94	[0.60,1.47]	0.79	0.94	[0.49,1.81]	0.86
Gestation		1.02	[0.88,1.19]	0.77	1.02	[0.88,1.19]	0.81	0.99	[0.80,1.24]	0.96
Birthweight z-score		1.18	[0.92,1.51]	0.19	1.16	[0.90,1.49]	0.25	1.15	[0.79,1.67]	0.47
Maternal age	Less than 25	1.13	[0.56,2.27]	0.74	1.2	[0.56,2.59]	0.64	1.34	[0.30,5.94]	0.70
	35+	1.38	[0.75,2.53]	0.30	1.43	[0.77,2.66]	0.26	1.93	[0.83,4.54]	0.13
First-born		0.97	[0.60,1.57]	0.91	0.96	[0.59,1.58]	0.88	1.11	[0.54,2.28]	0.77
RT		1.54	[0.97,2.45]	0.066	1.6	[1.00,2.56]	**0.05**	1.34	[0.69,2.63]	0.39
Adversity - Pre and post-natal		1.79	[1.02,3.14]	**0.044**	1.61	[0.88,2.94]	0.12	1.38	[0.56,3.41]	0.48
SES (birth)	Mid				1.64	[0.44,6.18]	0.46	1.41	[0.14,14.34]	0.77
	Most advantaged				1.48	[0.40,5.47]	0.56	0.78	[0.08,8.05]	0.83
SES (childhood)	Mid							1.05	[0.28,3.87]	0.94
	Most advantaged							1.13	[0.32,3.94]	0.85
n		527			498			270		
Pseudo R^2		0.02			0.02			0.026		

Coefficients (coef) and Odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p).Bold indicates p<Reference categories were; the least advantaged group for SES; and age 25-34 for maternal age. n, sample size; Bwtz, birthweight z-score; SES, socioeconomic status; RT, rapid thrive. The overall model p value was no longer significant for BMIz (age 17) and OWOW (age 7) after controlling for SES in childhood. There were indications that the models for OWOB at age 7 and 17 did not meet logistic regression criteria and were poorly specified models.

6.4.2.2 *Multivariable cohort-specific models*

In the previous section, models were estimated to harmonise with those in chapter 4. However, these models did not meet regression criteria, suggesting that using these variables did not result in a good predictive model for the ALSPAC cohort. Models were refitted using a stepwise procedure, informed by goodness of fit statistics (BIC) to create parsimonious models (Table 6.5 and Table 6.6). Model BIC provided strong support for models (BMIz7 and BMIz17) without SES at either time point. In the BMIz17 model, prenatal adversity had better model fit than when using the pre- and post-natal adversity variable, and postnatal adversity was not a significant predictor. For regression diagnostic plots for these models see Appendix E, Figure XVI.

In the parsimonious adjusted models (*Table 6.5* and *Table 6.6*), birthweight was positively associated with outcomes at age 7. RT was associated with increased BMIz (at both ages), and increased odds of OWOB at age 7. Pre-natal adversity was associated with BMIz at age 17, but not with outcomes at age 7 or OWOB at age 17.

There were few differences for the models using RWG, and similar to RT, RWG was the most significant factor associated with body composition (for RWG models see Appendix E).

Early life factors explained greater variation in body composition in childhood, particularly for BMIz at age 7 (R^2=0.16).

Table 6.5 Multivariable parsimonious linear regression models for BMIz at ages 7 and 17 in ALSPAC participants

	BMIz7			BMIz17		
	Coef	CI	p	Coef	CI	p
Female	-0.06	[-0.19,0.07]	0.36	-0.09	[-0.28,0.10]	0.36
Gestation length (weeks)	0.01	[-0.03,0.06]	0.57	0	[-0.07,0.07]	0.98
First-born	0.07	[-0.07,0.20]	0.32	0.03	[-0.18,0.24]	0.78
Maternal age (categories)						
Less than 25	-0.04	[-0.24,0.16]	0.68	0.2	[-0.12,0.51]	0.22
35+	-0.13	[-0.32,0.06]	0.17	-0.1	[-0.38,0.19]	0.51
Birthweight z-score	0.19	[0.12,0.26]	**<0.001**	0.05	[-0.06,0.16]	0.35
RT	0.77	[0.63,0.91]	**<0.001**	0.23	[0.02,0.45]	**0.03**
Adversity - prenatal				0.33	[0.12,0.54]	**0.002**
n	825			527		
Adjusted R^2	0.159			0.021		

Coefficients (coef) presented with 95% confidence intervals (CI) and the corresponding level of significance (p). The reference category was age 25-34 for maternal age. n, sample size; RT, rapid thrive. All variables included in the model are presented. Bold indicates p<0.05.

Table 6.6 Multivariable parsimonious logistic regression models for OWOB at ages 7 and 17 in ALSPAC study members

	OWOB7			OWOB17		
	OR	CI	p	OR	CI	p
Female	0.66	[0.42,1.02]	0.063	0.95	[0.50,1.82]	0.88
Gestation length (weeks)	1.01	[0.88,1.16]	0.91	1	[0.80,1.24]	0.96
First-born	1.15	[0.73,1.83]	0.54	1.1	[0.54,2.27]	0.79
Maternal age (categories)						
Less than 25	0.85	[0.43,1.69]	0.65	1.43	[0.33,6.30]	0.63
35+	0.62	[0.29,1.31]	0.21	1.98	[0.85,4.63]	0.12
Birthweight z-score	1.41	[1.11,1.80]	**0.005**	1.14	[0.79,1.66]	0.48
RT	3.93	[2.54,6.08]	**<0.001**	1.37	[0.70,2.67]	0.36
Adversity - prenatal				0.92	[0.46,1.87]	0.82
SES at birth						
Mid				1.47	[0.14,15.25]	0.75
Most advantaged				0.79	[0.08,8.32]	0.85
SES in childhood						
Mid				1.08	[0.29,3.98]	0.91
Most advantaged				1.14	[0.33,3.97]	0.83
n	825			270		
Pseudo R^2	0.091			0.024		

Odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Bold indicates p<0.05. Reference categories were; the least advantaged group for SES; and age 25-34 for maternal age. n, sample size; SES, socioeconomic status; RT, rapid thrive. All variables included in the model are presented.

6.4.2.3 Sensitivity analysis

Maternal BMI and maternal smoking during pregnancy are two factors known to influence offspring body composition and health. Data on these two important factors were not available in NTFS or GMS, therefore the ALSPAC dataset provided an opportunity to assess how these factors could influence the associations between early factors and BMI or OWOB (Table 6.7 and Table 6.8).

Adjusting for smoking made little difference to most estimates. Adjusting for smoking increased the OR for RT for OWOB (age 7), but the coefficient for BMI remained similar. Therefore, without adjustment the effect was underestimated.

Although the literature suggests that maternal BMI and smoking can affect birthweight, adjusting for these factors had no effect on the estimates for birthweight on the models at age 7.

Models including maternal BMI and smoking explained 8% and 14% more of the variation in BMIz, and 6% and 13% more variation in OWOB, at ages 7 and 17 respectively (*Table 6.7*).

Table 6.7 Sensitivity analyses including maternal factors for the multivariable linear regression (parsimonious) models for BMIz at ages 7 and 17 in ALSPAC study members

	BMIz (age 7)			BMIz (age 17)		
	coef	CI	p	coef	CI	p
Female	-0.06	[-0.19,0.06]	0.33	-0.06	[-0.25,0.13]	0.51
Gestation length (weeks)	0	[-0.04,0.05]	0.90	-0.03	[-0.09,0.03]	0.36
First-born	0.06	[-0.07,0.20]	0.35	0.08	[-0.13,0.28]	0.45
Maternal age (categories)						
Less than 25	-0.09	[-0.30,0.12]	0.39	0.11	[-0.21,0.44]	0.49
35+	-0.12	[-0.31,0.07]	0.21	-0.09	[-0.37,0.19]	0.52
Birthweight z-score	0.18	[0.11,0.25]	**<0.001**	0.02	[-0.09,0.13]	0.71
RT	0.76	[0.62,0.90]	**<0.001**	0.22	[0.02,0.43]	**0.033**
Adversity - prenatal	-	-	-	0.32	[0.12,0.52]	**0.002**
Pre-pregnancy BMI	0.06	[0.05,0.08]	**<0.001**	0.10	[0.08,0.12]	**<0.001**
Maternal smoking	0.07	[-0.11,0.24]	0.46	-0.23	[-0.51,0.06]	0.12
n	763			492		
Adjusted R^2	0.237			0.158		

Coefficients (coef) presented with 95% confidence intervals (CI) and the corresponding level of significance (p). The reference category was age 25-34 for maternal age. n, sample size; SES, socioeconomic status; RT, rapid thrive.

Table 6.8 Sensitivity analyses including maternal factors for the multivariable logistic regression (parsimonious) models for OWOB at ages 7 and 17 in ALSPAC study members

	OWOB (age 7)			OWOB (age 17)		
	OR	CI	p	OR	CI	p
Female	0.60	[0.37,0.98]	**0.039**	1.12	[0.54,2.31]	0.76
Gestation length (weeks)	0.95	[0.82,1.10]	0.47	0.91	[0.71,1.15]	0.43
First-born	1.18	[0.72,1.94]	0.51	1.01	[0.45,2.25]	0.98
Maternal age (categories)						
Less than 25	0.66	[0.30,1.46]	0.31	1.31	[0.23,7.48]	0.76
35+	0.65	[0.30,1.43]	0.29	1.35	[0.50,3.66]	0.55
Birthweight z-score	1.39	[1.07,1.80]	**0.012**	0.91	[0.60,1.39]	0.67
RT	4.53	[2.82,7.29]	**<0.001**	1.40	[0.66,2.95]	0.38
Adversity - prenatal				0.98	[0.44,2.16]	0.96
SES at birth				2.96	[0.17,52.75]	0.46
Mid				2.2	[0.12,39.43]	0.59
Most advantaged						
SES in childhood				1.13	[0.27,4.80]	0.86
Mid				0.9	[0.23,3.50]	0.88
Most advantaged						
Pre-pregnancy BMI	1.14	[1.09,1.20]	**<0.001**	1.27	[1.15,1.39]	**<0.001**
Maternal smoking	1.12	[0.58,2.16]	0.74	0.27	[0.04,1.70]	0.16
n	763			257		
pseudo R^2	0.147			0.154		

Odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Reference categories were; the least advantaged group for SES; and age 25-34 for maternal age. n, sample size; SES, socioeconomic status; RT, rapid thrive.

6.4.3 ARIES sample characteristics

Demographic characteristics of the ARIES sample are shown for the total sample and for each exposure. Sample sizes varied according to the exposure investigated (ranged from *n*=116 to 873, at age 7 with cells counts included) and were smaller at age 17 due to attrition.

Table 6.9 Descriptive statistics for all early life exposures in ARIES at ages 7 and 17, for models either with or without cell counts.

		Models adjusted for cell counts					**Models not adjusted for cell counts**				
Age	Variable	Total	No	%	Yes	%	Total	No	%	Yes	%
7	RWG	116	75	64.7	41	35.3	125	84	67.2	41	32.8
	RT	116	65	56.0	51	44.0	125	73	58.4	52	41.6
	Low SES (birth)	817	785	96.1	32	3.9	907	871	96.0	36	4.0
	AB	836	581	69.5	255	30.5	927	639	68.9	288	31.1
	Younger mother	860	634	73.7	105	12.2	822	707	86.0	115	14.0
	Older mother	860	634	73.7	115	13.4	829	707	85.3	122	14.7
	Prenatal adversity	873	569	65.2	304	34.8	966	620	64.2	346	35.8
	Postnatal adversity	873	618	70.8	255	29.2	966	676	70.0	290	30.0
	Pre and post-natal adversity	873	736	84.3	137	15.7	966	810	83.9	156	16.1
17	RWG	89	54	60.7	35	39.3	96	61	63.5	35	36.5
	RT	89	50	56.2	39	43.8	96	56	58.3	40	41.7
	Low SES (birth)	660	632	95.8	28	4.2	677	649	95.9	28	4.1
	AB	617	591	95.8	26	4.2	688	474	68.9	214	31.1
	Younger mother	592	506	85.5	86	14.5	540	458	84.8	82	15.2
	Older mother	606	506	83.5	100	16.5	551	458	83.1	93	16.9
	Prenatal adversity	651	420	64.5	231	35.5	713	457	64.1	256	35.9
	Postnatal adversity	651	457	70.2	194	29.8	713	494	69.3	219	30.7
	Pre and post-natal adversity	651	546	83.9	105	16.1	713	594	83.3	119	16.7

Exposed (yes) or not exposed (No) and proportion (%).
SES, socioeconomic status; RWG, rapid weight gain; RT, rapid thrive; AB, Antibiotics.

6.4.4 EWAS results

There were significant ($P_{FDR}<0.05$) associations at individual CpG sites for the exposures; RWG and pre-natal adversity (Table 6.10) at age 7. Across the various adjustment models there were 4 significant hits for RWG ($P_{FDR}<0.1$) corresponding to 3 unique CpG sites.

In the model which did not include cells counts, two of the loci (cg01379158, cg11531579) associated with RWG had p values that were also were also below the Bonferroni p value threshold ($1.04x10^{-7}$). The two significant loci (cg01379158, cg11531579) were investigated further. Methylation at these loci was also associated with RT, however the coefficient was

smaller and less significant than for RWG (Appendix E, Table XIX). The loci that was significantly associated with adversity (cg00397179) was located in the vicinity (7bp upstream) of a high frequency SNP (rs12811501) flagged in the Infinium HD Methylation SNP list and was not significant at P_{FDR}<0.05, and was therefore not analysed further. There were no other significant associations (P_{FDR} <0.1), and no associations with DNAm at age 17.

Table 6.10 Significant (FDR p<0.1) associations between individual CpG sites (*n*=482,855) and the early life exposures in models with or without cell counts.

Exposure	CpG name	Chr	Nearest gene	Gene region	CpG island name	Coef	P_{FDR}	P_{Bonf}	Model
With cell counts									
RWG	cg01379158	17	NT5M	TSS200	chr17:17206527-17207306	0.011	0.02	$2.91x10^{-7}$	ISVA
Without cell counts									
RWG	cg01379158	17	NT5M	TSS200	chr17:17206527-17207306	0.011	0.01	$2.33x10^{-8}$	ISVA
RWG	cg11531579	12	CHFR	Island	chr12:133484658-133485739	0.011	0.02	$4.16x10^{-8}$	SVA
RWG	cg11531579	12	CHFR	Island	chr12:133484658-133485739	0.011	0.03	$1.26 x10^{-7}$	ISVA
Pre-natal adversity	cg00397179	5	BTF3	3'UTR	-	0.009	0.09	$8.97x10^{-6}$	SVA

Bonferonni threshold= < 1.0355076×10^{-7}. Chr, chromosome; P_{FDR}, FDR p value; P_{Bonf}, Bonferonni p value, Coef, coefficient; TSS200, transcription start site; 3'UTR, 3' untranslated region; RWG, rapid weight gain.

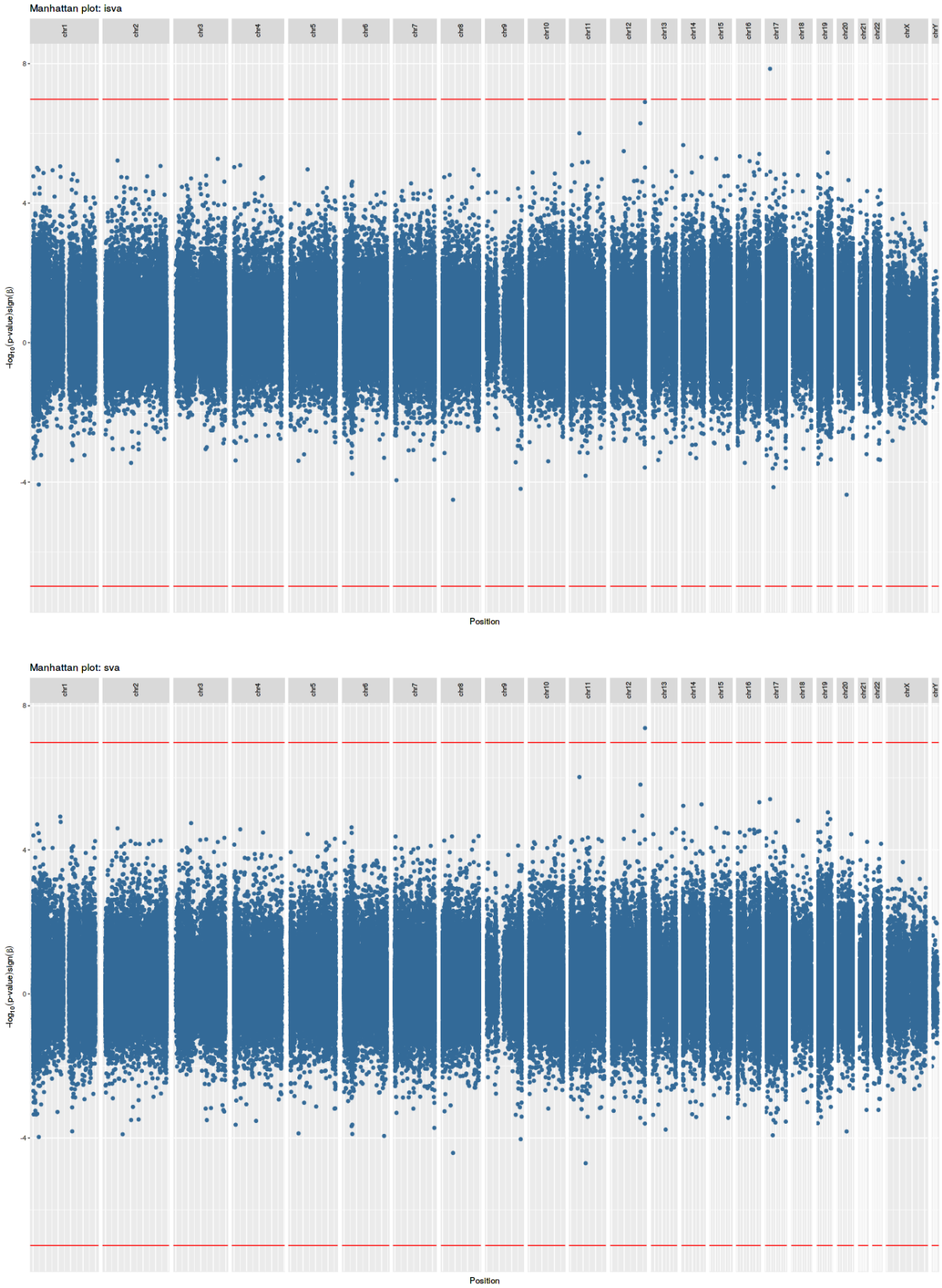


Figure 6.1 Bidirectional Manhattan plots for the EWAS linear regression models.
Models are ISVA (top) and SVA (bottom) for methylation at age 7 with RWG as the independent variable, without cell counts. The plots show the chromosomal location of significant CpG loci. The red line indicates the Bonferroni significance threshold.

6.4.4.1 Methylation changes at the cg01379158 locus

Those who experienced RWG on average had 1% greater methylation at cg01379158. RWG was associated with a 1% increase in methylation at this locus (FDR adjusted p value=0.02) in the ISVA model with cell counts, and also the ISVA model without cell counts (FDR adjusted p value=0.01) (Table 6.10). Methylation at this loci at age 17 was not associated with RWG.

The significant CpG locus; cg01379158 was located upstream of the transcriptional start site in a CpG island (chr17:17206527-17207306). There were 10 CpG loci on the 450K also located in this island (Figure 6.2).

Probes in the island were not necessarily concordantly correlated (Figure 6.2). In those who had RWG, the CpG of interest (cg01379158) demonstrated stronger (moderate) correlations with 3 CpG sites within the island (cg21614420, cg08693337, cg09810313), compared to those who did not experience RWG (Figure 6.2).

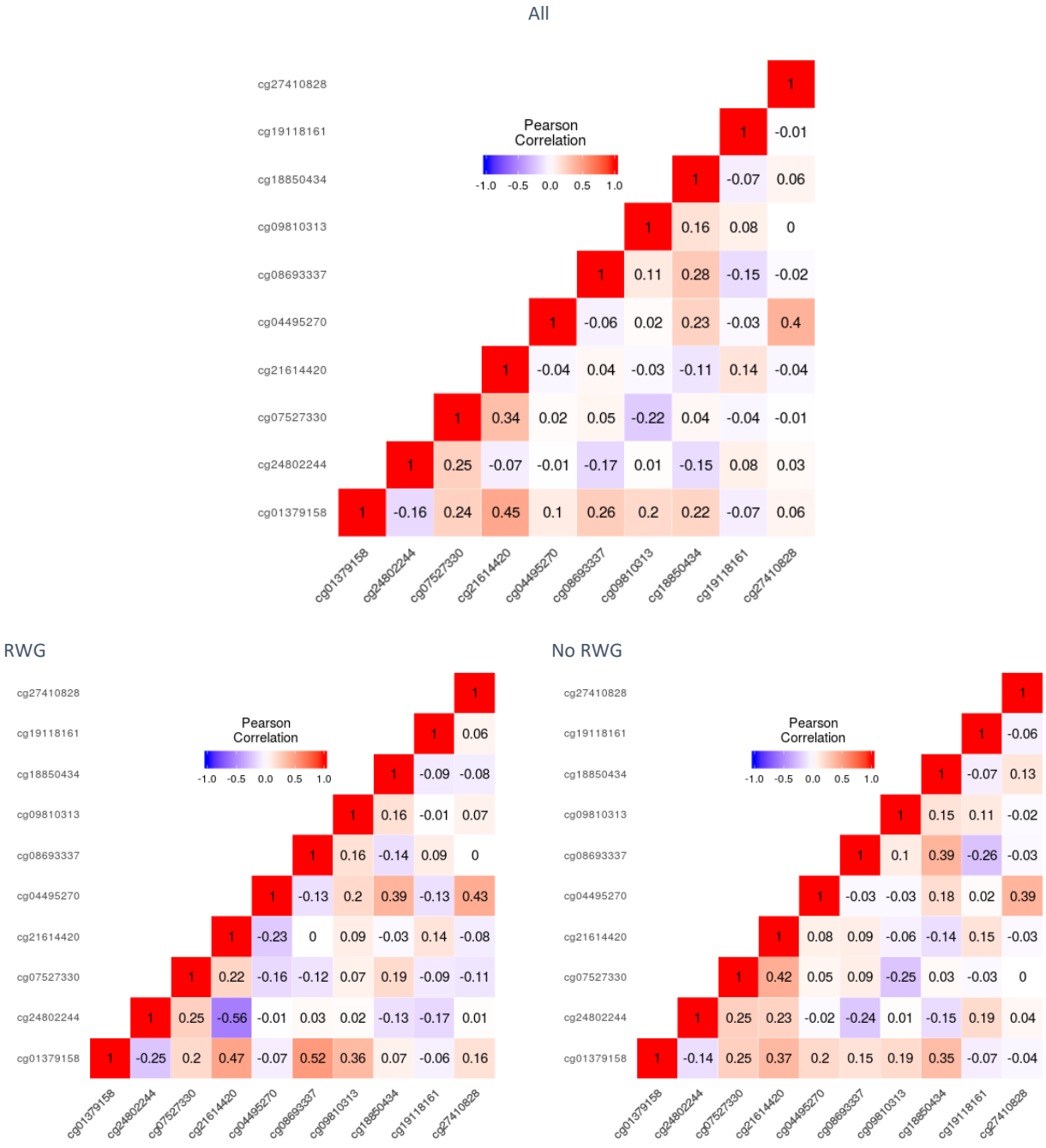


Figure 6.2 Correlations between CpG sites on the 450K array located within the CpG island (chr17:17,206,527-17,207,306). The CpG of interest (cg01379158), associated with RWG is located on the horizontal bottom row. Red indicates strong positive correlation, whilst purple indicates strong negative, proportionate to the colour intensity.

The nearest gene to cg01379158 is NT5M, also known as 5',3'-Nucleotidase, Mitochondrial. This probe is a type II probe with nearby SNPs. The other CpG loci which also mapped to the NT5M gene (*n*=12) and to the chr17:17,206,527-17,207,306 island (*n*=12) on the 450k array were also examined using linear regression (adjusted for age, sex and WBCs). There were 2 significant ($p<0.05$) CpG loci, however the beta coefficients were less than 0.01 and these

loci would not have been significant after correction for multiple comparisons (Appendix E, Table XVIII).

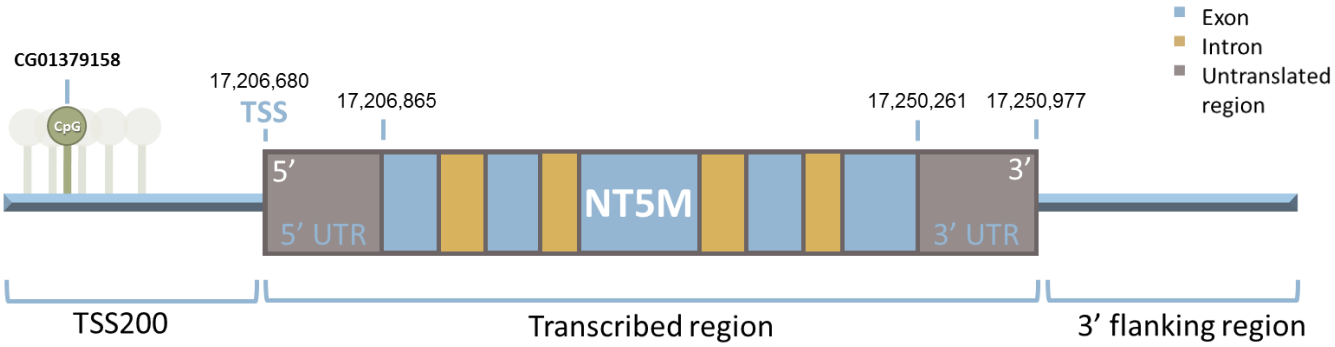


Figure 6.3 Annotated region of the CpG (cg01379158).
Chromosomal (17) position, the nearest gene and the downstream region for CpG (cg01379158) annotated using the (GRCh37/hg19) assembly.

6.4.4.1.1 Age-related changes in methylation

At the cg01379158 loci, within individuals who had RWG, methylation decreased from age 7 to 17. Whilst in those who did not have RWG, methylation increased (Figure 6.4).

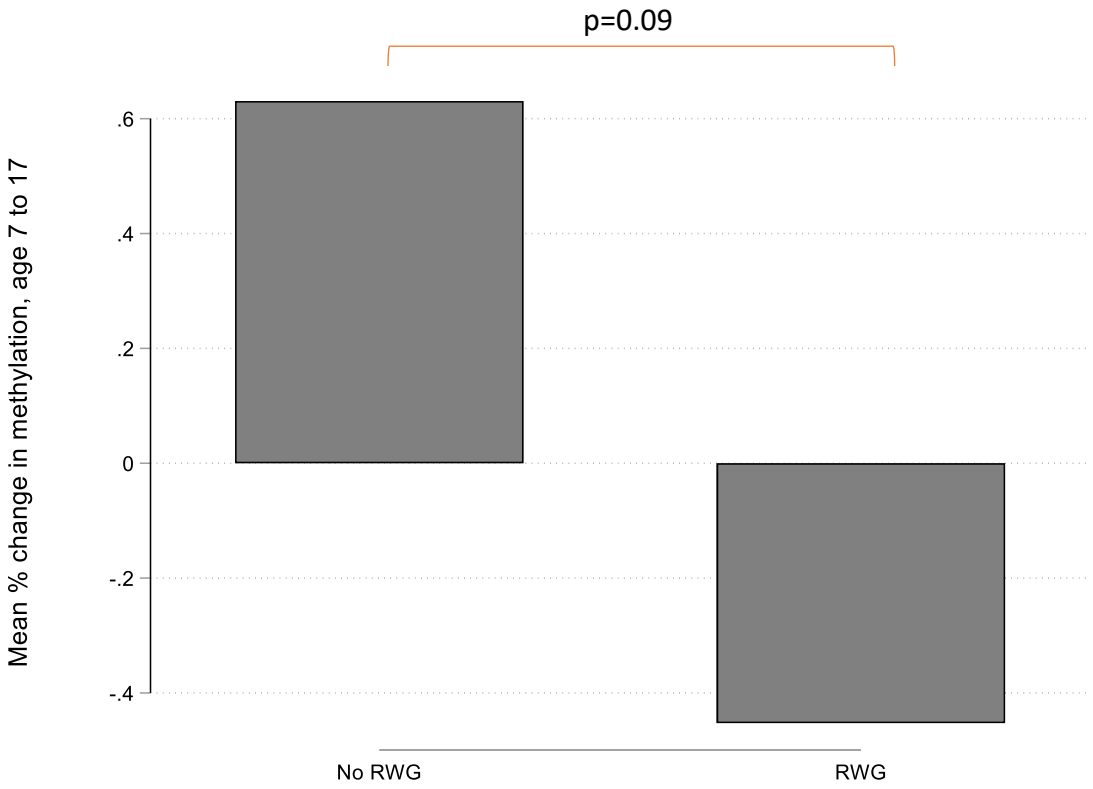


Figure 6.4 Change in methylation at the cg01379158 loci within individuals from age 7 to 17 by RWG.
Those who did not have RWG (*n*=60) demonstrated small mean increases (+0.63%) in methylation, whereas those who had RWG (*n*=34) demonstrated small (-0.45%) decreases in methylation between ages 7 and 17 (p=0.09, using the students t-test).

6.4.4.1.2 Relationship between methylation and adiposity outcomes

Methylation at this locus was not associated with BMIz at age 7 or 17. There were too few OWOB to sufficiently execute a logistic regression model (*n*=19 OWOB at age 17 for those with rapid weight gain data available), however phenotype group differences were examined. Highest methylation was in those who experienced RWG and were OWOB both in childhood (9.1% methylation) and adolescence (8.8% methylation) (Appendix E, Table XXI and Figure XIX).

6.4.4.2 Methylation changes at the cg11531579 loci

The other significant CpG loci (cg11531579), was positively associated with RWG in both the ISVA and SVA models (p_{FDR}<0.05) (Table 6.10). The association p values were not below the accepted Bonferroni threshold.

At age 7, cg11531579 methylation ranged from 0 to 12%, with median value around 3%. This CpG was also a significantly associated with RWG in the DMRcate analysis, however the fold change was very small (max beta fold change=0.01) and the Stouffer p value was non-significant (p=0.48).

This CpG (cg11531579) is located within a CpG island on chromosome 12 (Table 6.10). Correlations between the CpG of interest and other CpG sites in the island were generally positive and tended to be stronger in those who did not experience RWG (Figure 6.5). In those who did experience RWG, this particular CpG site was an anomaly in the island in that it demonstrated by very low correlation with the other CpG sites in the island, whereas generally speaking the rest of the CpG sites were mostly positively correlated to various degrees (Figure 6.5). Cg24459893 showed the greatest 'disagreement' with the rest of the island, with strong negative correlations in both those who had RWG and those who did not. Upon closer inspection, although this CpG maps to the island (using the 450K annotation file) located at chr12:133,484,658-133,485,739, its actual location is much further downstream at 133,488,122, which may explain this discordance.

The other CpG loci which also mapped to the chr12:133484658-133485739 island (*n*=10) on the 450k array were also examined using linear regression (adjusted for age, sex and WBCs). There was 1 significant CpG site, however the beta coefficient was less than 0.01 (Appendix E, Table XVIII).

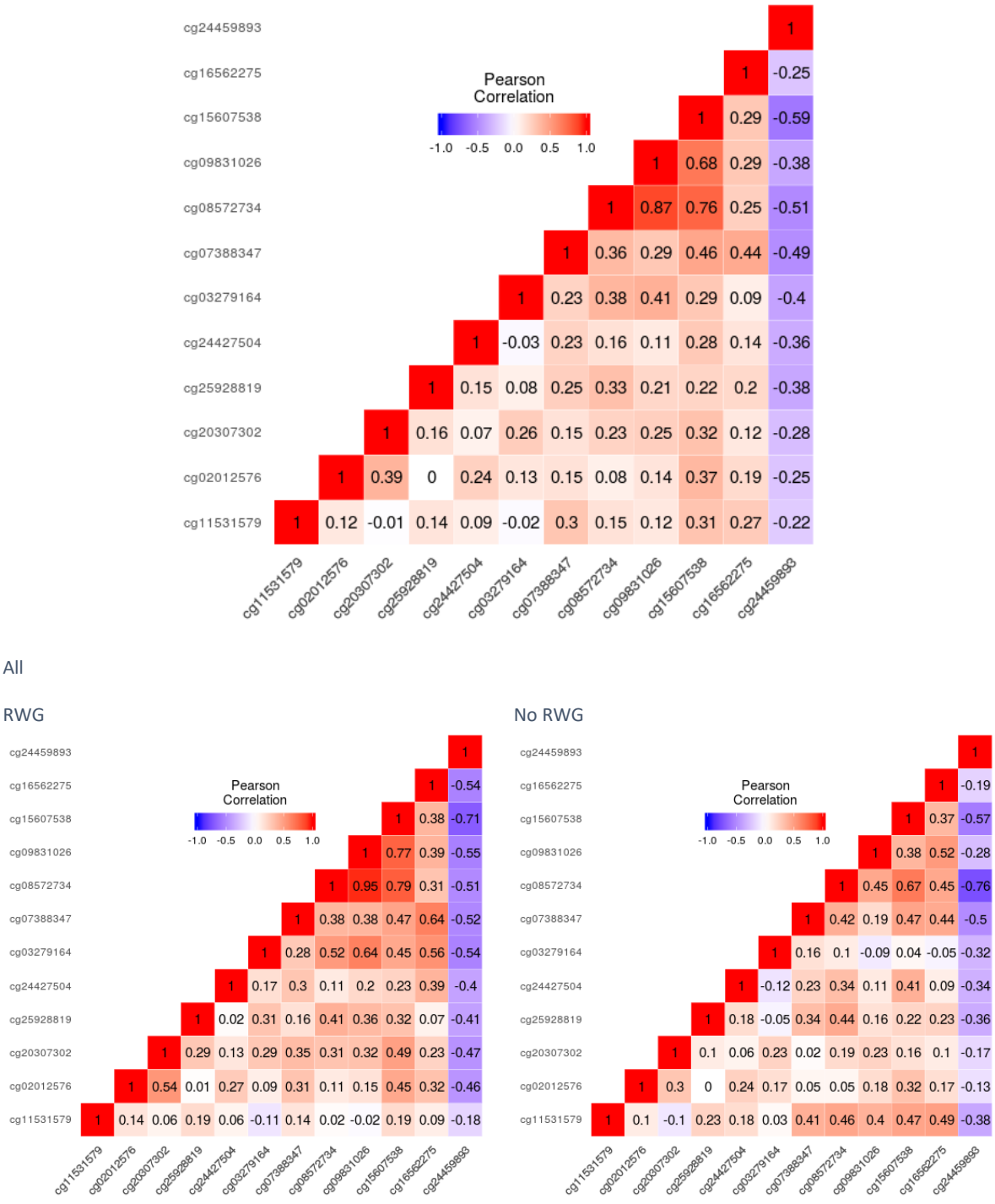


Figure 6.5 Correlations between CpG sites on the 450K array located within the CpG island (chr12:133484658-133485739).
The CpG of interest (cg11531579), associated with RWG is located on the horizontal bottom row. Red indicates strong positive correlation, whilst purple indicates strong negative, proportionate to the colour intensity.

The CpG of interest (cg11531579) was not located near to any protein-coding genes. Upstream 30,000+ bases is the protein coding gene; Checkpoint With Forkhead And Ring Finger Domains (CHFR) (Figure 6.6). Whilst 558 base pairs downstream, is a small (2 exons) non-coding region (AK055957), for which there is limited information.

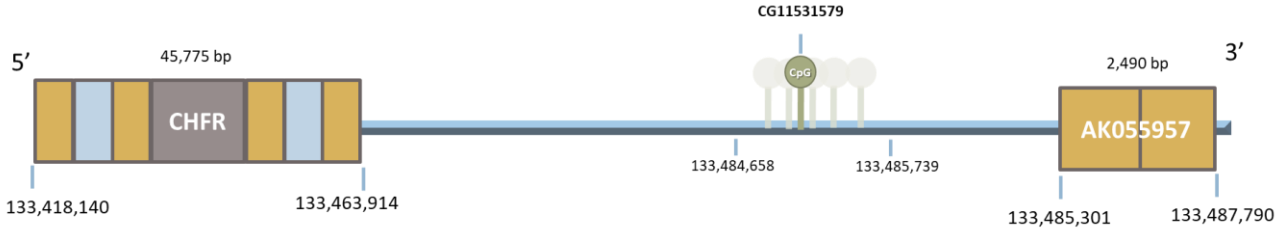


Figure 6.6 Annotated region of the CpG (cg11531579).
Chromosomal (12) position, the region, and nearest genes region for CpG (cg11531579) annotated using the (GRCh37/hg19) assembly.

6.4.4.2.1 Age-related phenotypic changes in methylation

Using methylation data at ages 7 and 17 shows that in those with RWG, on average methylation decreases from age 7 to 17 ($p<0.01$) (Figure 6.7). Whereas in those who did not experience RWG, on average methylation tends to increase between the time points.

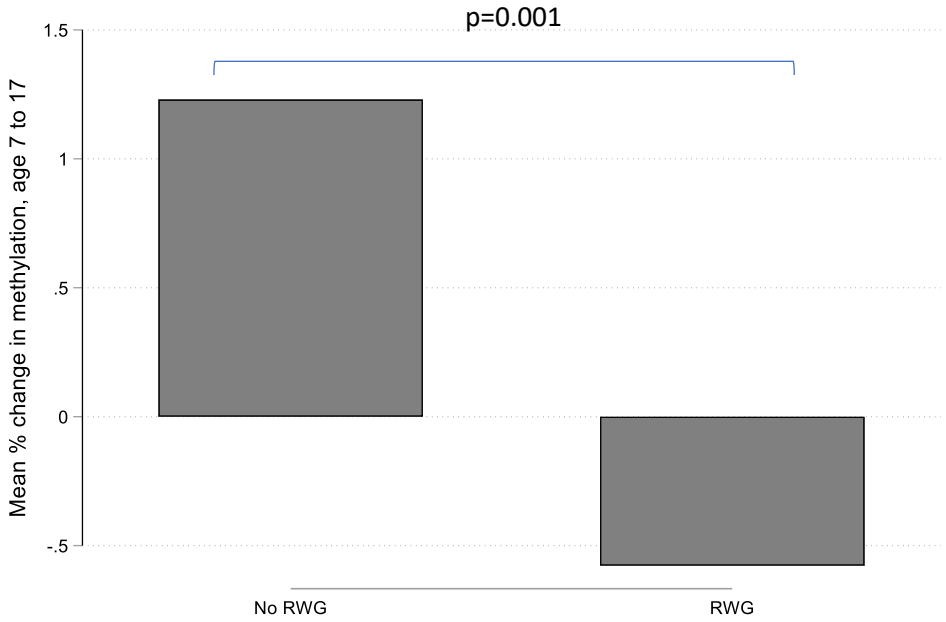


Figure 6.7 Change in methylation from age 7 to 17 within individuals by RWG.
Those who did not have RWG (*n*=60) demonstrated small mean increases (+1.3%) in methylation, whereas those who had RWG (*n*=34) demonstrated small (-0.58%) decreases in methylation between ages 7 and 17. This difference (1.8%) was significant (p=0.001, determined using the student t- test).

6.4.4.2.2 Relationship between methylation and adiposity outcomes

There were median differences in methylation level according to phenotype (Appendix E, Table XXII, Figure XX). There was some suggestion that methylation was higher in those who had RWG and were OWOB at the time (age 7) or subsequently (age 17). However, the sample size for these group were small and therefore these results are suggestive.

6.4.5 Candidate gene analysis results

As there were few significant probes found, which may have been due to the strict correction for multiple tests, an alternative approach was taken utilising a smaller, candidate set of epigenetic loci. The aim of the analyses was to isolate CpG loci already known to be associated with the outcome phenotype of interest (body composition). The candidate gene analysis utilised findings from a consortium, which integrated data from 4 discovery cohorts and replicated findings in 9 cohorts, and found 187 validated methylation markers associated with BMI (Wahl et al., 2017).

Using a smaller subset of loci as candidates has the advantage of reducing the stringent p values threshold imposed by correcting for multiple testing. The associations between these loci and the early life exposures were examined using the ALSPAC methylation childhood and adolescent data, however there were no significant associations identified (Bonferroni p value=$3x10^{-4}$).

6.4.6 Differentially methylated regions results

Whilst some significant individual probes were identified using DMRcate in unadjusted models for some exposures at age 7 (adversity all (no cells), low SES (cells and no cells), RWG (no cells) and younger mum (no cells)) and age 17 (low SES with and without cells), there were no overall DMRs identified. All Stouffer corrected p values were non-significant, which may be due to a lack of consistency in the direction of methylation in the region.

6.4.7 DNA methylation analysis

To investigate DNAm biomarkers of early life exposures, methylation changes in children and adolescents in response to early life exposures were investigated. To do so, epigenome-wide association studies (EWAS) were run for each exposure to determine methylation changes at individual CpG sites in response to the exposure(s). To further investigate the hypothesis that DNAm changes that are associated with adiposity are also associated with early life

exposures, a subset of CpG loci with confirmed associations with BMI were investigated with respect to each exposure. The EWAS methods and statistical methods are described in section 2.4.

6.4.7.1 **EWAS regression diagnostic plots**

Meffil simultaneously computes estimates for all 4 models (Figure 2.6). Q-Q plots were consulted, which allow graphical assessment of whether the data are derived from a normal distribution. Q-Q plots demonstrated that ISVA adjustment was most effective in correcting the data best so that it satisfies the assumptions of normality (Appendix E, Figure XVII). These plots indicate loci which deviate from the expected distribution, i.e. those that are above the p value threshold are significantly associated with the exposure.

All models were additionally run without adjustment for cell counts, for each exposure and for both methylation outcome time points. The p value distributions display greater deviation from normality than those from the adjusted models, however the SVA model performed well (Appendix E, Figure XVIII).

6.5 Discussion

The exposures that were important factors for childhood OWOB in the North East cohorts were investigated further in the larger, South West cohort; ALSPAC. The aim was to investigate exposures using the same models as done in the previous chapter (the cohort comparison) for consistency. However, in ALSPAC, this led to ill-fitting models that did not meet regression criteria, and instead the parsimonious models were more informative.

Birthweight, RWG, RT, and adversity were all positively associated with adiposity outcomes. Birthweight was consistently associated across all outcomes and time points. RT was consistently associated with body composition measures in childhood.

The sensitivity analyses revealed that maternal BMI was an important factor which predicted offspring body composition across all measures, similar to findings in many longitudinal birth cohorts (Baker et al., 2004, Harvey et al., 2007, Wright et al., 2010a, Bammann et al., 2014, Fairley et al., 2015b). Other analyses in the ALSPAC cohort found that children of obese parents had a greater fat mass at 7 years and larger increases in fat mass up to 11 years (Wright et al., 2010a), perhaps be due to a shared environment as well as genetic influences. Maternal obesity can affect multiple aspects of offspring health, which is accompanied by

changes in neuroendocrine, metabolic and immune system processes, increasing risk of multiple diseases and all-cause mortality, and thereby has great public health implications (Reynolds et al., 2013a, Godfrey et al., 2017).

Maternal smoking during pregnancy did not demonstrate direct associations with offspring body composition, however adjusting for smoking removed significant associations for birthweight for outcomes in adolescence, perhaps suggesting a latent indirect effect.

Factors that were associated with body composition across three cohorts; birthweight, RT, RWG, SES (at birth), and adversity, were all good candidates for further investigation. Birthweight was previously examined in this cohort with regards to DNAm, however no significant associations were identified, and therefore birthweight was not explored here. RWG was associated with BMIz in children in all three cohorts, and with OWOB in GMS and ALSPAC, and was therefore a robust exposure. As there were also some weak associations observed for maternal age and bivariate associations for antibiotic exposure (which were both also seen in GMS and NTFS respectively), these were also investigated further in the epigenetic analysis.

The exposures were examined with regards to DNAm in childhood and adolescence. Two CpG sites (cg1379158 and cg11531579) were identified which exhibited differential methylation in childhood, in association with RWG. No other early life factors were significantly associated with changes in DNAm.

Cg11531579 is located downstream of CHFR, an E3 ubiquitin-protein ligase that regulates the cell cycle at the antephase checkpoint, by delaying progression into mitosis in response to microtubule stress (Scolnick and Halazonetis, 2000). This achieved by preventing cyclin B1 access to the nucleus prior to chromosome condensation (Summers et al., 2005). Decreased expression has been exhibited in certain cancers as a result of promoter hypermethylation (Sanbhnani and Yeong, 2012, Derks et al., 2014), confirmed with treatment with the methyltransferase inhibitor; 5-aza-2'-deoxycytidine (Sakai et al., 2005). Silencing may also be a result of deacetylation of histones in the promoter region (Oh et al., 2009), however it is unclear whether silencing is a consequence or cause of cancer. However, as cg11531579 is located far (20,538 bases upstream) from the gene CHFR, it is perhaps speculative to discuss a role methylation at this locus could have in relation to the CHFR gene.

Upstream of cg11531579 is AK055957, a small non-coding RNA regulatory sequence. Its biological role is uncharacterised. Recently, this CpG (in combination with others) has been identified as a potential DNAm biomarker for use in detection panels for hepatocellular carcinoma (Kisiel et al., 2019) and pancreatic ductal adenocarcinoma (Majumder et al., 2019). The CpG site, AK055957, is located near a H3K27Ac histone mark, which is often found near regulatory elements in many cell types. The H3K27Ac histone mark is the acetylation of lysine 27 of the H3 histone protein and is thought to be a transcription enhancer. It is located near a DNAse I hypersensitivity cluster, which may suggest a transcription factor binding region. In a study in children diagnosed with acute myeloid leukaemia, cg11531579 was found to be differentially methylated in marrow after chemotherapy compared to baseline (-0.24 change in beta, p=0.004) (Gore et al., 2017). In summary, this locus may have a role in carcinogenesis, and therefore exists a tenuous link between RWG, DNAm and uncontrolled growth.

The cg01379158 locus is located in the TSS of the NT5M gene, which encodes a 5' nucleotidase (a hydrolytic enzyme that catalyses the hydrolysis of a nucleotide to a nucleoside and a phosphate) that is subcellularly located within the mitochondrion. This enzyme (dNT-2) dephosphorylates the 5'- and 2'(3')-phosphates of uracil and thymine deoxyribonucleotides, hence protecting mitochondrial DNA replication from excess dTTP (Rampazzo et al., 2000). It has associated pathways of pyrimidine metabolism and metabolism, and annotations relating to nucleotide binding and nucleotidase activity. The gene is located on chromosome 17 in the Smith-Magenis syndrome-critical region, and therefore could potentially play a role in this disease aetiology. Smith-Magenis syndrome is a rare condition characterised by inverse circadian rhythm and disturbed sleep, factors which have also been linked to obesity (Froy, 2010, Woo Baidal et al., 2016). Furthermore, an EWAS on sleep found a cluster of differentially methylated positions in this Smith-Magenis region of chromosome 17 (Lahtinen et al., 2019) suggesting methylation could play a role in regulation of sleep and circadian rhythm.

In the epidemiological analysis, RWG was a mediator between birthweight and adiposity outcomes, and therefore was closely linked with birthweight. DNAm differences related to birthweight are frequently related to growth control (Turan et al., 2012). For example, a panel of 23 genes explained 70-87% of the variation in birthweight in human or mouse models, and 6 of these genes had roles closely related to growth (Turan et al., 2012).

Therefore, as associations were stronger for DNAm and RWG (rather than RT), it is plausible that methylation at these loci reflect a combination of birthweight and postnatal growth.

Whole blood represents a mixed cell population with varying proportions of white blood cells. For cg01379158, the association remained with or without adjustment for cell composition. In the adjusted analysis, cg11531579 did not emerge as a significant CpG locus. Therefore, it is possible that RWG-linked differences in cell proportions could somewhat account for this association.

Whilst the cg01379158 was associated with RWG in both adjusted models, it did not demonstrate potential to be predictive of subsequent OWOB (age 17). However, the cg11531579 locus showed more promise as a predictive marker, as the highest levels of methylation were in those who had RWG and were subsequently OWOB.

Early life RWG was associated with changes in childhood methylation, however it is unclear how methylation at these CpG loci may change over the life course. There was not persistence of methylation differences through direct associations with RWG in adolescence, which could have been partly due to a smaller sample size. Alternatively, these methylation changes could exert effects throughout childhood, rather than persistence into adolescence. This is consistent with the epidemiological findings, whereby RWG was associated with changes in body composition in childhood, but not adolescence (OWOB in ALSPAC) or adulthood (BMI in NTFS). Individuals who had infancy RWG had higher childhood methylation (cg11531579), which decreased over time. The initial high methylation followed by the decrease over time mirrors the growth pattern seen with rapid infancy growth (Chapter 5, *Figure 5.2*), therefore the decrease over time may reflect the 'recovery' of DNAm levels similar to the normalisation of BMIz.

There were null findings for many exposures in this analysis. Recently, Houtepen *et al*,. ran a similar analyses looking at ACEs and maternal DNAm in ALSPAC and in a replication cohort, and also did not identify any individual CpG sites which replicated across cohorts (Houtepen et al., 2018a). However they did find some DMRs associated with parental health including parental mental or physical illness, and death (Houtepen et al., 2018a), suggesting that perhaps some of the more 'severe', hereditary or genetic factors could have persistent DNAm changes. Another recent ALSPAC study deemed very early childhood (0-3 years) as a critical period in establishment of DNAm patterns (Dunn et al., 2019), therefore perhaps the time period examined here (0-1 year) was too narrow.

In an ALSPAC EWAS, socioeconomic position during pregnancy was associated with DNAm signatures at three stages across the life course, with maternal education level as the most important socioeconomic variable (Alfano et al., 2018), whereas this study used occupational social class.

When studying exposures with small effect sizes there is less power to detect changes that span multiple CpG loci, which may explain why no significant DMRs were identified.

There were also no associations when using the consortium CpG loci as candidates, which could be for various reasons. Firstly, that these specific exposures are not associated with changes in DNAm (except for RWG). Secondly, the candidate loci mapped to genes with specific roles, which could be different to the roles and pathways of the exposures examined. The early life exposures studied here have been associated with subsequent changes in BMI (chapters 2-5). Whereas, in the consortium paper, using Mendelian Randomization it was determined that the majority of the significant CpGs were a consequence (rather than a cause) of changes in BMI (Wahl *et al*., 2016). Thereby BMI is the mediator in this hypothesis, and if any exposures were associated with DNA changes in this subset of CpG loci, this in theory would have been a result of BMI. Finally, although some of the associations have been replicated in pre-school children (Rzehak et al., 2017), primarily the candidate loci were relevant to an adult population, whereas this cohort were much younger.

The strengths and limitations of the Illumina 450K array should be considered. Whilst the array gives highly reproducible measures of DNAm at many loci, it only allows investigation of the predefined probes on the array. The CpG sites on the array cover 99% of RefSeq genes and 96% of CpG islands, which were selected because they were of particular interest and are not equally distributed across the genome (Bibikova et al., 2011). Whilst the array has good overall coverage for protein coding genes, it only covers 1.7% of CpG sites on the human genome, located mostly in promoter regions. This approach could neglect sites that could be markers for the investigated exposures.

Batch effects are a source of heterogeneity when analysing arrays and were dealt with using adjustment techniques. There are issues in probe design in that many of the probes cross hybridise to regions that were not the intended target loci, however this was dealt with by removing the probes known to cross-hybridise. The array does not encompass simultaneous analyses of SNPs and CpG sites, and loci which contain SNPs affect quantification of DNAm

levels. Furthermore, the relative contribution of genetics cannot be disregarded, for example, it is possible that some study members were genetically predisposed to obesity, and this could interact with epigenome.

The EWAS served as a starting point to identify potentially important candidate sites, which could be the focus of further investigation. Validation (in another independent cohort) would be required to determine if CpG loci are consistently, differentially methylated in different populations, time points, ages and disease-states.

Chapter 7. Investigating methylation in the Newcastle Thousand Families Study

7.1 Introduction

This chapter further addresses the third aim of the thesis and examines the gene-specific methylation differences identified (in children) in an adult population, many years after the exposure and when obesity has occurred.

The EWAS outlined in chapter 5 identified two differentially methylated CpG loci at age seven, in association with rapid weight gain (RWG) in the first year of life. RWG, was a focal exposure, as it was strongly associated with childhood body composition in all the cohorts (NTFS, GMS and ALSPAC). The Newcastle Thousand Families Study (NTFS) provided an opportunity to investigate if early life RWG has a transient or enduring impact on DNAm, in an older population.

This chapter describes the quantification of DNAm at these specific loci at age 50 in the NTFS cohort, to investigate if DNAm at these loci are biomarkers of early life RWG and subsequent adiposity in adults. If methylation at the RWG-associated locus is subsequently associated with obesity in adulthood, it is plausible that methylation could be used as a biomarker to predict those at higher risk of obesity in later life.

7.2 Aims

This chapter addresses the second part of the 3rd aim of the thesis; To investigate the methylation differences in relation to the early life exposure(s) found to influence obesity in an adult population.

The aims were addressed with the following objectives: i.) Develop locus-specific DNAm assays (Bisulfite polymerase chain reaction (PCR) and pyrosequencer-based) for the detection and quantification of methylation at identified CpG loci, and ii.) Analyse methylation at the CpG sites of interest in the NTFS cohort, and iii.) Determine if RWG in infancy is associated with altered DNAm at specific loci in adults.

7.3 Participants and methods

7.3.1 Study design and samples

DNAm (age 49-51) was investigated with respect to early life RWG, and both current (age 49-51) and subsequent (age 60) body composition (BMI). Early life data were collected prospectively for study members and participants of the NTFS were invited to the clinic to be measured around age 49-51 (*n*=412, including measurements of body composition and blood samples taken)(detailed in section 2.1.1.2). Participants had similar body composition measures taken again in a clinical assessment at age 60 (*n*=354). DNA was previously extracted (Pearce et al., 2012b) from the peripheral blood samples (age 49-51) and stored at -80^{o}C in the Newcastle University Biobank. Further details on the NTFS participants (at age 49-51) can be found in section 2.1.1.

Bisulfite-PCR and pyrosequencing assays were designed for the two epigenetic loci identified from the ALSPAC EWAS (see chapter 5). Of the two assays designed only one was viable (detailed in section 7.4.1), and therefore one target was analysed in peripheral blood DNA in NTS adults (age 50) NTFS.

The loci of interest were investigated in the DNA extracted from NTFS blood samples, and DNAm was quantified using pyrosequencing (section 2.5.9).

DNAm levels were analysed with respect to RWG in infancy, and weight-related outcomes (BMI and OWOB at ages 50 and 60) using statistical methods (detailed in section 2.5.12.1). The approaches taken to address the impact of outliers and potential SNPs are outlined in sections 2.5.11 and 2.5.13.

7.4 Results

7.4.1 Assay for cg01379158 (NT5M)

Primer design for the loci, cg01379158, was problematic. The region was CG rich and no primers could be identified for the genomic DNA, even with relaxed criteria. Therefore, the reverse complement sequence was trialled as the input sequence. This resulted in some primer matches, however it was still not possible to design a forward primer without any SNPs or of adequate length (the forward primer was 17bp). Multiple primer sets were examined (Appendix F, *Table XXIII*); either matching the original sequence (including SNPs) or

with degenerate primers using International Union of Pure and Applied Chemistry (IUPAC) nucleotide pairing in the forward primer. This would allow 'wobble' base pairing, in this case annealing whether the base was either Purine (A or G), by denoting an R nucleotide. No sets of primers gave clear, consistent bands of the expected size when visualised by gel electrophoresis, with either fragmentation or numerous primer dimers.

It was not possible to validate the assay for cg01379158 to a sufficient standard for sequencing. The presence of secondary bands meant that pyrosequencing reactions failed (due to low peak heights). Various PCR conditions were trialled in an attempt to increase the band size and reduce secondary structures. Firstly, Dimethyl sulfoxide (DMSO), used in reactions to inhibit formation of secondary products, at concentrations of 3% and 5%. Next, the number of PCR cycles was increased from 50 to 55. Finally, a touchdown PCR protocol (using a higher T_a, and gradually decreasing the T_a over successive cycles until 'touchdown' temperature is reached) was trialled. However, none of these methods were successful in inhibiting secondary structures or improving bands sufficiently for pyrosequencing. Overall, these results suggested that primer specificity was the issue. Primers were redesigned and all avenues of exploration re-trialled, however these also failed, and no there were further attempts to validate this assay.

There were also 3 SNPs within the vicinity of this CpG (cg01379158) (Table 7.1), which could have potentially impacted methylation array results if present in the population under study.

All further results will be in reference to the assay designed for cg11531579.

Table 7.1 Known common SNPs near to cg01379158, distance to CpG and the minor allele frequency of the SNP

SNP	Distance (bp) from CpG	Minor allele frequency
rs151107219	16	0.004600
rs8071972	18	0.020417
rs6502575	35	0.213047

7.4.2 Assay design for cg11531579

There were many SNPs in the region for the loci, cg11531579, which meant having to compromise on primer design. All viable primers contained a SNP (Appendix F, *Table XXIII*),

with the selected primer set containing a SNP of unknown frequency (observed A/G, rs867052755) in the forward primer. In order to ensure correct primer annealing, two forward primers were trialled during assay validation, one set with the wild-type base (A), and one set with IUPAC base pairing (R). The 'R' primer set produced the clearest, strongest bands visualised with gel electrophoresis. Furthermore, a benefit of using this set is that primer binding will occur whether the base is A or G (if the individual has the SNP), and thus was used in analyses.

Clear bands were observed at all the temperatures in the range of Ta (42-52oC) for the gradient PCR for cg11531579. The band at 52oC was marginally clearer (Appendix F, Figure XXII), therefore this was selected as the T_a.

7.4.3 Sample processing

Figure 7.1 shows a flow diagram for the samples included for methylation analysis in NTFS. There were 366 DNA samples remaining for NTFS study members at age 50 (Figure 7.1). Due to missing data, only 153 of these had early life RWG data. After quantifying DNA, six samples contained no DNA and were excluded. There were two samples which failed bisulfite conversion (incomplete conversion), and had no DNA remaining to repeat. Eleven samples failed pyrosequencing due to low peak heights. Repeating these samples did not lead to improvements and these samples were discarded. This left 134 samples for analysis.

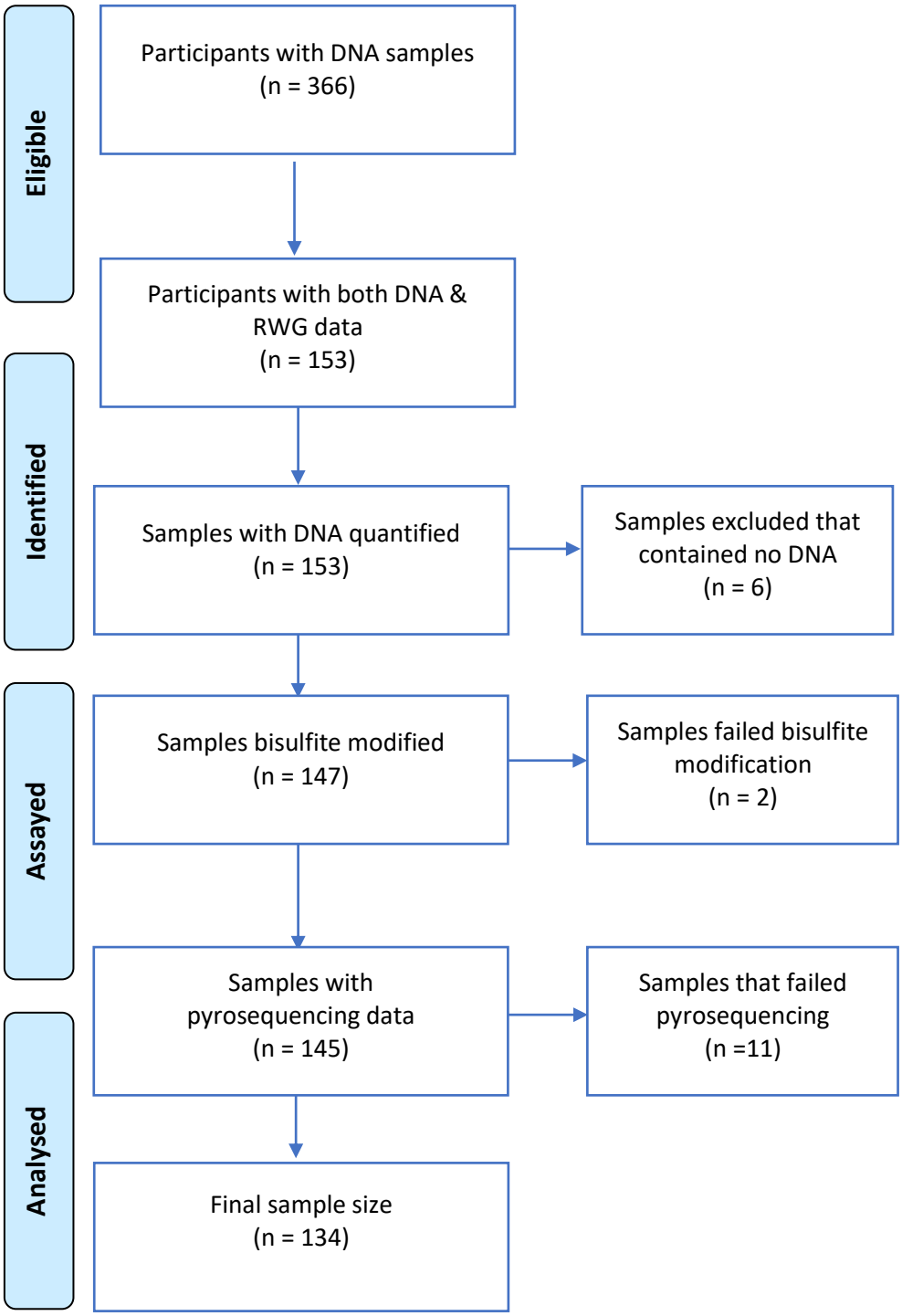


Figure 7.1 Flow diagram of sample processing and analysis for NTFS samples (age 50)

7.4.4 Pyrosequencing results

7.4.4.1 Validation of pyrosequencing assays

The validation results using the methylated controls are presented in Figure 7.3. Validation was performed in duplicate for 9 different methylation percentages ranging from 0-100%. The observed percentage methylation was plotted against the expected percentage

methylation, and the trend line and R^2 value determined. R^2 values for all CpG loci were close to 1, indicating that using this assay, observed values of methylation are in agreement with the actual values, and that the assay is able to accurately detect the range of methylation values. This provided confirmation that the assay was adequately detecting methylation levels in this region. Validation at different concentrations provided an opportunity to determine the optimum way in which perform duplicate experiments. Route A utilised 1 PCR reaction (total volume of 24µL), split over two pyrosequencing reactions (10µL in each). Whereas route B involved 2 PCR reactions, with 1 pyrosequencing reaction from each individual PCR (Figure 7.2). The hypothesis was that route B, dividing the bisulfite DNA across 2 PCRs, would lead to more variation in methylation due to the precision required when pipetting small volumes prior to amplification. However, both replication methods were comparable, with high R^2 values (for both methods the R^2 for mean methylation and individual CpG methylation was >0.99) (Figure 11). Therefore, route A was taken as it was faster and more cost effective.

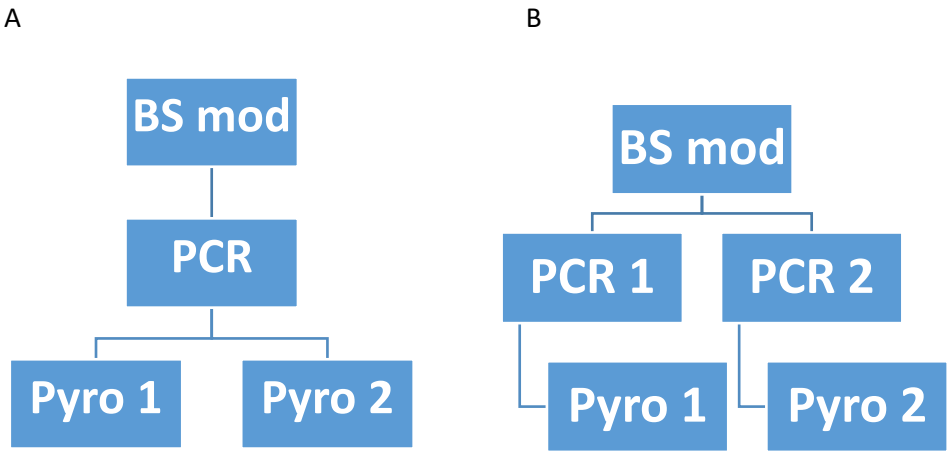


Figure 7.2 Replication methods tested to determine most accurate measure of DNA methylation. Route A used the same PCR products with duplicate pyrosequencing plates, whereas route B used two PCR reactions and separate pyrosequencing reactions from these PCR products. BS mod, bisulphite modification; Pyro, pyrosequencing run.

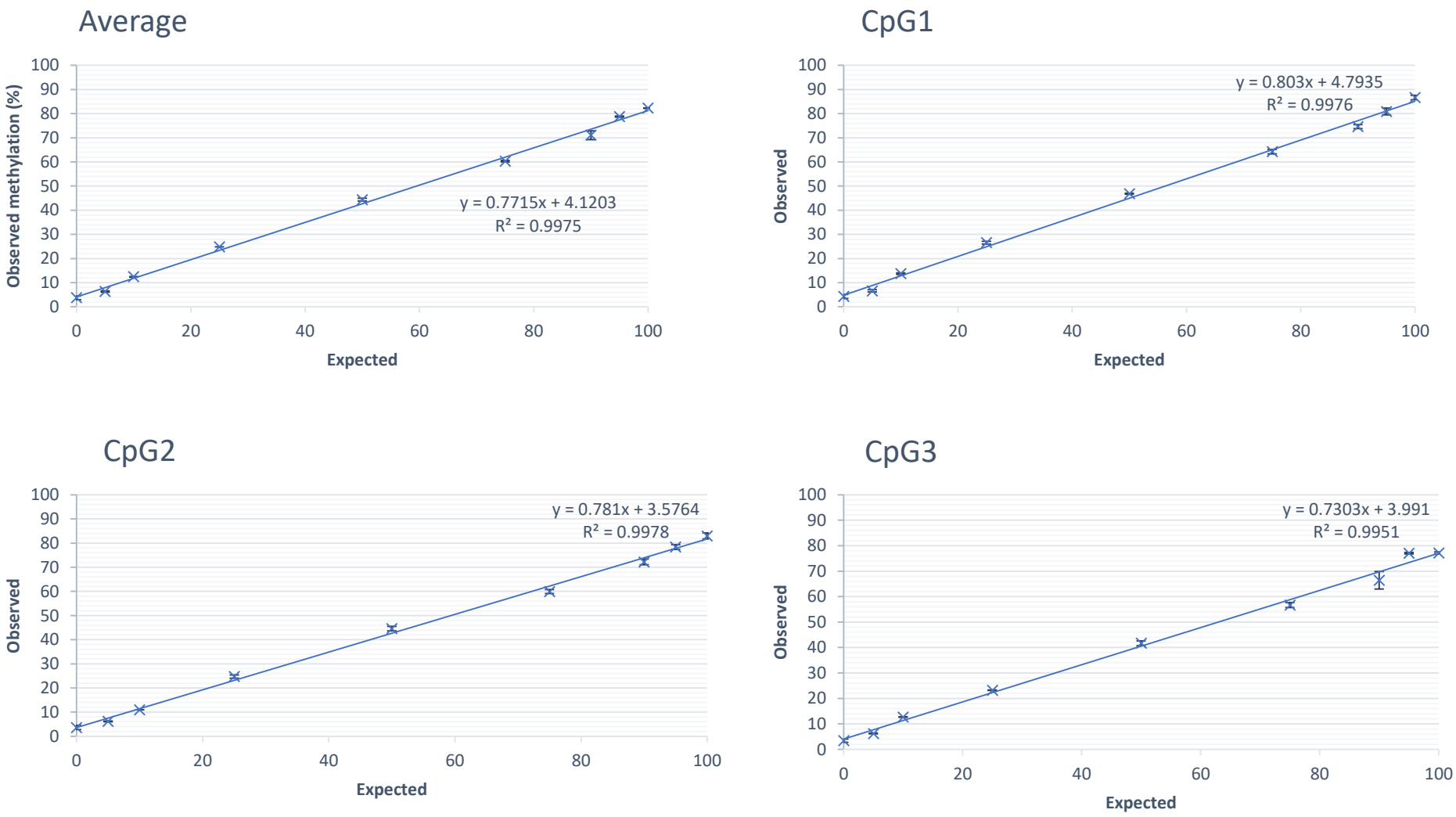


Figure 7.3 Validation (route A) curves for overall average methylation and each of the CpG loci in the pyroassay.
The solid lines represent the linear relationship, with the equations of the line shown, along with the R^2 values to demonstrate the degree of concordance between the observed and expected values (an R^2 close to 1 signifies exact correlation). Error bars represent standard error.

7.4.5 Distribution of DNAm data

DNAm was successfully quantified in 134 samples. The range of methylation detected in the NTFS samples by pyrosequencing was wide, ranging from 0-99% (Table 7.2). Average (median) methylation level across the 3 CpGs measured was 3.1%. From inspection of the histograms, mean methylation was much higher because it was skewed by high values (Figure 7.4). Most individuals had low methylation (<7%).

Table 7.2 Methylation levels at each CpG and the average methylation in the NTFS samples.

Mean methylation (%)	N	Mean	p50	p25	p75	Min	Max	Range	SD
Average	134	7.13	3.11	1	5.19	0	95.46	95.46	14.35
Cpg1	134	6.91	2.81	0	5.88	0	99.3	99.3	14.89
Cpg2	134	7.66	3.59	1.19	7.12	0	95.81	95.81	14.43
Cpg3	134	6.83	2.81	0	4.49	0	91.27	91.27	14.41

Median (p50), lower quartile (p25), upper quartile (p75), standard deviation (SD) minimum (min) and maximum (max) values shown for all CpG.

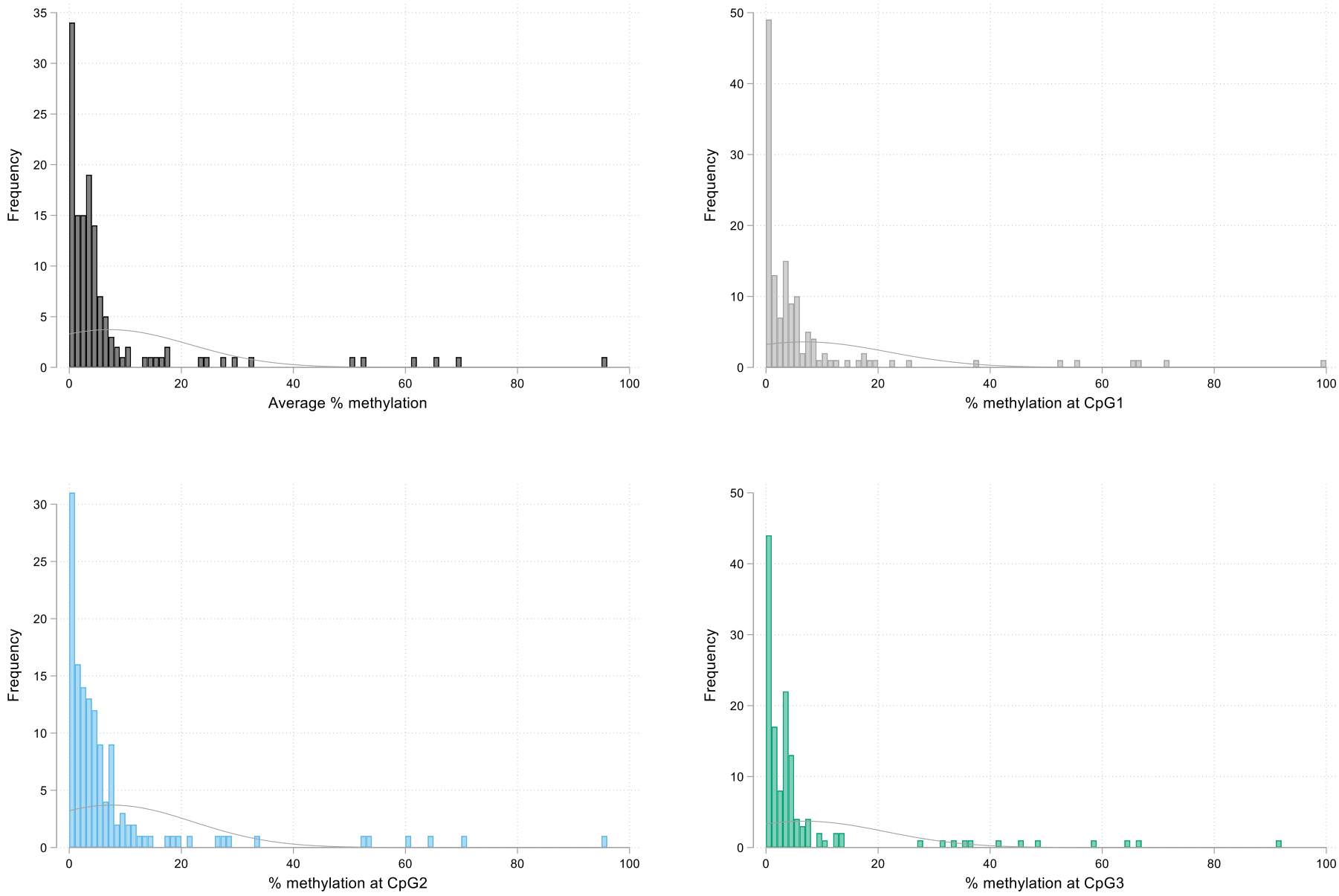


Figure 7.4 Distribution of average methylation (%) and methylation at each CpG loci determined by pyrosequencing

7.4.6 The relationship between RWG and DNAm (age 50)

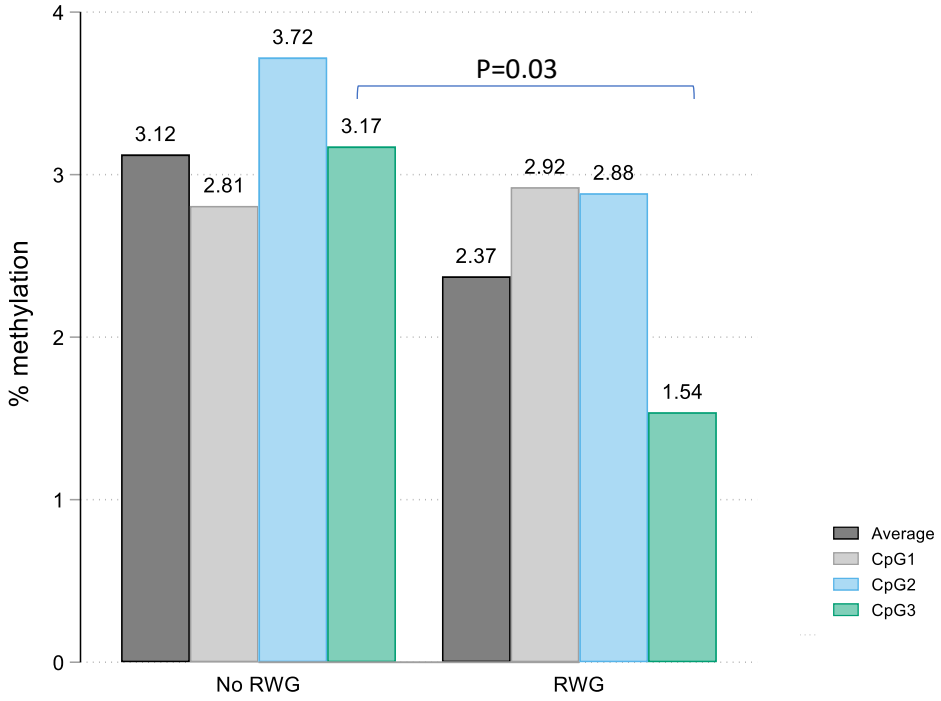
Of the 3 CpG sites present in the region examined, CpG3 was the original target (cg11531579). Correlations between the CpG sites were examined with Pearson correlation coefficients (*Table 7.3*), with all loci demonstrating strong, positive correlations (p<0.0001). Correlations were also similar between those who had RWG compared to those who did not (*Table 7.3*).

Table 7.3 Pearson correlation coefficients between the CpG loci examined in the NTFS samples stratified by RWG

	All			**No RWG**			**RWG**		
	Cpg1	Cpg2	Cpg3	Cpg1	Cpg2	Cpg3	Cpg1	Cpg2	Cpg3
Cpg1	1			1			1		
Cpg2	0.97	1		0.98			0.97	1	
Cpg3	0.92	0.97	1	0.92	0.97	1	0.94	0.97	1

Significance level for all p<0.0001

Overall, average (median) methylation was lower in those who had RWG (Figure 7.5). At CpG3, this difference was statistically significant (Wilcoxon ranksum, p=0.03).



Cpg loci	No RWG						RWG						
	n	p50	Min	max	p25	p75	n	p50	min	max	p25	p75	p
Average	78	3.12	0	95.46	1.22	6.05	56	2.37	0	69.53	0	4.98	0.077
Cpg1	78	2.81	0	99.3	0	5.86	56	2.92	0	72	0	5.89	0.55
Cpg2	78	3.72	0	95.81	1.5	7.49	56	2.88	0	70.59	0	6.15	0.082
Cpg3	78	3.17	0	91.27	0.91	5.33	56	1.54	0	66.01	0	3.87	**0.027**

Figure 7.5 DNA methylation (%) at the CpG loci by RWG.
Median (p50), lower quartile (p25),upper quartile (p75), minimum (min) and maximum (max) values shown for all CpG stratified by RWG. Comparison between the two groups performed by Wilcoxon rank-sum test (p).

Median regression was used to examine the relationship between methylation (age 50) and RWG, adjusted for confounders (sex and/or birthweight z-score) (Table 7.4). Methylation was significantly lower (-1.6%, p=0.03) for the target loci (CpG3, Table 7.4). RWG remained significantly associated after adjustment, with 1.9% lower methylation in the adjusted (age, sex and birthweight) model.

Table 7.4 Median regression models for RWG (exposure) and methylation (outcome) at loci adjusted for confounders in NTFS

Exposure	RWG			Adjusted for sex			Adjusted for birthweight z-score		
	Coef	CI	p	Coef	CI	p	Coef	CI	p
Average	-0.66	[-2.32,1.00]	0.43	-0.88	[-2.56,0.80]	0.30	-0.83	[-2.62,0.97]	0.37
CpG1	0.33	[-1.78,2.43]	0.76	0.81	[-1.38,2.99]	0.47	-0.22	[-2.60,2.17]	0.86
CpG2	-0.81	[-3.04,1.42]	0.48	-1.0	[-2.93,0.93]	0.31	-1.13	[-3.09,0.84]	0.26
CpG3	-1.51	[-3.11,0.09]	0.065	-1.84	[-3.44,-0.24]	**0.024**	-1.89	[-3.65,-0.13]	**0.035**

Methylation at each individual CpG loci was the outcome and RWG the exposure (*n*=134). Models were unadjusted, adjusted for sex, adjusted for sex & birthweight z-score (bwtz). Birthweight and sex were not-significant predictors of methylation. Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p).

The assay was not able detect methylation in the full range of 0-100%. The upper and lower levels of detection of percent methylation observed from the assay validation are presented in Table 7.5.

Table 7.5 Tobit upper and lower censoring cut-offs from the calibration curves for each CpG loci

CpG	Lower level (%)	Upper level (%)
1	4.2	86.7
2	3.6	83.0
3	3.4	77.2

A Tobit model was utilised to take into consideration the assay upper and lower detection limits in the observed values. There were no significant associations between RWG and DNAm in the Tobit models (Table 7.6). A large proportion of observations were left censored on the basis of their methylation values (left censored <3-4%, *n*=80; uncensored, *n*=53; right censored, *n*=1). The estimates were higher for the Tobit model (non-significant), which may be due to the censoring at 3%, which is the lower detection level whereby DNAm is not quantifiable. In this case, censoring should produce less biased estimates.

Table 7.6 Tobit regression models for RWG (exposure) and DNAm (outcome) at each CpG loci adjusted for confounders in NTFS

	RWG			Adjusted for sex			Adjusted for bwtz		
	coef	CI	p	Coef	CI	p	coef	CI	p
CpG 1	-3.8	[-15.44,7.83]	0.52	-3.61	[-15.20,7.98]	0.54	-4.96	[-17.52,7.59]	0.44
CpG 2	-4.65	[-12.98,3.68]	0.27	-4.63	[-12.94,3.68]	0.28	-5.49	[-14.41,3.43]	0.23
CpG 3	-5.85	[-15.91,4.22]	0.26	-5.85	[-15.93,4.22]	0.26	-5.90	[-16.75,4.96]	0.29

Upper and lower censoring values determined from calibration curves.
Coefficients for sex and birthweight z-score (bwtz) were not significant. *N*=134.
Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p).

Methylation at CpG3 was significant associated with RWG in the median regression model after adjustment (Table 7.4) but was not in the Tobit model (Table 7.6). As Tobit regression predicts the mean, and median regression the median, results are not directly comparable. Although estimates were in the same direction, median regression gave lower estimates, with concise confidence intervals that were significant, whilst Tobit regression had higher coefficients that did not reach statistical significance.

7.4.7 The relationship between DNAm and body composition (age 50)

Methylation levels were examined in participants with obesity, compared to those without obesity (i.e. a healthy weight or overweight BMI), and in study members with OWOB compared to healthy weight (Figure 7.6). Methylation levels at CpG1 and CpG3 were lower in study members with OWOB and OB, compared to healthy weight, whilst median methylation at CpG2 was similar across the body composition categories.

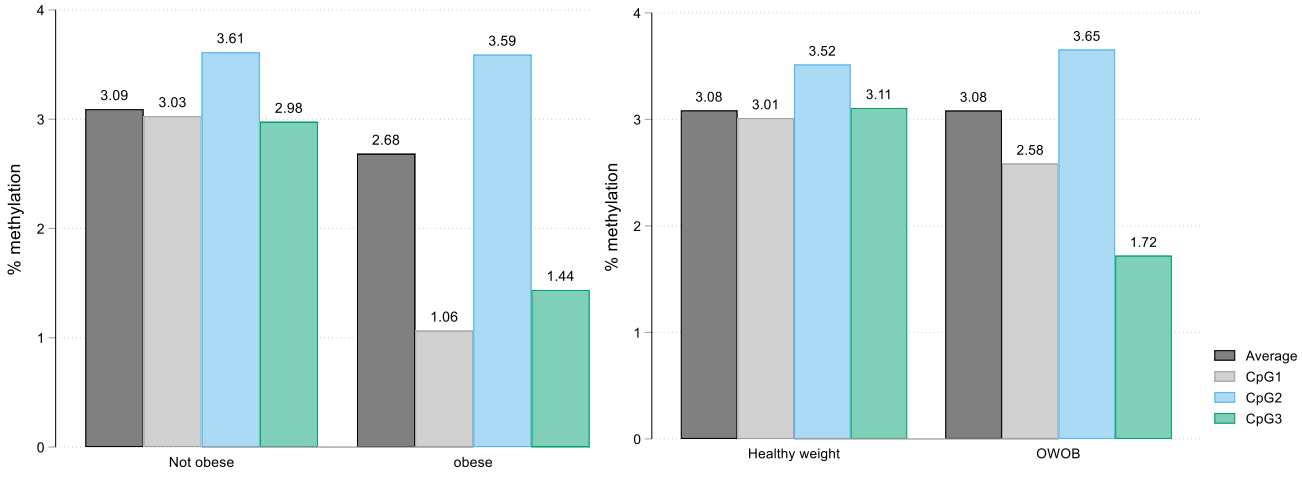


Figure 7.6 Median % DNA methylation at each CpG loci and the average, by body composition (age 50). There were no significant differences in methylation between body composition groups.

Although methylation levels were lower in participants with obesity compared to those not obese, there were no significant differences between the groups (Table 7.7). This could have been due to the small proportions with obesity in this sample. To include a larger sample, methylation levels in those OWOB were also analysed. Again, methylation was lower in OWOB participants, but these differences were not statistically significant. Likewise, there were no significant differences for age 60 body composition outcomes.

Table 7.7 Median % methylation levels at each CpG by body composition outcome (age 50) in NTFS

	Not Obese						Obese						
	N	p50	min	max	p25	p75	n	p50	min	max	p25	p75	Ranksum p value
Average	114	3.09	0	95.46	0.81	5.3	18	2.68	0	65.12	1.15	3.85	0.76
Cpg1	114	3.03	0	99.3	0	5.88	18	1.06	0	66.34	0	4.02	0.25
Cpg2	114	3.61	0	95.81	1.18	7.01	18	3.59	0	64.75	1.23	9.29	0.73
Cpg3	114	2.97	0	91.27	0	4.45	18	1.44	0	64.28	0	3.79	0.59
	Healthy weight						OWOB						
	N	p50	min	max	p25	p75	n	p50	min	max	p25	p75	Ranksum p value
Average	58	3.08	0	95.46	1.08	4.8	74	3.08	0	65.12	0.95	6.05	0.71
Cpg1	58	3.01	0	99.3	0	4.58	74	2.59	0	66.34	0	7.37	0.82
Cpg2	58	3.52	0	95.81	0.94	5.24	74	3.66	0	64.75	1.23	7.54	0.39
Cpg3	58	3.11	0	91.27	0	4.36	74	1.72	0	64.28	0	4.67	0.53

Median (p50), lower quartile (p25), upper quartile (p75), minimum (min) and maximum (max) values shown for all CpG stratified by body composition. Ranksum p value for differences in methylation between obese and not obese, or OWOB and healthy weight.

In multivariable analyses, methylation at either loci was not associated with weight outcomes: obesity, OWOB or BMI at age 50 (Table 7.8). Whilst methylation was lowest in those who had RWG and were OWOB (age 50) in the sub-phenotype groups, however these differences were not significant (Appendix F, Figure XXIII).

Table 7.8 16 Logistic/linear regression models for weight outcomes (age 50) and % DNAm (age 50) in NTFS

	OB			OWOB			BMI		
	OR	CI	p	OR	CI	p	coef	CI	p
Average	1	[0.96,1.03]	0.86	1	[0.98,1.03]	0.89	0.01	[-0.04,0.06]	0.76
Cpg1	0.99	[0.96,1.03]	0.76	1	[0.98,1.03]	0.70	0.01	[-0.04,0.05]	0.78
Cpg2	1	[0.97,1.04]	0.99	1	[0.98,1.03]	0.79	0.01	[-0.04,0.06]	0.68
Cpg3	1	[0.96,1.03]	0.83	1	[0.97,1.02]	0.81	0.01	[-0.04,0.05]	0.83

Weight outcome was the dependent variable and DNAm the independent variable. Healthy weight was the reference category for the OWOB and OB logistic regression models. Adjusted for sex. Odds ratios (OR) and coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p)

7.4.8 The relationship between DNA methylation and subsequent body composition (age 60)

There was a positive relationship between methylation (CpG3) at age 50 with BMI at age 60 (Pearson r=0.25, Figure 7.7). Additionally, there was a slightly stronger positive association between methylation (age 50) and the change in BMI from age 50 to 60 (Pearson r=0.31).

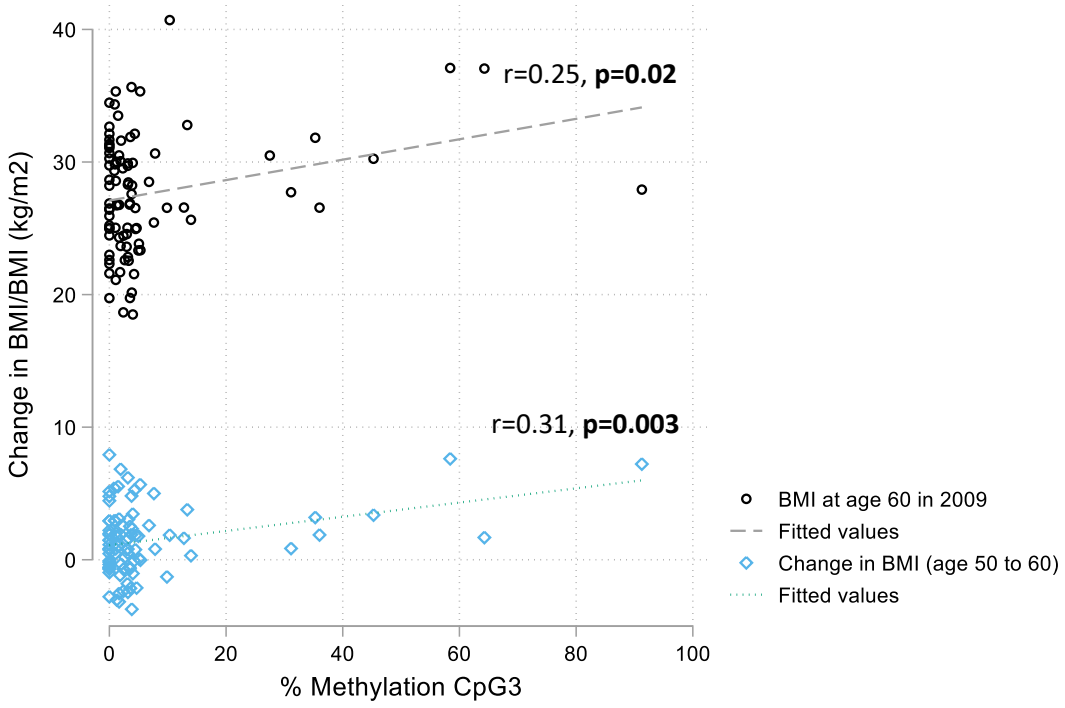


Figure 7.7 Scatter plot for DNAm (age 50) and BMI (age 60) and change in BMI (age 50-60).
Pearson correlation (r) and associated p value.

For outcomes at age 60, there were no significant differences in methylation between phenotype groups (RWG and healthy weight/OWOB) (Appendix F). CpG methylation at this locus explained an additional 3% of the variation in DNAm after controlling for current body composition (Table 7.9).

Table 7.9 Linear associations between BMI (age 60) with % methylation (age 50) and BMI (age 50) in NTFS study members

BMI (age 60)	% methylation			BMI (age 50)			Adjusted model		
	coef	CI	p	coef	CI	p	coef	CI	p
DNAm (%)	0.08	(0.156,0.14)	**0.015**				0.065	(0.02,0.09)	**0.03**
BMI (age 50)				0.97	(0.83,1.12)	**<0.001**	0.96	(0.81,1.10)	**<0.001**
n	91			90			90		
Adj R^2	0.05			0.67			0.70		

All models adjusted for sex (non-significant). Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p) and adjusted R^2 of the model.

Whilst there appeared to be a relationship between DNAm and BMI (age 60), because of the clustering of many values around 0% methylation and some high methylation values, the

nature (linear or non-linear) of the association was uncertain. Therefore, a fractional polynomial (FP) model was used to examine the relationship between methylation and subsequent BMI. The initial model was skewed by a recognisable outlier with methylation greater than 80% (circled, Figure 7.8). When fractional polynomial linear model was re-run without the outlier, there was no relationship between methylation and BMI at methylation levels <15%, but between 15-60% methylation there was a positive relationship between methylation and BMI (age 60) (Figure 7.8). This suggests that this association could be driven by high methylation outliers, with less certainty at higher values (wider confidence intervals, Figure 7.8B).

Generally, for a FP model, the best model is that with the lowest deviance. The best-fitting fractional polynomial had 1 exponent to the power 2 (squared term) and deviance 518.9, forming a curved relationship. There was no FP model that fit these methylation data for BMI at age 50.

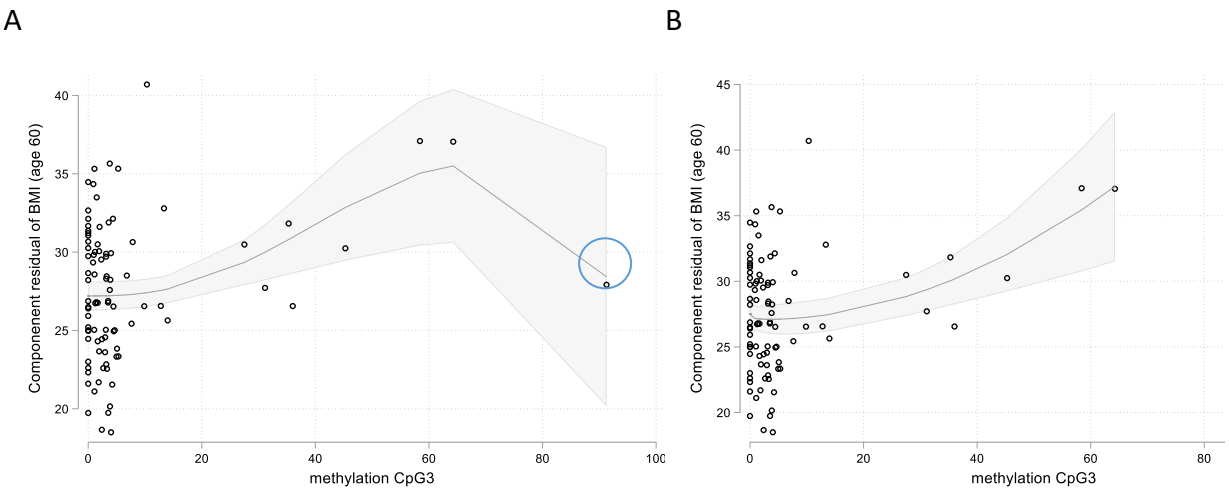


Figure 7.8 Fractional polynomial model plots for CpG3 methylation and BMI (age 60).
Models are adjusted for sex (not a -significant predictor), with all data points (A), and without the outlier (>80% methylation, circled) (B). For model A; Coefficient for methylation=0.08, CI 0.00-0.17; p=0.05, *n*=90. For model B; Coefficient for methylation=0.24, CI 0.09-0.38; p=0.001, *n*=90. Confidence intervals shown as shaded area around the line of best fit. *As there were few data points with methylation >20%, these results should be interpreted cautiously.*

7.4.9 Data validity

The range of methylation values obtained for the CpG loci of interest was unusual in that genome-wide methylation has a bi-modal distribution, whereby methylation levels at a

single locus are usually either high or low. In this case, this raises the question as to whether the high levels of methylation are valid.

Pyrosequencing assays were run in duplicate and replicates were within 5% methylation, which would imply the pyrosequencing results are precise. This phenomenon was investigated further to determine whether this represented measurement error or a SNP effect, and the extent of the effect of the outliers.

7.4.9.1 Assay measurement error

The high methylation samples all had low DNA concentrations (*n*=12, shaded points on Figure 7.9). It is likely that the low concentration samples had low total amounts of DNA. However, many samples had low DNA concentrations and there were some samples which had low methylation and a low DNA concentration (*n*=32). Therefore, it does not appear to be likely that the assay preferentially amplified based on methylation in low DNA concentration samples.

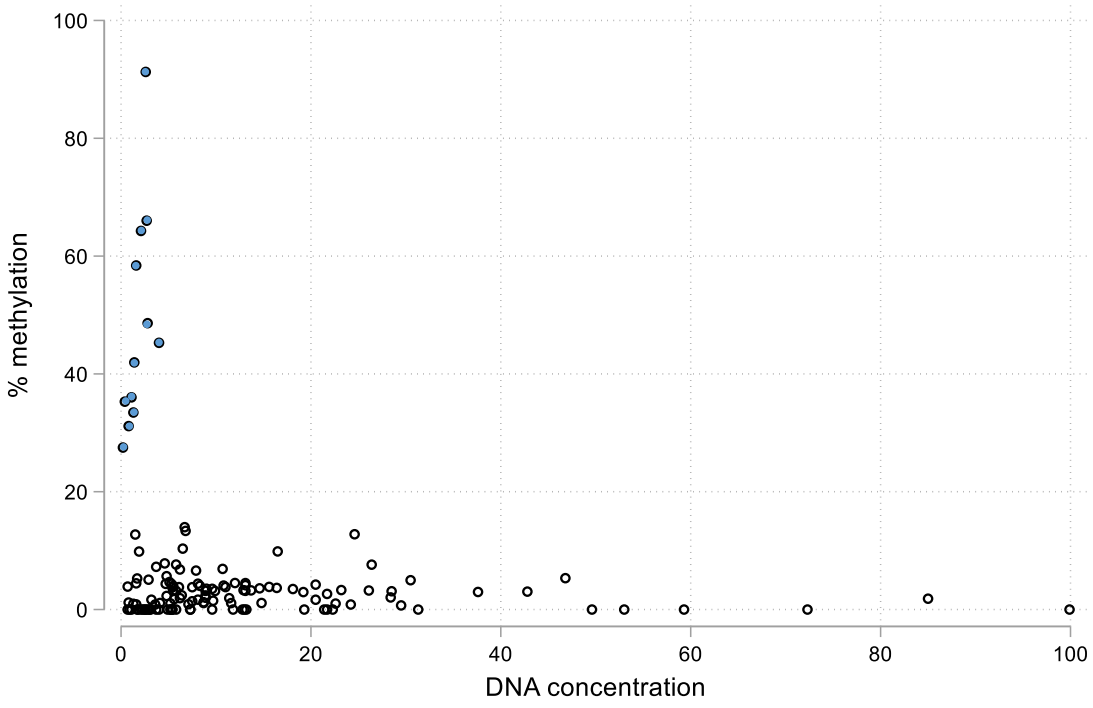


Figure 7.9 Plot of methylation at CpG3 (%) and DNA concentration (ng/µL) of the sample.
Shaded are the high methylation samples, which have a DNA concentration <5ng/µL.

7.4.9.2 Understanding any potential SNP effects on DNA methylation

The CpG loci of interest was located at the start of the 450K probe (Table 7.10). The region up- and down-stream of this loci was sequenced in a matched experimental design to highlight any nearby SNPs that could influence methylation.

Table 7.10 450K array probe characteristics for cg11531579

Probe details			Chromosome			
Probe ID	Forward genomic sequence	CpGs	No.	Start position	End position	site
cg11531579	CGAGTAGATGAACACATTTA AAGTTTGTAGTTAAGAGGA AAACAACGCCA	2	12	133,484,743	133,484,792	133,484,743

The CpG (cg11531579) is underlined in the probe sequence.

The right strand for each sequencing reaction was sequenced successfully, with some drop off in quality towards the end. There were a handful of SNPs within the sequences, however no clear pattern that distinguished high and low methylation samples, or paired samples. The CpG loci of interest was at position 110 in the sequence. There were some differences between high and low methylation samples at positions 90, 339 and 424 (Appendix F, Figure XXV). Pair B showed consistent SNP patterns for the matched samples, but there were differences in SNPs for the matched pairs A and C (Appendix F, Figure XXV).

There were nucleotide differences observed for the pairs (A and C) with characteristics of RWG and a BMI>25kg/m^2 (overweight category), whereas there was no difference between the pair (B) which had no RWG/healthy weight. This could suggest that genetic variation is occurring in a phenotypic-dependent manner. However, with the small sample size this cannot be said with great certainty.

Table 7.11 SNP differences in sequenced samples of matched pairs of high (>12%) and low methylation (<12%) in NTFS DNA samples

	Matched pairs	A		B		C	
Sample characteristics	Sample ID	TFS55	TFS314	TFS59	TFS296	TFS270	TFS152
	RWG	Yes		No		Yes	
	Sex	Female		Female		Female	
	BMI	29		25		27	
	Methylation	Low	High	Low	High	Low	Very high
	% methylation	2	13	3	13	4	45
Nucleotide at specific position in aligned	Position 90	G	A*	HZ	HZ	G	HZ
	Position 339	C *	T	C*	C*	C*	HZ
	Position 424	T*	C	C	C	T*	C
	Pair differences	✓		X		✓	

sequences							

Diff, differences in matched pairs; Pos, position in sequence read; %, percent methylation; HZ, heterozygous; * indicates major allele.

The SNPs at positions 90 and 424 were known SNPs, whereas position 339 was not a known SNP (Table 7.12). The known SNPs (rs2873193, rs4758916) were examined for linkage disequilibrium, however evidence suggested that these genetic variants occur independent of one other (Appendix F, *Figure XXVI*).

Table 7.12 SNPs identified in the sequenced region in NTFS paired samples and SNP characteristics

SNP name	Position in aligned sequence	Genome position	Allele	Frequency (in European population)	Study
rs2873193	90	chr12:133,484,722	A>G	A=0.724 G=0.276	1000 Genomes
Unknown	339	chr12:133,484,970	C>T	-	
rs4758916	424	chr12:133,485,056	T>C	T=0.048, C=0.952	1000 Genomes

Allele frequencies were determined from (phase3 release V3+)(Sudmant et al., 2015).

7.4.9.3 Sensitivity analysis

Sensitivity analysis was done to examine if the extreme values (>12% methylation) affected the results. This was done for CpG3 only, as this was the CpG loci of interest and the only loci that demonstrated any significant associations with exposures and outcomes. With outliers excluded, the mean was closer to median, and the standard deviation decreased (Table 7.13).

Table 7.13 Descriptive statistics for CpG3 methylation with (all) and without outliers, stratified by RWG

All	N	Mean	p50	p25	p75	Min	Max	SD
Total	134	6.83	2.81	0.00	4.49	0.00	91.27	14.41
No RWG	78	8.10	3.17	0.91	5.33	0.00	91.27	16.01
RWG	56	5.08	1.54	0.00	3.87	0.00	66.01	11.75
Excluding outliers	**N**	**Mean**	**p50**	**p25**	**p75**	**Min**	**Max**	**SD**
Total	118	2.40	1.88	0.00	3.82	0.00	10.34	2.45
No RWG	67	2.79	2.65	0.00	4.36	0.00	9.87	2.52
RWG	51	1.90	1.19	0.00	3.77	0.00	10.34	2.28

Outliers n=16. N, sample size; p50, median, p25, lower quartile; p75, upper quartile; min, minimum; max, maximum; SD, standard deviation.

In the median regression model when outliers were excluded, the coefficient for RWG increased and the p value was smaller (Table 7.14). Therefore, in the range of methylation from 0-12%, there was a stronger association between DNAm and RWG, than in the original

(wider) range of values. Most individuals exhibited methylation in the range 0-10%, suggesting that for the majority of the cohort, RWG was associated with -2% methylation.

Excluding the outliers did not result in any changes in significance in the Tobit model.

Table 7.14 Comparison between Tobit and median regression model results for the models with and without outliers excluded.

Model	Adjusted for sex and bwtz		
Tobit regression	coef	CI	p
Original	-5.85	[-15.91,4.22]	0.26
Outliers removed	-2.41	[-24.57,19.74]	0.83
Median regression	coef	CI	p
Original	-1.89	[-3.65,-0.13]	**0.035**
Outliers removed	-1.95	[-3.36,-0.54]	**0.007**

Methylation at CpG3 was the outcome and RWG was the main explanatory variable. Models were additionally adjusted for sex and birthweight z-score (bwtz). The original model included all the data points, whilst the outliers model excludes methylation values >12%. Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance

Methylation at CpG3 was not associated with body composition outcomes (at age 50) with or without outliers, in linear or logistic models (Table 7.15).

Table 7.15 Outlier analysis with body composition outcomes (age 50) using logistic and linear regression models, for all and stratified by RWG.

Outcome	OB			OWOB			BMI		
	OR	CI	p	OR	CI	p	Coef	CI	p
All study members, *n*=118									
Original	1	[0.96,1.03]	0.83	1	[0.97,1.02]	0.81	0.01	[-0.04,0.05]	0.82
Outliers removed	1.04	[0.85,1.28]	0.71	0.98	[0.85,1.15]	0.84	0.01	[-0.04,0.05]	0.83

Models are adjusted for sex. Sample sizes refer to the sample without the outliers. Odds ratios (OR) and coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p).

When outliers were excluded, there was an association between DNAm and subsequent body composition measures (age 60), as determined using a fractional polynomial model (Figure 7.10). The best fitting model was a with a cubic term (Coefficient = 5.5, CI 0.05-10.9, p=0.05). Both models had widening confidence intervals at higher values (Figure 7.10).

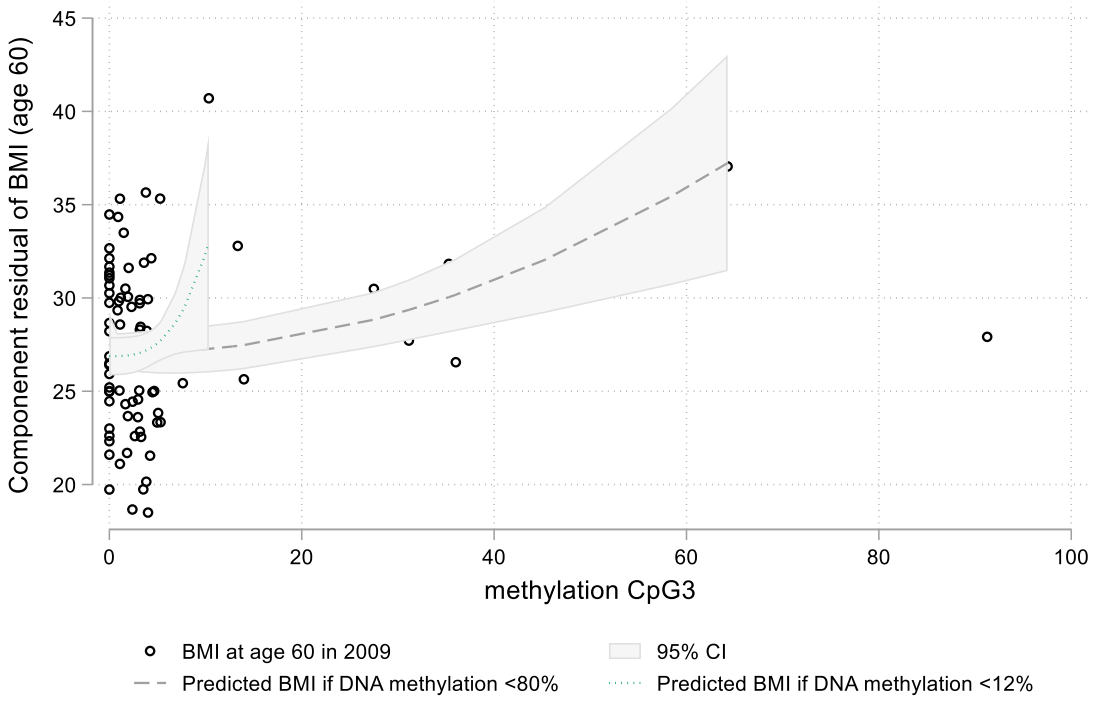


Figure 7.10 Fractional polynomial model for BMI (age 60) and methylation at CpG3 (age 50), for models with and without outliers.
The model with the original data is to power (2) whilst the model without outliers is to power (3).
The outlier model had coefficient = 5.5, CI 0.05-10.9, p=0.05.

Furthermore, from the FP model it appeared that BMI was only associated with the outliers (i.e. a change in slope at methylation >20%). To examine this, a linear regression model was run with BMI age 60 as the outcome, and outliers (>12% methylation) as a binary explanatory variable (study members with methylation >12% vs <12%) (Table 7.16). Having DNAm levels greater than 12% was associated with a 3.11kg/m^2 increase in BMI at age 60. High methylation remained associated with subsequent BMI after controlling for current BMI, however current BMI explained much more of the variation in BMI (age 60) and was therefore a better predictor.

Table 7.16 Linear associations between outlier DNAm (>12% methylated) and BMI (age 60)

BMI (age 60)	**Unadjusted**			**Adjusted for BMI (age 50)**		
	coef	CI	p	coef	CI	p
Outliers (>12% methylation)	3.11	[0.30,5.91]	**0.03**	1.96	[0.27-3.64]	**0.02**
n	91			90		
R^2	0.06			0.69		

Both models additionally adjusted for sex. Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p)

7.5 Discussion

An assay was successfully developed to measure DNAm at the CpG site associated with RWG in the ALSPAC cohort (cg11531579). DNAm levels in the region of interest were quantified in 134 viable samples from the NTFS cohort. Median DNAm at the cg11531579 locus was 2.8% and was lower on average in those who had RWG (3.2% vs 1.5%). Median regression revealed that RWG was associated with 2% lower methylation after adjustment for confounders. Whilst methylation tends to increase over time at this locus (BSGS, section 7.3.1), on average in NTFS adults those who had RWG had lower methylation.

There were differences when analysing the outliers, which suggested that RWG was more significantly associated with reduced methylation in the lower range of methylation values. There were also differences observed when utilising a censored regression model, which questions the accuracy of the results that lie outside the (lower) range of detection of the assay, perhaps due to the uncertainty introduced from many censored values, or due to reduced statistical power.

A wide range of methylation values were observed, including some very high values. It is possible that outliers could come from technical artefacts, however this seems unlikely in this as repeats were within 5% of each other, and some high methylation samples were re-run and values were similar. White *et al*,. found that low concentrations of template DNA (<10ng) can affect absolute quantification due to introduction of PCR bias (White et al., 2006). Increasing the number of technical replicates, or bisulfite sequencing could have been potential options to investigate further if there was abundant DNA remaining. Ideally if there was more DNA available, then all samples would have had greater starting material, and samples could have been run in PCR triplicate.

The DNAm changes identified here were small, particularly considering that technical replicates can often have methylation differences up to 10% (Dedeurwaerder et al., 2014). Here, replicates were only accepted if within 5% of one another. There is the possibility that methylation could be due to random technical variations rather than true biological differences (Dedeurwaerder et al., 2014). For individual CpG sites the pyrosequencer detection limit is approximately 5% (Mikeska et al., 2011), therefore meaningful differences may be too small to accurately quantify. Ideally differentially methylated loci should be confirmed using an independent assay and technique (i.e. next generation sequencing which is costly) in the same population.

Statistical methods were used which are appropriate for the data (without excluding points), as if this is a 'true' event, then removing the data points would generate a selection bias. On the other hand, an argument to exclude outliers would be in that the utility of a biomarker is that it applicable for most of the population, hence sensitivity analysis was done to examine the impact of outliers. Outliers could be a result of a difference in measurement method, genetic variation, cell composition, or disease state (discussed further in section 8.7).

The range of methylation values detected by pyrosequencing in the samples was higher than those observed in the epitect control DNA samples for a given methylation level. A possible explanation for this could be that the epitect control DNA, which uses whole-genome amplification (0% methylated) and SssI treatment (CpG Methyltransferase which methylates all cytosine residues), does not result in 0 or 100% methylation at all CpG loci (Choi et al., 2011). In another study which used the epitect control DNA, the values obtained for supposed 0% methylation at various CpG loci ranged from 0-23% (Nishitani et al., 2018). The methylation detection limit of the assay by the pyrosequencing machine also had a lower and upper threshold, and therefore relying on the control DNA to determine will impact on the specified detection ability of the assay and would invalidate the Tobit model results.

The high methylation levels were investigated further to determine if SNP patterns were driving differences between high and low methylation. There is literature on the two identified SNPs. Sequencing results were suggestive of a phenotypic (RWG and OWOB) pattern of methylation related to SNPs. As there were few samples with sequencing data and with very high methylation, these results could be spurious findings. If the results are not by chance, then high levels of methylation could be driven by SNP effects or be indicative of future body composition in those with high BMI. However, as there were few samples with high methylation, and only 3 matched pairs were successfully sequenced, these results are inconclusive, but this could be the focus of future research.

There was a wide range of methylation values could indicate a high degree of heterogeneity at this locus. For example, this could occur if the majority of cells (blood is comprised of numerous cell types) exhibit low levels of methylation but some exhibit very high levels. Outliers could arise from abnormal methylation patterns from white blood cells (e.g. from infection or cancer), or could be influenced by disease status or lifestyle (Alegría-Torres et al., 2011).

DNAm did not distinguish between healthy weight and OWOB (age 50), regardless of previous RWG. However, there was a positive association between DNAm and subsequent BMI (age 60), which on further inspection appeared to be non-linear. Those who had very high methylation also happened to have a greater BMI at age 60. The CpG methylation at loci improved prediction of subsequent BMI (age 60) over and above the use of current BMI (age 50). A 1% increase in CpG methylation at this 1 locus was associated with a 0.07 Kg/m^2 increase in subsequent BMI and explained around 3% of the variation in NTFS participants. However, this may have limited use as a biomarker in this age group, as BMI at age 50 was more predictive of BMI at age 60 than DNAm, but CpG methylation could be used to improve predictive models alongside BMI. This could also suggest that similar to findings in ALSPAC, relative increases in methylation are reflective of subsequent growth.

Chapter 8. Discussion

This final chapter is a discussion of the main findings of results chapters, and how the findings compare with the published literature. The chapter will also include the study strengths and limitations, and the public health implications and future directions.

8.1 Summary of the main findings

Early life exposures were primarily investigated in North East populations of children and adults, in two cohorts that commenced over 50 years apart. Over this time, the prevalence of obesity has vastly increased, from what were very low numbers in 1947.

Despite the change over time to a more obesogenic environment, early life weight gain was consistently associated with a higher BMIz in both NTFS and GMS cohorts, even when accounting for birthweight. RWG is characterised by initial rapid growth, and generally BMI z-scores decrease and 'normalise' over time. However, in the modern cohort those who had RWG had a higher, sustained BMIz throughout childhood.

The exposures that were important factors for childhood OWOB in the North East cohorts were investigated further in the larger, South West cohort; ALSPAC. Again, there were consistent associations for early life weight gain, both dependent and independent of birthweight. However, when examining long-term effects, the relationship between early life weight gain and BMI in adolescents weakened over time in ALSPAC and was non-existent in NTFS adults. Thereby questioning whether there are long-term effects, or if other factors become more important predictors of body composition.

The programming effect of early life risk factors on childhood BMI may arise through epigenetic marks laid down at an early developmental stage, which then elicit effects at a later stage (Mathers and McKay, 2009). Like the epidemiological findings with RWG and childhood BMI, there was a positive association between RWG and DNAm. DNAm changes were associated with RWG, but less so with RT, which may suggest that this includes some of the effects of birthweight. To examine if there were latent effects of RWG on DNAm, methylation at the significant locus (cg11531579) was verified by pyrosequencing in NTS adults. Interestingly, DNAm at this locus was negatively associated with RWG in NTFS (age 50), suggesting variability. In childhood, the highest levels of methylation were in those who had RWG and were subsequent OWOB. At age 50, methylation was positively associated

with subsequent BMI (age 60). Therefore, DNAm at this locus may reflect patterns of growth; increased methylation associated with increased subsequent growth. However, due to the differences in age, time period, and methods of measurement, direct comparisons are inadvisable. Furthermore, the methylation differences observed were small and findings need to be replicated.

There were other important early life risk factors identified in the epidemiological analyses, including birthweight, adversity, infection, maternal age and SES. There were more early life factors related to childhood body composition in the modern (GMS) cohort, which may suggest these factors have become more important over time, or that there is now greater variability in early life environments.

8.2 Discussion of the epidemiological findings

8.2.1 Birthweight

The positive associations between birthweight and adiposity outcomes in children in the modern cohorts agrees with the literature (Reilly et al., 2005a, Yu et al., 2011). In contrast, there was no association in NTFS children, and a protective effect of higher birthweights in adults, which could be due to time period effects. NTFS study members were *in utero* during a time of rationing in Britain, whereby it is generally accepted that the nation's health improved, which could have reduced numbers of LBW infants. Particularly rationing will have benefited less advantaged groups more and mothers were less likely to be undernourished. Similarly, theories around a mismatched intrauterine and postnatal environment do not apply to this cohort, as the period of rationing extended throughout early childhood, meaning a consistency in nutrient intake throughout pregnancy and early life. Furthermore, considering the time period there will have been fewer obese mothers due to lower prevalence of obesity generally, therefore findings in the modern cohort could reflect the cycle of higher birthweights and obesity across generations (Cnattingius et al., 2011).

8.2.2 Adversity

The components of adversity that were significantly associated with body composition in GMS had socioeconomic implications (Appendix C, Table IX). Whilst in NTFS, neither adversity nor early life SES were associated with any outcome measures in childhood or mid-life, which could indicate that the effects of adversity on BMI impact via SES.

An alternative explanation could be the timing, as in NTFS the adversity exposure was postnatal, whereas significant associations in ALSPAC and GMS were identified for prenatal adversity. This may support an intrauterine mechanism of altered stress responses, leading to disturbances in metabolism (Entringer et al., 2012, Johnson et al., 2013). However, early life stress could also impact on the development and calibration of the neuroendocrine-immune network, and have downstream effects on various social, economic, academic and behavioural aspects of a child's life, which could also influence likelihood of obesity (Johnson et al., 2013).

Another potential explanation could be that the effects of adversity do not persist to adulthood, although this was the case in ALSPAC, whereby associations were stronger over time (in adolescence). Furthermore, a large systematic review and meta-analysis found weak to moderate associations for adverse childhood experience (ACEs) and obesity in adulthood (OR 1·39 (95% CI 1·13–1·71) when examining multiple ACES (>4), which may be indicative of more severe childhood adversity (Hughes et al., 2017). A recent methodological approach has exploited the high dimensionality of the ALSPAC data to derive ACE scores and constructs from the numerous adversity questionnaires (Houtepen et al., 2018b). These scores which utilise multiple measures can also deal with missing data, and could be used in future analyses.

8.2.3 Infections

In NTFS bacterial infection was associated with OB and also with body fat in women. There was a lack of robust infection data in GMS, however, there were some indications in ALSPAC that antibiotic exposure in the first 6 months (as a proxy for bacterial infection) was associated with OWOB until adjustment for early life growth (correlated with antibiotic exposure). This could suggest a link between early life infection, antibiotics, and subsequent growth. The antibiotic growth effect is observed in farm animals who are given antibiotics to increase their weight, but has also been observed in humans (Haight and Pierce, 1955). Research seems to suggest that the mechanism linking antibiotics and growth could be via the microbiome (Cox and Blaser, 2014, Forrest et al., 2017).

A longitudinal study that compared infants with infections who either did or did not receive antibiotics found no difference in obesity risk up to 18 years (Li et al., 2017a). However, the number of infections was significantly associated with increased risk of obesity in those who

did not have antibiotics. Thereby the authors concluded that the exposure to infection rather than the antibiotics was the important factor.

The timing of antibiotic exposure could also be important, as early-life post-natal exposure was associated with increased child BMI, but not pre-natal (Poulsen et al., 2017). In support of this, in a large sample of mother-child pairs (*n*=53,320) also found that maternal antibiotic exposure during pregnancy was not associated on childhood BMI-z at 5 years (Heerman et al., 2019).

Furthermore, this study supports that bacterial, rather than viral infection is associated with increased BMI, however this could be due to the types of viruses and bacterial infections examined in NTFS, some of which are rare now.

8.2.4 Maternal age

Although maternal age has been related to detrimental birth outcomes, in these cohorts maternal age was not directly associated with an unfavourable body composition or birthweight. Instead, the association of older (NTFS) and younger (GMS) maternal age with decreased offspring BMI is an interesting observation and is opposite to previous findings (Myrskyla and Fenelon, 2012). However, maternal age may not be causally linked to body composition but could have indirect route (for example via parity, Figure 4.2), or could be confounded by SES.

This positive effect of older maternal age was not seen in ALSPAC, and therefore could be an oddity of the North East cohorts, or a spurious finding due to smaller sample sizes. In other studies, maternal age above 40 (Fuchs et al., 2018) or 45 (Myrskylä and Fenelon, 2012) has been associated with adverse offspring health outcomes. As the maximum maternal age in NTFS was 45, it is plausible that there were no any adverse associations, as the 'older' maternal age group were relatively young. Although, aside from for birthweight or length of gestational, other adverse offspring outcomes have been shown to remain fairly stable after a maternal age in the mid-30s (Carslake et al., 2017).

These results may suggest that in older mothers, whilst intrauterine conditions may be less optimum, the life experiences and enhanced childhood environment may confer certain advantages.

8.2.5 Infant feeding

In these cohorts, there was no evidence of a protective effect of breastfeeding on obesity, independently or after adjustment for socioeconomic factors. In contrast, breastfeeding was associated with small increases in BMI in NTFS adults. Whilst studies have found long-term benefits of breastfeeding (such as reductions in OWOB) in developed countries, studies in low and middle income countries have not shown such great effects, which could suggest differences in social and cultural structures, selection bias or confounding could in part explain some of the association (Kramer et al., 2009, Brion et al., 2011).

Previous analyses in the GMS cohort have shown there is a clear relationship between SES and breastfeeding duration (Wright *et al* 2005), with those in the most affluent quintile 3x more likely to breast-feed initially and 5x more likely to continue breastfeeding past 4 months (Wright et al., 2006a). Wright *et al*,. also found greater rapid weight gain (0-13 months) in those who stopped breastfeeding earliest, although this is likely a result of reverse causation; that starting solids and ceasing breastfeeding is a response to RWG (Wright et al., 2006b). They conclude that babies genetically destined to be larger make greater demands on their mother for breastmilk; and this greater demand increases likelihood of earlier cessation.

A study which utilised structural equation modelling found that shorter breastfeeding duration was associated with increased infant weight gain and subsequent higher childhood BMIz, however these estimates were not adjusted for SES (Lamb et al., 2010). Similar associations between weight gain and breastfeeding have been observed in other studies, with suggestion that any beneficial effect is obtained by 12 months of age (Scholtens et al., 2007). These analyses primarily examined the duration of breastmilk rather than the composition, however formula and other milks have been associated with greater dose-response in weight gain through infancy (first year) (Kramer et al., 2004).

There was a negative association between age at weaning and some outcomes in the GMS cohort (waist OB, BMIz). This is consistent with the literature on early weaning and could be related to the types of foods that are introduced (Pearce and Langley-Evans, 2013), and may explain a lack of association in NTFS. This association for waist OB was no longer significant after controlling for physical activity, which could suggest that this was operating through a shared pathway. Perhaps the families which introduce weaning foods later are generally healthier (more active) or with greater health literacy. Accordingly, a systematic review

determined that low maternal education was an important factor in early weaning (Wijndaele et al., 2009b), and is supported by findings in GMS (Wright et al., 2004).

8.2.6 Physical activity

In GMS and NTFS, there was a negative association between increased physical activity and most adiposity outcomes. These results are similar to other studies in adults (Reiner et al., 2013). Whereas studies on childhood obesity have found mixed results, with associations perhaps influenced by biological or social gender differences (Prentice-Dunn and Prentice-Dunn, 2012). In GMS, physical activity began to decline in both sexes from age 7 years (Farooq et al., 2018), therefore the effect sizes exhibited here may change over time. Females were also less likely to participate in activity and were more sedentary (King et al., 2011). The results here are in agreement (Figure 4.2), suggesting that girls are less likely to take part in MVPA, which could then impact on BMI.

There is also the possibility of bidirectionality between body composition and activity, which is supported by recent findings from GMS suggesting that adiposity influences levels of physical activity (Tanaka et al., 2018).

8.2.7 Socioeconomic status

As rationing reduced disparities in NTFS, we might expect more pronounced inequalities to be in GMS. However, SES at birth was not associated with outcomes in the modern cohorts, but higher social class at birth was negatively associated with BF% and WHR in NTFS males. The fact that SES was not associated with factors in females could be partly attributable to the fact that the UK Registrar General's classification was developed for male workers and may transfer poorly to females. Although in pooled analyses (both sexes), social mobility was associated with drastically decreased odds of obesity in NTFS adults, and those who were always in the most advantaged group had very low odds of obesity.

National level data find a relationship between childhood SES and childhood obesity (Health and Social Care Information Centre, 2016). Similarly, GMS results suggested that SES in childhood is more predictive of OWOB than SES at birth. There was no evidence for increasing upward mobility, which may be because it was basic measure using ownership and employment, rather than encompassing multiple aspects of social mobility. Furthermore, deprivation measured using home ownership is perhaps less informative over time with the resurgence of private renting (Kemp, 2015).

There were no socioeconomic differences in OWOB in ALSPAC either, which is in contrast to other studies. This could be due to the use of occupational social class or Townsend score, rather than parental education, which is most frequently associated with lower odds of obesity (Brophy et al., 2009, Shrewsbury and Wardle, 2012, Ruiz et al., 2016). However, the effects of maternal education appeared to be indirect in GMS, via other factors such as birthweight or parity or lifestyle (Figure 4.2). Similarly in NTFS, early life SES had indirect effects on BMI via later life SES and lifestyle factors (*Figure 3.3*). Brophy et *al*,. suggest that the relationship between maternal education and obesity could be a result of more protective health behaviours and empowerment for people to modify their own health. In pregnant women, this could influence health-related behaviours that affect birthweight, which may explain the relationships exhibited.

A potential limitation when comparing results is the difference in the socioeconomic variables available in each of the cohorts, which may limit interpretations.

8.2.8 Infant and childhood growth

This study is one of many to show a relationship between RWG and childhood adiposity (Stettler et al., 2003, Ong and Loos, 2006b, Druet et al., 2012). In these analyses, this association was observed across cohorts and time points, and therefore suggests this is a key exposure regardless of the modern obesogenic environment. Similarly, a 2005 systematic review investigating infant weight gain in populations born between 1945-94, found consistent associations across various time periods and ages (5-20 years)(Baird et al., 2005). Furthermore, findings from low and middle-income countries also support that early life weight gain is associated with subsequent BMI (Joglekar et al., 2007, Fernandes et al., 2009). Collectively these findings suggest that RWG is a consistent risk factor regardless of environmental influences.

Genetic markers support that early infancy gains in weight and length are on the pathway to adult obesity risk (Elks et al., 2010b). The path analyses attempted to disentangle the direct and indirect pathways and determined that a direct path between RWG and BMI existed, and that no factors predicted RWG, supporting a direct relationship. However, obesity is the result of the interplay and balance between multiple contributing and intermediary factors. Therefore, whilst this is a plausible risk factor, it is unlikely to be a sole determining factor.

An important consideration is that infancy RWG per se may not be causal, but could reflect an individual's predestined growth trajectory, or an antecedent to childhood obesity. However, previous systematic reviews have found weight gain in the first year (specifically, rather than periods greater or less than 1 year) to be most predictive of childhood obesity (Zheng et al., 2018). This does not rule out RWG as an antecedent, but suggests that the first year is a critical period.

However, many children with rapid infancy weight gain do not go on to have increased adiposity in childhood (Wright et al., 2012). Using data from three longitudinal growth cohorts, excess weight in in infancy had a moderate positive predictive value for becoming overweight in mid-childhood, but not necessarily for becoming obese (Wright et al., 2018). This therefore highlights the need for a means to detect those most at-risk (i.e. a biomarker). The entire childhood period, rather than the first year, could also be a critical period for growth and development of obesity (Cole, 2004).

Previous studies which have used BMI are unable to determine whether this reflects changes in fat mass. This study examined multiple outcomes and determined that RWG was also related to FMI, and the relationship was stronger in those with very high fat mass. RWG in infancy has been associated with higher concentrations of insulin-like growth factor I (Ong et al., 2002a) which could increase growth and lean mass. This may explain the relationship between RT and height (which contributes to lean mass), which was also noted in a Swedish cohort of young adults (age 17 years) (Ekelund et al., 2006). They found that RWG in the first 6 months was associated with both fat and fat free mass (measured using air-displacement plethysmography) and WC in adolescence (Ekelund et al., 2006).

In GMS, those with early RT had consistently higher weights throughout childhood; therefore infancy may be a key time to intervene. Children with early increasing BMI diverge from other growth trajectories as early as two years (Robinson et al., 2019), therefore rapid early growth may set infants on a higher weight trajectory over the life course. This also suggests that BMI measurements taken in the NCMP at school-age (age 4-5) will miss some high risk children (Robinson et al., 2019).

Although RWG is associated with increased childhood BMIz, in the historic cohort BMIz normalised over time and RWG was not associated with adult body composition. This could be interpreted as either; factors determining early life fat mass manifest in childhood BMI

and not adult BMI; the obesogenic environment is key; or that RWG is associated with adult BMI through changes in lean mass. Regarding the latter, a Finnish study examined the relationship between early life growth and body composition in adults aged 56-70, and determined that rapid gains in BMI (from 0-1 or 1-2 years) were associated with lean rather than fat mass (Ylihèrsilè et al., 2008).

Similar to findings from ALSPAC (Ong, 2006), early postnatal growth rates were a compensatory mechanism for lower birthweight or growth restriction. LBW infants are more likely to have catch-up growth, which increases CVD risk (Kelishadi et al., 2015). This supports the DOHaD and Barker hypotheses regarding the interaction between intrauterine conditions, environmental factors, and increased risk of metabolic disorders. Interestingly, although waist obesity has been specifically related to metabolic disorders (Hirschler et al., 2005), RWG was not significantly associated with waist OB.

Although LBW makes catch-up growth more likely, Druet et al. (2012) found that the effects of infant weight gain on childhood obesity were similar regardless of birthweight status. This was reiterated when analysing rapid thrive, which accounts for catch-up growth due to LBW. Factors which influence birthweight will impact on likelihood of RWG. For example, maternal smoking during pregnancy has been shown to be associated with increased risk of RWG (Mine et al., 2017). However, adjusting for maternal smoking in the ALSPAC analysis did not notably affect associations. Whilst it was not possible to adjust for maternal BMIz in GMS and NTFS, the associations for birthweight and rapid growth remained in the ALSPAC analyses, and in another ALSPAC study on childhood obesity which adjusted for parental obesity amongst other factors (Reilly et al., 2005a).

Aside from birthweight no factors predicted rapid weight gain in this study. Factors could be related to appetite regulation, feeding behaviour or breastmilk composition. Traffic pollution has also been considered a cause (Fleisch et al., 2015). In GMS, appetite at 6 weeks and 12 months was positively associated with weight gain at 12 months (Wright et al., 2006b). Accelerated growth in infancy could be due to increased intake of nutrients, in particular protein, which has been shown to increase levels of insulin-like growth factor 1 and promotes growth in the first 6 months of life (Socha et al., 2011). RWG is less likely in breastfed infants, which could also be related to the lower protein content of human breast milk compared to formula (Koletzko et al., 2013).

Overall, the findings in this thesis support previous findings that early life growth is an independent, persistent factor related to subsequent body composition.

8.3 The potential causal impact of RWG and future work

Before we can address the question of causation, it is important to establish that a valid association exists. Firstly, it must be determined that there is valid association between the exposure and the outcome. If the association is valid and is not due to bias or confounding, then causality can be addressed. The analysis undertaken here was not intended to determine causal factors, and thus this research design is not able to ascertain causality which is often only possible in randomised controlled trials. Although, in epidemiological studies, the Bradford Hill criteria are often used to assess the strength of a causal relationship and can be considered in this context. These criteria include plausibility, consistency, temporality, strength, specificity, change in risk factor, coherence, analogy and experiment.

Applying these criteria to the principle findings in this thesis, RWG would meet the criteria of plausibility, consistency and temporality. The strength of the association was not investigated here, however others have found that very rapid weight gain, equivalent to crossing 2 growth centile bands was associated with even higher risk of childhood obesity (Druet et al., 2012), which supports a dose-response relationship. There is plausibility in that there is a reasonable pathway to link excess weight gain and obesity. Epigenetics is also a plausible mechanism, however these findings would need to be replicated. There is temporality in that the exposure precedes the outcome. No single study is sufficient to determine causality, however this study is consistent with the previous literature and aetiology. There was also consistency when RWG was analysed in several populations, and across different outcome measures of obesity. Regarding specificity, thus far RWG has not been linked to other diseases and therefore its effects may be specific to obesity. Whether the mechanism is via epigenetic changes (specific), rather than another mechanism (non-specific), remains to be determined.

8.4 Strengths and limitations of the epidemiological analysis

There are several strengths to this study. The three birth cohorts had reasonable sample sizes and prospectively collected early life exposures and body composition measures. Data and methods were harmonised as thoroughly as possible to allow comparison between

studies. Trained professionals collected anthropometric data; therefore, there was no reliance on self-reported measures of adiposity outcomes. Another strength is the comprehensive analyses of multiple outcome measures allowing greater sensitivity to detect early life exposures associated with adiposity.

Whilst a range of exposures and factors were measured routinely in the cohorts, inevitably, the data collection for some variables was limited, and therefore there is potential for unmeasured confounding. Data were not available for other risk factors or confounding factors. For example, maternal smoking (Newnham, 1991) and maternal BMI (Harvey et al., 2007, Wright et al., 2010a, Bammann et al., 2014, Fairley et al., 2015a) are known to influence birthweight in opposing directions. Additionally, the associated risk factors for LGA infants include pre-pregnancy obesity, excessive GWG, maternal or GDM (Jolly et al., 2003). This was addressed partly in the sensitivity analyses in ALSPAC, which confirmed that the key associations remained.

In NTFS, the questionnaire used to collect retrospective data is vulnerable to recall bias and inaccurate responses. To reduce this bias, information was only taken on present habits in the age 50 follow-up questionnaire. Additionally, data considered unreliable (such as those on energy intake) were not utilised.

A main limitation of this study was the small sample sizes in the North East cohorts at follow-up relative to larger birth cohort studies. There was insufficient power to detect small effect sizes. However, a key strength of the project overall is through integrating results from several different approaches (triangulation). Each approach may have specific sources of bias, but by comparing cohorts from different times and geographical locations strengthens confidence in the findings. It is also reassuring that the main findings for rapid growth in infancy are in agreement with findings from large-scale meta-analyses.

8.4.1 Generalisability and attrition

Loss to follow-up is an important issue as it can result in biased estimates. As with most longitudinal population studies there has been attrition related to SES. Attrition could lead to underestimations, particularly of inequality when SES is the exposure of interest (Howe et al., 2013). However, in GMS, as more affluent families were initially underrepresented, this has led to a sample that is more now more representative of the region. There were missing

data for some exposures, notably for RWG, however this supports the utility of a biomarker when early life data are not available.

As cohorts situated in the North East of England, GMS and NTFS may not be representative of the rest of the UK. However, key findings were replicated using ALSPAC data (South West England). Furthermore, in NTFS the inclusion of study members who had moved away increased representativeness of the original cohort (shown for all early life factors bar gender (Lamont et al., 2000)).

Generalisability of these findings to other populations is limited by the predominantly white ethnicity of study members, and risk factors for other ethnic minorities may differ. However, research from the more diverse Born in Bradford birth cohort did not find strong evidence that risk factors for childhood obesity varied by ethnicity (Fairley et al., 2015b), and an Asian cohort also found similar risk for RWG (Aris et al., 2017). Other factors that may limit generalisability are inter-individual variation, transgenerational effects and genetic factors.

Due to the long-term nature of the Thousand Families cohort, findings may not be entirely generalisable to present-day populations. The influence of time-period effects cannot be overlooked, as the early years of NTFS study members will have been very different (section 5.5), even to adults age 50 today. However, despite these differences and the lesser impact of environmental factors, some early life exposures and lifestyle factors predicted adult body composition.

Adiposity is a complex trait and may not adequately captured by a single measurement, therefore multiple outcomes were analysed. Although a proxy measure, BMI was the only measure consistent across all of the cohorts in childhood. While there is a strong correlation between BMIz and fat mass, this alone is not enough to make inferences about individuals. For example, exercise can lead to reductions in fat mass and increases in lean mass (Prentice and Jebb, 2001), however if weight remains the same, BMIz will also stay the same. It could be argued that estimates obtained when using BF%, a direct measure of adiposity that is independent of bone mass, would more reliably depict adiposity. A study in GMS found that FMI was a more sensitive measure than BMI (Basterfield et al., 2012a), and proxy measures of adiposity were inferior (Basterfield et al., 2012b). In NTFS and GMS there was a high correlation between BMI and BF%/FMI, suggesting that BMI is a good proxy measure of body fat in these populations.

Risk factors for BMI were also risk factors for measures of body fat and waist obesity, which may imply that additional measures have little added value. Generally, there were more risk factors associated with BMI, which may suggest some risk factors may be related to total (both fat and lean) mass. However, due to the different cardiovascular consequences associated with waist obesity, and the consistent associations for birthweight and waist obesity (in both NTFS men and GMS children), there is definite value in this measure, and findings were in agreement with the DOHaD hypothesis (Barker, 1995). Furthermore, only analysing BMI misses those with normal weight but waist OB, who are at risk of higher mortality (Cerhan et al., 2014, Sun et al., 2019)

The validity of the 0.5 threshold for WHtR to identify cardiometabolic risk in children has been questioned (Hara et al., 2002, Yan et al., 2007). Although values vary by ethnicity, generally estimates are around the 0.5 value for children from: South Africa (Matsha et al., 2013), Australia (Nambiar et al., 2010) and the UK (McCarthy and Ashwell, 2006), and therefore this probably was an appropriate cut-off in this study. Tybor et al. (2008) found that there was a residual correlation between WHtR and height in children during periods of growth; therefore, it is possible that WHtR does not completely control for the effect of height and the effect sizes for WHtR here are underestimations. On the other hand, adjusting indices to be independent of height does not appreciate that taller children are fatter than their shorter peers (Metcalf et al., 2011), and that there is cross-over in genetic variants related to weight and height (Elks et al., 2010a).

The UK90 reference was consistently used to normalise BMI in all the cohorts, as prevalence of obesity should be relative to reference data at a fixed time point (Prentice, 1998). Although NTFS children were born before the reference, it is the most appropriate growth reference for these analyses. An alternative is the UK-WHO growth charts (Cole et al., 2011), which have been found to classify more children as obese, and could lead to misinterpretation of risk (Johnson et al., 2012). When growth in ALSPAC and GMS were compared to the WHO standard, more were classified as overweight in the pre-school years, and as it uses a breastfed child as a model for growth, may not be appropriate for a UK population (Wright et al., 2008a). The GMS and ALSPAC cohorts have been shown to be similar to the UK90 reference, with very similar weights and heights at times of overlap in childhood (Wright et al., 2008a).

Findings from this cohort support that when investigating early life risk factors of childhood obesity, BMI or categorical OWOB are adequate measures, but studies might be enhanced by the inclusion of a body fatness outcome measure (e.g. FMI) when examining SES.

8.5 Conclusion

In summary, this study provides evidence for early life weight gain as an important risk factor for multiple measures of childhood body composition, consistent across all three cohorts (time periods and regions). Some of these effects may be a consequence of lower birthweight, which can be influenced by many factors. The associations for rapid growth did not persist into adulthood and may represent a more important factor for childhood BMI and on the pathway to adult BMI.

Since early life, lifestyle and socioeconomic factors have emerged as important factors in the development of obesity, multi-component interventions have emerged (Lanigan, 2018). The path models support that a multi-component intervention may be the best approach and demonstrate the complexity and interrelated nature of risk factors (Figure 3.3, Figure 4.2).

8.6 Summary of the epigenetic findings

This study used a discovery cohort (ALSPAC) to determine loci associated with early life exposures, and a validation cohort (NTFS) to verify methylation. Robust associations were found for RWG and childhood and adult methylation, but for no other exposures. As described in the literature review (section 1.5.3), so far, only a handful of differentially methylated positions have been identified for early life risk factors of obesity. This study used stringent criteria and adjustment methods to identify specific CpG sites meaning that perhaps only the most robust associations were identified.

There were missing data for exposures, and smaller sample sizes reduce the likelihood of finding a statistically significant result, especially when the changes in methylation exhibited are small. Large differences in methylation are often found in cancer, but generally, small effect sizes (2-10%) are common when examining environmental influences (Breton et al., 2017). Both the sample size and effect sizes were small for RWG, which could support that this CpG is robustly associated, although there is always the possibility that the finding was a false positive.

Replication by an independent technique and dataset were used to examine methylation in adults. Although the direction of the association was different between the cohorts, the association was still significant. The findings may suggest that methylation is dynamic, has specific cohort or age effects, or that disparities are due to sample or phenotype differences.

8.6.1 Persistence of methylation over time

It is interesting that RWG was positively associated with DNA methylation in childhood at the specific loci identified, but was negatively associated with methylation in adults[1]. A similar phenomenon was seen in a paper by (Richmond et al., 2015) examining maternal smoking and offspring methylation. This 'recovery' of methylation could represent the adaptability of methylation to environmental cues. However, as there were no longitudinal data available (in the same individuals) it is unsure whether DNAm was also negatively associated in NTFS adults.

Simpkin et al. (2015) noted that lower birthweights and shorter gestational age were associated with a phase of rapid 'catch-up' in methylation differences. As RWG includes some of the impact of catch up growth from lower birthweights, this could be the case in ALSPAC, where catch-up growth (RWG) is mirrored by a 'catch-up' in methylation. Whilst the opposite, 'catch-down methylation', could be reflective of slowed growth and the normalisation of BMI over time in NTFS adults. Taken together, one hypothesis could be that methylation at this locus represents growth more generally; increases in methylation in response to growth, but relative decreases later in order to return to 'normal' methylation.

In terms of changes in methylation over the life course, this study identified that in ALSPAC, those who had RWG (ALSPAC) have higher levels of methylation in childhood (age 7), but then methylation decreases slightly in adolescence (age 17). If considering age 17 as adults, similar overall trends were observed in NTFS adults (age 50), in that those who had RWG

[1] It is granted that beta (β) values from the Infinium, and percentage methylation from pyrosequencing may not be interchangeable, therefore when comparing the two datasets the descriptions were focused on the size (small/large) and direction (increases/decreases) of change in methylation, in relation to the exposure.

have lower methylation levels compared to those who did not have RWG (Figure 8.1). This could suggest that the hypermethylation in those with RWG at age 7 may be 'corrected' over time from childhood to adolescence to middle age. Again, one could speculate that this could reflect growth patterns, as the majority of childhood growth will have happened by age 17.

These specific epigenetic changes detected in childhood might have downstream effects throughout childhood and adolescence rather than persistence across the life course, however currently there are a limited number of studies examining the variation and stability of methylation changes over the life course (Richmond et al., 2014, Simpkin et al., 2015). As there were no longitudinal methylation measurements, the patterns of methylation cannot be discerned in NTFS. There is also the possibility that other lifestyle factors, or pleiotropy (where one gene affects multiple characteristics) could influence DNAm at this locus.

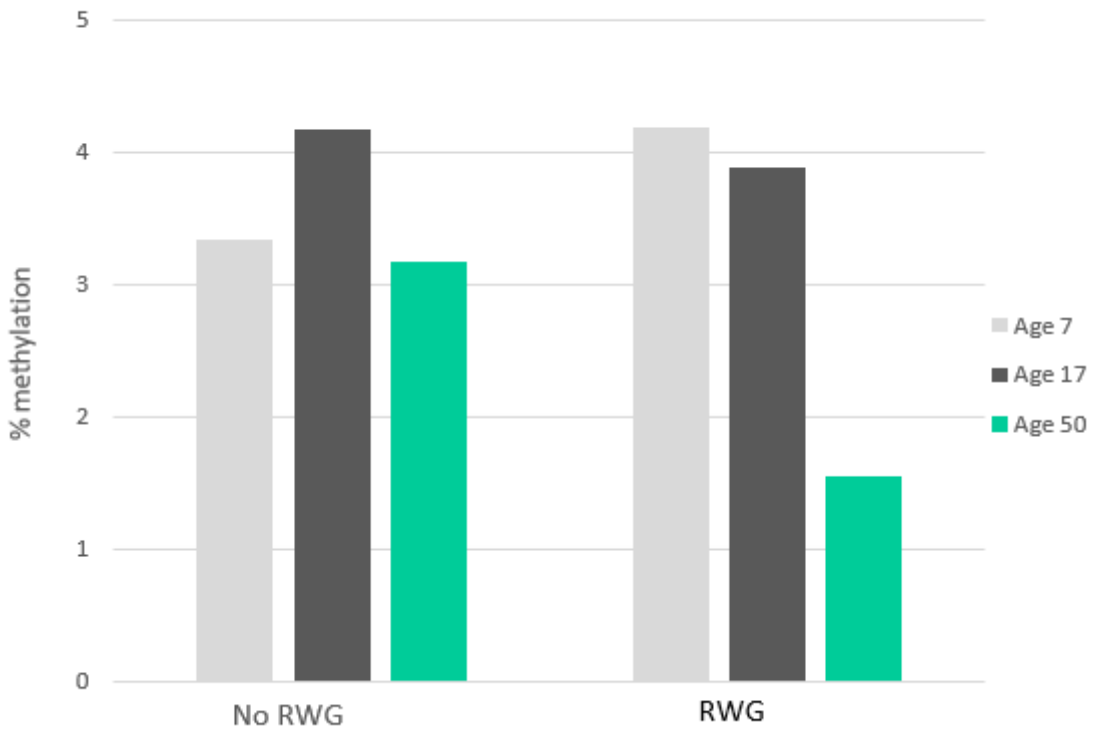


Figure 8.1 Summary of DNAm by RWG in ALSPAC AND NTFS.
Median DNAm (%) at ages 7 and 17 (ALSPAC) and age 50 (NTFS).

8.7 Strengths and Limitations of the epigenetic analysis

There are a number of strengths and limitations of the epigenetic analysis. These relate to sample representativeness, statistical power, cell heterogeneity, EWAS methodology, and specific limitations of the technologies and comparing methylation values across studies.

8.7.1 Sample representativeness

The ARIES sub-sample are more affluent than those not included (Alspac Study Team, 2001, Relton et al., 2015), and therefore there if mechanistic pathways in which SES impacts on DNAm exist this could introduce bias. However, the NTFS validation cohort was also biased in the same direction.

Some of the analyses will have been underpowered, it would be worth using other large cohorts with larger sample sizes, such as the PACE consortium, although despite the small sample size robust associations were still identified.

8.7.2 Cell types and tissue specificity

There is variability in methylation across cell types and the possibility for tissue specific methylation effects. Many studies have utilised blood DNAm in relation to obesity (van Dijk et al., 2015, Wahl et al., 2017), which is reflective of changes in other tissues (Dick et al., 2014). As biopsies from relevant tissues (such as adipose tissue) are more difficult to obtain, surrogate tissues (i.e. peripheral blood) are often used. However, for the determination of a biomarker, a surrogate, accessible tissue such as blood is ideal for diagnostic purposes and for use in a clinical setting.

Furthermore, blood may be a physiologically relevant tissue due to the close contact with all cells, tissues and organs, and therefore reflects the interactions of changes at the molecular level via internal (hormones) and external (such as nutrients or drug interactions) stimuli. This idea was proposed by Liew et al. and is referred to as the "Sentinel Principle", whereby blood cells respond to changes occurring in the micro- and macro-environments in the body's systems, therefore are acting as "sentinels" (Liew et al., 2006).

8.7.3 EWAS methodology

The EWAS serves as a starting point to identify potentially important candidate sites which can be the focus of further investigation, however this is essentially a 'needle in a haystack' approach. Methylation levels at time of measurement can only provide a snapshot, not how

methylation changes over the life course (without multiple measurements). Generally speaking, bar the limitations imposed by the technology (i.e. the CpG loci selected for the array), the EWAS approach is an unbiased method to identify novel CpG loci associated with exposures. An unbiased alternative is whole genome sequencing, however currently this technology is costly. Another approach could have been to use candidate genes associated with exposures from the literature, however there are no DNAm biomarkers associated with RWG and hence this would not have been feasible for all exposures. A limitation of the methodology employed here was using significance over effect size, i.e. using p values which do not consider the magnitude of the CpG-trait association. However, this is a trade off, and all associations with an FDR p value<0.1 were considered, and any less stringent criteria could increase the likelihood of false positive findings.

8.7.4 Differences in methods of measuring DNA methylation

The range of methylation values observed were vastly different when comparing values from pyrosequencing and the 450K array. Current methods to quantify DNAm are not always directly comparable, and the results presented here reflect fundamental difficulties with DNAm analyses. Even when comparing β values at the same CpG site and from the same platform (450K) from different studies (B-PROOF, BSGS), a wide range of DNAm values were observed (section 2.5.11), which may be due different normalisation methods. Others have observed that even when data are normalised using the same methods, there is disagreement in methylation values between the 450K and EPIC array platforms, and many (55%) sites exhibit low correlation due to low variability in methylation (Logue et al., 2017). They found that CpGs with a low range, had low correlation (r<0.2) across the arrays. Examining their data, cg11531579 was a locus with a low range, which could be because the ratio of the true variability relative to the measurement error is low.

The normalisation procedures utilised by these studies may in part explain some of the different methylation levels obtained in this study. Different normalisation procedures are a known source of heterogeneity in DNAm studies, particularly when changes observed are small or less significant (Wu et al., 2014a). The ALSPAC methylation levels were much lower than some of the NTFS samples. It is probable that if such high levels of methylation existed

in the 450K data, those points may be filtered out and removed as noise prior to statistical analysis (as per normalisation techniques), although they could be valid data points.

Differences could also arise from the technology utilised. Overall, there is good concordance between 450K and pyrosequencing data (Roessler et al., 2012). However, several validation studies show that β values are not equal to % methylation (e.g. 0.1 =10%) for the majority of loci examined, and although there is a trend of high and low between those data sets, differences between 450K assay values and sequencing values can vary wildly (Bibikova et al., 2011, Roessler et al., 2012). Roessler et al. (2012) similarly noted that in a cross validation study of the 450K array and whole-genome bisulfite sequencing (Bibikova et al., 2011), many measurements differed by 30-40%. As an unprocessed method that directly measures methylation, pyrosequencing may more closely reflect the true value. Furthermore, the use of bisulfite pyrosequencing for region-specific DNAm analysis is considered the gold standard.

Common SNPs could result in discrepancies between pyrosequencing and the 450k array, and the 'high methylation' samples could have underlying genetic differences. SNPs can account for 22-80% of the variability in DNAm between individuals (Birney et al., 2016). Further investigation of the region of interest using Sanger sequencing (section 7.4.9.2), indicated SNPs may be influencing methylation, perhaps in a phenotypic fashion (consistent patterns observed for RWG and high BMI). However, the effects of proximal or distal SNPs cannot be discerned using this small sample, but could be the focus of future work. There may still be purpose in biomarkers that can capturing both genetic and environmental influences on DNAm.

Additionally, for all methods, the influence of sampling, storage and processing on the methylation measures must be considered, as the variation due to these measures may well exceed the signal level.

8.7.5 The direction of the associations between DNAm and BMI

A recognised limitation is that many EWAS's are unable to distinguish causal or consequential epigenetic changes in the disease process (Baylin and Bestor, 2002, Bell et al., 2010). There is the potential for reverse causation, in that BMI alters methylation rather than vice versa (Wahl et al., 2017). Although childhood methylation was higher in ALSPAC participants who were OWOB in adolescence, these individuals could also have had a higher

BMI at age 7. When examining childhood methylation by phenotype and age (Figure), higher methylation levels were evident in those who had RWG, and then were OWOB in either childhood or adolescence, but also in those who were healthy weight in childhood and then OWOB in adolescence. This does not rule out reverse causality entirely, but suggests that regardless of childhood body size, if an individual had RWG and is then subsequently OWOB in adolescence their methylation was higher in childhood than someone of normal weight, supporting this locus as a predictive biomarker. The aim of the analysis was to determine biomarkers rather than causal DNAm changes. However, there are genotype data available for ALSPAC, and therefore causal inference methods such as Mendelian Randomization could be used to determine if changes were causal, if valid instrumental variables for RWG exist.

8.8 Remaining questions and future directions

These findings amongst others provide further evidence of the clinical importance of a fine balance between adequate weight gain to support neurodevelopment and ensuring infants with lower birthweights experience 'healthy' catch-up growth. Future developments could be to have closer monitoring of infant weight gain and personalised advice by health care professionals, to ensure weight gain is within accepted healthy limits for a given birthweight. In epidemiological studies, if the disease risk decreases following an intervention or removal of the exposure, this provides strong support for causal inference. There is limited capacity to assess causality in studies of early life exposures through randomised controlled trials, partly due to ethical issues of potentially causing harm. Instead, to further investigate causality, an intervention study or natural experiment could be used. When analysing the BMI trajectories in the two cohorts (chapter 4), it was the cohort exposed to an obesogenic environment in combination with RWG that had the most detrimental phenotype. Therefore, an intervention would need to alter the environment of those who had RWG (potentially through a diet or physical activity intervention, or perhaps in a rural population with reduced exposure to an obesogenic environment), or prevent the initial RWG.

There have been limited studies aimed at preventing RWG through interventions. Gungor et al. (2010) found that 1/3 of at-risk (those who had RWG) infants go on to become overweight children, whilst the 2/3 that do not: had parents that were more educated, had lower weight gain between 18-24 months, were exclusively breastfed for longer and had a later introduction of solid foods. Therefore interventions could focus on modifying these

factors. An intervention focusing on feeding alone, which utilised a theory-based behavioural intervention in formula milk-fed infants to promote responsive feeding, did not reduce the prevalence of RWG (Moorcroft et al., 2011). However, a multi-component intervention by Savage et al (2016) was more successful and therefore may be an effective strategy. The intervention was implemented via posted educational materials on infant feeding, sleep hygiene, active social play, emotional regulation and recording growth and involved numerous visits by research nurses between 3 and 48 weeks (Savage et al., 2016). Fewer than 6% of infants in the intervention group were overweight at age 1 year compared with 13% of control group infants, which is a relatively low prevalence (controls) compared to other studies (Zheng et al., 2018) therefore it would be interesting to see if these results extrapolate to populations with higher prevalence of RWG. If this critical period has passed and infancy weight is not monitored then there is utility for a biomarker of RWG to identify those at later risk.

An individual's current BMI alone will always be very predictive of future BMI, however DNAm could improve model prediction along with other known factors such as SNPs and environmental factors. For example, in older individuals from two European cohorts, methylation profiles explained 3.6 to 4.9% of the variation in BMI (Shah et al., 2015). Genome-wide association studies have found over 30 SNPs associated with BMI, which explain around 1.5% of the inter-individual variation in BMI when combined (Speliotes et al., 2010). The minor allele of FTO which is associated with obesity-related traits is associated with 1.2 fold risk of obesity and to 0·39 kg/m2 increases in BMI (Loos and Yeo, 2014). Therefore, as the effect sizes for individual CpG loci are much larger than effect sizes for individual SNPs, even from relatively small studies, there may be scope in future for larger studies to identify predictive changes in DNAm.

Our ability to identify methylation biomarkers will likely improve as the arrays increase in size and as more sophisticated bioinformatics methods are developed. The methylation differences identified were small, if biomarkers are to be useful then assays need to be able to detect these small differences accurately, which will likely become more commonplace with technological advancements. The investigation of methylation in this region was limited to a short region for pyrosequencing, however as whole genome sequencing becomes less expensive there will be future opportunities to examine larger regions.

These results suggest that the relationship between RWG and DNAm may change over the life course, or that it might be different in different cohorts or populations. Validation is required to determine if CpG loci are consistently differentially methylated in different populations, time points, ages and disease-states. The focus of future work could be to analyse DNAm at the locus of interest using pyrosequencing in a younger population. This was the original intent; however, the age 6-8 GMS DNA samples could not be traced. As there have been no other EWAS on RWG to date it was not possible to replicate the DNAm results. Additionally, if looking for causal epigenetic changes, gene expression and functionality will also need to be assessed.

8.9 Overall conclusions

Using two regional birth cohorts born 50 years apart, factors influencing obesity and how these have changed over time were investigated in the region with the highest rates of obesity in England (Health and Social Care Information Centre, 2014). The important finding from this study is that rapid infancy weight gain conditional on birthweight or not, was associated with increased childhood BMI. There is the potential for development of a CpG biomarker of RWG, which could be of use when early life data are not available in order to identify high-risk individuals for preventative intervention. However, whether this biomarker would be a useful predictor of future adiposity, or if methylation marks persist is unsure, and needs to be investigated in a younger population where a predictive biomarker of subsequent OWOB would be useful.

The life course path analysis approach allowed examination of the effects of exposures on prospective outcomes, and the results provide important insights into the multi-dimensional aetiology of obesity. Taken together, these results suggest that the combination of environmental insults at a critical developmental periods, combined with a multi-faceted, high-risk environment increases adiposity in childhood, which may track in adulthood. Understanding of the interactions and pathways between exposures and socioeconomic factors could be used to develop a risk-score for intervention purposes.

This work highlights the dynamic nature of methylation and contributes to the literature on life course changes in methylation in relation to early life exposures. This study provides a proof of concept and establishes a motive for further research to identify novel biomarkers of early life rapid weight gain in longitudinal studies.

References

ABARCA-GÓMEZ, L., ABDEEN, Z. A., HAMID, Z. A., ABU-RMEILEH, N. M., ACOSTA-CAZARES, B., ACUIN, C., ADAMS, R. J., AEKPLAKORN, W., AFSANA, K., AGUILAR-SALINAS, C. A., AGYEMANG, C., AHMADVAND, A., AHRENS, W., AJLOUNI, K., AKHTAEVA, N., AL-HAZZAA, H. M., AL-OTHMAN, A. R., AL-RADDADI, R., AL BUHAIRAN, F., AL DHUKAIR, S., ALI, M. M., ALI, O., ALKERWI, A. A., ALVAREZ-PEDREROL, M., ALY, E., AMARAPURKAR, D. N., AMOUYEL, P., AMUZU, A., ANDERSEN, L. B., ANDERSSEN, S. A., ANDRADE, D. S., ÄNGQUIST, L. H., ANJANA, R. M., AOUNALLAH-SKHIRI, H., ARAÚJO, J., ARIANSEN, I., ARIS, T., ARLAPPA, N., ARVEILER, D., ARYAL, K. K., ASPELUND, T., ASSAH, F. K., ASSUNÇÃO, M. C. F., AUNG, M. S., AVDICOVÁ, M., AZEVEDO, A., AZIZI, F., BABU, B. V., BAHIJRI, S., BAKER, J. L., BALAKRISHNA, N., BAMOSHMOOSH, M., BANACH, M., BANDOSZ, P., BANEGAS, J. R., BARBAGALLO, C. M., BARCELÓ, A., BARKAT, A., BARROS, A. J. D., BARROS, M. V. G., BATA, I., BATIEHA, A. M., BATISTA, R. L., BATYRBEK, A., BAUR, L. A., BEAGLEHOLE, R., ROMDHANE, H. B., BENEDICS, J., BENET, M., BENNETT, J. E., BERNABE-ORTIZ, A., BERNOTIENE, G., BETTIOL, H., BHAGYALAXMI, A., BHARADWAJ, S., BHARGAVA, S. K., BHATTI, Z., BHUTTA, Z. A., BI, H., BI, Y., BIEHL, A., BIKBOV, M., BISTA, B., BJELICA, D. J., BJERREGAARD, P., BJERTNESS, E., BJERTNESS, M. B., BJÖRKELUND, C., BLOKSTRA, A., BO, S., BOBAK, M., BODDY, L. M., BOEHM, B. O., BOEING, H., BOGGIA, J. G., BOISSONNET, C. P., BONACCIO, M., BONGARD, V., BOVET, P., BRAECKEVELT, L., et al. 2017. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *The Lancet*.

ACCOMANDO, W. P., WIENCKE, J. K., HOUSEMAN, E. A., NELSON, H. H. & KELSEY, K. T. 2014. Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome biology,* 15**,** R50.

ADAB, P., PALLAN, M. & WHINCUP, P. H. 2018. Is BMI the best measure of obesity? *BMJ,* 360**,** k1274.

ADALSTEINSSON, B. T., GUDNASON, H., ASPELUND, T., HARRIS, T. B., LAUNER, L. J., EIRIKSDOTTIR, G., SMITH, A. V. & GUDNASON, V. 2012. Heterogeneity in White Blood Cells Has Potential to Confound DNA Methylation Measurements. *PLOS ONE,* 7**,** e46705.

ADKINS, R. M., THOMAS, F., TYLAVSKY, F. A. & KRUSHKAL, J. 2011. Parental ages and levels of DNA methylation in the newborn are correlated. *BMC medical genetics,* 12**,** 47-47.

AGOSTONI, C., DECSI, T., FEWTRELL, M., GOULET, O., KOLACEK, S., KOLETZKO, B., MICHAELSEN, K. F., MORENO, L., PUNTIS, J. & RIGO, J. 2008. Complementary feeding: a commentary by the ESPGHAN Committee on Nutrition. *Journal of pediatric gastroenterology and nutrition,* 46**,** 99-110.

AJSLEV, T. A., ANDERSEN, C. S., GAMBORG, M., SORENSEN, T. I. A. & JESS, T. 2011. Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics. *International Journal of Obesity,* 35**,** 522-529.

ALAMIAN, A., WANG, L., HALL, A. M., PITTS, M. & IKEKWERE, J. 2016. Infant sleep problems and childhood overweight: Effects of three definitions of sleep problems. *Preventive Medicine Reports,* 4**,** 463-468.

ALEGRÍA-TORRES, J. A., BACCARELLI, A. & BOLLATI, V. 2011. Epigenetics and lifestyle. *Epigenomics,* 3**,** 267-277.

ALFANO, R., GUIDA, F., GALOBARDES, B., CHADEAU-HYAM, M., DELPIERRE, C., GHANTOUS, A., HENDERSON, J., HERCEG, Z., JAIN, P., NAWROT, T. S., RELTON, C., VINEIS, P., CASTAGNÉ, R. & PLUSQUIN, M. 2018. Socioeconomic position during pregnancy and DNA methylation signatures at three stages across early life: epigenome-wide association studies in the ALSPAC birth cohort. *International Journal of Epidemiology,* 48**,** 30-44.

ALIKHANI-KOOPAEI, R., FOULADKOU, F., FREY, F. J. & FREY, B. M. 2004. Epigenetic regulation of 11β-hydroxysteroid dehydrogenase type 2 expression. *The Journal of clinical investigation,* 114**,** 1146-1157.

ALSPAC STUDY TEAM 2001. ALSPAC–the avon longitudinal study of parents and children. *Paediatric and perinatal epidemiology,* 15**,** 74-87.

ANDA, R. F., FELITTI, V. J., BREMNER, J. D., WALKER, J. D., WHITFIELD, C., PERRY, B. D., DUBE, S. R. & GILES, W. H. 2006. The enduring effects of abuse and related adverse experiences in childhood. *European archives of psychiatry and clinical neuroscience,* 256**,** 174-186.

ANDERSON, O. S., SANT, K. E. & DOLINOY, D. C. 2012. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. *The Journal of Nutritional Biochemistry,* 23**,** 853-859.

ANDREAS, N. J., HYDE, M. J., GALE, C., PARKINSON, J. R., JEFFRIES, S., HOLMES, E. & MODI, N. 2014. Effect of maternal body mass index on hormones in breast milk: a systematic review. *PLoS One,* 9**,** e115043.

ANGRIST, J. D. & IMBENS, G. W. 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association,* 90**,** 431-442.

ANTONAKIS, J., BENDAHAN, S., JACQUART, P. & LALIVE, R. 2014. Causality and Endogeneity: Problems and Solutions. *The Oxford handbook of leadership and organizations***,** 93.

ARENZ, S., RÜCKERL, R., KOLETZKO, B. & VON KRIES, R. 2004. Breast-feeding and childhood obesity—a systematic review. *International journal of obesity,* 28**,** 1247-1256.

ARIS, I. M., CHEN, L.-W., TINT, M. T., PANG, W. W., SOH, S. E., SAW, S.-M., SHEK, L. P.-C., TAN, K.-H., GLUCKMAN, P. D., CHONG, Y.-S., YAP, F., GODFREY, K. M., KRAMER, M. S. & LEE, Y. S. 2017. Body mass index trajectories in the first two years and subsequent childhood cardio-metabolic outcomes: a prospective multi-ethnic Asian cohort study. *Scientific Reports,* 7**,** 8424.

ARMSTRONG, J. & REILLY, J. J. 2002. Breastfeeding and lowering the risk of childhood obesity. *Lancet,* 359.

ARMSTRONG, J., REILLY, J. J. & CHILD HLTH INFORMATION, T. 2002. Breastfeeding and lowering the risk of childhood obesity. *Lancet,* 359**,** 2003-2004.

ARMSTRONG, K. L., QUINN, R. A. & DADDS, M. R. 1994. The sleep patterns of normal children. *The Medical journal of Australia,* 161**,** 202-206.

ARYEE, M. J., JAFFE, A. E., CORRADA-BRAVO, H., LADD-ACOSTA, C., FEINBERG, A. P., HANSEN, K. D. & IRIZARRY, R. A. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics,* 30**,** 1363-1369.

ASHWELL, M., GUNN, P. & GIBSON, S. 2012. Waist-to-height ratio is a better screening tool than waist circumference and BMI for adult cardiometabolic risk factors: systematic review and meta-analysis. *Obes Rev,* 13**,** 275-86.

ASSI, N., FAGES, A., VINEIS, P., CHADEAU-HYAM, M., STEPIEN, M., DUARTE-SALLES, T., BYRNES, G., BOUMAZA, H., KNÜPPEL, S., KÜHN, T., PALLI, D., BAMIA, C., BOSHUIZEN, H., BONET, C., OVERVAD, K., JOHANSSON, M., TRAVIS, R., GUNTER, M. J., LUND, E.,

DOSSUS, L., ELENA-HERRMANN, B., RIBOLI, E., JENAB, M., VIALLON, V. & FERRARI, P. 2015. A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. *Mutagenesis,* 30**,** 743-753.

ATKIN, A. J., SHARP, S. J., HARRISON, F. L. O., BRAGE, S. & VAN SLUIJS, E. M. F. 2016. Seasonal Variation in Children's Physical Activity and Sedentary Time. *Medicine and Science in Sports and Exercise,* 48**,** 449-456.

ATKINSON, R., DHURANDHAR, N., ALLISON, D., BOWEN, R., ISRAEL, B., ALBU, J. & AUGUSTUS, A. 2005. Human adenovirus-36 is associated with increased body weight and paradoxical reduction of serum lipids. *International journal of obesity,* 29**,** 281.

ATKINSON, R. L. Viruses as an etiology of obesity. Mayo Clinic Proceedings, 2007. Elsevier, 1192-1198.

AUNE, D., SEN, A., PRASAD, M., NORAT, T., JANSZKY, I., TONSTAD, S., ROMUNDSTAD, P. & VATTEN, L. J. 2016. BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. *Bmj,* 353**,** i2156.

AUSTIN, M. K., CHEN, E., ROSS, K. M., MCEWEN, L. M., MACLSAAC, J. L., KOBOR, M. S. & MILLER, G. E. 2018. Early-life socioeconomic disadvantage, not current, predicts accelerated epigenetic aging of monocytes. *Psychoneuroendocrinology,* 97**,** 131-134.

AZAD, M. B., BRIDGMAN, S. L., BECKER, A. B. & KOZYRSKYJ, A. L. 2014. Infant antibiotic exposure and the development of childhood overweight and central adiposity. *International Journal of Obesity,* 38**,** 1290-1298.

BAILEY, L., FORREST, C. B., ZHANG, P., RICHARDS, T. M., LIVSHITS, A. & DERUSSO, P. A. 2014. Association of antibiotics in infancy with early childhood obesity. *JAMA Pediatrics,* 168**,** 1063-1069.

BAIRD, J., FISHER, D., LUCAS, P., KLEIJNEN, J., ROBERTS, H. & LAW, C. 2005. Being big or growing fast: systematic review of size and growth in infancy and later obesity. *Bmj,* 331**,** 929.

BAKER, J. L., MICHAELSEN, K. F., RASMUSSEN, K. M. & SORENSEN, T. I. 2004. Maternal prepregnant body mass index, duration of breastfeeding, and timing of complementary food introduction are associated with infant weight gain. *Am J Clin Nutr,* 80**,** 1579-88.

BALL, K. & CRAWFORD, D. 2005. Socioeconomic status and weight change in adults: a review. *Soc Sci Med,* 60**,** 1987-2010.

BAMMANN, K., PEPLIES, J., DE HENAUW, S., HUNSBERGER, M., MOLNAR, D., MORENO, L. A., TORNARITIS, M., VEIDEBAUM, T., AHRENS, W. & SIANI, A. 2014. Early life course risk factors for childhood obesity: the IDEFICS case-control study. *PLoS One,* 9**,** e86914.

BANN, D., JOHNSON, W., LI, L., KUH, D. & HARDY, R. 2018. Socioeconomic inequalities in childhood and adolescent body-mass index, weight, and height from 1953 to 2015: an analysis of four longitudinal, observational, British birth cohort studies. *The Lancet Public Health,* 3**,** e194-e203.

BANN, D., WILLS, A., COOPER, R., HARDY, R., AIHIE SAYER, A., ADAMS, J. & KUH, D. 2014. Birth weight and growth from infancy to late adolescence in relation to fat and lean mass in early old age: findings from the MRC National Survey of Health and Development. *Int J Obes (Lond),* 38**,** 69-75.

BARCLAY, K. & MYRSKYLÄ, M. 2016. Maternal age and offspring health and health behaviours in late adolescence in Sweden. *SSM - Population Health,* 2**,** 68-76.

BARKER, D., OSMOND, C., GOLDING, J., KUH, D. & WADSWORTH, M. E. 1989. Growth in utero, blood pressure in childhood and adult life, and mortality from cardiovascular disease. *Bmj,* 298**,** 564-567.

BARKER, D. J. 1995. Fetal origins of coronary heart disease. *Bmj,* 311**,** 171-4.

BARKER, D. J. P. 1994. *Mothers, babies, and disease in later life*, BMJ Publishing Group London.

BARROS, A. J. D., SANTOS, L. P., WEHRMEISTER, F., MOTTA, J. V. D. S., MATIJASEVICH, A., SANTOS, I. S., MENEZES, A. M. B., GONÇALVES, H., ASSUNÇÃO, M. C. F., HORTA, B. L. & BARROS, F. C. 2017. Caesarean section and adiposity at 6, 18 and 30 years of age: results from three Pelotas (Brazil) birth cohorts. *BMC Public Health,* 17**,** 256.

BARTSCH, E., MEDCALF, K. E., PARK, A. L., RAY, J. G. & HIGH RISK, P.-E. 2016. Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. *Bmj-British Medical Journal,* 353.

BASTARD, J.-P., MAACHI, M., LAGATHU, C., KIM, M. J., CARON, M., VIDAL, H., CAPEAU, J. & FEVE, B. 2006. Recent advances in the relationship between obesity, inflammation, and insulin resistance. *European cytokine network,* 17**,** 4-12.

BASTERFIELD, L., ADAMSON, A. J., PARKINSON, K. N., MAUTE, U., LI, P. X. & REILLY, J. J. 2008. Surveillance of physical activity in the UK is flawed: validation of the Health Survey for England Physical Activity Questionnaire. *Archives of Disease in Childhood,* 93**,** 1054.

BASTERFIELD, L., ADAMSON, A. J., PEARCE, M. S. & REILLY, J. J. 2011. Stability of habitual physical activity and sedentary behavior monitoring by accelerometry in 6-to 8-year-olds. *Journal of Physical Activity and Health,* 8**,** 543-547.

BASTERFIELD, L., PEARCE, M. S., ADAMSON, A. J., FRARY, J. K., PARKINSON, K. N., WRIGHT, C. M. & REILLY, J. J. 2012a. Physical Activity, Sedentary Behavior, and Adiposity in English Children. *American Journal of Preventive Medicine,* 42**,** 445-451.

BASTERFIELD, L., PEARCE, M. S., ADAMSON, A. J., REILLY, J. K., PARKINSON, K. N. & REILLY, J. J. 2012b. Effect of choice of outcome measure on studies of the etiology of obesity in children. *Annals of Epidemiology,* 22**,** 888-891.

BAYER, O., KRÜGER, H., VON KRIES, R. & TOSCHKE, A. M. 2011a. Factors Associated With Tracking of BMI: A Meta-Regression Analysis on BMI Tracking*. *Obesity,* 19**,** 1069-1076.

BAYER, O., KRÜGER, H., VON KRIES, R. & TOSCHKE, A. M. 2011b. Factors associated with tracking of BMI: a meta-regression analysis on BMI tracking. *Obesity,* 19**,** 1069-1076.

BAYLIN, S. & BESTOR, T. H. 2002. Altered methylation patterns in cancer cell genomes: cause or consequence? *Cancer cell,* 1**,** 299-305.

BAYOL, S. A., FARRINGTON, S. J. & STICKLAND, N. C. 2007. A maternal 'junk food' diet in pregnancy and lactation promotes an exacerbated taste for 'junk food' and a greater propensity for obesity in rat offspring. *British Journal of Nutrition,* 98**,** 843-851.

BEDOGNI, G., IUGHETTI, L., FERRARI, M., MALAVOLTI, M., POLI, M., BERNASCONI, S. & BATTISTINI, N. 2003. Sensitivity and specificity of body mass index and skinfold thicknesses in detecting excess adiposity in children aged 8-12 years. *Annals of human biology,* 30**,** 132-139.

BELFORT, M. B., RIFAS-SHIMAN, S. L., SULLIVAN, T., COLLINS, C. T., MCPHEE, A. J., RYAN, P., KLEINMAN, K. P., GILLMAN, M. W., GIBSON, R. A. & MAKRIDES, M. 2011. Infant growth before and after term: effects on neurodevelopment in preterm infants. *Pediatrics,* 128**,** e899-906.

BELL, C. G., TESCHENDORFF, A. E., RAKYAN, V. K., MAXWELL, A. P., BECK, S. & SAVAGE, D. A. 2010. Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC medical genomics,* 3**,** 33.

BELL, J. F. & ZIMMERMAN, F. J. 2010. SHortened nighttime sleep duration in early life and subsequent childhood obesity. *Archives of Pediatrics & Adolescent Medicine,* 164**,** 840-845.

BELLIS, M. A., LOWEY, H., LECKENBY, N., HUGHES, K. & HARRISON, D. 2013. Adverse childhood experiences: retrospective study to determine their impact on adult health behaviours and health outcomes in a UK population. *Journal of Public Health,* 36**,** 81-91.

BEN-SHLOMO, Y., COOPER, R. & KUH, D. 2016. The last two decades of life course epidemiology, and its relevance for research on ageing. *International Journal of Epidemiology,* 45**,** 973-988.

BEN-SHLOMO, Y. & KUH, D. 2002. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International journal of epidemiology,* 31**,** 285-293.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological),* 57**,** 289-300.

BEYERLEIN, A. & VON KRIES, R. 2011. Breastfeeding and body composition in children: will there ever be conclusive empirical evidence for a protective effect against overweight? *The American journal of clinical nutrition,* 94**,** 1772S-1775S.

BIBIKOVA, M., BARNES, B., TSAN, C., HO, V., KLOTZLE, B., LE, J. M., DELANO, D., ZHANG, L., SCHROTH, G. P., GUNDERSON, K. L., FAN, J.-B. & SHEN, R. 2011. High density DNA methylation array with single CpG site resolution. *Genomics,* 98**,** 288-295.

BIRNEY, E., SMITH, G. D. & GREALLY, J. M. 2016. Epigenome-wide Association Studies and the Interpretation of Disease -Omics. *PLoS genetics,* 12**,** e1006105-e1006105.

BISPHAM, J., GOPALAKRISHNAN, G. S., DANDREA, J., WILSON, V., BUDGE, H., KEISLER, D. H., BROUGHTON PIPKIN, F., STEPHENSON, T. & SYMONDS, M. E. 2003. Maternal endocrine adaptation throughout pregnancy to nutritional manipulation: consequences for maternal plasma leptin and cortisol and the programming of fetal adipose tissue development. *Endocrinology,* 144**,** 3575-85.

BOHN, B., MÜLLER, M. J., SIMIC-SCHLEICHER, G., KIESS, W., SIEGFRIED, W., OELERT, M., TUSCHY, S., BERGHEM, S., HOLL, R. W., FOR THE, A. P. V. I., THE GERMAN, B. C. N. O., A GERMAN/AUSTRIAN/SWISS MULTICENTER APV ANALYSIS OF, C. & ADOLESCENTS 2015. BMI or BIA: Is Body Mass Index or Body Fat Mass a Better Predictor of Cardiovascular Risk in Overweight or Obese Children and Adolescents? *Obesity Facts,* 8**,** 156-165.

BONEY, C. M., VERMA, A., TUCKER, R. & VOHR, B. R. 2005. Metabolic Syndrome in Childhood: Association With Birth Weight, Maternal Obesity, and Gestational Diabetes Mellitus. *Pediatrics,* 115**,** e290.

BORGHOL, N., SUDERMAN, M., MCARDLE, W., RACINE, A., HALLETT, M., PEMBREY, M., HERTZMAN, C., POWER, C. & SZYF, M. 2012. Associations with early-life socio-economic position in adult DNA methylation. *International Journal of Epidemiology,* 41**,** 62-74.

BOUWLAND-BOTH, M. I., VAN MIL, N. H., STOLK, L., EILERS, P. H., VERBIEST, M. M., HEIJMANS, B. T., TIEMEIER, H., HOFMAN, A., STEEGERS, E. A. & JADDOE, V. W. 2013. DNA methylation of IGF2DMR and H19 is associated with fetal and infant growth: the generation R study. *PLoS One,* 8**,** e81731.

BOYD, A., GOLDING, J., MACLEOD, J., LAWLOR, D. A., FRASER, A., HENDERSON, J., MOLLOY, L., NESS, A., RING, S. & DAVEY SMITH, G. 2013. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol,* 42**,** 111-27.

BRETON, C. V., MARSIT, C. J., FAUSTMAN, E., NADEAU, K., GOODRICH, J. M., DOLINOY, D. C., HERBSTMAN, J., HOLLAND, N., LASALLE, J. M. & SCHMIDT, R. 2017. Small-magnitude effect sizes in epigenetic end points are important in children's environmental health studies: the children's environmental health and disease prevention research center's epigenetics working group. *Environmental health perspectives,* 125**,** 511-526.

BRION, M. J., LAWLOR, D. A., MATIJASEVICH, A., HORTA, B., ANSELMI, L., ARAUJO, C. L., MENEZES, A. M., VICTORA, C. G. & SMITH, G. D. 2011. What are the causal effects of breastfeeding on IQ, obesity and blood pressure? Evidence from comparing high-income with middle-income cohorts. *Int J Epidemiol,* 40**,** 670-80.

BRISBOIS, T. D., FARMER, A. P. & MCCARGAR, L. J. 2012. Early markers of adult obesity: a review. *obesity reviews,* 13**,** 347-367.

BROPHY, S., COOKSEY, R., GRAVENOR, M., MISTRY, R., THOMAS, N., LYONS, R. & WILLIAMS, R. 2009. Risk factors for childhood obesity at age 5: Analysis of the Millennium Cohort Study. *BMC Public Health,* 9**,** 467.

BURDGE, G. C., SLATER-JEFFERIES, J., TORRENS, C., PHILLIPS, E. S., HANSON, M. A. & LILLYCROP, K. A. 2007. Dietary protein restriction of pregnant rats in the F0 generation induces altered methylation of hepatic gene promoters in the adult male offspring in the F1 and F2 generations. *British Journal of Nutrition,* 97**,** 435-439.

BUSCOT, M.-J., THOMSON, R. J., JUONALA, M., SABIN, M. A., BURGNER, D. P., LEHTIMÄKI, T., HUTRI-KÄHÖNEN, N., VIIKARI, J. S., JOKINEN, E. & TOSSAVAINEN, P. 2018. BMI trajectories associated with resolution of elevated youth BMI and incident adult obesity. *Pediatrics,* 141**,** e20172003.

BUTLAND, B., JEBB, S., KOPELMAN, P., MCPHERSON, K., THOMAS, S., MARDELL, J. & PARRY, V. 2007. Foresight. Tackling obesities: future choices. Project report. *Foresight. Tackling obesities: future choices. Project report*.

BUYKEN, A. E., KARAOLIS-DANCKERT, N., REMER, T., BOLZENIUS, K., LANDSBERG, B. & KROKE, A. 2008. Effects of breastfeeding on trajectories of body fat and BMI throughout childhood. *Obesity (Silver Spring),* 16**,** 389-95.

CAMERON, A. C. & TRIVEDI, P. K. 2010. Microeconometrics using Stata (revised ed.). *Number musr in Stata Press books. StataCorp LP*.

CANOY, D., WAREHAM, N., LUBEN, R., WELCH, A., BINGHAM, S., DAY, N. & KHAW, K. T. 2005. Cigarette Smoking and Fat Distribution in 21, 828 British Men and Women: A Population-based Study. *Obesity research,* 13**,** 1466-1475.

CAROLAN-OLAH, M., DUARTE-GARDEA, M. & LECHUGA, J. 2015. A critical review: early life nutrition and prenatal programming for adult disease. *Journal of Clinical Nursing,* 24**,** 3716-3729.

CARR, I. M., CAMM, N., TAYLOR, G. R., CHARLTON, R., ELLARD, S., SHERIDAN, E. G., MARKHAM, A. F. & BONTHRON, D. T. 2011. GeneScreen: a program for high-throughput mutation detection in DNA sequence electropherograms. *J Med Genet,* 48**,** 123-30.

CARSLAKE, D., TYNELIUS, P., VAN DEN BERG, G., DAVEY SMITH, G. & RASMUSSEN, F. 2017. Associations of parental age with health and social factors in adult offspring. Methodological pitfalls and possibilities. *Scientific Reports,* 7**,** 45278.

CAVACO, S., ERIKSSON, T. & SKALLI, A. 2014. Life cycle development of obesity and its determinants in six European countries. *Economics & Human Biology,* 14**,** 62-78.

CERHAN, J. R., MOORE, S. C., JACOBS, E. J., KITAHARA, C. M., ROSENBERG, P. S., ADAMI, H.-O., EBBERT, J. O., ENGLISH, D. R., GAPSTUR, S. M. & GILES, G. G. A pooled analysis of waist circumference and mortality in 650,000 adults. Mayo Clinic Proceedings, 2014. Elsevier, 335-345.

CHADEAU-HYAM, M., ATHERSUCH, T. J., KEUN, H. C., DE IORIO, M., EBBELS, T. M., JENAB, M., SACERDOTE, C., BRUCE, S. J., HOLMES, E. & VINEIS, P. 2011. Meeting-in-the-middle using metabolic profiling–a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers,* 16**,** 83-88.

CHANDLER-LANEY, P. C., GOWER, B. A. & FIELDS, D. A. 2013. Gestational and early life influences on infant body composition at 1 year. *Obesity,* 21**,** 144-148.

CHEN, X., BEYDOUN, M. A. & WANG, Y. 2008. Is Sleep Duration Associated With Childhood Obesity? A Systematic Review and Meta-analysis. *Obesity,* 16**,** 265-274.

CHEN, Y.-A., LEMIRE, M., CHOUFANI, S., BUTCHER, D. T., GRAFODATSKAYA, D., ZANKE, B. W., GALLINGER, S., HUDSON, T. J. & WEKSBERG, R. 2013. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics,* 8**,** 203-209.

CHIOLERO, A., FAEH, D., PACCAUD, F. & CORNUZ, J. 2008. Consequences of smoking for body weight, body fat distribution, and insulin resistance. *The American Journal of Clinical Nutrition,* 87**,** 801-809.

CHOI, S. H., HEO, K., BYUN, H.-M., AN, W., LU, W. & YANG, A. S. 2011. Identification of preferential target sites for human DNA methyltransferases. *Nucleic acids research,* 39**,** 104-118.

CHULA DE CASTRO, J. A., LIMA, T. R. & SILVA, D. A. S. 2018. Body composition estimation in children and adolescents by bioelectrical impedance analysis: A systematic review. *J Bodyw Mov Ther,* 22**,** 134-146.

CHUMLEA, W. C., GUO, S. S., KUCZMARSKI, R. J., FLEGAL, K. M., JOHNSON, C. L., HEYMSFIELD, S. B., LUKASKI, H. C., FRIEDL, K. & HUBBARD, V. S. 2002. Body composition estimates from NHANES III bioelectrical impedance data. *International Journal of Obesity,* 26**,** 1596-1609.

CIARK, S., J , HARRISON, J., PAUL, C. L. & FROMMER, M. 1994. High sensitivity mapping of methylated cytosines. *Nucleic acids research,* 22**,** 2990-2997.

CLARK, S. J., STATHAM, A., STIRZAKER, C., MOLLOY, P. L. & FROMMER, M. 2006. DNA methylation: bisulphite modification and analysis. *Nature protocols,* 1**,** 2353.

CLEARY-GOLDMAN, J., MALONE, F. D., VIDAVER, J., BALL, R. H., NYBERG, D. A., COMSTOCK, C. H., SAADE, G. R., EDDLEMAN, K. A., KLUGMAN, S. & DUGOFF, L. 2005. Impact of maternal age on obstetric outcome. *Obstetrics & Gynecology,* 105**,** 983-990.

CNATTINGIUS, S. 2004. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine & Tobacco Research,* 6**,** S125-S140.

CNATTINGIUS, S., VILLAMOR, E., LAGERROS, Y. T., WIKSTRÖM, A. K. & GRANATH, F. 2011. High birth weight and obesity—a vicious circle across generations. *International Journal Of Obesity,* 36**,** 1320.

COCOROS, N. M., LASH, T. L., NORGAARD, M., FARKAS, D. K., DEMARIA, A., JR. & SORENSEN, H. T. 2013. Hospitalized prenatal and childhood infections and obesity in Danish male conscripts. *Ann Epidemiol,* 23**,** 307-13.

COLE, T. 2004. Children grow and horses race: is the adiposity rebound a critical period for later obesity? *BMC pediatrics,* 4**,** 6.

COLE, T. J. 1990. The LMS method for constructing normalized growth standards. *Eur J Clin Nutr,* 44**,** 45-60.

COLE, T. J., FREEMAN, J. V. & PREECE, M. A. 1995. Body mass index reference curves for the UK, 1990. *Archives of Disease in Childhood,* 73**,** 25.

COLE, T. J., WILLIAMS, A. F. & WRIGHT, C. M. 2011. Revised birth centiles for weight, length and head circumference in the UK-WHO growth charts. *Ann Hum Biol,* 38.

COOK, D. 2015. CCMATCH: Stata module to match cases and controls using specified variables.

COUSINEAU, D. & CHARTIER, S. 2010. Outliers detection and treatment: a review. *International Journal of Psychological Research,* 3**,** 58-67.

COX, L. M. & BLASER, M. J. 2014. Antibiotics in early life and obesity. *Nature Reviews Endocrinology,* 11**,** 182.

CRAIG, R., MINDELL, J. & BOODHNA, G. 2014. *Health Survey for England, 2013*.

CRAIGIE, A. M., LAKE, A. A., KELLY, S. A., ADAMSON, A. J. & MATHERS, J. C. 2011. Tracking of obesity-related behaviours from childhood to adulthood: A systematic review. *Maturitas,* 70**,** 266-284.

CROWTHER, C. A., HILLER, J. E., MOSS, J. R., MCPHEE, A. J., JEFFRIES, W. S. & ROBINSON, J. S. 2005. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *New England Journal of Medicine,* 352**,** 2477-2486.

CURHAN, G. C., CHERTOW, G. M., WILLETT, W. C., SPIEGELMAN, D., COLDITZ, G. A., MANSON, J. E., SPEIZER, F. E. & STAMPFER, M. J. 1996. Birth weight and adult hypertension and obesity in women. *Circulation,* 94**,** 1310-5.

CUTLER, D. M., LLERAS-MUNEY, A. & VOGL, T. 2008. Socioeconomic status and health: dimensions and mechanisms. National Bureau of Economic Research.

CYPESS, A. M., LEHMAN, S., WILLIAMS, G., TAL, I., RODMAN, D., GOLDFINE, A. B., KUO, F. C., PALMER, E. L., TSENG, Y.-H. & DORIA, A. 2009. Identification and importance of brown adipose tissue in adult humans. *New England Journal of Medicine,* 360**,** 1509-1517.

DABELEA, D. & CRUME, T. 2011. Maternal Environment and the Transgenerational Cycle of Obesity and Diabetes. *Diabetes,* 60**,** 1849-1855.

DABELEA, D., HANSON, R. L., LINDSAY, R. S., PETTITT, D. J., IMPERATORE, G., GABIR, M. M., ROUMAIN, J., BENNETT, P. H. & KNOWLER, W. C. 2000. Intrauterine exposure to diabetes conveys risks for type 2 diabetes and obesity: a study of discordant sibships. *Diabetes,* 49**,** 2208-2211.

DALTON, M., CAMERON, A. J., ZIMMET, P. Z., SHAW, J. E., JOLLEY, D., DUNSTAN, D. W. & WELBORN, T. A. 2003. Waist circumference, waist-hip ratio and body mass index and their correlation with cardiovascular disease risk factors in Australian adults. *J Intern Med,* 254**,** 555-63.

DANCAUSE, K. N., LAPLANTE, D. P., FRASER, S., BRUNET, A., CIAMPI, A., SCHMITZ, N. & KING, S. 2012. Prenatal exposure to a natural disaster increases risk for obesity in 5 [frac12]-year-old children. *Pediatric research,* 71**,** 126-131.

DANCAUSE, K. N., LAPLANTE, D. P., HART, K. J., X, HARA, M. W., ELGBEILI, G., BRUNET, A. & KING, S. 2015. Prenatal Stress due to a Natural Disaster Predicts Adiposity in Childhood: The Iowa Flood Study. *Journal of Obesity,* 2015**,** 10.

DANESE, A. & J LEWIS, S. 2017. Psychoneuroimmunology of Early-Life Stress: The Hidden Wounds of Childhood Trauma? *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology,* 42**,** 99-114.

DARMASSEELANE, K., HYDE, M. J., SANTHAKUMARAN, S., GALE, C. & MODI, N. 2014. Mode of Delivery and Offspring Body Mass Index, Overweight and Obesity in Adult Life: A Systematic Review and Meta-Analysis. *PLOS ONE,* 9**,** e87896.

DAVEY-SMITH, G. 2008. Assessing Intrauterine Influences on Offspring Health Outcomes: Can Epidemiological Studies Yield Robust Findings? *Basic & Clinical Pharmacology & Toxicology,* 102**,** 245-256.

DAVIS, L., BARNES, A. J., GROSS, A. C., RYDER, J. R. & SHLAFER, R. J. 2019. Adverse Childhood Experiences and Weight Status among Adolescents. *The Journal of Pediatrics,* 204**,** 71-76.e1.

DE VRIES, S. I., BAKKER, I., HOPMAN-ROCK, M., HIRASING, R. A. & VAN MECHELEN, W. 2006. Clinimetric review of motion sensors in children and adolescents. *J Clin Epidemiol,* 59**,** 670-80.

DEDEURWAERDER, S., DEFRANCE, M., BIZET, M., CALONNE, E., BONTEMPI, G. & FUKS, F. 2014. A comprehensive overview of Infinium HumanMethylation450 data processing. *Briefings in bioinformatics,* 15**,** 929-941.

DEDEURWAERDER, S., DEFRANCE, M., CALONNE, E., DENIS, H., SOTIRIOU, C. & FUKS, F. 2011. Evaluation of the Infinium Methylation 450K technology. *Epigenomics,* 3**,** 771-784.

DEHGHAN, M. & MERCHANT, A. T. 2008. Is bioelectrical impedance accurate for use in large epidemiological studies? *Nutrition Journal,* 7**,** 26-26.

DEMETRIOU, C. A., VAN VELDHOVEN, K., RELTON, C., STRINGHINI, S., KYRIACOU, K. & VINEIS, P. 2015. Biological embedding of early-life exposures and disease risk in humans: a role for DNA methylation. *European journal of clinical investigation,* 45**,** 303-332.

DEMILLE, D. & GROSE, J. H. 2013. PAS kinase: a nutrient sensing regulator of glucose homeostasis. *IUBMB life,* 65**,** 921-929.

DERKS, S., CLEVEN, A. H., MELOTTE, V., SMITS, K. M., BRANDES, J. C., AZAD, N., VAN CRIEKINGE, W., DE BRUÏNE, A. P., HERMAN, J. G. & VAN ENGELAND, M. 2014. Emerging evidence for CHFR as a cancer biomarker: from tumor biology to precision medicine. *Cancer and Metastasis Reviews,* 33**,** 161-171.

DESAI, M., CROWTHER, N. J., LUCAS, A. & HALES, C. N. 1996. Organ-selective growth in the offspring of protein-restricted mothers. *Br J Nutr,* 76**,** 591-603.

DESAI, M., GAYLE, D., BABU, J. & ROSS, M. G. 2005. Programmed obesity in intrauterine growth-restricted newborns: modulation by newborn nutrition. *Am J Physiol Regul Integr Comp Physiol,* 288**,** R91-6.

DESAI, M. & ROSS, M. G. 2011. Fetal Programming of Adipose Tissue: Effects of IUGR and Maternal Obesity/High Fat Diet. *Seminars in reproductive medicine,* 29**,** 237-245.

DEWEY, K. G. 2003. Is breastfeeding protective against child obesity? *Journal of human lactation : official journal of International Lactation Consultant Association,* 19**,** 9-18.

DHURANDHAR, N. V. 2011. A framework for identification of infections that contribute to human obesity. *The Lancet Infectious Diseases,* 11**,** 963-969.

DHURANDHAR, N. V., KULKARNI, P. R., AJINKYA, S. M., SHERIKAR, A. A. & ATKINSON, R. L. 1997. Association of adenovirus infection with human obesity. *Obesity research,* 5**,** 464-469.

DICK, K. J., NELSON, C. P., TSAPROUNI, L., SANDLING, J. K., AISSI, D., WAHL, S., MEDURI, E., MORANGE, P. E., GAGNON, F., GRALLERT, H., WALDENBERGER, M., PETERS, A., ERDMANN, J., HENGSTENBERG, C., CAMBIEN, F., GOODALL, A. H., OUWEHAND, W. H., SCHUNKERT, H., THOMPSON, J. R., SPECTOR, T. D., GIEGER, C., TREGOUET, D. A., DELOUKAS, P. & SAMANI, N. J. 2014. DNA methylation and body-mass index: a genome-wide analysis. *Lancet,* 383**,** 1990-8.

DIETHELM, K., BOLZENIUS, K., CHENG, G., REMER, T. & BUYKEN, A. E. 2011. Longitudinal associations between reported sleep duration in early childhood and the development of body mass index, fat mass index and fat free mass index until age 7. *Int J Pediatr Obes,* 6**,** e114-23.

DIETZ, P. M., HOMA, D., ENGLAND, L. J., BURLEY, K., TONG, V. T., DUBE, S. R. & BERNERT, J. T. 2010. Estimates of nondisclosure of cigarette smoking among pregnant and nonpregnant women of reproductive age in the United States. *American journal of epidemiology,* 173**,** 355-359.

DIETZ, W. H. 1994. Critical periods in childhood for the development of obesity. *The American Journal of Clinical Nutrition,* 59**,** 955-9.

DINNO, A. 2015. dunn. test: Dunn's test of multiple comparisons using rank sums. *R package version,* 1.

DINSDALE, H., RIDLER, C. & ELLS, L. 2011. A simple guide to classifying body mass index in children. *National Obesity Observatory: Oxford*.

DODD, J. M., TURNBULL, D., MCPHEE, A. J., DEUSSEN, A. R., GRIVELL, R. M., YELLAND, L. N., CROWTHER, C. A., WITTERT, G., OWENS, J. A. & ROBINSON, J. S. 2014. Antenatal lifestyle advice for women who are overweight or obese: LIMIT randomised trial. *BMJ : British Medical Journal,* 348**,** g1285.

DRAKE, A. J., WALKER, B. R. & SECKL, J. R. 2005. Intergenerational consequences of fetal programming by in utero exposure to glucocorticoids in rats. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology,* 288**,** R34-R38.

DRUET, C., STETTLER, N., SHARP, S., SIMMONS, R. K., COOPER, C., DAVEY SMITH, G., EKELUND, U., LÉVY-MARCHAL, C., JARVELIN, M.-R., KUH, D. & ONG, K. K. 2012. Prediction of childhood obesity by infancy weight gain: an individual-level meta-analysis. *Paediatric and Perinatal Epidemiology,* 26**,** 19-26.

DU, P., ZHANG, X., HUANG, C.-C., JAFARI, N., KIBBE, W. A., HOU, L. & LIN, S. M. 2010. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics,* 11**,** 587.

DUNN, E. C., SOARE, T. W., ZHU, Y., SIMPKIN, A. J., SUDERMAN, M. J., KLENGEL, T., SMITH, A. D. A. C., RESSLER, K. J. & RELTON, C. L. 2019. Sensitive Periods for the Effect of Childhood Adversity on DNA Methylation: Results From a Prospective, Longitudinal Study. *Biological Psychiatry,* 85**,** 838-849.

DUNN, O. J. 1964. Multiple comparisons using rank sums. *Technometrics,* 6**,** 241-252.

EGELAND, G. M., IGLAND, J., VOLLSET, S. E., SULO, G., EIDE, G. E. & TELL, G. S. 2016. High population attributable fractions of myocardial infarction associated with waist–hip ratio. *Obesity,* 24**,** 1162-1169.

EISEN, S. A., LYONS, M. J., GOLDBERG, J. & TRUE, W. R. 1993. The impact of cigarette and alcohol consumption on weight and obesity: an analysis of 1911 monozygotic male twin pairs. *Archives of internal medicine,* 153**,** 2457-2463.

EKELUND, U., ONG, K., LINNÉ, Y., NEOVIUS, M., BRAGE, S., DUNGER, D. B., WAREHAM, N. J. & RÖSSNER, S. 2006. Upward weight percentile crossing in infancy and early childhood independently predicts fat mass in young adults: the Stockholm Weight Development Study (SWEDES). *The American Journal of Clinical Nutrition,* 83**,** 324-330.

ELKS, C. E., LOOS, R. J., SHARP, S. J., LANGENBERG, C., RING, S. M., TIMPSON, N. J., NESS, A. R., DAVEY SMITH, G., DUNGER, D. B., WAREHAM, N. J. & ONG, K. K. 2010a. Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. *PLoS Med,* 7**,** e1000284.

ELKS, C. E., LOOS, R. J. F., SHARP, S. J., LANGENBERG, C., RING, S. M., TIMPSON, N. J., NESS, A. R., DAVEY SMITH, G., DUNGER, D. B., WAREHAM, N. J. & ONG, K. K. 2010b. Genetic Markers of Adult Obesity Risk Are Associated with Greater Early Infancy Weight Gain and Growth. *PLOS Medicine,* 7**,** e1000284.

ENTRINGER, S., BUSS, C., SWANSON, J. M., COOPER, D. M., WING, D. A., WAFFARN, F. & WADHWA, P. D. 2012. Fetal programming of body composition, obesity, and

metabolic function: the role of intrauterine stress and stress biology. *Journal of nutrition and metabolism,* 2012.

ENTRINGER, S., BUSS, C. & WADHWA, P. D. 2010. Prenatal stress and developmental programming of human health and disease risk: concepts and integration of empirical findings. *Current Opinion in Endocrinology Diabetes and Obesity,* 17**,** 507-516.

EOM, G. H., KIM, K., KIM, S.-M., KEE, H. J., KIM, J.-Y., JIN, H. M., KIM, J.-R., KIM, J. H., CHOE, N. & KIM, K.-B. 2009. Histone methyltransferase PRDM8 regulates mouse testis steroidogenesis. *Biochemical and biophysical research communications,* 388**,** 131-136.

ESPOSITO, E. A., JONES, M. J., DOOM, J. R., MACISAAC, J. L., GUNNAR, M. R. & KOBOR, M. S. 2016. Differential DNA methylation in peripheral blood mononuclear cells in adolescents exposed to significant early but not later childhood adversity. *Development and Psychopathology,* 28**,** 1385-1399.

ESPOSITO, S., PRETI, V., CONSOLO, S., NAZZARI, E. & PRINCIPI, N. 2012. Adenovirus 36 infection and obesity. *Journal of Clinical Virology,* 55**,** 95-100.

FAIRLEY, L., SANTORELLI, G., LAWLOR, D., BRYANT, M., BHOPAL, R., PETHERICK, E., SAHOTA, P., GREENWOOD, D., HILL, A., CAMERON, N., BALL, H., BARBER, S. & WRIGHT, J. 2015a. The relationship between early life modifiable risk factors for childhood obesity, ethnicity and body mass index at age 3years: findings from the Born in Bradford birth cohort study. *BMC Obesity,* 2**,** 9.

FAIRLEY, L., SANTORELLI, G., LAWLOR, D. A., BRYANT, M., BHOPAL, R., PETHERICK, E. S., SAHOTA, P., GREENWOOD, D. C., HILL, A. J., CAMERON, N., BALL, H., BARBER, S. & WRIGHT, J. 2015b. The relationship between early life modifiable risk factors for childhood obesity, ethnicity and body mass index at age 3 years: findings from the Born in Bradford birth cohort study. *BMC obesity,* 2**,** 9-9.

FALL, C. H. D., SACHDEV, H. S., OSMOND, C., RESTREPO-MENDEZ, M. C., VICTORA, C., MARTORELL, R., STEIN, A. D., SINHA, S., TANDON, N., ADAIR, L., BAS, I., NORRIS, S. & RICHTER, L. M. 2015. Association between maternal age at childbirth and child and adult outcomes in the offspring: a prospective study in five low-income and middle-income countries (COHORTS collaboration). *The Lancet Global Health,* 3**,** e366-e377.

FANG, F., HODGES, E., MOLARO, A., DEAN, M., HANNON, G. J. & SMITH, A. D. 2012. Genomic landscape of human allele-specific DNA methylation. *Proc Natl Acad Sci U S A,* 109**,** 7332-7.

FARHANGI, M. A., KESHAVARZ, S.-A., ESHRAGHIAN, M., OSTADRAHIMI, A. & SABOOR-YARAGHI, A.-A. 2013. White Blood Cell Count in Women: Relation to Inflammatory Biomarkers, Haematological Profiles, Visceral Adiposity, and Other Cardiovascular Risk Factors. *Journal of Health, Population, and Nutrition,* 31**,** 58-64.

FAROOQ, M. A., PARKINSON, K. N., ADAMSON, A. J., PEARCE, M. S., REILLY, J. K., HUGHES, A. R., JANSSEN, X., BASTERFIELD, L. & REILLY, J. J. 2018. Timing of the decline in physical activity in childhood and adolescence: Gateshead Millennium Cohort Study. *British Journal of Sports Medicine,* 52**,** 1002.

FAROOQI, I. S. & O'RAHILLY, S. 2004. Monogenic Obesity in Humans. *Annual Review of Medicine,* 56**,** 443-458.

FENTON, T. R., PREMJI, S. S., AL-WASSIA, H. & SAUVE, R. S. 2014. Higher versus lower protein intake in formula-fed low birth weight infants. *Cochrane Database Syst Rev,* 4**,** CD003959.

FERNANDES, M., FERRARO, A., BASSOTTO, A., SANTOS, E. & SCHWARTSMAN, C. 2009. Association between catch-up in the first years of life and overweight/obesity at 10

years of age, in a low-income community in Sao Paulo, Brazil. *J Dev Orig Health Dis,* 1**,** s201-s205.

FIELDS, D. A. & DEMERATH, E. W. 2012. Relationship of insulin, glucose, leptin, IL-6 and TNF-α in human breast milk with infant growth and body composition. *Pediatric obesity,* 7**,** 304-312.

FILIBERTO, A. C., MACCANI, M. A., KOESTLER, D., WILHELM-BENARTZI, C., AVISSAR-WHITING, M., BANISTER, C. E., GAGNE, L. A. & MARSIT, C. J. 2011. Birthweight is associated with DNA promoter methylation of the glucocorticoid receptor in human placenta. *Epigenetics,* 6**,** 566-72.

FINKELHOR, D., SHATTUCK, A., TURNER, H. & HAMBY, S. 2013. Improving the adverse childhood experiences study scale. *JAMA pediatrics,* 167**,** 70-75.

FIORITO, G., POLIDORO, S., DUGUÉ, P.-A., KIVIMAKI, M., PONZI, E., MATULLO, G., GUARRERA, S., ASSUMMA, M. B., GEORGIADIS, P., KYRTOPOULOS, S. A., KROGH, V., PALLI, D., PANICO, S., SACERDOTE, C., TUMINO, R., CHADEAU-HYAM, M., STRINGHINI, S., SEVERI, G., HODGE, A. M., GILES, G. G., MARIONI, R., KARLSSON LINNÉR, R., O'HALLORAN, A. M., KENNY, R. A., LAYTE, R., BAGLIETTO, L., ROBINSON, O., MCCRORY, C., MILNE, R. L. & VINEIS, P. 2017. Social adversity and epigenetic aging: a multi-cohort study on socioeconomic differences in peripheral blood DNA methylation. *Scientific Reports,* 7**,** 16266.

FLEISCH, A. F., RIFAS-SHIMAN, S. L., KOUTRAKIS, P., SCHWARTZ, J. D., KLOOG, I., MELLY, S., COULL, B. A., ZANOBETTI, A., GILLMAN, M. W., GOLD, D. R. & OKEN, E. 2015. Prenatal exposure to traffic pollution: associations with reduced fetal growth and rapid infant weight gain. *Epidemiology (Cambridge, Mass.),* 26**,** 43-50.

FOG, C. K., GALLI, G. G. & LUND, A. H. 2012. PRDM proteins: important players in differentiation and disease. *Bioessays,* 34**,** 50-60.

FORREST, C. B., BLOCK, J. P. & BAILEY, L. C. 2017. Antibiotics, infections, and childhood obesity. *The Lancet Diabetes & Endocrinology,* 5**,** 2-3.

FORREST, L. F., HODGSON, S., PARKER, L. & PEARCE, M. S. 2011. The influence of childhood IQ and education on social mobility in the Newcastle Thousand Families birth cohort. *BMC Public Health,* 11**,** 895.

FORTIN, J.-P., LABBE, A., LEMIRE, M., ZANKE, B. W., HUDSON, T. J., FERTIG, E. J., GREENWOOD, C. M. & HANSEN, K. D. 2014. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome biology,* 15**,** 503.

FRANKENFIELD, D. C., ROWE, W. A., COONEY, R. N., SMITH, J. S. & BECKER, D. 2001. Limits of body mass index to detect obesity and predict body composition. *Nutrition,* 17**,** 26-30.

FRAYLING, T. M., TIMPSON, N. J., WEEDON, M. N., ZEGGINI, E., FREATHY, R. M., LINDGREN, C. M., PERRY, J. R. B., ELLIOTT, K. S., LANGO, H., RAYNER, N. W., SHIELDS, B., HARRIES, L. W., BARRETT, J. C., ELLARD, S., GROVES, C. J., KNIGHT, B., PATCH, A.-M., NESS, A. R., EBRAHIM, S., LAWLOR, D. A., RING, S. M., BEN-SHLOMO, Y., JARVELIN, M.-R., SOVIO, U., BENNETT, A. J., MELZER, D., FERRUCCI, L., LOOS, R. J. F., BARROSO, I., WAREHAM, N. J., KARPE, F., OWEN, K. R., CARDON, L. R., WALKER, M., HITMAN, G. A., PALMER, C. N. A., DONEY, A. S. F., MORRIS, A. D., SMITH, G. D., HATTERSLEY, A. T. & MCCARTHY, M. I. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science (New York, N.Y.),* 316**,** 889-894.

FRAZIER, C. R. M., MASON, P., ZHUANG, X. & BEELER, J. A. 2008. Sucrose Exposure in Early Life Alters Adult Motivation and Weight Gain. *PLoS ONE,* 3**,** e3221.

FREEDMAN, D. S., KHAN, L. K., DIETZ, W. H., SRINIVASAN, S. R. & BERENSON, G. S. 2001. Relationship of childhood obesity to coronary heart disease risk factors in adulthood: the Bogalusa Heart Study. *Pediatrics,* 108**,** 712-718.

FREEDMAN, D. S., KHAN, L. K., SERDULA, M. K., DIETZ, W. H., SRINIVASAN, S. R. & BERENSON, G. S. 2002. Relation of age at menarche to race, time period, and anthropometric dimensions: the Bogalusa Heart Study. *Pediatrics,* 110**,** e43-e43.

FREEDMAN, D. S., KHAN, L. K., SERDULA, M. K., DIETZ, W. H., SRINIVASAN, S. R. & BERENSON, G. S. 2005a. Racial Differences in the Tracking of Childhood BMI to Adulthood. *Obesity Research,* 13**,** 928-935.

FREEDMAN, D. S., WANG, J., MAYNARD, L. M., THORNTON, J. C., MEI, Z., PIERSON, R. N., DIETZ, W. H. & HORLICK, M. 2005b. Relation of BMI to fat and fat-free mass among children and adolescents. *International Journal of Obesity,* 29**,** 1-8.

FRIAS, A. E. & GROVE, K. L. Obesity: a transgenerational problem linked to nutrition during pregnancy. Seminars in reproductive medicine, 2012. NIH Public Access, 472.

FROMMER, M., MCDONALD, L. E., MILLAR, D. S., COLLIS, C. M., WATT, F., GRIGG, G. W., MOLLOY, P. L. & PAUL, C. L. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences,* 89**,** 1827-1831.

FROY, O. 2010. Metabolism and Circadian Rhythms—Implications for Obesity. *Endocrine Reviews,* 31**,** 1-24.

FUCHS, F., MONET, B., DUCRUET, T., CHAILLET, N. & AUDIBERT, F. 2018. Effect of maternal age on the risk of preterm birth: A large cohort study. *PLOS ONE,* 13**,** e0191002.

GAGNON-BARTSCH, J. A. & SPEED, T. P. 2012. Using control genes to correct for unwanted variation in microarray data. *Biostatistics,* 13**,** 539-552.

GALLAGHER, D., HEYMSFIELD, S. B., HEO, M., JEBB, S. A., MURGATROYD, P. R. & SAKAMOTO, Y. 2000. Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index. *Am J Clin Nutr,* 72**,** 694-701.

GALOBARDES, B., SHAW, M., LAWLOR, D. A., LYNCH, J. W. & SMITH, G. D. 2006. Indicators of socioeconomic position (part 1). *Journal of Epidemiology & Community Health,* 60**,** 7-12.

GAO, Y. J., HOLLOWAY, A. C., ZENG, Z. H., LIM, G. E., PETRIK, J. J., FOSTER, W. G. & LEE, R. M. 2005. Prenatal exposure to nicotine causes postnatal obesity and altered perivascular adipose tissue function. *Obes Res,* 13**,** 687-92.

GARDNER, D. S., HOSKING, J., METCALF, B. S., JEFFERY, A. N., VOSS, L. D. & WILKIN, T. J. 2009. Contribution of early weight gain to childhood overweight and metabolic health: a longitudinal study (EarlyBird 36). *Pediatrics,* 123**,** e67-73.

GARSON, G. D. 2008. Path analysis. *from Statnotes: Topics in Multivariate Analysis. Retrieved,* 9**,** 2009.

GARSON, G. D. 2017. Mediation & moderation: Partial correlation and regression approaches. *Asheboro, NC: Statistical Associates Publishers*.

GAYLOR, E. E., GOODLIN-JONES, B. L. & ANDERS, T. F. 2001. Classification of young children's sleep problems: a pilot study. *J Am Acad Child Adolesc Psychiatry,* 40**,** 61-7.

GIBBS, B. G. & FORSTE, R. 2014. Socioeconomic status, infant feeding practices and early childhood obesity. *Pediatr Obes,* 9**,** 135-46.

GIBBS, J. R., VAN DER BRUG, M. P., HERNANDEZ, D. G., TRAYNOR, B. J., NALLS, M. A., LAI, S.-L., AREPALLI, S., DILLMAN, A., RAFFERTY, I. P., TRONCOSO, J., JOHNSON, R., ZIELKE, H. R., FERRUCCI, L., LONGO, D. L., COOKSON, M. R. & SINGLETON, A. B. 2010. Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain. *PLOS Genetics,* 6**,** e1000952.

GILLMAN, M. W. & LUDWIG, D. S. 2013. How early should obesity prevention start? *New England Journal of Medicine,* 369**,** 2173-2175.

GILLMAN, M. W., OAKEY, H., BAGHURST, P. A., VOLKMER, R. E., ROBINSON, J. S. & CROWTHER, C. A. 2010. Effect of treatment of gestational diabetes mellitus on obesity in the next generation. *Diabetes Care,* 33**,** 964-8.

GILLMAN, M. W., RICH-EDWARDS, J. W., HUH, S., MAJZOUB, J. A., OKEN, E., TAVERAS, E. M. & RIFAS-SHIMAN, S. L. 2006. Maternal Corticotropin-Releasing Hormone Levels during Pregnancy and Offspring Adiposity. *Obesity (Silver Spring, Md.),* 14**,** 1647-1653.

GILLMAN, M. W., RIFAS-SHIMAN, S. L., KLEINMAN, K., OKEN, E., RICH-EDWARDS, J. W. & TAVERAS, E. M. 2008a. Developmental Origins of Childhood Overweight: Potential Public Health Impact. *Obesity,* 16**,** 1651-1656.

GILLMAN, M. W., RIFAS-SHIMAN, S. L., KLEINMAN, K., OKEN, E., RICH-EDWARDS, J. W. & TAVERAS, E. M. 2008b. Developmental origins of childhood overweight: potential public health impact. *Obesity,* 16**,** 1651-1656.

GLOTZBACH, S. F., EDGAR, D. M., BOEDDIKER, M. & ARIAGNO, R. L. 1994. Biological rhythmicity in normal infants during the first 3 months of life. *Pediatrics,* 94**,** 482-488.

GODFREY, K. M., REYNOLDS, R. M., PRESCOTT, S. L., NYIRENDA, M., JADDOE, V. W. V., ERIKSSON, J. G. & BROEKMAN, B. F. P. 2017. Influence of maternal obesity on the long-term health of off spring. *Lancet Diabetes & Endocrinology,* 5**,** 53-64.

GODFREY, K. M., SHEPPARD, A., GLUCKMAN, P. D., LILLYCROP, K. A., BURDGE, G. C., MCLEAN, C., RODFORD, J., SLATER-JEFFERIES, J. L., GARRATT, E., CROZIER, S. R., EMERALD, B. S., GALE, C. R., INSKIP, H. M., COOPER, C. & HANSON, M. A. 2011. Epigenetic gene promoter methylation at birth is associated with child's later adiposity. *Diabetes,* 60**,** 1528-34.

GOISIS, A., REMES, H., BARCLAY, K., MARTIKAINEN, P. & MYRSKYLÄ, M. 2017. Advanced Maternal Age and the Risk of Low Birth Weight and Preterm Delivery: a Within-Family Analysis Using Finnish Population Registers. *American Journal of Epidemiology,* 186**,** 1219-1226.

GONZALEZ-NAHM, S., MENDEZ, M. A., BENJAMIN-NEELON, S. E., MURPHY, S. K., HOGAN, V. K., ROWLEY, D. L. & HOYO, C. 2018. DNA methylation of imprinted genes at birth is associated with child weight status at birth, 1 year, and 3 years. *Clinical Epigenetics,* 10**,** 90.

GORE, L., TRICHE, T. J., FARRAR, J. E., WAI, D., LEGENDRE, C., GOODEN, G. C., LIANG, W. S., CARPTEN, J., LEE, D. & ALVARO, F. 2017. A multicenter, randomized study of decitabine as epigenetic priming with induction chemotherapy in children with AML. *Clinical epigenetics,* 9**,** 108.

GRONDAHL, M. L., YDING ANDERSEN, C., BOGSTAD, J., NIELSEN, F. C., MEINERTZ, H. & BORUP, R. 2010. Gene expression profiles of single human mature oocytes in relation to age. *Hum Reprod,* 25**,** 957-68.

GROOM, A., POTTER, C., SWAN, D. C., FATEMIFAR, G., EVANS, D. M., RING, S. M., TURCOT, V., PEARCE, M. S., EMBLETON, N. D., SMITH, G. D., MATHERS, J. C. & RELTON, C. L. 2012. Postnatal growth and DNA methylation are associated with differential gene expression of the TACSTD2 gene and childhood fat mass. *Diabetes,* 61**,** 391-400.

GROSSMAN, M. 1972. The demand for health: a theoretical and empirical investigation. *NBER Books*.

GUILLEMETTE, L., DURKSEN, A., RABBANI, R., ZARYCHANSKI, R., ABOU-SETTA, A. M., DUHAMEL, T. A., MCGAVOCK, J. M. & WICKLOW, B. 2017. Intensive gestational

glycemic management and childhood obesity: a systematic review and meta-analysis. *International Journal Of Obesity,* 41**,** 999.
GUNGOR, D., PAUL, I., BIRCH, L. & BARTOK, C. 2010. Risky vs Rapid Growth in Infancy. *Archives of pediatrics & adolescent medicine,* 164**,** 1091-7.
HAIGHT, T. H. & PIERCE, W. E. 1955. Effect of Prolonged Antibiotic Administration on the Weight of Healthy Young Males: One Figure. *The Journal of nutrition,* 56**,** 151-161.
HALES, C. N. & BARKER, D. J. 1992. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia,* 35**,** 595-601.
HANSEN, K. 2015. IlluminaHumanMethylation450kanno. ilmn12. hg19: annotation for illumina's 450k methylation arrays. *R package, version 0.2,* 1.
HANSON, M. A. & GLUCKMAN, P. 2014. Early developmental conditioning of later health and disease: physiology or pathophysiology? *Physiological reviews,* 94**,** 1027-1076.
HANSON, M. D. & CHEN, E. 2007. Socioeconomic status and health behaviors in adolescence: a review of the literature. *Journal of behavioral medicine,* 30**,** 263-285.
HARA, M., SAITOU, E., IWATA, F., OKADA, T. & HARADA, K. 2002. Waist-to-height ratio is the best predictor of cardiovascular disease risk factors in Japanese schoolchildren. *Journal of atherosclerosis and thrombosis,* 9**,** 127-132.
HARDER, T., BERGMANN, R., KALLISCHNIGG, G. & PLAGEMANN, A. 2005. Duration of breastfeeding and risk of overweight: A meta-analysis. *American Journal of Epidemiology,* 162**,** 397-403.
HARRIS, J. A. & BENEDICT, F. G. 1918. A Biometric Study of Human Basal Metabolism. *Proceedings of the National Academy of Sciences of the United States of America,* 4**,** 370-373.
HARTWIG, F. P., DE MOLA, C. L., DAVIES, N. M., VICTORA, C. G. & RELTON, C. L. 2017. Breastfeeding effects on DNA methylation in the offspring: A systematic literature review. *PloS one,* 12**,** e0173070.
HARVEY, N. C., POOLE, J. R., JAVAID, M. K., DENNISON, E. M., ROBINSON, S., INSKIP, H. M., GODFREY, K. M., COOPER, C., AIHIE SAYER, A. & THE, S. W. S. S. G. 2007. PARENTAL DETERMINANTS OF NEONATAL BODY COMPOSITION. *The Journal of clinical endocrinology and metabolism,* 92**,** 523-526.
HAWKINS, S. S., OKEN, E. & GILLMAN, M. W. 2018. Early in the life course: Time for obesity prevention. *Handbook of Life Course Health Development.* Springer, Cham.
HEALTH AND SOCIAL CARE INFORMATION CENTRE. 2014. Health Survey for England - 2013, Trend tables. . Available: www.hscic.gov.uk/catalogue/PUB16077.
HEALTH AND SOCIAL CARE INFORMATION CENTRE. 2016. National Child Measurement Programme - England, 2015-16. Available: http://digital.nhs.uk/catalogue/PUB22269.
HEERMAN, W. J., DALEY, M. F., BOONE-HEINONEN, J., RIFAS-SHIMAN, S. L., BAILEY, L. C., FORREST, C. B., YOUNG, J. G., GILLMAN, M. W., HORGAN, C. E., JANICKE, D. M., JENTER, C., KHARBANDA, E. O., LUNSFORD, D., MESSITO, M. J., TOH, S., BLOCK, J. P., ARTERBURN, D., CLEVELAND, L. P., FINKELSTEIN, J., FITZPATRICK, S. L., GOODMAN, A., HORBERG, M., INGBER, J., MURPHY, K., PEAY, H. L., RIVERA, P., REYNOLDS, J. S., STURTEVANT, J. L., TORRES, I., ON BEHALF OF THE, P. A. & CHILDHOOD GROWTH STUDY, G. 2019. Maternal antibiotic use during pregnancy and childhood obesity at age 5 years. *International Journal of Obesity,* 43**,** 1202-1209.
HEERWAGEN, M. J. R., MILLER, M. R., BARBOUR, L. A. & FRIEDMAN, J. E. 2010. Maternal obesity and fetal metabolic programming: a fertile epigenetic soil. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology,* 299**,** R711-R722.
HEIJMANS, B. T., TOBI, E. W., STEIN, A. D., PUTTER, H., BLAUW, G. J., SUSSER, E. S., SLAGBOOM, P. E. & LUMEY, L. 2008a. Persistent epigenetic differences associated

with prenatal exposure to famine in humans. *Proceedings of the National Academy of Sciences,* 105**,** 17046-17049.

HEIJMANS, B. T., TOBI, E. W., STEIN, A. D., PUTTER, H., BLAUW, G. J., SUSSER, E. S., SLAGBOOM, P. E. & LUMEY, L. H. 2008b. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 17046-17049.

HEPPE, D. H. M., KIEFTE-DE JONG, J. C., DURMUŞ, B., MOLL, H. A., RAAT, H., HOFMAN, A. & JADDOE, V. W. V. 2012. Parental, fetal, and infant risk factors for preschool overweight: the Generation R Study. *Pediatric Research,* 73**,** 120.

HERAS, V. L., CLOONEY, A. G., RYAN, F. J., CABRERA-RUBIO, R., CASEY, P. G., HUESTON, C. M., PINHEIRO, J., RUDKIN, J. K., MELGAR, S., COTTER, P. D., HILL, C. & GAHAN, C. G. M. 2019. Short-term consumption of a high-fat diet increases host susceptibility to Listeria monocytogenes infection. *Microbiome,* 7**,** 7.

HERMAN-GIDDENS, M. E., SLORA, E. J., WASSERMAN, R. C., BOURDONY, C. J., BHAPKAR, M. V., KOCH, G. G. & HASEMEIER, C. M. 1997. Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings network. *Pediatrics,* 99**,** 505-512.

HINKLE, S. N., ALBERT, P. S., MENDOLA, P., SJAARDA, L. A., YEUNG, E., BOGHOSSIAN, N. S. & LAUGHON, S. K. 2014. The Association between Parity and Birthweight in a Longitudinal Consecutive Pregnancy Cohort. *Paediatric and Perinatal Epidemiology,* 28**,** 106-115.

HINRICHS, A. S., ZWEIG, A. S., RANEY, B. J., LEE, B. T., TYNER, C., LEE, C. M., GIBSON, D., BARBER, G. P., CLAWSON, H., GONZALEZ, J. N., CASPER, J., ROSENBLOOM, K. R., DIEKHANS, M., SPEIR, M. L., KUHN, R. M., KENT, W J., HAEUSSLER, M. & HAUSSLER, D. 2018. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Research,* 47**,** D853-D858.

HIRSCHLER, V., ARANDA, C., DE LUJÁN CALCAGNO, M., MACCALINI, G. & JADZINSKY, M. 2005. Can waist circumference identify children with the metabolic syndrome? *Archives of pediatrics & adolescent medicine,* 159**,** 740-744.

HISCOCK, H., SCALZO, K., CANTERFORD, L. & WAKE, M. 2011. Sleep duration and body mass index in 0-7-year olds. *Arch Dis Child,* 96**,** 735-9.

HOCHBERG, Z., FEIL, R., CONSTANCIA, M., FRAGA, M., JUNIEN, C., CAREL, J. C., BOILEAU, P., LE BOUC, Y., DEAL, C. L., LILLYCROP, K., SCHARFMANN, R., SHEPPARD, A., SKINNER, M., SZYF, M., WATERLAND, R. A., WAXMAN, D. J., WHITELAW, E., ONG, K. & ALBERTSSON-WIKLAND, K. 2011. Child health, developmental plasticity, and epigenetic programming. *Endocrine reviews,* 32**,** 159-224.

HOHWÜ, L., LI, J., OLSEN, J., SØRENSEN, T. I. A. & OBEL, C. 2014. Severe Maternal Stress Exposure Due to Bereavement before, during and after Pregnancy and Risk of Overweight and Obesity in Young Adult Men: A Danish National Cohort Study. *PLoS ONE,* 9**,** e97490.

HOLMES, E. E., JUNG, M., MELLER, S., LEISSE, A., SAILER, V., ZECH, J., MENGDEHL, M., GARBE, L.-A., UHL, B., KRISTIANSEN, G. & DIETRICH, D. 2014. Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PloS one,* 9**,** e93933-e93933.

HORVATH, S. 2013. DNA methylation age of human tissues and cell types. *Genome Biol,* 14**,** R115.

HORVATH, S. & LEVINE, A. J. 2015. HIV-1 Infection Accelerates Age According to the Epigenetic Clock. *J Infect Dis,* 212**,** 1563-73.

HOUSEMAN, E. A., ACCOMANDO, W. P., KOESTLER, D. C., CHRISTENSEN, B. C., MARSIT, C. J., NELSON, H. H., WIENCKE, J. K. & KELSEY, K. T. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics,* 13**,** 86.

HOUSEMAN, E. A., KILE, M. L., CHRISTIANI, D. C., INCE, T. A., KELSEY, K. T. & MARSIT, C. J. 2016. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *BMC bioinformatics,* 17**,** 259.

HOUSEMAN, E. A., MOLITOR, J. & MARSIT, C. J. 2014. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics,* 30**,** 1431-1439.

HOUTEPEN, L. C., HARDY, R., MADDOCK, J., KUH, D., ANDERSON, E. L., RELTON, C. L., SUDERMAN, M. J. & HOWE, L. D. 2018a. Childhood adversity and DNA methylation in two population-based cohorts. *Translational Psychiatry,* 8**,** 266.

HOUTEPEN, L. C., HERON, J., SUDERMAN, M. J., FRASER, A., CHITTLEBOROUGH, C. R. & HOWE, L. D. 2019. Adverse childhood experiences: associations with educational attainment and adolescent health, and the role of family and socioeconomic factors. Analysis of a prospective cohort study. *bioRxiv***,** 612390.

HOUTEPEN, L. C., HERON, J., SUDERMAN, M. J., TILLING, K. & HOWE, L. D. 2018b. Adverse childhood experiences in the children of the Avon Longitudinal Study of Parents and Children (ALSPAC). *Wellcome open research,* 3**,** 106-106.

HOWE, L. D. 2012. Childhood obesity: socioeconomic inequalities and consequences for later cardiovascular health. *Longitudinal and Life Course Studies,* 4**,** 4-16.

HOWE, L. D., MATIJASEVICH, A., TILLING, K., BRION, M.-J., LEARY, S. D., SMITH, G. D. & LAWLOR, D. A. 2012. Maternal smoking during pregnancy and offspring trajectories of height and adiposity: comparing maternal and paternal associations. *International Journal of Epidemiology,* 41**,** 722-732.

HOWE, L. D., TILLING, K., GALOBARDES, B. & LAWLOR, D. A. 2013. Loss to Follow-up in Cohort Studies Bias in Estimates of Socioeconomic Inequalities. *Epidemiology,* 24**,** 1-9.

HOWE, L. D., TILLING, K., GALOBARDES, B., SMITH, G. D., NESS, A. R. & LAWLOR, D. A. 2011. Socioeconomic disparities in trajectories of adiposity across childhood. *Int J Pediatr Obes,* 6.

HUGHES, K., BELLIS, M. A., HARDCASTLE, K. A., SETHI, D., BUTCHART, A., MIKTON, C., JONES, L. & DUNNE, M. P. 2017. The effect of multiple adverse childhood experiences on health: a systematic review and meta-analysis. *Lancet Public Health,* 2**,** e356-e366.

HUMPHREYS, B. R. 2013. Dealing with zeros in economic data. *University of Alberta, Department of Economics***,** 1-27.

HUXLEY, R., MENDIS, S., ZHELEZNYAKOV, E., REDDY, S. & CHAN, J. 2009. Body mass index, waist circumference and waist:hip ratio as predictors of cardiovascular risk—a review of the literature. *European Journal Of Clinical Nutrition,* 64**,** 16.

ILLINGWORTH, R., KERR, A., DESOUSA, D., JØRGENSEN, H., ELLIS, P., STALKER, J., JACKSON, D., CLEE, C., PLUMB, R., ROGERS, J., HUMPHRAY, S., COX, T., LANGFORD, C. & BIRD, A. 2008. A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci. *PLOS Biology,* 6**,** e22.

IP, S., CHUNG, M., RAMAN, G., CHEW, P., MAGULA, N., DEVINE, D., TRIKALINOS, T. & LAU, J. 2007. Breastfeeding and maternal and infant health outcomes in developed countries. *Evid Rep Technol Assess (Full Rep)***,** 1-186.

ITO, S., D'ALESSIO, A. C., TARANOVA, O. V., HONG, K., SOWERS, L. C. & ZHANG, Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature,* 466**,** 1129-U151.

IZQUIERDO, A. G. & CRUJEIRAS, A. B. 2019. Obesity-Related Epigenetic Changes After Bariatric Surgery. *Frontiers in endocrinology,* 10**,** 232-232.
JACOB, C., BAIRD, J., BARKER, M., COOPER, C. & HANSON, M. 2015. The importance of a life course approach to health: chronic disease risk from preconception through adolescence and adulthood. *Southampton: University of Southampton*.
JACOBSSON, B., LADFORS, L. & MILSOM, I. 2004. Advanced maternal age and adverse perinatal outcome. *Obstetrics & Gynecology,* 104**,** 727-733.
JAENISCH, R. & BIRD, A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics,* 33**,** 245-254.
JAFFE, A. E. & IRIZARRY, R. A. 2014a. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biology,* 15**,** R31.
JAFFE, A. E. & IRIZARRY, R. A. 2014b. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol,* 15**,** R31.
JAVED, A., JUMEAN, M., MURAD, M., OKORODUDU, D., KUMAR, S., SOMERS, V., SOCHOR, O. & LOPEZ-JIMENEZ, F. 2015. Diagnostic performance of body mass index to identify obesity as defined by body adiposity in children and adolescents: a systematic review and meta-analysis. *Pediatric obesity,* 10**,** 234-244.
JENKINS, T. G., ASTON, K. I., PFLUEGER, C., CAIRNS, B. R. & CARRELL, D. T. 2014. Age-Associated Sperm DNA Methylation Alterations: Possible Implications in Offspring Disease Susceptibility. *PLOS Genetics,* 10**,** e1004458.
JÉQUIER, E. & TAPPY, L. 1999. Regulation of body weight in humans. *Physiological reviews,* 79**,** 451-480.
JERNBERG, C., LOFMARK, S., EDLUND, C. & JANSSON, J. K. 2010. Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology,* 156**,** 3216-23.
JESPERSEN, N. Z., LARSEN, T. J., PEIJS, L., DAUGAARD, S., HOMØE, P., LOFT, A., DE JONG, J., MATHUR, N., CANNON, B. & NEDERGAARD, J. 2013. A classical brown adipose tissue mRNA signature partly overlaps with brite in the supraclavicular region of adult humans. *Cell metabolism,* 17**,** 798-805.
JOGLEKAR, C., FALL, C., DESHPANDE, V., JOSHI, N., BHALERAO, A., SOLAT, V., DEOKAR, T., CHOUGULE, S., LEARY, S. & OSMOND, C. 2007. Newborn size, infant and childhood growth, and body composition and cardiovascular disease risk factors at the age of 6 years: the Pune Maternal Nutrition Study. *International journal of obesity,* 31**,** 1534.
JOHANSSON, G., WIKMAN, Å., ÅHRÉN, A.-M., HALLMANS, G. & JOHANSSON, I. 2007. Underreporting of energy intake in repeated 24-hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living. *Public Health Nutrition,* 4**,** 919-927.
JOHANSSON, G., WIKMAN, A., ÅHRÉN, A. M., HALLMANS, G. & JOHANSSON, I. 2001. Underreporting of energy intake in repeated 24-hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living. *Public Health Nutrition,* 4**,** 919-927.
JOHNSON, R. K., SOULTANAKIS, R. P. & MATTHEWS, D. E. 1998. Literacy and body fatness are associated with underreporting of energy intake in US low-income women using the multiple-pass 24-hour recall: a doubly labeled water study. *J Am Diet Assoc,* 98**,** 1136-40.
JOHNSON, S. B., RILEY, A. W., GRANGER, D. A. & RIIS, J. 2013. The science of early life toxic stress for pediatric practice and advocacy. *Pediatrics,* 131**,** 319-27.
JOHNSON, W., LI, L., KUH, D. & HARDY, R. 2015. How Has the Age-Related Process of Overweight or Obesity Development Changed over Time? Co-ordinated Analyses of

Individual Participant Data from Five United Kingdom Birth Cohorts. *PLOS Medicine,* 12**,** e1001828.

JOHNSON, W., WRIGHT, J. & CAMERON, N. 2012. The risk of obesity by assessing infant growth against the UK-WHO charts compared to the UK90 reference: findings from the Born in Bradford birth cohort study. *BMC Pediatrics,* 12**,** 1-9.

JOHNSON, W. E., LI, C. & RABINOVIC, A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics,* 8**,** 118-127.

JOLLY, M. C., SEBIRE, N. J., HARRIS, J. P., REGAN, L. & ROBINSON, S. 2003. Risk factors for macrosomia and its clinical consequences: a study of 350,311 pregnancies. *European Journal of Obstetrics & Gynecology and Reproductive Biology,* 111**,** 9-14.

JONES, P. A. 1999. The DNA methylation paradox. *Trends Genet,* 15**,** 34-7.

JONES, P. A. & BAYLIN, S. B. 2002. The fundamental role of epigenetic events in cancer. *Nature reviews genetics,* 3**,** 415.

JOUBERT, BONNIE R., FELIX, JANINE F., YOUSEFI, P., BAKULSKI, KELLY M., JUST, ALLAN C., BRETON, C., REESE, S. E., MARKUNAS, CHRISTINA A., RICHMOND, REBECCA C., XU, C.-J., KÜPERS, LEANNE K., OH, SAM S., HOYO, C., GRUZIEVA, O., SÖDERHÄLL, C., SALAS, LUCAS A., BAÏZ, N., ZHANG, H., LEPEULE, J., RUIZ, C., LIGTHART, S., WANG, T., TAYLOR, JACK A., DUIJTS, L., SHARP, GEMMA C., JANKIPERSADSING, SOESMA A., NILSEN, ROY M., VAEZ, A., FALLIN, M. D., HU, D., LITONJUA, AUGUSTO A., FUEMMELER, BERNARD F., HUEN, K., KERE, J., KULL, I., MUNTHE-KAAS, MONICA C., GEHRING, U., BUSTAMANTE, M., SAUREL-COUBIZOLLES, MARIE J., QURAISHI, BILAL M., REN, J., TOST, J., GONZALEZ, JUAN R., PETERS, MARJOLEIN J., HÅBERG, SIRI E., XU, Z., VAN MEURS, JOYCE B., GAUNT, TOM R., KERKHOF, M., CORPELEIJN, E., FEINBERG, ANDREW P., ENG, C., BACCARELLI, ANDREA A., BENJAMIN NEELON, SARA E., BRADMAN, A., MERID, SIMON K., BERGSTRÖM, A., HERCEG, Z., HERNANDEZ-VARGAS, H., BRUNEKREEF, B., PINART, M., HEUDE, B., EWART, S., YAO, J., LEMONNIER, N., FRANCO, OSCAR H., WU, MICHAEL C., HOFMAN, A., MCARDLE, W., VAN DER VLIES, P., FALAHI, F., GILLMAN, MATTHEW W., BARCELLOS, LISA F., KUMAR, A., WICKMAN, M., GUERRA, S., CHARLES, M.-A., HOLLOWAY, J., AUFFRAY, C., TIEMEIER, HENNING W., SMITH, GEORGE D., POSTMA, D., HIVERT, M.-F., ESKENAZI, B., VRIJHEID, M., ARSHAD, H., ANTÓ, JOSEP M., DEHGHAN, A., KARMAUS, W., ANNESI-MAESANO, I., SUNYER, J., GHANTOUS, A., PERSHAGEN, G., HOLLAND, N., MURPHY, SUSAN K., DEMEO, DAWN L., BURCHARD, ESTEBAN G., LADD-ACOSTA, C., SNIEDER, H., NYSTAD, W., et al. 2016. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *The American Journal of Human Genetics,* 98**,** 680-696.

KAHVECI, B., MELEKOGLU, R., EVRUKE, I. C. & CETIN, C. 2018. The effect of advanced maternal age on perinatal outcomes in nulliparous singleton pregnancies. *BMC Pregnancy and Childbirth,* 18**,** 343.

KANG, H. 2013. The prevention and handling of the missing data. *Korean journal of anesthesiology,* 64**,** 402-406.

KAPLOWITZ, P. B. 2008. Link between body fat and the timing of puberty. *Pediatrics,* 121**,** S208-S217.

KATHIRVEL, M. & MAHADEVAN, S. 2016. The role of epigenetics in tuberculosis infection. *Epigenomics,* 8**,** 537-549.

KAUFMAN, D., BANERJI, M. A., SHORMAN, I., SMITH, E. L., COPLAN, J. D., ROSENBLUM, L. A. & KRAL, J. G. 2007. Early-life stress and the development of obesity and insulin resistance in juvenile bonnet macaques. *Diabetes,* 56**,** 1382-6.

KAUFMAN, J., MONTALVO-ORTIZ, J. L., HOLBROOK, H., O'LOUGHLIN, K., ORR, C., KEARNEY, C., YANG, B.-Z., WANG, T., ZHAO, H., ALTHOFF, R., GARAVAN, H., GELERNTER, J. & HUDZIAK, J. 2018. Adverse Childhood Experiences, Epigenetic Measures, and Obesity in Youth. *The Journal of Pediatrics,* 202**,** 150-156.e3.

KAUSHAL, A., ZHANG, H., KARMAUS, W. J., RAY, M., TORRES, M. A., SMITH, A. K. & WANG, S.-L. 2017. Comparison of different cell type correction methods for genome-scale epigenetics studies. *BMC bioinformatics,* 18**,** 216.

KEHOE, S. H., KRISHNAVENI, G. V., LUBREE, H. G., WILLS, A. K., GUNTUPALLI, A. M., VEENA, S. R., BHAT, D. S., KISHORE, R., FALL, C. H. D., YAJNIK, C. S. & KURPAD, A. 2011. Prediction of body-fat percentage from skinfold and bio-impedance measurements in Indian school children. *European journal of clinical nutrition,* 65**,** 1263-1270.

KELISHADI, R., HAGHDOOST, A. A., JAMSHIDI, F., ALIRAMEZANY, M. & MOOSAZADEH, M. 2015. Low birthweight or rapid catch-up growth: which is more associated with cardiovascular disease and its risk factors in later life? A systematic review and cryptanalysis. *Paediatrics and International Child Health,* 35**,** 110-123.

KEMP, P. A. 2015. Private Renting After the Global Financial Crisis. *Housing Studies,* 30**,** 601-620.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome research,* 12**,** 996-1006.

KHANDELWAL, P., JAIN, V., GUPTA, A. K., KALAIVANI, M. & PAUL, V. K. 2014. Association of early postnatal growth trajectory with body composition in term low birth weight infants. *Journal of developmental origins of health and disease,* 5**,** 189-196.

KILPELÄINEN, T. O., QI, L., BRAGE, S., SHARP, S. J., SONESTEDT, E., DEMERATH, E., AHMAD, T., MORA, S., KAAKINEN, M., SANDHOLT, C. H., HOLZAPFEL, C., AUTENRIETH, C. S., HYPPÖNEN, E., CAUCHI, S., HE, M., KUTALIK, Z., KUMARI, M., STANČÁKOVÁ, A., MEIDTNER, K., BALKAU, B., TAN, J. T., MANGINO, M., TIMPSON, N. J., SONG, Y., ZILLIKENS, M. C., JABLONSKI, K. A., GARCIA, M. E., JOHANSSON, S., BRAGG-GRESHAM, J. L., WU, Y., VAN VLIET-OSTAPTCHOUK, J. V., ONLAND-MORET, N. C., ZIMMERMANN, E., RIVERA, N. V., TANAKA, T., STRINGHAM, H. M., SILBERNAGEL, G., KANONI, S., FEITOSA, M. F., SNITKER, S., RUIZ, J. R., METTER, J., LARRAD, M. T. M., ATALAY, M., HAKANEN, M., AMIN, N., CAVALCANTI-PROENÇA, C., GRØNTVED, A., HALLMANS, G., JANSSON, J.-O., KUUSISTO, J., KÄHÖNEN, M., LUTSEY, P. L., NOLAN, J. J., PALLA, L., PEDERSEN, O., PÉRUSSE, L., RENSTRÖM, F., SCOTT, R. A., SHUNGIN, D., SOVIO, U., TAMMELIN, T. H., RÖNNEMAA, T., LAKKA, T. A., UUSITUPA, M., RIOS, M. S., FERRUCCI, L., BOUCHARD, C., MEIRHAEGHE, A., FU, M., WALKER, M., BORECKI, I. B., DEDOUSSIS, G. V., FRITSCHE, A., OHLSSON, C., BOEHNKE, M., BANDINELLI, S., VAN DUIJN, C. M., EBRAHIM, S., LAWLOR, D. A., GUDNASON, V., HARRIS, T. B., SØRENSEN, T. I. A., MOHLKE, K. L., HOFMAN, A., UITTERLINDEN, A. G., TUOMILEHTO, J., LEHTIMÄKI, T., RAITAKARI, O., ISOMAA, B., NJØLSTAD, P. R., FLOREZ, J. C., LIU, S., NESS, A., SPECTOR, T. D., TAI, E. S., FROGUEL, P., BOEING, H., LAAKSO, M., MARMOT, M., et al. 2011. Physical Activity Attenuates the Influence of FTO Variants on Obesity Risk: A Meta-Analysis of 218,166 Adults and 19,268 Children. *PLOS Medicine,* 8**,** e1001116.

KING, A. C., PARKINSON, K. N., ADAMSON, A. J., MURRAY, L., BESSON, H., REILLY, J. J., BASTERFIELD, L. & THE GATESHEAD MILLENNIUM STUDY CORE, T. 2011. Correlates of objectively measured physical activity and sedentary behaviour in English children. *European Journal of Public Health,* 21**,** 424-431.

KINT, S., DE SPIEGELAERE, W., DE KESEL, J., VANDEKERCKHOVE, L. & VAN CRIEKINGE, W. 2018. Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. *PloS one,* 13**,** e0199091-e0199091.

KISIEL, J. B., DUKEK, B. A., R, V. S. R. K., GHOZ, H. M., YAB, T. C., BERGER, C. K., TAYLOR, W. R., FOOTE, P. H., GIAMA, N. H., ONYIRIOHA, K., ABDALLAH, M. A., BURGER, K. N., SLETTEDAHL, S. W., MAHONEY, D. W., SMYRK, T. C., LEWIS, J. T., GIAKOUMOPOULOS, M., ALLAWI, H. T., LIDGARD, G. P., ROBERTS, L. R. & AHLQUIST, D. A. 2019. Hepatocellular Carcinoma Detection by Plasma Methylated DNA: Discovery, Phase I Pilot, and Phase II Clinical Validation. *Hepatology,* 69**,** 1180-1192.

KLEINBAUM, D. G., KUPPER, L. L. & MORGENSTERN, H. 1982. *Epidemiologic research: principles and quantitative methods*, John Wiley & Sons.

KLEINMAN, J. C. & MADANS, J. H. 1985. The effects of maternal smoking, physical stature, and educational attainment on the incidence of low birth weight. *Am J Epidemiol,* 121**,** 843-55.

KLINGENBERG, L., CHRISTENSEN, L. B., HJORTH, M. F., ZANGENBERG, S., CHAPUT, J. P., SJÖDIN, A., MØLGAARD, C. & MICHAELSEN, K. F. 2013. No relation between sleep duration and adiposity indicators in 9–36 months old children: the SKOT cohort. *Pediatric obesity,* 8**,** e14-e18.

KNAI, C., LOBSTEIN, T., DARMON, N., RUTTER, H. & MCKEE, M. 2012. Socioeconomic Patterning of Childhood Overweight Status in Europe. *International Journal of Environmental Research and Public Health,* 9**,** 1472.

KOENKER, R. & BASSETT JR, G. 1978. Regression quantiles. *Econometrica: journal of the Econometric Society***,** 33-50.

KOK, D. E. G., DHONUKSHE-RUTTEN, R. A. M., LUTE, C., HEIL, S. G., UITTERLINDEN, A. G., VAN DER VELDE, N., VAN MEURS, J. B. J., VAN SCHOOR, N. M., HOOIVELD, G. J. E. J., DE GROOT, L. C. P. G. M., KAMPMAN, E. & STEEGENGA, W. T. 2015. The effects of long-term daily folic acid and vitamin B12 supplementation on genome-wide DNA methylation in elderly subjects. *Clinical Epigenetics,* 7**,** 121.

KOLETZKO, B., BEYER, J., BRANDS, B., DEMMELMAIR, H., GROTE, V., HAILE, G., GRUSZFELD, D., RZEHAK, P., SOCHA, P. & WEBER, M. 2013. Early influences of nutrition on postnatal growth. *Recent Advances in Growth Research: Nutritional, Molecular and Endocrine Perspectives.* Karger Publishers.

KOLETZKO, B., BROEKAERT, I., DEMMELMAIR, H., FRANKE, J., HANNIBAL, I., OBERLE, D., SCHIESS, S., BAUMANN, B. T. & VERWIED-JORKY, S. 2005. Protein intake in the first year of life: a risk factor for later obesity? The E.U. childhood obesity project. *Adv Exp Med Biol,* 569**,** 69-79.

KOLETZKO, B., VON KRIES, R., MONASTEROLO, R. C., SUBIAS, J. E., SCAGLIONI, S., GIOVANNINI, M., BEYER, J., DEMMELMAIR, H., ANTON, B. & GRUSZFELD, D. 2009. Infant feeding and later obesity risk. *Adv Exp Med Biol,* 646.

KOUPIL, I. & TOIVANEN, P. 2007. Social and early-life determinants of overweight and obesity in 18-year-old Swedish men. *Int J Obes,* 32**,** 73-81.

KRAMER, M. S., GUO, T., PLATT, R. W., VANILOVICH, I., SEVKOVSKAYA, Z., DZIKOVICH, I., MICHAELSEN, K. F. & DEWEY, K. 2004. Feeding effects on growth during infancy. *J Pediatr,* 145**,** 600-5.

KRAMER, M. S., MATUSH, L., VANILOVICH, I., PLATT, R. W., BOGDANOVICH, N., SEVKOVSKAYA, Z., DZIKOVICH, I., SHISHKO, G., COLLET, J.-P., MARTIN, R. M., SMITH, G. D., GILLMAN, M. W., CHALMERS, B., HODNETT, E. & SHAPIRO, S. 2009. A

Randomized Breast-feeding Promotion Intervention Did Not Reduce Child Obesity in Belarus. *The Journal of Nutrition,* 139**,** 417S-421S.
KUH, D., HARDY, R., CHATURVEDI, N. & WADSWORTH, M. E. J. 2002. Birth weight, childhood growth and abdominal obesity in adult life. *International journal of obesity,* 26**,** 40.
KUH, D. J. & COOPER, C. 1992. Physical activity at 36 years: patterns and childhood predictors in a longitudinal study. *J Epidemiol Community Health,* 46**,** 114-9.
KUHLE, S., TONG, O. S. & WOOLCOTT, C. G. 2015. Association between caesarean section and childhood obesity: a systematic review and meta-analysis. *Obesity Reviews,* 16**,** 295-303.
KUHLE, S. & WOOLCOTT, C. G. 2017. Caesarean section is associated with offspring obesity in childhood and young adulthood. *BMJ Evidence-Based Medicine,* 22**,** 111-111.
KULESHOV, V., XIE, D., CHEN, R., PUSHKAREV, D., MA, Z., BLAUWKAMP, T., KERTESZ, M. & SNYDER, M. 2014. Whole-genome haplotyping using long reads and statistical methods. *Nat Biotechnol,* 32**,** 261-266.
KÜPERS, L. K., MONNEREAU, C., SHARP, G. C., YOUSEFI, P., SALAS, L. A., GHANTOUS, A., PAGE, C. M., REESE, S. E., WILCOX, A. J., CZAMARA, D., STARLING, A. P., NOVOLOACA, A., LENT, S., ROY, R., HOYO, C., BRETON, C. V., ALLARD, C., JUST, A. C., BAKULSKI, K. M., HOLLOWAY, J. W., EVERSON, T. M., XU, C.-J., HUANG, R.-C., VAN DER PLAAT, D. A., WIELSCHER, M., MERID, S. K., ULLEMAR, V., REZWAN, F. I., LAHTI, J., VAN DONGEN, J., LANGIE, S. A. S., RICHARDSON, T. G., MAGNUS, M. C., NOHR, E. A., XU, Z., DUIJTS, L., ZHAO, S., ZHANG, W., PLUSQUIN, M., DEMEO, D. L., SOLOMON, O., HEIMOVAARA, J. H., JIMA, D. D., GAO, L., BUSTAMANTE, M., PERRON, P., WRIGHT, R. O., HERTZ-PICCIOTTO, I., ZHANG, H., KARAGAS, M. R., GEHRING, U., MARSIT, C. J., BEILIN, L. J., VONK, J. M., JARVELIN, M.-R., BERGSTRÖM, A., ÖRTQVIST, A. K., EWART, S., VILLA, P. M., MOORE, S. E., WILLEMSEN, G., STANDAERT, A. R. L., HÅBERG, S. E., SØRENSEN, T. I. A., TAYLOR, J. A., RÄIKKÖNEN, K., YANG, I. V., KECHRIS, K., NAWROT, T. S., SILVER, M. J., GONG, Y. Y., RICHIARDI, L., KOGEVINAS, M., LITONJUA, A. A., ESKENAZI, B., HUEN, K., MBAREK, H., MAGUIRE, R. L., DWYER, T., VRIJHEID, M., BOUCHARD, L., BACCARELLI, A. A., CROEN, L. A., KARMAUS, W., ANDERSON, D., DE VRIES, M., SEBERT, S., KERE, J., KARLSSON, R., ARSHAD, S. H., HÄMÄLÄINEN, E., ROUTLEDGE, M. N., BOOMSMA, D. I., FEINBERG, A. P., NEWSCHAFFER, C. J., GOVARTS, E., MOISSE, M., FALLIN, M. D., MELÉN, E., PRENTICE, A. M., et al. 2019. Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nature Communications,* 10**,** 1893.
KURDYUKOV, S. & BULLOCK, M. 2016. DNA Methylation Analysis: Choosing the Right Method. *Biology,* 5**,** 3.
KVAAVIK, E., TELL, G. S. & KLEPP, K.-I. 2003. Predictors and Tracking of Body Mass Index From Adolescence Into Adulthood: Follow-up of 18 to 20 Years in the Oslo Youth Study. *Archives of Pediatrics & Adolescent Medicine,* 157**,** 1212-1218.
LABONTÉ, B., SUDERMAN, M., MAUSSION, G., NAVARO, L., YERKO, V., MAHAR, I., BUREAU, A., MECHAWAR, N., SZYF, M., MEANEY, M. J. & TURECKI, G. 2012. Genome-wide Epigenetic Regulation by Early-Life TraumaGenome Epigenetic Regulation by Early-Life Trauma. *JAMA Psychiatry,* 69**,** 722-731.
LAHTINEN, A., PUTTONEN, S., VANTTOLA, P., VIITASALO, K., SULKAVA, S., PERVJAKOVA, N., JOENSUU, A., SALO, P., TOIVOLA, A., HÄRMÄ, M., MILANI, L., PEROLA, M. & PAUNIO, T. 2019. A distinctive DNA methylation pattern in insufficient sleep. *Scientific Reports,* 9**,** 1193.

LAKDAWALLA, D. & PHILIPSON, T. 2009. The growth of obesity and technological change. *Economics and Human Biology,* 7**,** 283-293.
LAKE, A., TOWNSHEND, T. G. & ALVANIDES, S. 2011. *Obesogenic environments: complexities, perceptions and objective measures*, John Wiley & Sons.
LAM, L. L., EMBERLY, E., FRASER, H. B., NEUMANN, S. M., CHEN, E., MILLER, G. E. & KOBOR, M. S. 2012. Factors underlying variable DNA methylation in a human community cohort. *Proceedings of the National Academy of Sciences,* 109**,** 17253.
LAMB, M. M., DABELEA, D., YIN, X., OGDEN, L. G., KLINGENSMITH, G. J., REWERS, M. & NORRIS, J. M. 2010. Early-life predictors of higher body mass index in healthy children. *Annals of Nutrition and Metabolism,* 56**,** 16-22.
LAMONT, D., PARKER, L., WHITE, M., UNWIN, N., BENNETT, S. M. A., COHEN, M., RICHARDSON, D., DICKINSON, H. O., ADAMSON, A., ALBERTI, K. G. M. & CRAFT, A. W. 2000. Risk of cardiovascular disease measured by carotid intima-media thickness at age 49-51: lifecourse study. *BMJ,* 320**,** 273-278.
LANIGAN, J. 2018. Prevention of overweight and obesity in early life. *Proceedings of the Nutrition Society,* 77**,** 247-256.
LAURENT, L., WONG, E., LI, G., HUYNH, T., TSIRIGOS, A., ONG, C. T., LOW, H. M., SUNG, K. W. K., RIGOUTSOS, I. & LORING, J. 2010. Dynamic changes in the human methylome during differentiation. *Genome research,* 20**,** 320-331.
LEDDY, M. A., POWER, M. L. & SCHULKIN, J. 2008. The Impact of Maternal Obesity on Maternal and Fetal Health. *Reviews in Obstetrics and Gynecology,* 1**,** 170-178.
LEE, S. Y. & GALLAGHER, D. 2008. Assessment methods in human body composition. *Current opinion in clinical nutrition and metabolic care,* 11**,** 566.
LEE, Y. S. 2009. Consequences of childhood obesity. *Ann Acad Med Singapore,* 38**,** 75-7.
LEEK, J. T., SCHARPF, R. B., BRAVO, H. C., SIMCHA, D., LANGMEAD, B., JOHNSON, W. E., GEMAN, D., BAGGERLY, K. & IRIZARRY, R. A. 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics,* 11**,** 733.
LEEK, J. T. & STOREY, J. D. 2007. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. *PLOS Genetics,* 3**,** e161.
LEVINE, J. A. 2005. Measurement of energy expenditure. *Public health nutrition,* 8**,** 1123-1132.
LI, D.-K., CHEN, H., FERBER, J. & ODOULI, R. 2017a. Infection and antibiotic use in infancy and risk of childhood obesity: a longitudinal birth cohort study. *The lancet Diabetes & endocrinology,* 5**,** 18-25.
LI, H. T., ZHOU, Y. B. & LIU, J. M. 2012. The impact of cesarean section on offspring overweight and obesity: a systematic review and meta-analysis. *International Journal Of Obesity,* 37**,** 893.
LI, J., OLSEN, J., VESTERGAARD, M., OBEL, C., BAKER, J. L. & SØRENSEN, T. I. A. 2010. Prenatal Stress Exposure Related to Maternal Bereavement and Risk of Childhood Overweight. *PLoS ONE,* 5**,** e11896.
LI, L.-C. & DAHIYA, R. 2002. MethPrimer: designing primers for methylation PCRs. *Bioinformatics,* 18**,** 1427-1431.
LI, L., HARDY, R., KUH, D., LO CONTE, R. & POWER, C. 2008. Child-to-adult body mass index and height trajectories: a comparison of 2 British birth cohorts. *Am J Epidemiol,* 168**,** 1008-15.
LI, W., LIU, Q., DENG, X., CHEN, Y., LIU, S. & STORY, M. 2017b. Association between Obesity and Puberty Timing: A Systematic Review and Meta-Analysis. *International journal of environmental research and public health,* 14**,** 1266.

LIEBERMAN, E., GREMY, I., LANG, J. M. & COHEN, A. P. 1994. Low Birthweight at Term and the Timing of Fetal Exposure to Maternal Smoking. *American Journal of Public Health,* 84**,** 1127-1131.

LIEW, C.-C., MA, J., TANG, H.-C., ZHENG, R. & DEMPSEY, A. A. 2006. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *Journal of Laboratory and Clinical Medicine,* 147**,** 126-132.

LINDGREN, G., STRANDELL, A., COLE, T., HEALY, M. & TANNER, J. 1995. Swedish population reference standards for height, weight and body mass index attained at 6 to 16 years (girls) or 19 years (boys). *Acta Paediatrica,* 84**,** 1019-1028.

LIU, Y., ARYEE, M. J., PADYUKOV, L., FALLIN, M. D., HESSELBERG, E., RUNARSSON, A., REINIUS, L., ACEVEDO, N., TAUB, M., RONNINGER, M., SHCHETYNSKY, K., SCHEYNIUS, A., KERE, J., ALFREDSSON, L., KLARESKOG, L., EKSTRÖM, T. J. & FEINBERG, A. P. 2013a. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature Biotechnology,* 31**,** 142.

LIU, Y., HOYO, C., MURPHY, S., HUANG, Z., OVERCASH, F., THOMPSON, J., BROWN, H. & MURTHA, A. P. 2013b. DNA methylation at imprint regulatory regions in preterm birth and infection. *American journal of obstetrics and gynecology,* 208**,** 395.e1-395.e3957.

LIU, Y. J., GUO, Y. F., ZHANG, L. S., PEI, Y. F., YU, N., YU, P., PAPASIAN, C. J. & DENG, H. W. 2010. Biological pathway-based genome-wide association analysis identified the vasoactive intestinal peptide (VIP) pathway important for obesity. *Obesity,* 18**,** 2339-2346.

LLEWELLYN, A., SIMMONDS, M., OWEN, C. G. & WOOLACOTT, N. 2016. Childhood obesity as a predictor of morbidity in adulthood: a systematic review and meta-analysis. *Obesity Reviews,* 17**,** 56-67.

LOBSTEIN, T., BAUR, L. & UAUY, R. 2004. Obesity in children and young people: a crisis in public health. *Obes Rev,* 5 Suppl 1**,** 4-104.

LOCKE, A. E., KAHALI, B., BERNDT, S. I., JUSTICE, A. E., PERS, T. H., DAY, F. R., POWELL, C., VEDANTAM, S., BUCHKOVICH, M. L. & YANG, J. 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature,* 518**,** 197.

LOGUE, M. W., SMITH, A. K., WOLF, E. J., MANIATES, H., STONE, A., SCHICHMAN, S. A., MCGLINCHEY, R. E., MILBERG, W. & MILLER, M. W. 2017. The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. *Epigenomics,* 9**,** 1363-1371.

LOOPSTRA, R. & LALOR, D. 2017. Financial insecurity, food insecurity, and disability: The profile of people receiving emergency food assistance from The Trussell Trust Foodbank Network in Britain. *London: the Trussell trust*.

LOOS, R. J. & YEO, G. S. 2014. The bigger picture of FTO—the first GWAS-identified obesity gene. *Nature Reviews Endocrinology,* 10**,** 51.

LOVASI, G. S., HUTSON, M. A., GUERRA, M. & NECKERMAN, K. M. 2009. Built Environments and Obesity in Disadvantaged Populations. *Epidemiologic Reviews,* 31**,** 7-20.

LOZOFF, B., WOLF, A. W. & DAVIS, N. S. 1985. Sleep problems seen in pediatric practice. *Pediatrics,* 75**,** 477-83.

MACHIELA, M. J. & CHANOCK, S. J. 2015. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics,* 31**,** 3555-7.

MAIER, H. E., LOPEZ, R., SANCHEZ, N., NG, S., GRESH, L., OJEDA, S., BURGER-CALDERON, R., KUAN, G., HARRIS, E., BALMASEDA, A. & GORDON, A. 2018. Obesity Increases the

Duration of Influenza A Virus Shedding in Adults. *The Journal of Infectious Diseases,* 218**,** 1378-1382.
MAJUMDER, S., TAYLOR, W. R., FOOTE, P. H., BERGER, C. K., WU, C. W., YAB, T. C., MAHONEY, D. W., BAMLET, W. R., BURGER, K. N., POSTIER, N., DOERING, K., LIDGARD, G. P., ALLAWI, H. T., KISIEL, J. B., PETERSEN, G. M., CHARI, S. T. & AHLQUIST, D. A. 2019. Mo1370 - Pancreatic Cancer Detection by Plasma Assay of Novel Methylated Dna Markers: A Case-Control Study. *Gastroenterology,* 156**,** S-754-S-755.
MAKSIMOVIC, J., GORDON, L. & OSHLACK, A. 2012. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome biology,* 13**,** R44.
MALASPINA, D., GILMAN, C. & KRANZ, T. M. 2015. Paternal age and mental health of offspring. *Fertility and sterility,* 103**,** 1392-1396.
MANN, K. D. 2017. *Whole grain intake, diet quality and cardio-metabolic health in two UK cohorts.* Newcastle University.
MARKUNAS, C. A., WILCOX, A. J., XU, Z., JOUBERT, B. R., HARLID, S., PANDURI, V., HÅBERG, S. E., NYSTAD, W., LONDON, S. J., SANDLER, D. P., LIE, R. T., WADE, P. A. & TAYLOR, J. A. 2016. Maternal Age at Delivery Is Associated with an Epigenetic Signature in Both Newborns and Adults. *PloS one,* 11**,** e0156361-e0156361.
MARMOT, M., ALLEN, J., GOLDBLATT, P., BOYCE, T., MCNEISH, D. & GRADY, M. 2010. Strategic review of health inequalities in England post 2010 (Marmot review). *Global health equity group, UCL research department of epidemiology and public health [Internet]*.
MARSEGLIA, L., MANTI, S., D'ANGELO, G., CUPPARI, C., SALPIETRO, V., FILIPPELLI, M., TROVATO, A., GITTO, E., SALPIETRO, C. & ARRIGO, T. 2015. Obesity and breastfeeding: The strength of association. *Women and Birth,* 28**,** 81-86.
MARSHALL, W. A. & TANNER, J. M. 1969. Variations in pattern of pubertal changes in girls. *Archives of disease in childhood,* 44**,** 291.
MARSHALL, W. A. & TANNER, J. M. 1970. Variations in the pattern of pubertal changes in boys. *Archives of disease in childhood,* 45**,** 13-23.
MARSIT, C. J., MACCANI, M. A., PADBURY, J. F. & LESTER, B. M. 2012. Placental 11-beta hydroxysteroid dehydrogenase methylation is associated with newborn growth and a measure of neurobehavioral outcome. *PloS one,* 7**,** e33794.
MARTIN, M. A. & SELA, D. A. 2013. Infant gut microbiota: developmental influences and health outcomes. *Building babies.* Springer.
MARTIN, R. M., KRAMER, M. S., PATEL, R., RIFAS-SHIMAN, S. L., THOMPSON, J., YANG, S., VILCHUCK, K., BOGDANOVICH, N., HAMEZA, M., TILLING, K. & OKEN, E. 2017. Effects of Promoting Long-term, Exclusive Breastfeeding on Adolescent Adiposity, Blood Pressure, and Growth Trajectories: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Pediatr,* 171**,** e170698.
MARTINO, D., KESPER, D. A., AMARASEKERA, M., HARB, H., RENZ, H. & PRESCOTT, S. 2014. Epigenetics in immune development and in allergic and autoimmune diseases. *Journal of Reproductive Immunology,* 104-105**,** 43-48.
MARTORELL-MARUGAN, J., GONZALEZ-RUMAYOR, V. & CARMONA-SAEZ, P. 2018. mCSEA: Detecting subtle differentially methylated regions. *bioRxiv***,** 293381.
MASSART, R., FREYBURGER, M., SUDERMAN, M., PAQUET, J., EL HELOU, J., BELANGER-NELSON, E., RACHALSKI, A., KOUMAR, O. C., CARRIER, J., SZYF, M. & MONGRAIN, V. 2014. The genome-wide landscape of DNA methylation and hydroxymethylation in

response to sleep deprivation impacts on synaptic plasticity genes. *Translational Psychiatry,* 4.
MAST, M., KORTZINGER, I., KONIG, E. & MULLER, M. J. 1998. Gender differences in fat mass of 5-7-year old children. *Int J Obes Relat Metab Disord,* 22**,** 878-84.
MAST, M., LANGNÄSE, K., LABITZKE, K., BRUSE, U., PREUSS, U. & MÜLLER, M. 2002. Use of BMI as a measure of overweight and obesity in a field study on 5–7 year old children. *European journal of nutrition,* 41**,** 61-67.
MASUKUME, G., O'NEILL, S. M., BAKER, P. N., KENNY, L. C., MORTON, S. M. B. & KHASHAN, A. S. 2018. The Impact of Caesarean Section on the Risk of Childhood Overweight and Obesity: New Evidence from a Contemporary Cohort Study. *Scientific reports,* 8**,** 15113-15113.
MATHERS, J. C. & MCKAY, J. A. 2009. Epigenetics - potential contribution to fetal programming. *Advances in experimental medicine and biology,* 646**,** 119-123.
MATSHA, T. E., KENGNE, A.-P., YAKO, Y. Y., HON, G. M., HASSAN, M. S. & ERASMUS, R. T. 2013. Optimal waist-to-height ratio values for cardiometabolic risk screening in an ethnically diverse sample of South African urban and rural school boys and girls. *PloS one,* 8**,** e71133-e71133.
MCANDREW, F., THOMPSON, J., FELLOWS, L., LARGE, A., SPEED, M. & RENFREW, M. J. 2012. Infant feeding survey 2010. *Leeds: Health and Social Care Information Centre*.
MCCARTHY, H. D. & ASHWELL, M. 2006. A study of central fatness using waist-to-height ratios in UK children and adolescents over two decades supports the simple message – 'keep your waist circumference to less than half your height'. *International Journal Of Obesity,* 30**,** 988.
MCERLEAN, P., FAVORETO, S., COSTA, F. F., SHEN, J., QURAISHI, J., BIYASHEVA, A., COOPER, J. J., SCHOLTENS, D. M., VANIN, E. F. & DE BONALDO, M. F. 2014. Human rhinovirus infection causes different DNA methylation changes in nasal epithelial cells from healthy and asthmatic subjects. *BMC medical genomics,* 7**,** 37.
MCGOWAN, P. O., SASAKI, A., D'ALESSIO, A. C., DYMOV, S., LABONTE, B., SZYF, M., TURECKI, G. & MEANEY, M. J. 2009. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci,* 12**,** 342-8.
MCGREGOR, K., BERNATSKY, S., COLMEGNA, I., HUDSON, M., PASTINEN, T., LABBE, A. & GREENWOOD, C. M. T. 2016. An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. *Genome Biology,* 17**,** 84.
MCKAY, J. A., XIE, L., MANUS, C., LANGIE, S. A., MAXWELL, R. J., FORD, D. & MATHERS, J. C. 2014. Metabolic effects of a high-fat diet post-weaning after low maternal dietary folate during pregnancy and lactation. *Mol Nutr Food Res,* 58**,** 1087-97.
MCLAREN, L. 2007a. Socioeconomic Status and Obesity. *Epidemiologic Reviews,* 29**,** 29-48.
MCLAREN, L. 2007b. Socioeconomic status and obesity. *Epidemiol Rev,* 29**,** 29-48.
MCLAREN, L. & KUH, D. 2004. Women's body dissatisfaction, social class, and social mobility. *Social Science & Medicine,* 58**,** 1575-1584.
MEANEY, M. J. & SZYF, M. 2005. Environmental programming of stress responses through DNA methylation: life at the interface between a dynamic environment and a fixed genome. *Dialogues in clinical neuroscience,* 7**,** 103-23.
MEREDITH-JONES, K. A., WILLIAMS, S. M. & TAYLOR, R. W. 2015. Bioelectrical impedance as a measure of change in body composition in young children. *Pediatr Obes,* 10**,** 252-9.
MERSKY, J. P., JANCZEWSKI, C. E. & TOPITZES, J. 2017. Rethinking the measurement of adversity: Moving toward second-generation research on adverse childhood experiences. *Child maltreatment,* 22**,** 58-68.

METCALF, B. S., HOSKING, J., FREMEAUX, A. E., JEFFERY, A. N., VOSS, L. D. & WILKIN, T. J. 2011. BMI was right all along: taller children really are fatter (implications of making childhood BMI independent of height) EarlyBird 48. *Int J Obes (Lond),* 35**,** 541-7.

MIFFLIN, M. D., ST JEOR, S. T., HILL, L. A., SCOTT, B. J., DAUGHERTY, S. A. & KOH, Y. O. 1990. A new predictive equation for resting energy expenditure in healthy individuals. *The American Journal of Clinical Nutrition,* 51**,** 241-7.

MIKESKA, T., FELSBERG, J., HEWITT, C. A. & DOBROVIC, A. 2011. Analysing DNA methylation using bisulphite pyrosequencing. *Epigenetics Protocols.* Springer.

MILL, J. & HEIJMANS, B. T. 2013. From promises to practical strategies in epigenetic epidemiology. *Nature Reviews Genetics,* 14**,** 585.

MILLION, M., LAGIER, J. C., YAHAV, D. & PAUL, M. 2013. Gut bacterial microbiota and obesity. *Clin Microbiol Infect,* 19**,** 305-13.

MIN, J., HEMANI, G., DAVEY SMITH, G., RELTON, C. L. & SUDERMAN, M. 2017. Meffil: efficient normalisation and analysis of very large DNA methylation samples. *bioRxiv*.

MINE, T., TANAKA, T., NAKASONE, T., ITOKAZU, T., YAMAGATA, Z. & NISHIWAKI, Y. 2017. Maternal smoking during pregnancy and rapid weight gain from birth to early infancy. *Journal of Epidemiology,* 27**,** 112-116.

MISCHKE, M. & PLOSCH, T. 2013. More than just a gut instinct-the potential interplay between a baby's nutrition, its gut microbiome, and the epigenome. *Am J Physiol Regul Integr Comp Physiol,* 304**,** R1065-9.

MONASTA, L., BATTY, G. D., CATTANEO, A., LUTJE, V., RONFANI, L., VAN LENTHE, F. J. & BRUG, J. 2010. Early-life determinants of overweight and obesity: a review of systematic reviews. *Obes Rev,* 11**,** 695-708.

MONTAGUE, C. T., FAROOQI, I. S., WHITEHEAD, J. P., SOOS, M. A., RAU, H., WAREHAM, N. J., SEWTER, C. P., DIGBY, J. E., MOHAMMED, S. N., HURST, J. A., CHEETHAM, C. H., EARLEY, A. R., BARNETT, A. H., PRINS, J. B. & O'RAHILLY, S. 1997. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature,* 387**,** 903-908.

MOORCROFT, K. E., MARSHALL, J. L. & MCCORMICK, F. M. 2011. Association between timing of introducing solid foods and obesity in infancy and childhood: A systematic review. *Maternal & Child Nutrition,* 7**,** 3-26.

MOORE, A. M., XU, Z., KOLLI, R. T., WHITE, A. J., SANDLER, D. P. & TAYLOR, J. A. 2019. Persistent epigenetic changes in adult daughters of older mothers. *Epigenetics,* 14**,** 467-476.

MOORE, S. E. 2016. Early life nutritional programming of health and disease in The Gambia. *Journal of Developmental Origins of Health and Disease,* 7**,** 123-131.

MORALES, E., GROOM, A., LAWLOR, D. A. & RELTON, C. L. 2014. DNA methylation signatures in cord blood associated with maternal gestational weight gain: results from the ALSPAC cohort. *BMC Res Notes,* 7**,** 278.

MORALES, E., RODRIGUEZ, A., VALVI, D., INIGUEZ, C., ESPLUGUES, A., VIOQUE, J., MARINA, L. S., JIMENEZ, A., ESPADA, M., DEHLI, C. R., FERNANDEZ-SOMOANO, A., VRIJHEID, M. & SUNYER, J. 2015. Deficit of vitamin D in pregnancy and growth and overweight in the offspring. *Int J Obes,* 39**,** 61-68.

MORTENSEN, L. H., DIDERICHSEN, F., ARNTZEN, A., GISSLER, M., CNATTINGIUS, S., SCHNOR, O., DAVEY-SMITH, G. & ANDERSEN, A. N. 2008. Social inequality in fetal growth: a comparative study of Denmark, Finland, Norway and Sweden in the period 1981–2000. *Journal of Epidemiology & Community Health,* 62**,** 325-331.

MOSTAZIR, M., JEFFERY, A., VOSS, L. & WILKIN, T. 2015. Childhood obesity: evidence for distinct early and late environmental determinants a 12-year longitudinal cohort study (EarlyBird 62). *Int J Obes,* 39**,** 1057-1062.

MUELLER, N. T., WHYATT, R., HOEPNER, L., OBERFIELD, S., DOMINGUEZ-BELLO, M. G., WIDEN, E., HASSOUN, A., PERERA, F. & RUNDLE, A. 2015a. Prenatal exposure to antibiotics, cesarean section and risk of childhood obesity. *International journal of obesity,* 39**,** 665.

MUELLER, N. T., WHYATT, R., HOEPNER, L., OBERFIELD, S., DOMINGUEZ-BELLO, M. G., WIDEN, E. M., HASSOUN, A., PERERA, F. & RUNDLE, A. 2015b. Prenatal exposure to antibiotics, cesarean section and risk of childhood obesity. *International Journal of Obesity,* 39**,** 665-670.

MUHLHAUSLER, B. S., GIBSON, R. A. & MAKRIDES, M. 2010. Effect of long-chain polyunsaturated fatty acid supplementation during pregnancy or lactation on infant and child body composition: a systematic review. *The American Journal of Clinical Nutrition,* 92**,** 857-863.

MULLIGAN, A. A., LUBEN, R. N., BHANIANI, A., PARRY-SMITH, D. J., O'CONNOR, L., KHAWAJA, A. P., FOROUHI, N. G., KHAW, K.-T., DICKINSON, A., WAREHAM, N. & NATOUR, J. 2014. A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability. *BMJ Open,* 4.

MUMBY, H. S., ELKS, C. E., LI, S., SHARP, S. J., KHAW, K.-T., LUBEN, R. N., WAREHAM, N. J., LOOS, R. J. F. & ONG, K. K. 2011. Mendelian Randomisation Study of Childhood BMI and Early Menarche. *Journal of obesity,* 2011**,** 180729-180729.

MYINT, P. K., KWOK, C. S., LUBEN, R. N., WAREHAM, N. J. & KHAW, K.-T. 2014. Body fat percentage, body mass index and waist-to-hip ratio as predictors of mortality and cardiovascular disease. *Heart,* 100**,** 1613.

MYRSKYLA, M. & FENELON, A. 2012. Maternal Age and Offspring Adult Health: Evidence From the Health and Retirement Study. *Demography,* 49**,** 1231-1257.

MYRSKYLÄ, M. & FENELON, A. 2012. Maternal Age and Offspring Adult Health: Evidence From the Health and Retirement Study. *Demography,* 49**,** 10.1007/s13524-012-0132-x.

NAMBIAR, S., HUGHES, I. & DAVIES, P. S. 2010. Developing waist-to-height ratio cut-offs to define overweight and obesity in children and adolescents. *Public Health Nutr,* 13**,** 1566-74.

NEEL, J. V. 1962. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *American journal of human genetics,* 14**,** 353.

NELSON, K. N., HUI, Q., RIMLAND, D., XU, K., FREIBERG, M. S., JUSTICE, A. C., MARCONI, V. C. & SUN, Y. V. 2017. Identification of HIV infection-related DNA methylation sites and advanced epigenetic aging in HIV-positive, treatment-naive U.S. veterans. *AIDS (London, England),* 31**,** 571-575.

NEWMAN, A. M., LIU, C. L., GREEN, M. R., GENTLES, A. J., FENG, W., XU, Y., HOANG, C. D., DIEHN, M. & ALIZADEH, A. A. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nature methods,* 12**,** 453.

NEWNHAM, J. P. 1991. Smoking in pregnancy. *Fetal and Maternal Medicine Review,* 3**,** 115-132.

NHS DIGITAL 2016. National Child Measurement Programme, England, 2015/16 school year.

NHS DIGITAL 2017. National Child Measurement Programme England 2016/7 school year,. 19/10/2017 ed. https://files.digital.nhs.uk/publication/j/n/nati-chil-meas-prog-eng-2016-2017-rep.pdf.

NILSSON, E. K., BOSTROM, A. E., MWINYI, J. & SCHIOTH, H. B. 2016. Epigenomics of Total Acute Sleep Deprivation in Relation to Genome-Wide DNA Methylation Profiles and RNA Expression. *Omics-a Journal of Integrative Biology,* 20**,** 334-342.

NISHITANI, S., PARETS, S. E., HAAS, B. W. & SMITH, A. K. 2018. DNA methylation analysis from saliva samples for epidemiological studies. *Epigenetics,* 13**,** 352-362.

NOO. 2015. *UK and Ireland prevalence and trends* [Online]. Available: http://www.noo.org.uk/NOO_about_obesity/inequalities#d6888 [Accessed 31/03/2016].

NORRIS, T., BANN, D., HARDY, R. & JOHNSON, W. 2019. Socioeconomic inequalities in childhood-to-adulthood BMI tracking in three British birth cohorts. *International Journal of Obesity*.

NÖTHLING, J., MALAN-MÜLLER, S., ABRAHAMS, N., JOANNA HEMMINGS, S. M. & SEEDAT, S. 2019. Epigenetic alterations associated with childhood trauma and adult mental health outcomes: A systematic review. *World Journal of Biological Psychiatry*.

NYBO ANDERSEN, A.-M. & URHOJ, S. K. 2017. Is advanced paternal age a health risk for the offspring? *Fertility and Sterility,* 107**,** 312-318.

O'NEILL, D. 2015. Measuring obesity in the absence of a gold standard. *Econ Hum Biol,* 17**,** 116-28.

OAKLEY, L. L., RENFREW, M. J., KURINCZUK, J. J. & QUIGLEY, M. A. 2013. Factors associated with breastfeeding in England: an analysis by primary care trust. *BMJ Open,* 3.

OBERMANN-BORST, S. A., EILERS, P. H. C., TOBI, E. W., DE JONG, F. H., SLAGBOOM, P. E., HEIJMANS, B. T. & STEEGERS-THEUNISSEN, R. P. M. 2013. Duration of breastfeeding and gender are associated with methylation of the LEPTIN gene in very young children. *Pediatric Research,* 74**,** 344.

OBERMANN-BORST, S. A., HEIJMANS, B. T., EILERS, P. H., TOBI, E. W., STEEGERS, E. A., SLAGBOOM, P. E. & STEEGERS-THEUNISSEN, R. P. 2012. Periconception maternal smoking and low education are associated with methylation of INSIGF in children at the age of 17 months. *J Dev Orig Health Dis,* 3**,** 315-20.

ODOM, G. J., MALLIK, S., CHEN, X., WANG, L., GAO, Z. & GOMEZ, L. 2018. An evaluation of supervised methods for identifying differentially methylated regions in Illumina methylation arrays.

OFFICE FOR NATIONAL STATISTICS 2017. Birth characteristics in England and Wales: 2016. *Office for National Statistics*.

OH, Y. M., KWON, Y. E., KIM, J. M., BAE, S. J., LEE, B. K., YOO, S. J., CHUNG, C. H., DESHAIES, R. J. & SEOL, J. H. 2009. Chfr is linked to tumour metastasis through the downregulation of HDAC1. *Nature Cell Biology,* 11**,** 295.

OKEN, E., LEVITAN, E. B. & GILLMAN, M. W. 2008. Maternal smoking during pregnancy and child overweight: systematic review and meta-analysis. *International Journal of Obesity,* 32**,** 201-210.

ONG, K., KRATZSCH, J., KIESS, W., DUNGER, D. & TEAM, A. S. 2002a. Circulating IGF-I levels in childhood are related to both current body composition and early postnatal growth rate. *The Journal of Clinical Endocrinology & Metabolism,* 87**,** 1041-1044.

ONG, K. K. 2006. Size at birth, postnatal growth and risk of obesity. *Hormone Research in Paediatrics,* 65**,** 65-69.

ONG, K. K. & LOOS, R. J. 2006a. Rapid infancy weight gain and subsequent obesity: systematic reviews and hopeful suggestions. *Acta Paediatr,* 95.

ONG, K. K. & LOOS, R. J. F. 2006b. Rapid infancy weight gain and subsequent obesity: Systematic reviews and hopeful suggestions. *Acta Pædiatrica,* 95**,** 904-908.

ONG, K. K., PREECE, M. A., EMMETT, P. M., AHMED, M. L. & DUNGER, D. B. 2002b. Size at birth and early childhood growth in relation to maternal smoking, parity and infant breast-feeding: longitudinal birth cohort study and analysis. *Pediatr Res,* 52.

ONG, K. K. L., AHMED, M. L., EMMETT, P. M., PREECE, M. A. & DUNGER, D. B. 2000. Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. *BMJ,* 320**,** 967-971.

ONG, K. K. L., PREECE, M. A., EMMETT, P. M., AHMED, M. L. & DUNGER, D. B. 2002c. Size at Birth and Early Childhood Growth in Relation to Maternal Smoking, Parity and Infant Breast-Feeding: Longitudinal Birth Cohort Study and Analysis. *Pediatric Research,* 52**,** 863.

ONS 2014. Adult smoking habits in Great Britain: 2014. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2014: Office for National Statistics.

ONS 2015. Birth characteristics in England and Wales: 2015. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthcharacteristicsinenglandandwales/2015: Office for National Statistics.

ORTIZ, L., NAKAMURA, B., LI, X., BLUMBERG, B. & LUDERER, U. 2013. In Utero Exposure to Benzo[a]pyrene Increases Adiposity and Causes Hepatic Steatosis in Female Mice, and Glutathione Deficiency Is Protective. *Toxicology letters,* 223**,** 10.1016/j.toxlet.2013.09.017.

OWEN, C. G., MARTIN, R. M., WHINCUP, P. H., DAVEY-SMITH, G., GILLMAN, M. W. & COOK, D. G. 2005a. The effect of breastfeeding on mean body mass index throughout life: a quantitative review of published and unpublished observational evidence(1-3). *American Journal of Clinical Nutrition,* 82**,** 1298-1307.

OWEN, C. G., MARTIN, R. M., WHINCUP, P. H., SMITH, G. D. & COOK, D. G. 2005b. Effect of infant feeding on the risk of obesity across the life course: a quantitative review of published evidence. *Pediatrics,* 115**,** 1367-77.

PAINTER, R. C., OSMOND, C., GLUCKMAN, P., HANSON, M., PHILLIPS, D. I. & ROSEBOOM, T. J. 2008. Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life. *Bjog,* 115**,** 1243-9.

PAPAS, M. A., ALBERG, A. J., EWING, R., HELZLSOUER, K. J., GARY, T. L. & KLASSEN, A. C. 2007. The built environment and obesity. *Epidemiologic Reviews,* 29**,** 129-143.

PARKER, J. D., SCHOENDORF, K. C. & KIELY, J. L. 1994. Associations between measures of socioeconomic status and low birth weight, small for gestational age, and premature delivery in the United States. *Annals of epidemiology,* 4**,** 271-278.

PARKINSON, K. N., PEARCE, M. S., DALE, A., REILLY, J. J., DREWETT, R. F., WRIGHT, C. M., RELTON, C. L., MCARDLE, P., LE COUTEUR, A. S. & ADAMSON, A. J. 2011. Cohort profile: the gateshead millennium study. *International journal of epidemiology,* 40**,** 308-317.

PARKINSON, K. N., REILLY, J. J., BASTERFIELD, L., REILLY, J. K., JANSSEN, X., JONES, A. R., CUTLER, L. R., LE COUTEUR, A. & ADAMSON, A. J. 2017. Mothers' perceptions of child weight status and the subsequent weight gain of their children: a population-based longitudinal study. *Int J Obes (Lond),* 41**,** 801-806.

PARMELEE, A. H., WENNER, W. H. & SCHULZ, H. R. 1964. Infant sleep patterns: from birth to 16 weeks of age. *The Journal of pediatrics,* 65**,** 576-582.

PARSONS, T. J., POWER, C., LOGAN, S. & SUMMERBELT, C. D. 1999. Childhood predictors of adult obesity: a systematic review. *International Journal of Obesity,* 23.

PASCHOS, K. & ALLDAY, M. J. 2010. Epigenetic reprogramming of host genes in viral and microbial pathogenesis. *Trends in Microbiology,* 18**,** 439-447.
PATEL, S. R. & HU, F. B. 2008. Short sleep duration and weight gain: a systematic review. *Obesity (Silver Spring),* 16**,** 643-53.
PATTERSON, K., MOLLOY, L., QU, W. & CLARK, S. 2011. DNA methylation: bisulphite modification and analysis. *Journal of visualized experiments: JoVE*.
PAUWELS, S., SYMONS, L., VANAUTGAERDEN, E.-L., GHOSH, M., DUCA, R. C., BEKAERT, B., FRESON, K., HUYBRECHTS, I., LANGIE, S. A. & KOPPEN, G. 2019. The Influence of the Duration of Breastfeeding on the Infant's Metabolic Epigenome. *Nutrients,* 11**,** 1408.
PEARCE, J. & LANGLEY-EVANS, S. C. 2013. The types of food introduced during complementary feeding and risk of childhood obesity: a systematic review. *International Journal Of Obesity,* 37**,** 477.
PEARCE, M. S., BASTERFIELD, L., MANN, K. D., PARKINSON, K. N., ADAMSON, A. J. & JOHN, J. R. O. B. O. T. G. M. S. C. T. 2012a. Early Predictors of Objectively Measured Physical Activity and Sedentary Behaviour in 8–10 Year Old Children: The Gateshead Millennium Study. *PLOS ONE,* 7**,** e37975.
PEARCE, M. S., MANN, K. D., MARTIN-RUIZ, C., PARKER, L., WHITE, M., VON ZGLINICKI, T. & ADAMS, J. 2012b. Childhood Growth, IQ and Education as Predictors of White Blood Cell Telomere Length at Age 49–51 Years: The Newcastle Thousand Families Study. *PLOS ONE,* 7**,** e40116.
PEARCE, M. S., MCCONNELL, J. C., POTTER, C., BARRETT, L. M., PARKER, L., MATHERS, J. C. & RELTON, C. L. 2012c. Global LINE-1 DNA methylation is associated with blood glycaemic and lipid profiles. *International journal of epidemiology,* 41**,** 210-217.
PEARCE, M. S., UNWIN, N. C., PARKER, L. & CRAFT, A. W. 2009. Cohort profile: the Newcastle Thousand Families 1947 birth cohort. *International journal of epidemiology,* 38**,** 932-937.
PENPRAZE, V., REILLY, J. J., MACLEAN, C. M., MONTGOMERY, C., KELLY, L. A., PATON, J. Y., AITCHISON, T. & GRANT, S. 2006. Monitoring of physical activity in young children: how much is enough? *Pediatric Exercise Science,* 18**,** 483-491.
PÉREZ-ESCAMILLA, R. & KAC, G. 2013. Childhood obesity prevention: a life-course framework. *International journal of obesity supplements,* 3**,** S3-S5.
PETERS, T., PETERS, M. T., BIOCVIEWS DIFFERENTIALMETHYLATION, G., GENOMEANNOTATION, D., ONECHANNEL, T. & MULTIPLECOMPARISON, Q. 2016. Package 'DMRcate'.
PETERS, T. J., BUCKLEY, M. J., STATHAM, A. L., PIDSLEY, R., SAMARAS, K., V LORD, R., CLARK, S. J. & MOLLOY, P. L. 2015. De novo identification of differentially methylated regions in the human genome. *Epigenetics & Chromatin,* 8**,** 6.
PETTITT, D. J., ALECK, K. A., BAIRD, H. R., CARRAHER, M. J., BENNETT, P. H. & KNOWLER, W. C. 1988. Congenital susceptibility to NIDDM: role of intrauterine environment. *Diabetes,* 37**,** 622-628.
PETTITT, D. J., BAIRD, H. R., ALECK, K. A., BENNETT, P. H. & KNOWLER, W. C. 1983. Excessive obesity in offspring of Pima Indian women with diabetes during pregnancy. *New England Journal of Medicine,* 308**,** 242-245.
PHE 2013. The Health Survey for England - 2012 trend tables. London: Health and Social Care Information Centre.
PHE PUBLISHING. 2018. *Trends in childhood BMI between 2006/07 and 2016/17* [Online]. Available: https://app.box.com/s/og3q86aqejc99okxe9xyvpfvo21xai21/file/306723044116 [Accessed].

PIDSLEY, R., Y WONG, C. C., VOLTA, M., LUNNON, K., MILL, J. & SCHALKWYK, L. C. 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics,* 14**,** 293.

PIDSLEY, R., ZOTENKO, E., PETERS, T. J., LAWRENCE, M. G., RISBRIDGER, G. P., MOLLOY, P., VAN DJIK, S., MUHLHAUSLER, B., STIRZAKER, C. & CLARK, S. J. 2016. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome biology,* 17**,** 208.

PIKHOLZ, C., SWINBURN, B. & METCALF, P. 2004. Under-reporting of energy intake in the 1997 National Nutrition Survey. *N Z Med J,* 117**,** U1079.

PISCHON, T., BOEING, H., HOFFMANN, K., BERGMANN, M., SCHULZE, M. B., OVERVAD, K., VAN DER SCHOUW, Y. T., SPENCER, E., MOONS, K. G., TJONNELAND, A., HALKJAER, J., JENSEN, M. K., STEGGER, J., CLAVEL-CHAPELON, F., BOUTRON-RUAULT, M. C., CHAJES, V., LINSEISEN, J., KAAKS, R., TRICHOPOULOU, A., TRICHOPOULOS, D., BAMIA, C., SIERI, S., PALLI, D., TUMINO, R., VINEIS, P., PANICO, S., PEETERS, P. H., MAY, A. M., BUENO-DE-MESQUITA, H. B., VAN DUIJNHOVEN, F. J., HALLMANS, G., WEINEHALL, L., MANJER, J., HEDBLAD, B., LUND, E., AGUDO, A., ARRIOLA, L., BARRICARTE, A., NAVARRO, C., MARTINEZ, C., QUIROS, J. R., KEY, T., BINGHAM, S., KHAW, K. T., BOFFETTA, P., JENAB, M., FERRARI, P. & RIBOLI, E. 2008. General and abdominal adiposity and risk of death in Europe. *N Engl J Med,* 359**,** 2105-20.

PLAGEMANN, A., HARDER, T., FRANKE, K. & KOHLHOFF, R. 2002. Long-Term Impact of Neonatal Breast-Feeding on Body Weight and Glucose Tolerance in Children of Diabetic Mothers. *Diabetes Care,* 25**,** 16-22.

PODSAKOFF, P. M., MACKENZIE, S. B. & PODSAKOFF, N. P. 2011. Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology,* 63**,** 539-569.

POPKIN, B. M. 2001. The nutrition transition and obesity in the developing world. *J Nutr,* 131**,** 871s-873s.

POPKIN, B. M., ADAIR, L. S. & NG, S. W. 2012. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev,* 70**,** 3-21.

PORTHA, B., CHAVEY, A. & MOVASSAT, J. 2011. Early-Life Origins of Type 2 Diabetes: Fetal Programming of the Beta-Cell Mass. *Experimental Diabetes Research,* 2011**,** 16.

POULSEN, M. N., POLLAK, J., BAILEY-DAVIS, L., HIRSCH, A. G., GLASS, T. A. & SCHWARTZ, B. S. 2017. Associations of prenatal and childhood antibiotic use with child body mass index at age 3 years. *Obesity,* 25**,** 438-444.

POULTON, R., CASPI, A., MILNE, B. J., THOMSON, W. M., TAYLOR, A., SEARS, M. R. & MOFFITT, T. E. 2002. Association between children's experience of socioeconomic disadvantage and adult health: a life-course study. *The lancet,* 360**,** 1640-1645.

POWELL, J. E., HENDERS, A. K., MCRAE, A. F., CARACELLA, A., SMITH, S., WRIGHT, M. J., WHITFIELD, J. B., DERMITZAKIS, E. T., MARTIN, N. G. & VISSCHER, P. M. 2012. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One,* 7**,** e35430.

POWER, C., LI, L., MANOR, O. & DAVEY SMITH, G. 2003. Combination of low birth weight and high adult body mass index: at what age is it established and what are its determinants? *Journal of Epidemiology and Community Health,* 57**,** 969-973.

POWER, C., MANOR, O. & MATTHEWS, S. 1999. The duration and timing of exposure: effects of socioeconomic environment on adult health. *American Journal of Public Health,* 89**,** 1059-1065.

PREGIBON, D. J. 1980. Data analytic methods for generalized linear models.

PRENTICE-DUNN, H. & PRENTICE-DUNN, S. 2012. Physical activity, sedentary behavior, and childhood obesity: A review of cross-sectional studies. *Psychology, Health & Medicine,* 17**,** 255-273.
PRENTICE, A. M. 1998. Body mass index standards for children: Are useful for clinicians but not yet for epidemiologists. British Medical Journal Publishing Group.
PRENTICE, A. M. & JEBB, S. A. 2001. Beyond body mass index. *Obesity Reviews,* 2**,** 141-147.
PROSPECTIVE STUDIES, C. 2009. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *The Lancet,* 373**,** 1083-1096.
PROSPECTIVE STUDIES COLLABORATION 2009. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *The Lancet,* 373**,** 1083-1096.
PUYAU, M. R., ADOLPH, A. L., VOHRA, F. A. & BUTTE, N. F. 2002. Validation and calibration of physical activity monitors in children. *Obesity research,* 10**,** 150-157.
QIAGEN 2010. Pyrosequencing – the synergy of sequencing and quantification. https://www.qiagen.com/gb/resources/.
RADTKE, K. M., RUF, M., GUNTER, H. M., DOHRMANN, K., SCHAUER, M., MEYER, A. & ELBERT, T. 2011. Transgenerational impact of intimate partner violence on methylation in the promoter of the glucocorticoid receptor. *Transl Psychiatry,* 1**,** e21.
RAMPAZZO, C., GALLINARO, L., MILANESI, E., FRIGIMELICA, E., REICHARD, P. & BIANCHI, V. 2000. A deoxyribonucleotidase in mitochondria: involvement in regulation of dNTP pools and possible link to genetic disease. *Proceedings of the National Academy of Sciences,* 97**,** 8239-8244.
RAMSEY, J. B. 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological),* 31**,** 350-371.
RASMUSSEN, S. H., SHRESTHA, S., BJERREGAARD, L. G., ÄNGQUIST, L. H., BAKER, J. L., JESS, T. & ALLIN, K. H. 2018. Antibiotic exposure in early life and childhood overweight and obesity: A systematic review and meta-analysis. *Diabetes, Obesity and Metabolism,* 20**,** 1508-1514.
RATHNAYAKE, K., SATCHITHANANTHAN, A., MAHAMITHAWA, S. & JAYAWARDENA, R. 2013. Early life predictors of preschool overweight and obesity: a case-control study in Sri Lanka. *BMC Public Health,* 13**,** 994.
RAVELLI, A. C. J., VAN DER MEULEN, J. H. P., MICHELS, R. P. J., OSMOND, C., BARKER, D. J. P., HALES, C. N. & BLEKER, O. P. 1998. Glucose tolerance in adults after prenatal exposure to famine. *Lancet,* 351**,** 173-177.
RAVELLI, A. C. J., VAN DER MEULEN, J. H. P., OSMOND, C., BARKER, D. J. P. & BLEKER, O. P. 1999. Obesity at the age of 50 y in men and women exposed to famine prenatally. *American Journal of Clinical Nutrition,* 70**,** 811-816.
RAYCO-SOLON, P., FULFORD, A. J. & PRENTICE, A. M. 2005. Differential effects of seasonality on preterm birth and intrauterine growth restriction in rural Africans. *Am J Clin Nutr,* 81**,** 134-9.
RAYFIELD, S. & PLUGGE, E. 2017. Systematic review and meta-analysis of the association between maternal smoking in pregnancy and childhood overweight and obesity. *J Epidemiol Community Health,* 71**,** 162-173.
REDDON, H., PATEL, Y., TURCOTTE, M., PIGEYRE, M. & MEYRE, D. 2018. Revisiting the evolutionary origins of obesity: lazy versus peppy-thrifty genotype hypothesis. *Obesity Reviews,* 0.
REDSELL, S. A., EDMONDS, B., SWIFT, J. A., SIRIWARDENA, A. N., WENG, S., NATHAN, D. & GLAZEBROOK, C. 2016. Systematic review of randomised controlled trials of

interventions that aim to reduce the risk, either directly or indirectly, of overweight and obesity in infancy and early childhood. *Maternal & child nutrition,* 12**,** 24-38.
REIFSNYDER, P. C., CHURCHILL, G. & LEITER, E. H. 2000. Maternal Environment and Genotype Interact to Establish Diabesity in Mice. *Genome Research,* 10**,** 1568-1578.
REILLY, J. J., ARMSTRONG, J., DOROSTY, A. R., EMMETT, P. M., NESS, A., ROGERS, I., STEER, C. & SHERRIFF, A. 2005a. Early life risk factors for obesity in childhood: cohort study. *BMJ,* 330.
REILLY, J. J., ARMSTRONG, J., DOROSTY, A. R., EMMETT, P. M., NESS, A., ROGERS, I., STEER, C. & SHERRIFF, A. 2005b. Early life risk factors for obesity in childhood: cohort study. *BMJ,* 330**,** 1357.
REILLY, J. J. & KELLY, J. 2010. Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. *International Journal Of Obesity,* 35**,** 891.
REILLY, J. J., METHVEN, E., MCDOWELL, Z. C., HACKING, B., ALEXANDER, D., STEWART, L. & KELNAR, C. J. 2003. Health consequences of obesity. *Archives of disease in childhood,* 88**,** 748-752.
REILLY, J. J., WILSON, J., MCCOLL, J. H., CARMICHAEL, M. & DURNIN, J. V. G. A. 1996. Ability of Biolectric Impedance to Predict Fat-Free Mass in Prepubertal Children. *Pediatric Research,* 39**,** 176.
REINER, M., NIERMANN, C., JEKAUC, D. & WOLL, A. 2013. Long-term health benefits of physical activity – a systematic review of longitudinal studies. *BMC Public Health,* 13**,** 813.
RELTON, C. L., GAUNT, T., MCARDLE, W., HO, K., DUGGIRALA, A., SHIHAB, H., WOODWARD, G., LYTTLETON, O., EVANS, D. M., REIK, W., PAUL, Y.-L., FICZ, G., OZANNE, S. E., WIPAT, A., FLANAGAN, K., LISTER, A., HEIJMANS, B. T., RING, S. M. & DAVEY SMITH, G. 2015. Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). *International Journal of Epidemiology,* 44**,** 1181-1190.
REYNOLDS, R. M., ALLAN, K. M., RAJA, E. A., BHATTACHARYA, S., MCNEILL, G., HANNAFORD, P. C., SARWAR, N., LEE, A. J., BHATTACHARYA, S. & NORMAN, J. E. 2013a. Maternal obesity during pregnancy and premature mortality from cardiovascular event in adult offspring: follow-up of 1 323 275 person years. *Bmj-British Medical Journal,* 347.
REYNOLDS, R. M., JACOBSEN, G. H. & DRAKE, A. J. 2013b. What is the evidence in humans that DNA methylation changes link events in utero and later life disease? *Clinical Endocrinology,* 78**,** 814-822.
REYNOLDS, R. M., OSMOND, C., PHILLIPS, D. I. W. & GODFREY, K. M. 2010. Maternal BMI, Parity, and Pregnancy Weight Gain: Influences on Offspring Adiposity in Young Adulthood. *Journal of Clinical Endocrinology & Metabolism,* 95**,** 5365-5369.
RICHMAN, N. 1981. A community survey of characteristics of one- to two- year-olds with sleep disruptions. *J Am Acad Child Psychiatry,* 20**,** 281-91.
RICHMOND, R. C., SHARP, G. C., WARD, M. E., FRASER, A., LYTTLETON, O., MCARDLE, W. L., RING, S. M., GAUNT, T. R., LAWLOR, D. A., DAVEY SMITH, G. & RELTON, C. L. 2016. DNA Methylation and BMI: Investigating Identified Methylation Sites at HIF3A in a Causal Framework. *Diabetes,* 65**,** 1231.
RICHMOND, R. C., SIMPKIN, A. J., WOODWARD, G., GAUNT, T. R., LYTTLETON, O., MCARDLE, W. L., RING, S. M., SMITH, A. D. A. C., TIMPSON, N. J., TILLING, K., DAVEY SMITH, G. & RELTON, C. L. 2014. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Human Molecular Genetics,* 24**,** 2201-2217.

RICHMOND, R. C., SIMPKIN, A. J., WOODWARD, G., GAUNT, T. R., LYTTLETON, O., MCARDLE, W. L., RING, S. M., SMITH, A. D. A. C., TIMPSON, N. J., TILLING, K., DAVEY SMITH, G. & RELTON, C. L. 2015. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Human molecular genetics,* 24**,** 2201-2217.

RICHMOND, R. C., SUDERMAN, M., LANGDON, R., RELTON, C. L. & DAVEY SMITH, G. 2018. DNA methylation as a marker for prenatal smoke exposure in adults. *International Journal of Epidemiology,* 47**,** 1120-1130.

RICHMOND, R. C., TIMPSON, N. J., FELIX, J. F., PALMER, T., GAILLARD, R., MCMAHON, G., DAVEY SMITH, G., JADDOE, V. W. & LAWLOR, D. A. 2017. Using Genetic Variation to Explore the Causal Effect of Maternal Pregnancy Adiposity on Future Offspring Adiposity: A Mendelian Randomisation Study. *PLOS Medicine,* 14**,** e1002221.

RIDLER, C., TOWNSEND, N., DINSDALE, H., MULHALL, C. & RUTTER, H. 2009. National child measurement programme: detailed analysis of the 2007/08 national dataset. *National Obesity Observatory, London***,** 41.

RIFAS-SHIMAN, S. L., GILLMAN, M. W., HAWKINS, S. S., OKEN, E., TAVERAS, E. M. & KLEINMAN, K. P. 2018. Association of Cesarean Delivery With Body Mass Index z Score at Age 5 YearsWithin-Family Analysis of Cesarean Delivery and BMI z Score at Age 5 YearsLetters. *JAMA Pediatrics,* 172**,** 777-779.

RILEY, J. C. 1989. *Sickness, recovery, and death: A history and forecast of ill health*, University of Iowa Press.

ROBERTSON, K. D. 2005. DNA methylation and human disease. *Nature Reviews Genetics,* 6**,** 597-610.

ROBINSON, C. J. & YOUNG, V. B. 2010. Antibiotic administration alters the community structure of the gastrointestinal micobiota. *Gut Microbes,* 1**,** 279-284.

ROBINSON, H. A., DAM, R., HASSAN, L., JENKINS, D., BUCHAN, I. & SPERRIN, M. 2019. Post-2000 growth trajectories in children aged 4–11 years: A review and quantitative analysis. *Preventive Medicine Reports,* 14**,** 100834.

ROBINSON, M. D., KAHRAMAN, A., LAW, C. W., LINDSAY, H., NOWICKA, M., WEBER, L. M. & ZHOU, X. 2014. Statistical methods for detecting differentially methylated loci and regions. *Frontiers in genetics,* 5**,** 324.

ROBINSON, S. M., CROZIER, S. R., HARVEY, N. C., BARTON, B. D., LAW, C. M., GODFREY, K. M., COOPER, C. & INSKIP, H. M. 2015. Modifiable early-life risk factors for childhood adiposity and overweight: an analysis of their combined impact and potential for prevention. *The American Journal of Clinical Nutrition*.

ROESSLER, J., AMMERPOHL, O., GUTWEIN, J., HASEMEIER, B., ANWAR, S. L., KREIPE, H. & LEHMANN, U. 2012. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. *BMC research notes,* 5**,** 210-210.

ROGERS, N. T., POWER, C. & PINTO PEREIRA, S. M. 2019. Birthweight, lifetime obesity and physical functioning in mid-adulthood: a nationwide birth cohort study. *International Journal of Epidemiology*.

ROHDE, K., KELLER, M., LA COUR POULSEN, L., BLÜHER, M., KOVACS, P. & BÖTTCHER, Y. 2018. Genetics and epigenetics in obesity. *Metabolism*.

ROLLAND-CACHERA, M. F., DEHEEGER, M., MAILLOT, M. & BELLISLE, F. 2006. Early adiposity rebound: causes and consequences for obesity in children and adults. *International Journal of Obesity,* 30**,** S11-S17.

ROSEBOOM, T. J., VAN DER MEULEN, J. H. P., OSMOND, C., BARKER, D. J. P., RAVELLI, A. C. J. & BLEKER, O. P. 2000. Plasma lipid profiles in adults after prenatal exposure to the Dutch famine. *The American Journal of Clinical Nutrition,* 72**,** 1101-1106.

ROUSSEEUW, P. J. & LEROY, A. M. 2005. *Robust regression and outlier detection*, John wiley & sons.

ROYSTON, P. & ALTMAN, D. G. 1994. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 43**,** 429-453.

RUIZ, M., GOLDBLATT, P., MORRISON, J., PORTA, D., FORASTIERE, F., HRYHORCZUK, D., ANTIPKIN, Y., SAUREL-CUBIZOLLES, M. J., LIORET, S., VRIJHEID, M., TORRENT, M., INIGUEZ, C., LARRANAGA, I., BAKOULA, C., VELTSISTA, A., VAN EIJSDEN, M., VRIJKOTTE, T. G., ANDRYSKOVA, L., DUSEK, L., BARROS, H., CORREIA, S., JARVELIN, M. R., TAANILA, A., LUDVIGSSON, J., FARESJO, T., MARMOT, M. & PIKHART, H. 2016. Impact of Low Maternal Education on Early Childhood Overweight and Obesity in Europe. *Paediatr Perinat Epidemiol,* 30**,** 274-84.

RYCHLIK, W., SPENCER, W. J. & RHOADS, R. E. 1990. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res,* 18**,** 6409-12.

RZEHAK, P., COVIC, M., SAFFERY, R., REISCHL, E., WAHL, S., GROTE, V., WEBER, M., XHONNEUX, A., LANGHENDRIES, J.-P., FERRE, N., CLOSA-MONASTEROLO, R., ESCRIBANO, J., VERDUCI, E., RIVA, E., SOCHA, P., GRUSZFELD, D. & KOLETZKO, B. 2017. DNA-Methylation and Body Composition in Preschool Children: Epigenome-Wide-Analysis in the European Childhood Obesity Project (CHOP)-Study. *Scientific Reports,* 7**,** 14349.

SACN. 2012. Consideration of issues around the use of BMI centile thresholds for defining underweight, overweight and obesity in children aged 2-18 years in the UK Available: http://www.sacn.gov.uk/pdfs/sacnrcpch_position_statement_bmi_thresholds.pdf

SADEH, A., LAVIE, P., SCHER, A., TIROSH, E. & EPSTEIN, R. 1991. Actigraphic home-monitoring sleep-disturbed and control infants and young children: a new method for pediatric assessment of sleep-wake patterns. *Pediatrics,* 87**,** 494-499.

SADEH, A., MINDELL, J. & RIVERA, L. 2011. "My child has a sleep problem": a cross-cultural comparison of parental definitions. *Sleep Med,* 12**,** 478-82.

SAHAKYAN, K. R., SOMERS, V. K., RODRIGUEZ-ESCUDERO, J. P. & ET AL. 2015. Normal-weight central obesity: Implications for total and cardiovascular mortality. *Annals of Internal Medicine,* 163**,** 827-835.

SAKAI, M., HIBI, K., KANAZUMI, N., NOMOTO, S., INOUE, S., TAKEDA, S. & NAKAO, A. 2005. Aberrant methylation of the CHFR gene in advanced hepatocellular carcinoma. *Hepatogastroenterology,* 52**,** 1854-7.

SALAM, R. A., DAS, J. K. & BHUTTA, Z. A. 2014. Impact of intrauterine growth restriction on long-term health. *Curr Opin Clin Nutr Metab Care,* 17**,** 249-54.

SAMUELSSON, A.-M., MATTHEWS, P. A., ARGENTON, M., CHRISTIE, M. R., MCCONNELL, J. M., JANSEN, E. H. J. M., PIERSMA, A. H., OZANNE, S. E., TWINN, D. F., REMACLE, C., ROWLERSON, A., POSTON, L. & TAYLOR, P. D. 2008. Diet-induced obesity in female mice leads to offspring hyperphagia, adiposity, hypertension, and insulin resistance - A novel murine model of developmental programming. *Hypertension,* 51**,** 383-392.

SANBHNANI, S. & YEONG, F. M. 2012. CHFR: a key checkpoint component implicated in a wide range of cancers. *Cell Mol Life Sci,* 69**,** 1669-87.

SANTOS-SILVA, A. P., OLIVEIRA, E., PINHEIRO, C. R., SANTANA, A. C., NASCIMENTO-SABA, C. C., ABREU-VILLAÇA, Y., MOURA, E. G. & LISBOA, P. C. 2013. Endocrine effects of tobacco smoke exposure during lactation in weaned and adult male offspring. *Journal of Endocrinology,* 218**,** 13-24.

SANTOS, S., VOERMAN, E., AMIANO, P., BARROS, H., BEILIN, L. J., BERGSTROM, A., CHARLES, M. A., CHATZI, L., CHEVRIER, C., CHROUSOS, G. P., CORPELEIJN, E., COSTA, O.,

COSTET, N., CROZIER, S., DEVEREUX, G., DOYON, M., EGGESBO, M., FANTINI, M. P., FARCHI, S., FORASTIERE, F., GEORGIU, V., GODFREY, K. M., GORI, D., GROTE, V., HANKE, W., HERTZ-PICCIOTTO, I., HEUDE, B., HIVERT, M. F., HRYHORCZUK, D., HUANG, R. C., INSKIP, H., KARVONEN, A. M., KENNY, L. C., KOLETZKO, B., KUPERS, L. K., LAGSTROM, H., LEHMANN, I., MAGNUS, P., MAJEWSKA, R., MAKELA, J., MANIOS, Y., MCAULIFFE, F. M., MCDONALD, S. W., MEHEGAN, J., MELEN, E., MOMMERS, M., MORGEN, C. S., MOSCHONIS, G., MURRAY, D., NI CHAOIMH, C., NOHR, E. A., NYBO ANDERSEN, A. M., OKEN, E., OOSTVOGELS, A., PAC, A., PAPADOPOULOU, E., PEKKANEN, J., PIZZI, C., POLANSKA, K., PORTA, D., RICHIARDI, L., RIFAS-SHIMAN, S. L., ROELEVELD, N., RONFANI, L., SANTOS, A. C., STANDL, M., STIGUM, H., STOLTENBERG, C., THIERING, E., THIJS, C., TORRENT, M., TOUGH, S. C., TRNOVEC, T., TURNER, S., VAN GELDER, M., VAN ROSSEM, L., VON BERG, A., VRIJHEID, M., VRIJKOTTE, T., WEST, J., WIJGA, A. H., WRIGHT, J., ZVINCHUK, O., SORENSEN, T., LAWLOR, D. A., GAILLARD, R. & JADDOE, V. 2019. Impact of maternal body mass index and gestational weight gain on pregnancy complications: an individual participant data meta-analysis of European, North American and Australian cohorts. *Bjog,* 126**,** 984-995.

SARDINA, J. L., COLLOMBET, S., TIAN, T. V., GÓMEZ, A., DI STEFANO, B., BERENGUER, C., BRUMBAUGH, J., STADHOUDERS, R., SEGURA-MORALES, C. & GUT, M. 2018. Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. *Cell stem cell,* 23**,** 727-741. e9.

SARKER, G., BERRENS, R., VON ARX, J., PELCZAR, P., REIK, W., WOLFRUM, C. & PELEG-RAIBSTEIN, D. 2018. Transgenerational transmission of hedonic behaviors and metabolic phenotypes induced by maternal overnutrition. *Translational psychiatry,* 8**,** 195.

SAVAGE, J. S., BIRCH, L. L., MARINI, M., ANZMAN-FRASCA, S. & PAUL, I. M. 2016. Effect of the INSIGHT Responsive Parenting Intervention on Rapid Infant Weight Gain and Overweight Status at Age 1 Year: A Randomized Clinical Trial. *JAMA Pediatr,* 170**,** 742-9.

SAVAGE, T., DERRAIK, J. G. B., MILES, H. L., MOUAT, F., HOFMAN, P. L. & CUTFIELD, W. S. 2013. Increasing Maternal Age Is Associated with Taller Stature and Reduced Abdominal Fat in Their Children. *Plos One,* 8**,** 7.

SCHELLONG, K., SCHULZ, S., HARDER, T. & PLAGEMANN, A. 2012. Birth weight and long-term overweight risk: systematic review and a meta-analysis including 643,902 persons from 66 studies and 26 countries globally.

SCHOLTENS, S., GEHRING, U., BRUNEKREEF, B., SMIT, H. A., DE JONGSTE, J. C., KERKHOF, M., GERRITSEN, J. & WIJGA, A. H. 2007. Breastfeeding, Weight Gain in Infancy, and Overweight at Seven Years of AgeThe Prevention and Incidence of Asthma and Mite Allergy Birth Cohort Study. *American Journal of Epidemiology,* 165**,** 919-926.

SCHWARTZ, B. S., POLLAK, J., BAILEY-DAVIS, L., HIRSCH, A. G., COSGROVE, S. E., NAU, C., KRESS, A. M., GLASS, T. A. & BANDEEN-ROCHE, K. 2016. Antibiotic use and childhood body mass index trajectory. *Int J Obes (Lond),* 40**,** 615-21.

SCHWARZ, G. 1978. Estimating the dimension of a model. *The annals of statistics,* 6**,** 461-464.

SCOLNICK, D. M. & HALAZONETIS, T. D. 2000. Chfr defines a mitotic stress checkpoint that delays entry into metaphase. *Nature,* 406**,** 430.

SECKL, J. R. 2004. Prenatal glucocorticoids and long-term programming. *European journal of endocrinology,* 151**,** U49.

SEIDELL, J. C. 2009. Waist circumference and waist/hip ratio in relation to all-cause mortality, cancer and sleep apnea. *European Journal Of Clinical Nutrition,* 64**,** 35.

SHAH, S., BONDER, M. J., MARIONI, R. E., ZHU, Z., MCRAE, A. F., ZHERNAKOVA, A., HARRIS, S. E., LIEWALD, D., HENDERS, A. K. & MENDELSON, M. M. 2015. Improving phenotypic prediction by combining genetic and epigenetic associations. *The American Journal of Human Genetics,* 97**,** 75-85.

SHAO, X., DING, X., WANG, B., LI, L., AN, X., YAO, Q., SONG, R. & ZHANG, J.-A. 2017. Antibiotic Exposure in Early Life Increases Risk of Childhood Obesity: A Systematic Review and Meta-Analysis. *Frontiers in Endocrinology,* 8.

SHARP, G. C., LAWLOR, D. A., RICHMOND, R. C., FRASER, A., SIMPKIN, A., SUDERMAN, M., SHIHAB, H. A., LYTTLETON, O., MCARDLE, W., RING, S. M., GAUNT, T. R., DAVEY SMITH, G. & RELTON, C. L. 2015a. Maternal pre-pregnancy BMI and gestational weight gain, offspring DNA methylation and later offspring adiposity: findings from the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology,* 44**,** 1288-1304.

SHARP, G. C., LAWLOR, D. A., RICHMOND, R. C., FRASER, A., SIMPKIN, A., SUDERMAN, M., SHIHAB, H. A., LYTTLETON, O., MCARDLE, W., RING, S. M., GAUNT, T. R., DAVEY SMITH, G. & RELTON, C. L. 2015b. Maternal pre-pregnancy BMI and gestational weight gain, offspring DNA methylation and later offspring adiposity: findings from the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*.

SHARP, G. C., SALAS, L. A., MONNEREAU, C., ALLARD, C., YOUSEFI, P., EVERSON, T. M., BOHLIN, J., XU, Z., HUANG, R.-C., REESE, S. E., XU, C.-J., BAÏZ, N., HOYO, C., AGHA, G., ROY, R., HOLLOWAY, J. W., GHANTOUS, A., MERID, S. K., BAKULSKI, K. M., KÜPERS, L. K., ZHANG, H., RICHMOND, R. C., PAGE, C. M., DUIJTS, L., LIE, R. T., MELTON, P. E., VONK, J. M., NOHR, E. A., WILLIAMS-DEVANE, C., HUEN, K., RIFAS-SHIMAN, S. L., RUIZ-ARENAS, C., GONSETH, S., REZWAN, F. I., HERCEG, Z., EKSTRÖM, S., CROEN, L., FALAHI, F., PERRON, P., KARAGAS, M. R., QURAISHI, B. M., SUDERMAN, M., MAGNUS, M. C., JADDOE, V. W. V., TAYLOR, J. A., ANDERSON, D., ZHAO, S., SMIT, H. A., JOSEY, M. J., BRADMAN, A., BACCARELLI, A. A., BUSTAMANTE, M., HÅBERG, S. E., PERSHAGEN, G., HERTZ-PICCIOTTO, I., NEWSCHAFFER, C., CORPELEIJN, E., BOUCHARD, L., LAWLOR, D. A., MAGUIRE, R. L., BARCELLOS, L. F., DAVEY SMITH, G., ESKENAZI, B., KARMAUS, W., MARSIT, C. J., HIVERT, M.-F., SNIEDER, H., FALLIN, M. D., MELÉN, E., MUNTHE-KAAS, M. C., ARSHAD, H., WIEMELS, J. L., ANNESI-MAESANO, I., VRIJHEID, M., OKEN, E., HOLLAND, N., MURPHY, S. K., SØRENSEN, T. I. A., KOPPELMAN, G. H., NEWNHAM, J. P., WILCOX, A. J., NYSTAD, W., LONDON, S. J., FELIX, J. F. & RELTON, C. L. 2017. Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: findings from the pregnancy and childhood epigenetics (PACE) consortium. *Human Molecular Genetics,* 26**,** 4067-4085.

SHEN, L., GUO, Y., CHEN, X., AHMED, S. & ISSA, J.-P. J. 2007. Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis. *Biotechniques,* 42**,** 48-58.

SHIN, S. H. & MILLER, D. P. 2012. A longitudinal examination of childhood maltreatment and adolescent obesity: Results from the National Longitudinal Study of Adolescent Health (AddHealth) Study. *Child Abuse & Neglect,* 36**,** 84-94.

SHREWSBURY, V. & WARDLE, J. 2008. Socioeconomic status and adiposity in childhood: a systematic review of cross-sectional studies 1990-2005. *Obesity (Silver Spring),* 16**,** 275-84.

SHREWSBURY, V. & WARDLE, J. 2012. Socioeconomic Status and Adiposity in Childhood: A Systematic Review of Cross-sectional Studies 1990–2005. *Obesity,* 16**,** 275-284.

SILVERMAN, B. L., RIZZO, T. A., CHO, N. H. & METZGER, B. E. 1998. Long-term effect of the intrauterine environment - The Northwestern University Diabetes in Pregnancy Center. *Diabetes Care,* 21**,** B142-B149.

SIMMONDS, M., BURCH, J., LLEWELLYN, A., GRIFFITHS, C., YANG, H., OWEN, C., DUFFY, S. & WOOLACOTT, N. 2015a. The use of measures of obesity in childhood for predicting obesity and the development of obesity-related diseases in adulthood: a systematic review and meta-analysis. *Health Technology Assessment (Winchester, England),* 19**,** 1-336.

SIMMONDS, M., LLEWELLYN, A., OWEN, C. G. & WOOLACOTT, N. 2015b. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. *Obesity Reviews,* 17**,** 95-107.

SIMMONDS, M., LLEWELLYN, A., OWEN, C. G. & WOOLACOTT, N. 2016. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. *Obesity Reviews,* 17**,** 95-107.

SIMPKIN, A. J., HOWE, L. D., TILLING, K., GAUNT, T. R., LYTTLETON, O., MCARDLE, W. L., RING, S. M., HORVATH, S., SMITH, G. D. & RELTON, C. L. 2017. The epigenetic clock and physical development during childhood and adolescence: longitudinal analysis from a UK birth cohort. *International journal of epidemiology,* 46**,** 549-558.

SIMPKIN, A. J., SUDERMAN, M., GAUNT, T. R., LYTTLETON, O., MCARDLE, W. L., RING, S. M., TILLING, K., DAVEY SMITH, G. & RELTON, C. L. 2015. Longitudinal analysis of DNA methylation associated with birth weight and gestational age. *Human molecular genetics,* 24**,** 3752-3763.

SINGH, A. S., MULDER, C., TWISK, J. W. R., VAN MECHELEN, W. & CHINAPAW, M. J. M. 2008. Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obesity Reviews,* 9**,** 474-488.

SMITH, G. 2018. Step away from stepwise. *Journal of Big Data,* 5**,** 32.

SMITH, G. D., HART, C., BLANE, D. & HOLE, D. 1998. Adverse socioeconomic conditions in childhood and cause specific adult mortality: prospective observational study. *Bmj,* 316**,** 1631-1635.

SOBEL, M. E. 2008. Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics,* 33**,** 230-251.

SOCHA, P., GROTE, V., GRUSZFELD, D., JANAS, R., DEMMELMAIR, H., CLOSA-MONASTEROLO, R., SUBIAS, J. E., SCAGLIONI, S., VERDUCI, E., DAIN, E., LANGHENDRIES, J. P., PERRIN, E. & KOLETZKO, B. 2011. Milk protein intake, the metabolic-endocrine response, and growth in infancy: data from a randomized clinical trial. *Am J Clin Nutr,* 94**,** 1776s-1784s.

SOHN, K. 2017. The trend in the relationship of advanced maternal age to preterm birth and low birthweight. *Eur J Contracept Reprod Health Care,* 22**,** 363-368.

SONESTEDT, E., ROOS, C., GULLBERG, B., ERICSON, U., WIRFÄLT, E. & ORHO-MELANDER, M. 2009. Fat and carbohydrate intake modify the association between genetic variation in the FTO genotype and obesity. *The American Journal of Clinical Nutrition,* 90**,** 1418-1425.

SONG, Q., DECATO, B., HONG, E. E., ZHOU, M., FANG, F., QU, J., GARVIN, T., KESSLER, M., ZHOU, J. & SMITH, A. D. 2013. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One,* 8**,** e81148.

SOUBRY, A., MURPHY, S. K., WANG, F., HUANG, Z., VIDAL, A. C., FUEMMELER, B. F., KURTZBERG, J., MURTHA, A., JIRTLE, R. L., SCHILDKRAUT, J. M. & HOYO, C. 2013. Newborns of obese parents have altered DNA methylation patterns at imprinted genes. *International Journal Of Obesity,* 39**,** 650.

SPEAKMAN, J. R. 2008. Thrifty genes for obesity, an attractive but flawed idea, and an alternative perspective: the 'drifty gene' hypothesis. *Int J Obes (Lond),* 32**,** 1611-7.

SPEAKMAN, J. R. & WESTERTERP, K. R. 2013. A mathematical model of weight loss under total starvation: evidence against the thrifty-gene hypothesis. *Disease models & mechanisms,* 6**,** 236-251.

SPELIOTES, E. K., WILLER, C. J., BERNDT, S. I., MONDA, K. L., THORLEIFSSON, G., JACKSON, A. U., ALLEN, H. L., LINDGREN, C. M., MÄGI, R. & RANDALL, J. C. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics,* 42**,** 937-948.

STAMATAKIS, E., WARDLE, J. & COLE, T. J. 2009. Childhood obesity and overweight prevalence trends in England: evidence for growing socioeconomic disparities. *Int J Obes,* 34**,** 41-47.

STEIN, A. D., KAHN, H. S., RUNDLE, A., ZYBERT, P. A., VAN DER PAL–DE BRUIN, K. & LUMEY, L. H. 2007. Anthropometric measures in middle age after exposure to famine during gestation: evidence from the Dutch famine. *The American journal of clinical nutrition,* 85**,** 869-876.

STEINBERGER, J., JACOBS, D. R., RAATZ, S., MORAN, A., HONG, C. P. & SINAIKO, A. R. 2005. Comparison of body fatness measurements by BMI and skinfolds vs dual energy X-ray absorptiometry and their relation to cardiovascular risk factors in adolescents. *Int J Obes (Lond),* 29**,** 1346-52.

STETTLER, N., KUMANYIKA, S. K., KATZ, S. H., ZEMEL, B. S. & STALLINGS, V. A. 2003. Rapid weight gain during infancy and obesity in young adulthood in a cohort of African Americans. *The American Journal of Clinical Nutrition,* 77**,** 1374-1378.

STETTLER, N., STALLINGS, V. A., TROXEL, A. B., ZHAO, J., SCHINNAR, R., NELSON, S. E., ZIEGLER, E. E. & STROM, B. L. 2005. Weight gain in the first week of life and overweight in adulthood: a cohort study of European American subjects fed infant formula. *Circulation,* 111**,** 1897-903.

STOUFFER, S. A., LUMSDAINE, A. A., LUMSDAINE, M. H., WILLIAMS JR, R. M., SMITH, M. B., JANIS, I. L., STAR, S. A. & COTTRELL JR, L. S. 1949. The American soldier: Combat and its aftermath.(Studies in social psychology in World War II), Vol. 2.

STRINGHINI, S., POLIDORO, S., SACERDOTE, C., KELLY, R. S., VAN VELDHOVEN, K., AGNOLI, C., GRIONI, S., TUMINO, R., GIURDANELLA, M. C., PANICO, S., MATTIELLO, A., PALLI, D., MASALA, G., GALLO, V., CASTAGNÉ, R., PACCAUD, F., CAMPANELLA, G., CHADEAU-HYAM, M. & VINEIS, P. 2015. Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *International Journal of Epidemiology,* 44**,** 1320-1330.

STRUNK, T., JAMIESON, S. E. & BURGNER, D. 2013. Genetic and epigenetic susceptibility to early life infection. *Current opinion in infectious diseases,* 26**,** 241-247.

STUNKARD, A. J. & SORENSEN, T. I. 1993. Obesity and socioeconomic status--a complex relation. *N Engl J Med,* 329**,** 1036-7.

SUBRAMANYAM, M. A., DIEZ-ROUX, A. V., PILSNER, J. R., VILLAMOR, E., DONOHUE, K. M., LIU, Y. & JENNY, N. S. 2013. Social Factors and Leukocyte DNA Methylation of Repetitive Sequences: The Multi-Ethnic Study of Atherosclerosis. *PLoS ONE,* 8**,** e54018.

SUDMANT, P. H., RAUSCH, T., GARDNER, E. J., HANDSAKER, R. E., ABYZOV, A., HUDDLESTON, J., ZHANG, Y., YE, K., JUN, G. & FRITZ, M. H.-Y. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature,* 526**,** 75.

SULLIVAN, A. & BROWN, M. 2013. Overweight and obesity in mid-life: Evidence from the 1970 birth cohort study at age 42. Centre for Longitudinal Studies, Institute of Education, University of London.

SUMMERS, M. K., BOTHOS, J. & HALAZONETIS, T. D. 2005. The CHFR mitotic checkpoint protein delays cell cycle progression by excluding Cyclin B1 from the nucleus. *Oncogene,* 24**,** 2589.

SUN, B., PURCELL, R. H., TERRILLION, C. E., YAN, J., MORAN, T. H. & TAMASHIRO, K. L. K. 2012. Maternal High-Fat Diet During Gestation or Suckling Differentially Affects Offspring Leptin Sensitivity and Obesity. *Diabetes,* 61**,** 2833-2841.

SUN, Y., LIU, B., SNETSELAAR, L. G., WALLACE, R. B., CAAN, B. J., ROHAN, T. E., NEUHOUSER, M. L., SHADYAB, A. H., CHLEBOWSKI, R. T., MANSON, J. E. & BAO, W. 2019. Association of Normal-Weight Central Obesity With All-Cause and Cause-Specific Mortality Among Postmenopausal WomenAssociation of Normal-Weight Central Obesity With Mortality Among Postmenopausal WomenAssociation of Normal-Weight Central Obesity With Mortality Among Postmenopausal Women. *JAMA Network Open,* 2**,** e197337-e197337.

SUTCLIFFE, A. G., BARNES, J., BELSKY, J., GARDINER, J. & MELHUISH, E. 2012. The health and development of children born to older mothers in the United Kingdom: observational study using longitudinal cohort data. *BMJ : British Medical Journal,* 345.

TAHILIANI, M., KOH, K. P., SHEN, Y., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R., ARAVIND, L. & RAO, A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science,* 324**,** 930-5.

TALMA, H., CHINAPAW, M. J., BAKKER, B., HIRASING, R. A., TERWEE, C. B. & ALTENBURG, T. M. 2013. Bioelectrical impedance analysis to estimate body composition in children and adolescents: a systematic review and evidence appraisal of validity, responsiveness, reliability and measurement error. *Obes Rev,* 14**,** 895-905.

TAMASHIRO, K. L. K., TERRILLION, C. E., HYUN, J., KOENIG, J. I. & MORAN, T. H. 2009. Prenatal Stress or High-Fat Diet Increases Susceptibility to Diet-Induced Obesity in Rat Offspring. *Diabetes,* 58**,** 1116-1125.

TAMAYO, T., HERDER, C. & RATHMANN, W. 2010. Impact of early psychosocial factors (childhood socioeconomic factors and adversities) on future risk of type 2 diabetes, metabolic disturbances and obesity: a systematic review. *BMC Public Health,* 10**,** 525.

TANAKA, C., JANSSEN, X., PEARCE, M., PARKINSON, K., BASTERFIELD, L., ADAMSON, A. & REILLY, J. J. 2018. Bidirectional Associations Between Adiposity, Sedentary Behavior, and Physical Activity: A Longitudinal Study in Children. *Journal of Physical Activity and Health,* 15**,** 918-926.

TANNER, J. M. & WHITEHOUSE, R. H. 1975. Revised standards for triceps and subscapular skinfolds in British children. *Archives of disease in childhood,* 50**,** 142-145.

TARLOV, A. R. 1999. Public Policy Frameworks for Improving Population Health. *Annals of the New York Academy of Sciences,* 896**,** 281-293.

TAVERAS, E. M., GILLMAN, M. W., PEÑA, M.-M., REDLINE, S. & RIFAS-SHIMAN, S. L. 2014. Chronic Sleep Curtailment and Adiposity. *Pediatrics,* 133**,** 1013.

TAVERAS, E. M., RIFAS-SHIMAN, S. L., OKEN, E., GUNDERSON, E. P. & GILLMAN, M. W. 2008. Short sleep duration in infancy and risk of childhood overweight. *Arch Pediatr Adolesc Med,* 162**,** 305-11.

TEHRANIFAR, P., WU, H.-C., FAN, X., FLOM, J. D., FERRIS, J. S., CHO, Y. H., GONZALEZ, K., SANTELLA, R. M. & TERRY, M. B. 2013. Early life socioeconomic factors and genomic DNA methylation in mid-life. *Epigenetics,* 8**,** 23-27.

TESCHENDORFF, A. E., MENON, U., GENTRY-MAHARAJ, A., RAMUS, S. J., GAYTHER, S. A., APOSTOLIDOU, S., JONES, A., LECHNER, M., BECK, S., JACOBS, I. J. & WIDSCHWENDTER, M. 2009. An Epigenetic Signature in Peripheral Blood Predicts Active Ovarian Cancer. *PLOS ONE,* 4**,** e8274.

TESCHENDORFF, A. E. & ZHENG, S. C. 2017. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics,* 9**,** 757-768.

TESCHENDORFF, A. E., ZHUANG, J. & WIDSCHWENDTER, M. 2011. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics,* 27**,** 1496-505.

THE UK HEALTH FORUM AND CANCER RESEARCH UK 2015. Tipping the scales: Why preventing obesity makes economic sense. Current Awareness Service for Health.

THIEDKE, C. C. 2001. Sleep disorders and sleep problems in childhood. *American family physician,* 63**,** 277-287.

THOMPSON, A. L. 2012. Developmental origins of obesity: early feeding environments, infant growth, and the intestinal microbiome. *Am J Hum Biol,* 24**,** 350-60.

THORLEIFSSON, G., WALTERS, G. B., GUDBJARTSSON, D. F., STEINTHORSDOTTIR, V., SULEM, P., HELGADOTTIR, A., STYRKARSDOTTIR, U., GRETARSDOTTIR, S., THORLACIUS, S., JONSDOTTIR, I., JONSDOTTIR, T., OLAFSDOTTIR, E. J., OLAFSDOTTIR, G. H., JONSSON, T., JONSSON, F., BORCH-JOHNSEN, K., HANSEN, T., ANDERSEN, G., JORGENSEN, T., LAURITZEN, T., ABEN, K. K., VERBEEK, A. L., ROELEVELD, N., KAMPMAN, E., YANEK, L. R., BECKER, L. C., TRYGGVADOTTIR, L., RAFNAR, T., BECKER, D. M., GULCHER, J., KIEMENEY, L. A., PEDERSEN, O., KONG, A., THORSTEINSDOTTIR, U. & STEFANSSON, K. 2009. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet,* 41**,** 18-24.

THULIER, D. & MERCER, J. 2009. Variables associated with breastfeeding duration. *Journal of Obstetric, Gynecologic, & Neonatal Nursing,* 38**,** 259-268.

TIKOTZKY, L., DE MARCAS, G., HAR-TOOV, J., DOLLBERG, S., BAR-HAIM, Y. & SADEH, A. V. I. 2010a. Sleep and physical growth in infants during the first 6 months. *Journal of Sleep Research,* 19**,** 103-110.

TIKOTZKY, L., DE MARCAS, G., HAR-TOOV, J., DOLLBERG, S., BAR-HAIM, Y. & SADEH, A. 2010b. Sleep and physical growth in infants during the first 6 months. *Journal of sleep research,* 19**,** 103-110.

TOBI, E. W., SLIEKER, R. C., LUIJK, R., DEKKERS, K. F., STEIN, A. D., XU, K. M., SLAGBOOM, P. E., VAN ZWET, E. W., LUMEY, L. & HEIJMANS, B. T. 2018. DNA methylation as a mediator of the association between prenatal adversity and risk factors for metabolic disease in adulthood. *Science advances,* 4**,** eaao4364.

TOBIN, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society***,** 24-36.

TORRES, S. J. & NOWSON, C. A. 2007. Relationship between stress, eating behavior, and obesity. *Nutrition,* 23**,** 887-894.

TOST, J. & GUT, I. G. 2007. DNA methylation analysis by pyrosequencing. *Nature Protocols,* 2**,** 2265-2275.

TOULEIMAT, N. & TOST, J. 2012. Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics,* 4**,** 325-341.

TOWNSEND, P., PHILLIMORE, P. & BEATTIE, A. 1988. *Health and deprivation: inequality and the North*, Routledge.

TRASANDE, L., BLUSTEIN, J., LIU, M., CORWIN, E., COX, L. M. & BLASER, M. J. 2013. Infant antibiotic exposures and early-life body mass. *Int J Obes (Lond),* 37**,** 16-23.

TSAI, S. A., LV, N., XIAO, L. & MA, J. 2016. Gender Differences in Weight-Related Attitudes and Behaviors Among Overweight and Obese Adults in the United States. *Am J Mens Health,* 10**,** 389-98.

TU, Y.-K., WEST, R., ELLISON, G. T. H. & GILTHORPE, M. S. 2005. Why Evidence for the Fetal Origins of Adult Disease Might Be a Statistical Artifact: The "Reversal Paradox" for the Relation between Birth Weight and Blood Pressure in Later Life. *American Journal of Epidemiology,* 161**,** 27-32.

TUN, H. M., BRIDGMAN, S. L., CHARI, R., FIELD, C. J., GUTTMAN, D. S., BECKER, A. B., MANDHANE, P. J., TURVEY, S. E., SUBBARAO, P., SEARS, M. R., SCOTT, J. A., KOZYRSKYJ, A. L. & CANADIAN HEALTHY INFANT LONGITUDINAL DEVELOPMENT STUDY, I. 2018. Roles of Birth Mode and Infant Gut Microbiota in Intergenerational Transmission of Overweight and Obesity From Mother to Offspring. *JAMA pediatrics,* 172**,** 368-377.

TURAN, N., GHALWASH, M. F., KATARI, S., COUTIFARIS, C., OBRADOVIC, Z. & SAPIENZA, C. 2012. DNA methylation differences at growth related genes correlate with birth weight: a molecular signature linked to developmental origins of adult disease? *Bmc Medical Genomics,* 5.

TWISK, J. W. 2003. The problem of evaluating the magnitude of tracking coefficients. *European journal of epidemiology,* 18**,** 1025.

TYBOR, D. J., LICHTENSTEIN, A. H., DALLAL, G. E. & MUST, A. 2008. Waist-to-height ratio is correlated with height in US children and adolescents aged 2-18 years. *International journal of pediatric obesity : IJPO : an official journal of the International Association for the Study of Obesity,* 3**,** 148-151.

TYRRELL, J., RICHMOND, R. C., PALMER, T. M., FEENSTRA, B., RANGARAJAN, J., METRUSTRY, S., CAVADINO, A., PATERNOSTER, L., ARMSTRONG, L. L., DE SILVA, N. M. G., WOOD, A. R., HORIKOSHI, M., GELLER, F., MYHRE, R., BRADFIELD, J. P., KREINER-MOLLER, E., HUIKARI, V., PAINTER, J. N., HOTTENGA, J. J., ALLARD, C., BERRY, D. J., BOUCHARD, L., DAS, S., EVANS, D. M., HAKONARSON, H., HAYES, M. G., HEIKKINEN, J., HOFMAN, A., KNIGHT, B., LIND, P. A., MCCARTHY, M. I., MCMAHON, G., MEDLAND, S. E., MELBYE, M., MORRIS, A. P., NODZENSKI, M., REICHETZEDER, C., RING, S. M., SEBERT, S., SENGPIEL, V., SORENSEN, T. I. A., WILLEMSEN, G., DE GEUS, E. J. C., MARTIN, N. G., SPECTOR, T. D., POWER, C., JARVELIN, M. R., BISGAARD, H., GRANT, S. F. A., NOHR, E. A., JADDOE, V. W., JACOBSSON, B., MURRAY, J. C., HOCHER, B., HATTERSLEY, A. T., SCHOLTENS, D. M., SMITH, G. D., HIVERT, M. F., FELIX, J. F., HYPPONEN, E., LOWE, W. L., FRAYLING, T. M., LAWLOR, D. A., FREATHY, R. M. & CONSORTIUM, E. G. G. 2016. Genetic Evidence for Causal Relationships Between Maternal Obesity-Related Traits and Birth Weight. *Jama-Journal of the American Medical Association,* 315**,** 1129-1140.

UMER, A., KELLEY, G. A., COTTRELL, L. E., GIACOBBI, P., INNES, K. E. & LILLY, C. L. 2017. Childhood obesity and adult cardiovascular disease risk factors: a systematic review with meta-analysis. *BMC Public Health,* 17**,** 683.

VAN DER SLUIS, I., DE RIDDER, M., BOOT, A., KRENNING, E. & DE MUINCK KEIZER-SCHRAMA, S. 2002. Reference data for bone density and body composition measured with dual energy x ray absorptiometry in white children and young adults. *Archives of disease in childhood,* 87**,** 341-347.

VAN DIJK, S. J., TELLAM, R. L., MORRISON, J. L., MUHLHAUSLER, B. S. & MOLLOY, P. L. 2015. Recent developments on the role of epigenetics in obesity and metabolic disease. *Clinical epigenetics,* 7**,** 66-66.

VEENENDAAL, M. V., PAINTER, R. C., DE ROOIJ, S. R., BOSSUYT, P. M., VAN DER POST, J. A., GLUCKMAN, P. D., HANSON, M. A. & ROSEBOOM, T. J. 2013. Transgenerational effects of prenatal exposure to the 1944–45 Dutch famine. *BJOG: An International Journal of Obstetrics & Gynaecology,* 120**,** 548-554.

VENDITTI, S., MATTHEWS, T. J., ISGANAITIS, E., LERIN, C., DEMERATH, E. W. & FIELDS, D. A. 2019. Maternal obesity and the human milk metabolome: associations with infant body composition and postnatal weight gain.

VENU, L., HARISHANKAR, N., PRASANNA KRISHNA, T. & RAGHUNATH, M. 2004. Maternal dietary vitamin restriction increases body fat content but not insulin resistance in WNIN rat offspring up to 6 months of age. *Diabetologia,* 47**,** 1493-501.

VICKERS, M. H., BREIER, B. H., CUTFIELD, W. S., HOFMAN, P. L. & GLUCKMAN, P. D. 2000. Fetal origins of hyperphagia, obesity, and hypertension and postnatal amplification by hypercaloric nutrition. *American Journal of Physiology - Endocrinology and Metabolism,* 279**,** E83-E87.

VICKERS, M. H. & SLOBODA, D. M. 2012. Leptin as mediator of the effects of developmental programming. *Best Pract Res Clin Endocrinol Metab,* 26**,** 677-87.

VICTORA, C. G., BAHL, R., BARROS, A. J. D., FRANÇA, G. V. A., HORTON, S., KRASEVEC, J., MURCH, S., SANKAR, M. J., WALKER, N. & ROLLINS, N. C. 2016. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *The Lancet,* 387**,** 475-490.

VIDAL, A. C., MURPHY, S. K., MURTHA, A. P., SCHILDKRAUT, J. M., SOUBRY, A., HUANG, Z., NEELON, S. E. B., FUEMMELER, B., IVERSEN, E., WANG, F., KURTZBERG, J., JIRTLE, R. L. & HOYO, C. 2013. Associations between antibiotic exposure during pregnancy, birth weight and aberrant methylation at imprinted genes among offspring. *International Journal Of Obesity,* 37**,** 907.

VIDMAR, S., CARLIN, J., HESKETH, K. & COLE, T. 2004. Standardizing anthropometric measures in children and adolescents with new functions for egen. *Stata J,* 4**,** 50-55.

VILLAR, J., CARROLI, G., WOJDYLA, D., ABALOS, E., GIORDANO, D., BA'AQEEL, H., FARNOT, U., BERGSJØ, P., BAKKETEIG, L., LUMBIGANON, P., CAMPODÓNICO, L., AL-MAZROU, Y., LINDHEIMER, M. & KRAMER, M. 2006. Preeclampsia, gestational hypertension and intrauterine growth restriction, related or independent conditions? *American Journal of Obstetrics and Gynecology,* 194**,** 921-931.

VINEIS, P. & PERERA, F. 2007. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiology and Prevention Biomarkers,* 16**,** 1954-1965.

VINKERS, C. H., KALAFATELI, A. L., RUTTEN, B. P., KAS, M. J., KAMINSKY, Z., TURNER, J. D. & BOKS, M. P. 2015. Traumatic stress and human DNA methylation: a critical review. *Epigenomics,* 7**,** 593-608.

VISSER, M., BOUTER, L. M., MCQUILLAN, G. M., WENER, M. H. & HARRIS, T. B. 1999a. Elevated C-reactive protein levels in overweight and obese adults. *Jama,* 282**,** 2131-5.

VISSER, M., LAUNER, L. J., DEURENBERG, P. & DEEG, D. J. H. 1999b. Past and Current Smoking in Relation to Body Fat Distribution in Older Men and Women. *The Journals of Gerontology: Series A,* 54**,** M293-M298.

WAHL, S., DRONG, A., LEHNE, B., LOH, M., SCOTT, W. R., KUNZE, S., TSAI, P.-C., RIED, J. S., ZHANG, W., YANG, Y., TAN, S., FIORITO, G., FRANKE, L., GUARRERA, S., KASELA, S., KRIEBEL, J., RICHMOND, R. C., ADAMO, M., AFZAL, U., ALA-KORPELA, M., ALBETTI, B., AMMERPOHL, O., APPERLEY, J. F., BEEKMAN, M., BERTAZZI, P. A., BLACK, S. L., BLANCHER, C., BONDER, M.-J., BROSCH, M., CARSTENSEN-KIRBERG, M., DE CRAEN, A. J. M., DE LUSIGNAN, S., DEHGHAN, A., ELKALAAWY, M., FISCHER, K., FRANCO, O. H.,

GAUNT, T. R., HAMPE, J., HASHEMI, M., ISAACS, A., JENKINSON, A., JHA, S., KATO, N., KROGH, V., LAFFAN, M., MEISINGER, C., MEITINGER, T., MOK, Z. Y., MOTTA, V., NG, H. K., NIKOLAKOPOULOU, Z., NTELIOPOULOS, G., PANICO, S., PERVJAKOVA, N., PROKISCH, H., RATHMANN, W., RODEN, M., ROTA, F., ROZARIO, M. A., SANDLING, J. K., SCHAFMAYER, C., SCHRAMM, K., SIEBERT, R., SLAGBOOM, P. E., SOININEN, P., STOLK, L., STRAUCH, K., TAI, E. S., TARANTINI, L., THORAND, B., TIGCHELAAR, E. F., TUMINO, R., UITTERLINDEN, A. G., VAN DUIJN, C., VAN MEURS, J. B. J., VINEIS, P., WICKREMASINGHE, A. R., WIJMENGA, C., YANG, T.-P., YUAN, W., ZHERNAKOVA, A., BATTERHAM, R. L., SMITH, G. D., DELOUKAS, P., HEIJMANS, B. T., HERDER, C., HOFMAN, A., LINDGREN, C. M., MILANI, L., VAN DER HARST, P., PETERS, A., ILLIG, T., RELTON, C. L., WALDENBERGER, M., JÄRVELIN, M.-R., BOLLATI, V., SOONG, R., SPECTOR, T. D., SCOTT, J., MCCARTHY, M. I., et al. 2016. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature,* 541**,** 81.

WAHL, S., DRONG, A., LEHNE, B., LOH, M., SCOTT, W. R., KUNZE, S., TSAI, P.-C., RIED, J. S., ZHANG, W., YANG, Y., TAN, S., FIORITO, G., FRANKE, L., GUARRERA, S., KASELA, S., KRIEBEL, J., RICHMOND, R. C., ADAMO, M., AFZAL, U., ALA-KORPELA, M., ALBETTI, B., AMMERPOHL, O., APPERLEY, J. F., BEEKMAN, M., BERTAZZI, P. A., BLACK, S. L., BLANCHER, C., BONDER, M.-J., BROSCH, M., CARSTENSEN-KIRBERG, M., DE CRAEN, A. J. M., DE LUSIGNAN, S., DEHGHAN, A., ELKALAAWY, M., FISCHER, K., FRANCO, O. H., GAUNT, T. R., HAMPE, J., HASHEMI, M., ISAACS, A., JENKINSON, A., JHA, S., KATO, N., KROGH, V., LAFFAN, M., MEISINGER, C., MEITINGER, T., MOK, Z. Y., MOTTA, V., NG, H. K., NIKOLAKOPOULOU, Z., NTELIOPOULOS, G., PANICO, S., PERVJAKOVA, N., PROKISCH, H., RATHMANN, W., RODEN, M., ROTA, F., ROZARIO, M. A., SANDLING, J. K., SCHAFMAYER, C., SCHRAMM, K., SIEBERT, R., SLAGBOOM, P. E., SOININEN, P., STOLK, L., STRAUCH, K., TAI, E. S., TARANTINI, L., THORAND, B., TIGCHELAAR, E. F., TUMINO, R., UITTERLINDEN, A. G., VAN DUIJN, C., VAN MEURS, J. B. J., VINEIS, P., WICKREMASINGHE, A. R., WIJMENGA, C., YANG, T.-P., YUAN, W., ZHERNAKOVA, A., BATTERHAM, R. L., SMITH, G. D., DELOUKAS, P., HEIJMANS, B. T., HERDER, C., HOFMAN, A., LINDGREN, C. M., MILANI, L., VAN DER HARST, P., PETERS, A., ILLIG, T., RELTON, C. L., WALDENBERGER, M., JÄRVELIN, M.-R., BOLLATI, V., SOONG, R., SPECTOR, T. D., SCOTT, J., MCCARTHY, M. I., et al. 2017. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature,* 541**,** 81-86.

WANG, G. & SPEAKMAN, JOHN R. 2016. Analysis of Positive Selection at Single Nucleotide Polymorphisms Associated with Body Mass Index Does Not Support the “Thrifty Gene” Hypothesis. *Cell Metabolism,* 24**,** 531-541.

WANG, L., JANSEN, W., BOERE-BOONEKAMP, M. M., VLASBLOM, E., L'HOIR, M. P., BELTMAN, M., VAN GRIEKEN, A. & RAAT, H. 2019. Sleep and body mass index in infancy and early childhood (6-36 mo): a longitudinal study. *Pediatric obesity,* 14**,** e12506-e12506.

WARD, D. S., EVENSON, K. R., VAUGHN, A., RODGERS, A. B. & TROIANO, R. P. 2005. Accelerometer use in physical activity: best practices and research recommendations. *Medicine and science in sports and exercise,* 37**,** S582-8.

WARDLE, J., WALLER, J. & JARVIS, M. J. 2002. Sex differences in the association of socioeconomic status with obesity. *American journal of public health,* 92**,** 1299-1304.

WEAVER, I. C., CERVONI, N., CHAMPAGNE, F. A., D'ALESSIO, A. C., SHARMA, S., SECKL, J. R., DYMOV, S., SZYF, M. & MEANEY, M. J. 2004a. Epigenetic programming by maternal behavior. *Nat Neurosci,* 7**,** 847-54.

WEAVER, I. C. G., CERVONI, N., CHAMPAGNE, F. A., D'ALESSIO, A. C., SHARMA, S., SECKL, J. R., DYMOV, S., SZYF, M. & MEANEY, M. J. 2004b. Epigenetic programming by maternal behavior. *Nature Neuroscience,* 7**,** 847.

WEBER, D. R., MOORE, R. H., LEONARD, M. B. & ZEMEL, B. S. 2013. Fat and lean BMI reference curves in children and adolescents and their utility in identifying excess adiposity compared with BMI and percentage body fat. *The American journal of clinical nutrition,* 98**,** 49-56.

WELLS, J. C., WILLIAMS, J. E., CHOMTHO, S., DARCH, T., GRIJALVA-ETERNOD, C., KENNEDY, K., HAROUN, D., WILSON, C., COLE, T. J. & FEWTRELL, M. S. 2012. Body-composition reference data for simple and reference techniques and a 4-component model: a new UK reference child. *The American Journal of Clinical Nutrition,* 96**,** 1316-1326.

WELLS, J. C. K. 2000. A Hattori chart analysis of body mass index in infants and children. *International Journal of Obesity,* 24**,** 325-329.

WENG, S. F., REDSELL, S. A., SWIFT, J. A., YANG, M. & GLAZEBROOK, C. P. 2012. Systematic review and meta-analyses of risk factors for childhood overweight identifiable during infancy. *Archives of disease in childhood,* 97**,** 1019-1026.

WHINCUP, P. H., KAYE, S. J., OWEN, C. G., HUXLEY, R., COOK, D. G., ANAZAWA, S., BARRETT-CONNOR, E., BHARGAVA, S. K., BIRGISDOTTIR, B. E. & CARLSSON, S. 2008. Birth weight and risk of type 2 diabetes: a systematic review. *Jama,* 300**,** 2886-2897.

WHITE, H. E., DURSTON, V. J., HARVEY, J. F. & CROSS, N. C. 2006. Quantitative analysis of SRNPN gene methylation by Pyrosequencing as a diagnostic test for prader–willi syndrome and angelman syndrome. *Clinical chemistry,* 52**,** 1005-1013.

WIENCH, M., JOHN, S., BAEK, S., JOHNSON, T. A., SUNG, M. H., ESCOBAR, T., SIMMONS, C. A., PEARCE, K. H., BIDDIE, S. C., SABO, P. J., THURMAN, R. E., STAMATOYANNOPOULOS, J. A. & HAGER, G. L. 2011. DNA methylation status predicts cell type-specific enhancer activity. *Embo j,* 30**,** 3028-39.

WIJNDAELE, K., LAKSHMAN, R., LANDSBAUGH, J. R., ONG, K. K. & OGILVIE, D. 2009a. Determinants of early weaning and use of unmodified cow's milk in infants: a systematic review. *Journal of the American Dietetic Association,* 109**,** 2017-2028.

WIJNDAELE, K., LAKSHMAN, R., LANDSBAUGH, J. R., ONG, K. K. & OGILVIE, D. 2009b. Determinants of early weaning and use of unmodified cow's milk in infants: a systematic review. *J Am Diet Assoc,* 109**,** 2017-28.

WILLER, C. J., SPELIOTES, E. K., LOOS, R. J., LI, S., LINDGREN, C. M., HEID, I. M., BERNDT, S. I., ELLIOTT, A. L., JACKSON, A. U. & LAMINA, C. 2009. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature genetics,* 41**,** 25.

WOLFF, G. L., KODELL, R. L., MOORE, S. R. & COONEY, C. A. 1998. Maternal epigenetics and methyl supplements affect agouti gene expression in A(vy)/a mice. *Faseb Journal,* 12**,** 949-957.

WOO BAIDAL, J. A., LOCKS, L. M., CHENG, E. R., BLAKE-LAMB, T. L., PERKINS, M. E. & TAVERAS, E. M. 2016. Risk Factors for Childhood Obesity in the First 1,000 Days: A Systematic Review. *American Journal of Preventive Medicine,* 50**,** 761-779.

WORLD HEALTH ORGANIZATION 2000. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. *World Health Organ Tech Rep Ser,* 894**,** i-xii, 1-253.

WORLD HEALTH ORGANIZATION 2003. Recommendations for preventing excess weight gain and obesity. *Diet, Nutrition and the Prevention of Chronic Diseases. Joint WHO/FAO Expert Consultation. WHO Technical Report Series***,** 61-71.

WORLD HEALTH ORGANIZATION 2011a. Adverse childhood experiences international questionnaire. *Violence and Injury Prevention*.
WORLD HEALTH ORGANIZATION 2011b. Exclusive breastfeeding for six months best for babies everywhere. *World Health Organization*.
WORLD HEALTH ORGANIZATION 2011c. Waist circumference and waist-hip ratio: Report of a WHO expert consultation, Geneva, 8-11 December 2008.
WORLD HEALTH ORGANIZATION. 2019. *Obesity and overweight: World Health Organization* [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight [Accessed 20/08/2019].
WRIGHT, C., LAKSHMAN, R., EMMETT, P. & ONG, K. K. 2008a. Implications of adopting the WHO 2006 Child Growth Standard in the UK: two prospective cohort studies. *Arch Dis Child,* 93.
WRIGHT, C., MATTHEWS, J., WATERSTON, A. & AYNSLEY-GREEN, A. 1994. What is a normal rate of weight gain in infancy? *Acta Paediatrica,* 83**,** 351-356.
WRIGHT, C. M., BOOTH, I. W., BUCKLER, J. M. H., CAMERON, N., COLE, T. J., HEALY, M. J. R., HULSE, J. A., PREECE, M. A., REILLY, J. J. & WILLIAMS, A. F. 2002. Growth reference charts for use in the United Kingdom. *Archives of Disease in Childhood,* 86**,** 11.
WRIGHT, C. M., COX, K. M., SHERRIFF, A., FRANCO-VILLORIA, M., PEARCE, M. S. & ADAMSON, A. J. 2012. To what extent do weight gain and eating avidity during infancy predict later adiposity? *Public Health Nutrition,* 15**,** 656-662.
WRIGHT, C. M., EMMETT, P. M., NESS, A. R., REILLY, J. J. & SHERRIFF, A. 2010a. Tracking of obesity and body fatness through mid-childhood. *Archives of Disease in Childhood,* 95**,** 612-617.
WRIGHT, C. M., MARRYAT, L., MCCOLL, J., HARJUNMAA, U. & COLE, T. J. 2018. Pathways into and out of overweight and obesity from infancy to mid-childhood. *Pediatric Obesity,* 13**,** 621-627.
WRIGHT, C. M. & PARKER, L. 2004. Forty years on: the effect of deprivation on growth in two Newcastle birth cohorts. *International journal of epidemiology,* 33**,** 147-152.
WRIGHT, C. M., PARKINSON, K. & SCOTT, J. 2006a. Breast-feeding in a UK urban context: who breast-feeds, for how long and does it matter? *Public Health Nutr,* 9**,** 686-91.
WRIGHT, C. M., PARKINSON, K. N. & DREWETT, R. F. 2004. Why are babies weaned early? Data from a prospective population based cohort study. *Archives of Disease in Childhood,* 89**,** 813.
WRIGHT, C. M., PARKINSON, K. N. & DREWETT, R. F. 2006b. How does maternal and child feeding behavior relate to weight gain and failure to thrive? Data from a prospective birth cohort. *Pediatrics,* 117**,** 1262-1269.
WRIGHT, C. M., SHERRIFF, A., WARD, S. C., MCCOLL, J. H., REILLY, J. J. & NESS, A. R. 2008b. Development of bioelectrical impedance-derived indices of fat and fat-free mass for assessment of nutritional status in childhood. *Eur J Clin Nutr,* 62**,** 210-7.
WRIGHT, C. M., STONE, D. H. & PARKINSON, K. N. 2010b. Undernutrition in British Haredi infants within the gateshead millennium cohort study. *Archives of disease in childhood,* 95**,** 630-633.
WU, M. C., JOUBERT, B. R., KUAN, P.-F., HÅBERG, S. E., NYSTAD, W., PEDDADA, S. D. & LONDON, S. J. 2014a. A systematic assessment of normalization approaches for the Infinium 450K methylation platform. *Epigenetics,* 9**,** 318-329.
WU, Y., PATCHEV, A. V., DANIEL, G., ALMEIDA, O. F. X. & SPENGLER, D. 2014b. Early-Life Stress Reduces DNA Methylation of the Pomc Gene in Male Mice. *Endocrinology,* 155**,** 1751-1762.

YAN, J., LIU, L., ZHU, Y., HUANG, G. & WANG, P. P. 2014. The association between breastfeeding and childhood obesity: a meta-analysis. *BMC Public Health,* 14**,** 1.

YAN, W., BINGXIAN, H., HUA, Y., JIANGHONG, D., JUN, C., DONGLIANG, G., YUJIAN, Z., LING, L., YANYING, G. & KAITI, X. 2007. Waist-to-height ratio is an accurate and easier index for evaluating obesity in children and adolescents. *Obesity,* 15**,** 748-752.

YANG, W., KELLY, T. & HE, J. 2007. Genetic Epidemiology of Obesity. *Epidemiologic Reviews,* 29**,** 49-61.

YANG, Z., ZHAO, W., ZHANG, X., MU, R., ZHAI, Y., KONG, L. & CHEN, C. 2008. Impact of famine during pregnancy and infancy on health in adulthood. *Obesity Reviews,* 9**,** 95-99.

YLIHÈRSILÈ, H., KAJANTIE, E., OSMOND, C., FORSÉN, T., BARKER, D. J. & ERIKSSON, J. G. 2008. Body mass index during childhood and adult body composition in men and women aged 56–70 y. *The American Journal of Clinical Nutrition,* 87**,** 1769-1775.

YODER, J. A., WALSH, C. P. & BESTOR, T. H. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends in genetics,* 13**,** 335-340.

YOUNG, B. E., JOHNSON, S. L. & KREBS, N. F. 2012. Biological determinants linking infant weight gain and child obesity: current knowledge and future directions. *Advances in Nutrition: An International Review Journal,* 3**,** 675-686.

YU, Z. B., HAN, S. P., ZHU, G. Z., ZHU, C., WANG, X. J., CAO, X. G. & GUO, X. R. 2011. Birth weight and subsequent risk of obesity: a systematic review and meta-analysis. *Obesity Reviews,* 12**,** 525-542.

ZHANG, D.-D., ZHANG, J.-G., WANG, Y.-Z., LIU, Y., LIU, G.-L. & LI, X.-Y. 2015. Per-Arnt-Sim kinase (PASK): an emerging regulator of mammalian glucose and lipid metabolism. *Nutrients,* 7**,** 7437-7450.

ZHENG, M., LAMB, K. E., GRIMES, C., LAWS, R., BOLTON, K., ONG, K. K. & CAMPBELL, K. 2018. Rapid weight gain during infancy and subsequent adiposity: a systematic review and meta-analysis of evidence. *Obesity Reviews,* 19**,** 321-332.

ZUCKERMAN, B., STEVENSON, J. & BAILEY, V. 1987. Sleep problems in early childhood: continuities, predictive factors, and behavioral correlates. *Pediatrics,* 80**,** 664-71.

ZWEINIGER-BARGIELOWSKA, I. 2000. *Austerity in Britain: rationing, controls, and consumption, 1939-1955*, OUP Oxford.

Appendices

Appendix A

Directed Acyclic Graph (DAG) of the hypothesised relationships from the literature review

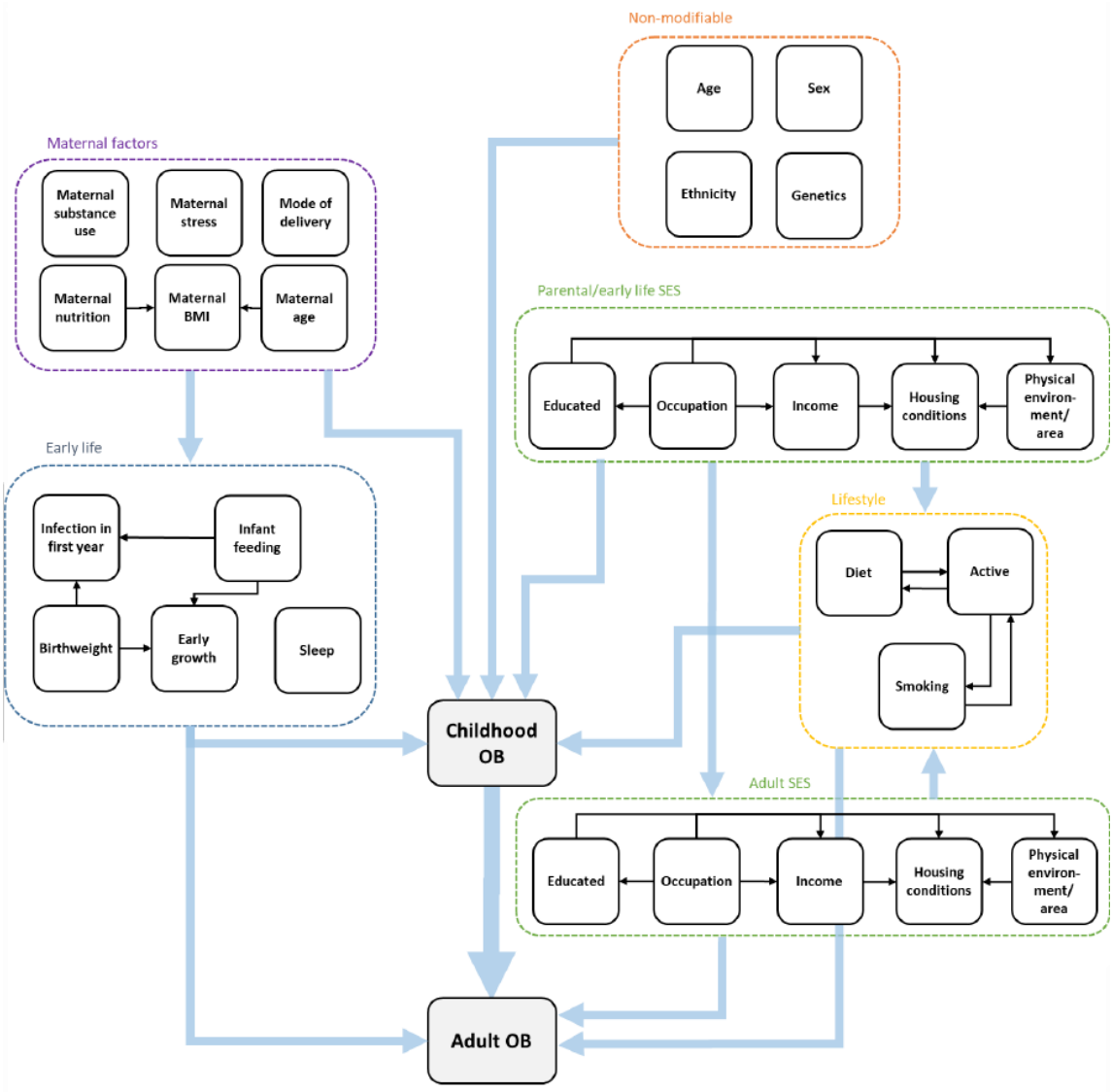


Figure I The hypothesised pathways between exposures, covariates and obesity outcomes across the lifecourse

Energy intake data for NTFS participants (age 50)

Overall, there was a small negative correlation (r=~-0.1) between daily energy intake and BMI. When stratified by sex there was no relationship for males, and a negative correlation between energy intake and BMI in females (Figure II).

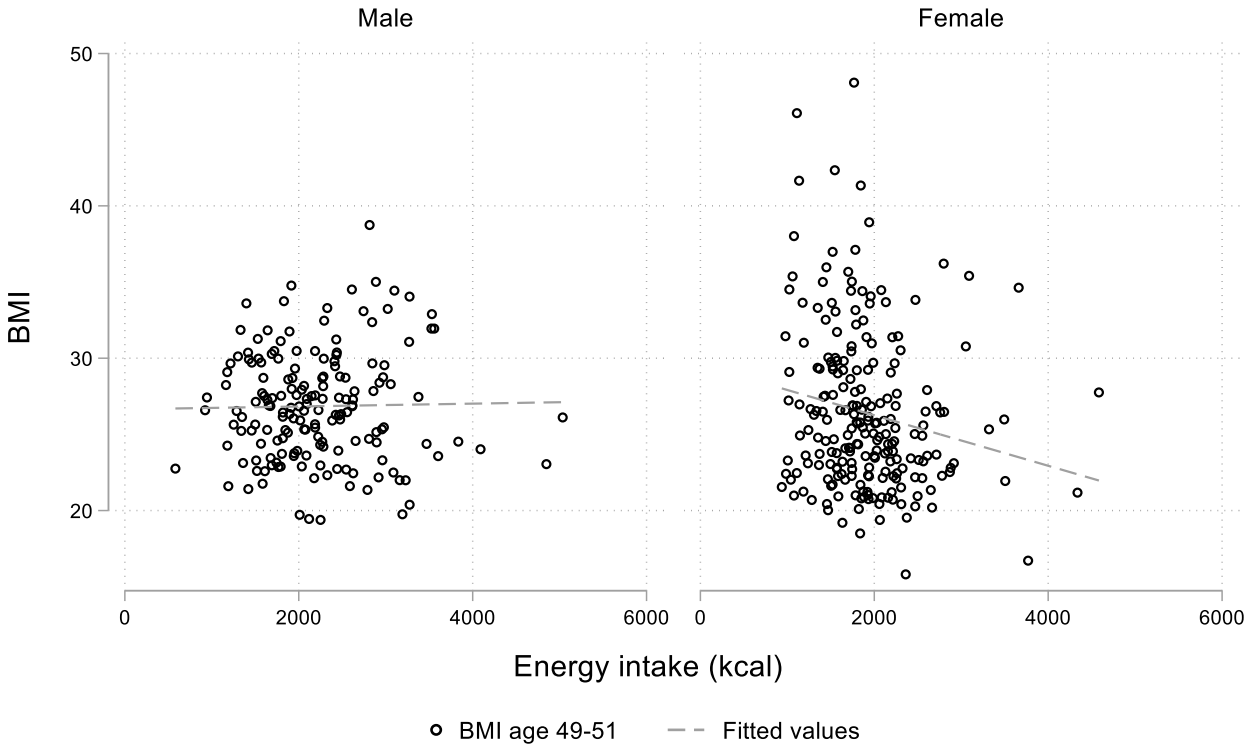


Figure II Scatter plot showing the relationship between BMI and energy intake in males and females. Correlation in males, r=0.02, p=0.8; females, r=-0.18, p=0.01.

In women, mean energy intake decreased as weight category progressed from healthy weight to overweight to obese. Mean energy intake in overweight men was lower compared to healthy weight men (Table I). These results state that those who are overweight or obese on average have a lower calorie intake according to FFQ data. This may be due to the cross-sectional nature of the data, which could suggest that those with the highest BMI or BF% are on weight loss diets. Underreporting has also been noted in smokers (Johansson et al., 2007), however in this cohort there was no correlation between energy intake and smoking (Appendix B).

Table I Energy intake (EI) by weight category stratified by sex in NTFS (age 50)

Weight category	Males			Females		
	N	Mean EI	SD	N	Mean EI	SD
Healthy	54	2319	801	103	1947	543
Overweight	84	2100	626	68	1921	586
Obese	31	2346	708	46	1789	563

Standard deviation (SD) and group size (N)

In order to investigate further, basal metabolic rate (BMR) was calculated using the Harris Benedict equation (Harris and Benedict, 1918) which takes into account height, weight, sex and age, revised by Mifflin (1990).

$$For\ men,$$

$$BMR\ =\ 9.99\ \text{x}\ \text{weight (kg)} +\ 6.25\ \text{x height (cm)} -\ 4.92\ \text{x age (years)}\ +\ 5$$

$$For\ women,$$

$$BMR\ =\ 9.99\ \text{x}\ \text{weight(kg)}\ +\ 6.25\ \text{x height (cm)}\ -\ 4.92\ \text{x age (years)}\ -\ 161$$

Equation I BMR equations for men and women revised by (Mifflin et al., 1990)

Total daily energy expenditure (TDEE) was calculated using the physical activity data according to the calculation factors in Table II.

Table II Calculation factor for varying levels of physical activity

Little to no exercise	Daily kilocalories needed = BMR x 1.2
Light exercise	Daily kilocalories needed = BMR x 1.375
Moderate exercise	Daily kilocalories needed = BMR x 1.55
Heavy exercise	Daily kilocalories needed = BMR x 1.725

Using these equations, TDEE was determined and the difference between this value and the value calculated from the food intake data was calculated. Negative values indicate a calorie deficit (i.e. more calories expended than ingested) and positive values an energy surplus (more energy ingested than expended). These results show that those with the highest BMI report an energy deficit (Figure III). Similar results were seen for BF% (Figure IV) and WHR

(Figure V). For each outcome there was a negative correlation in females, which is moderate and significant for BMI, with a weak negative correlation and borderline significant for BF% and WHR. Relationships between the difference in energy intake and anthropometric outcomes were less prominent for males.

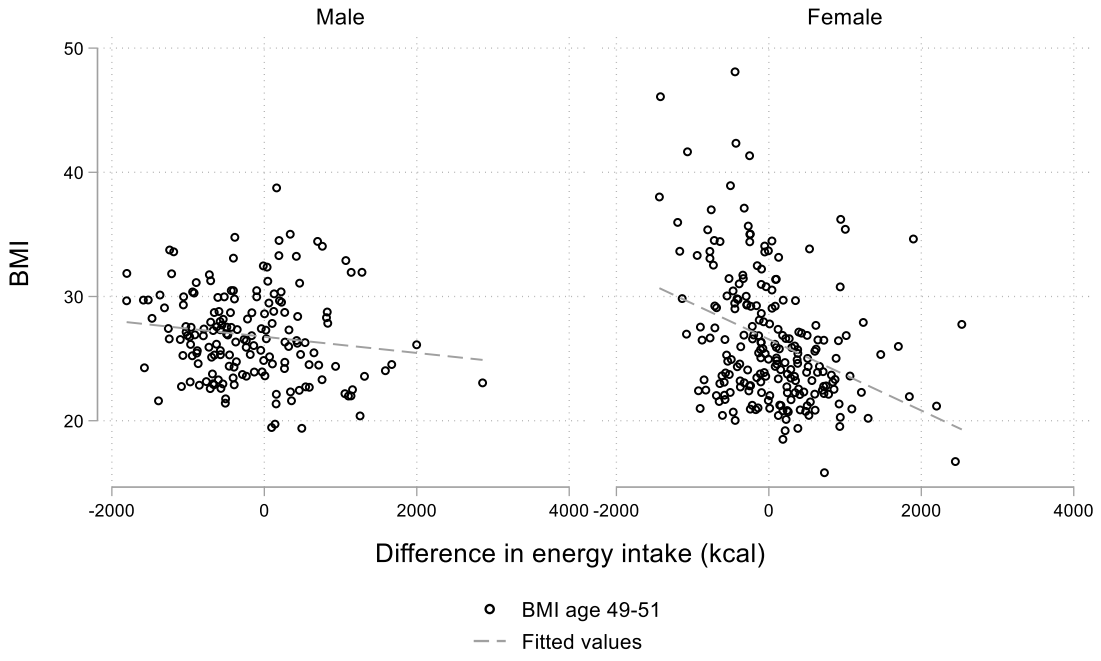


Figure III Scatter plot of the relationship between BMI and differences in energy intake.
Scatter plot with linear fit showing the relationship between BMI and the differences in energy intake (reported calorie intake-TDEE) in males and females. Correlation in males, r=-0.14, p=0.07; females, r=-0.35, p<0.001.

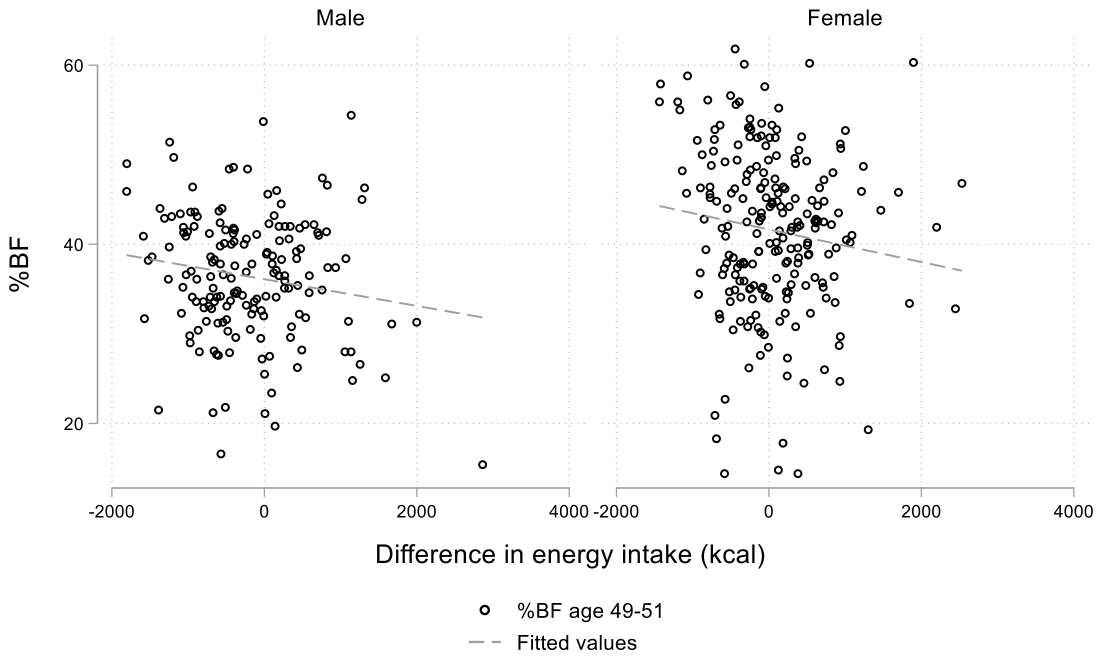


Figure IV Scatter plot of the relationship between BF% and difference in energy intake.
Scatter plot with linear fit showing the relationship between BF% and the differences in energy intake (reported calorie intake-TDEE) in males and females. Correlation in males, r=-0.16, p=0.04; females, r=-0.13, p=0.06.

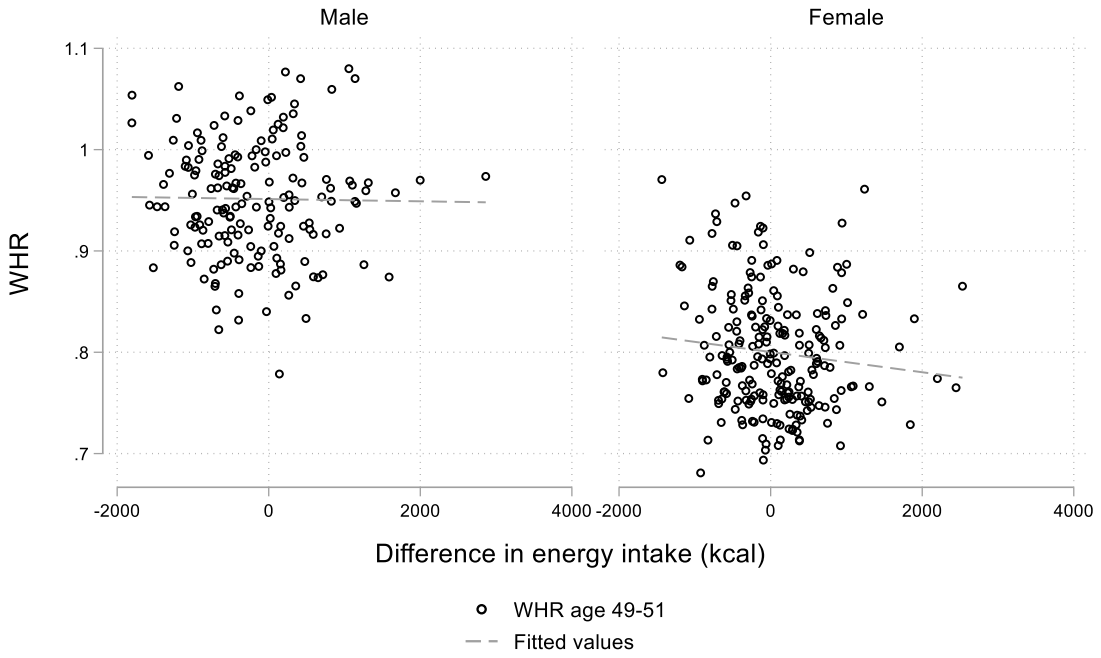


Figure V Scatter plot of the relationship between WHR and difference in energy intake.
Scatter plot with linear fit showing the relationship between WHR and the differences in energy intake (reported calorie intake-TDEE) in males and females. Correlation in males, r=-0.01, p=0.8; females, r=-0.11, p=0.1.

These differences could have been due to those with higher body weights endeavouring to reduce their calorie intake. Many studies report that that overweight and obesity are factors in underreporting of energy intake (Johnson et al., 1998, Johansson et al., 2001). Additionally there is evidence that underreporting is more likely in women (Pikholz et al., 2004).

Reversing the calculations in Table II (using energy intake divided by BMR as an estimate of physical activity level) gives values ranging from 0.42 to 3.92 for study participants. Taking a physical activity level (PAL) value of 1.2 to indicate bed-bound and a value of 2.4 as strenuous activity, values that fall outside of this range would be outliers. Removing these observations would lead to the exclusion of 135 study members from analyses.

On this basis of these issues with the data, the energy intake data were not used in these analyses. Firstly, it would give biased estimates if the variables were inaccurately measured, and secondly the direction of the association (between energy intake and body weight) cannot be discerned with these cross-sectional data and endogeneity bias. As energy intake was not analysed in NTFS, for coherence this was also not analysed in GMS.

Appendix B

Correlations between explanatory variables in NTFS

	Female	Maternal age	Gestation length (weeks)	Birthweight (z-score)	RWG	RT	Breastfed (days)	Exclusively breastfed (days)	Wean age	Bacterial infection	Viral infection	No. of infections	Adversity	Housing problems	Social class (birth)	Social class (age 50)	Education level	PAL (age 50)	Energy intake (age 50)	Smoker (age 50)	Married (age 50)
Female	1.00																				
Maternal age	0.02 0.555	1.00																			
Gestation	-0.04 0.158	0.02 0.605	1.00																		
Birthweight (z-score)	0.08 0.008	0.08 0.009	-0.17 <0.001	1.00																	
RWG	0.00 0.969	-0.14 0.010	0.11 0.031	-0.36 <0.001	1.00																
RT	0.07 0.164	-0.08 0.127	0.07 0.185	0.00 0.982	0.67 <0.001	1.00															
Breastfed (days)	0.03 0.528	0.04 0.388	-0.03 0.603	0.11 0.022	-0.29 <0.001	-0.24 0.002	1.00														
Exclusively breastfed	0.06 0.214	-0.05 0.318	-0.04 0.418	0.08 0.092	-0.20 0.006	-0.22 0.004	0.66 <0.001	1.00													
Wean age	0.03 0.588	0.07 0.162	-0.09 0.076	-0.05 0.349	0.05 0.550	0.03 0.679	-0.05 0.336	0.10 0.034	1.00												
Bacterial infection	-0.04 0.256	0.05 0.166	0.00 0.912	0.00 0.926	-0.01 0.854	-0.06 0.342	-0.12 0.026	-0.09 0.103	-0.02 0.665	1.00											
Viral infection	-0.02 0.596	0.04 0.306	0.00 0.891	-0.01 0.877	-0.07 0.271	-0.05 0.427	-0.08 0.152	-0.08 0.151	0.01 0.864	0.02 0.513	1.00										
No. of infections	-0.06 0.077	0.01 0.688	0.06 0.057	0.03 0.413	-0.12 0.023	-0.08 0.111	-0.12 0.013	-0.17 0.001	-0.03 0.551	0.29 <0.001	0.23 <0.001	1.00									
Adversity	0.04 0.495	-0.01 0.839	0.02 0.740	-0.09 0.076	-0.05 0.584	0.02 0.807	-0.01 0.867	-0.10 0.223	0.12 0.165	-0.01 0.926	0.08 0.191	0.10 0.067	1.00								
Housing problems	0.00 0.951	-0.08 0.015	0.04 0.169	-0.01 0.690	-0.11 0.035	-0.12 0.028	0.02 0.631	-0.01 0.909	-0.13 0.009	0.06 0.117	0.02 0.542	0.08 0.012	0.07 0.172	1.00							
Social class (birth)	-0.02 0.605	0.08 0.009	0.04 0.280	0.02 0.522	0.03 0.602	0.06 0.268	0.07 0.188	0.04 0.434	0.09 0.077	-0.07 0.049	-0.03 0.385	-0.10 0.004	-0.14 0.011	-0.31 <0.001	1.00						
Social class (age 50)	0.00 0.938	-0.03 0.564	-0.02 0.664	0.06 0.180	-0.03 0.613	0.08 0.218	0.05 0.351	0.03 0.591	0.06 0.215	-0.04 0.395	0.05 0.306	-0.05 0.263	-0.06 0.452	-0.19 <0.001	0.32 <0.001	1.00					
Education level	-0.21 <0.001	-0.01 0.823	0.04 0.347	0.03 0.427	0.00 0.989	0.03 0.614	0.15 0.003	0.12 0.017	-0.02 0.744	-0.11 0.026	-0.08 0.115	-0.09 0.053	-0.16 0.028	-0.20 <0.001	0.35 <0.001	0.42 <0.001	1.00				
PAL (age 50)	-0.01 0.846	0.03 0.532	-0.02 0.680	0.01 0.889	-0.10 0.224	-0.04 0.642	0.00 0.944	-0.01 0.810	0.02 0.746	-0.10 0.080	0.03 0.580	-0.06 0.209	-0.15 0.078	-0.09 0.068	0.12 0.024	0.24 <0.001	0.20 <0.001	1.00			
Energy intake (age	-0.18 <0.001	0.06 0.140	0.03 0.481	0.01 0.810	0.08 0.229	-0.03 0.615	0.02 0.671	0.00 0.967	-0.09 0.086	0.00 0.963	0.01 0.872	0.00 0.919	-0.01 0.863	-0.02 0.640	0.01 0.738	-0.11 0.015	0.01 0.782	-0.02 0.769	1.00		
Smoker (age 50)	0.05 0.206	0.02 0.623	0.07 0.104	-0.03 0.548	0.10 0.124	0.04 0.526	-0.01 0.858	0.03 0.595	-0.06 0.274	-0.03 0.546	0.03 0.513	0.00 0.950	0.14 0.053	0.13 0.003	-0.19 <0.001	-0.22 <0.001	-0.25 <0.001	-0.26 <0.001	0.02 0.731	1.00	
Married (age 50)	-0.02 0.615	-0.05 0.255	0.00 0.931	-0.04 0.311	-0.01 0.912	0.06 0.361	0.04 0.458	0.00 0.948	-0.07 0.140	0.01 0.797	0.00 0.981	0.04 0.327	-0.09 0.192	-0.02 0.724	-0.03 0.558	0.09 0.037	-0.07 0.138	0.08 0.096	-0.06 0.199	-0.17 <0.001	1.00

Distributions of outcome variables in NTFS (age 50)

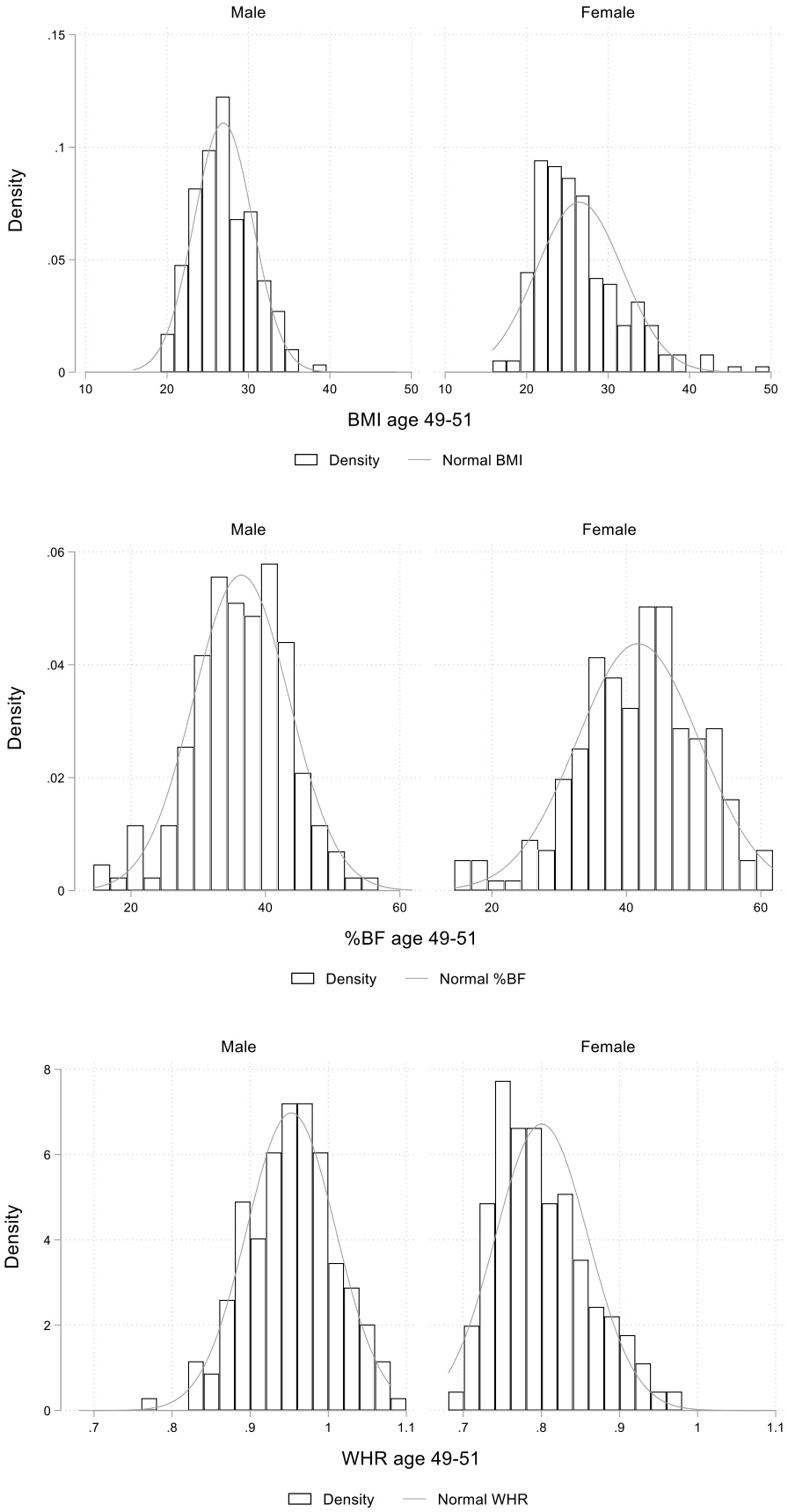


Figure VII Distribution of continuous outcome measures (from top: BMI, BF%, WHR) stratified by sex.

Social mobility (birth to age 50) and body composition (age 50) in NTFS

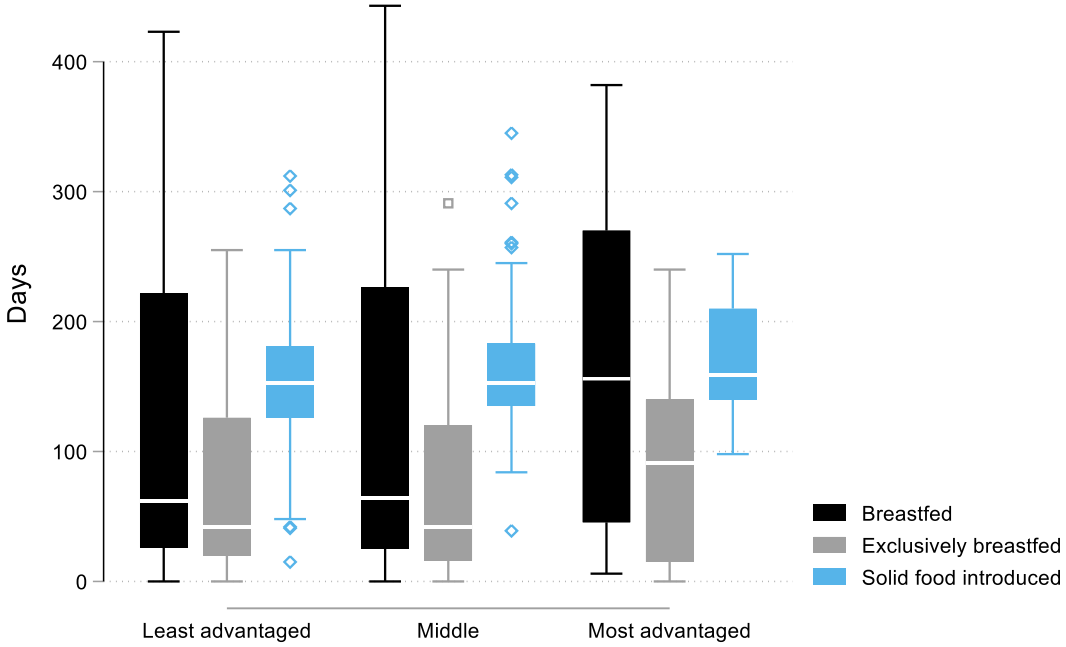
Table III Linear (BMI) and logistic (OB) regression models for social mobility from birth to age 50 in NTFS

			BMI			Obesity		
	n	%	Coef	CI	p	OR	CI	p
Always least advantaged	24	7.0	Ref		.	Ref		.
Least--> mid	43	12.5	-0.15	[-2.34,2.04]	0.892	0.68	[0.24,1.90]	0.457
Least--> most advantaged	40	11.6	-1.26	[-3.48,0.96]	0.265	0.20	[0.06,0.69]	0.011
Always mid	77	22.4	-2.55	[-4.56,-0.54]	0.013	0.10	[0.05,0.48]	0.001
Mid--> most advantaged	128	37.2	-1.19	[-3.10,0.72]	0.223	0.36	[0.14,0.89]	0.028
Always most advantaged	32	9.3	-2.12	[-4.44,0.20]	0.073	0.09	[0.02,0.48]	0.005
n	344		344			344		
Adjusted R^2			0.021			0.063		

Coefficients (coef) and odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p)

Socioeconomic differences in breastfeeding in NTFS

In the UK, socioeconomic differences in breastfeeding are often described. In NTFS, the most advantaged group had higher median values for duration breastfeeding, exclusive breastfeeding and introduction of solids compared to lower groups however these differences were not significant (Figure VIII).



	Least advantaged			Middle			Most advantaged		
Statistics	p25	p50	p75	p25	p50	p75	p25	p50	p75
Duration breastfed	12.5	74	163	21	83	237	17.5	137	257
Exclusively breastfed	20	42	126	16	42	120	15	91	140
Solid foods	126	153	181	135	153	183	140	159	210

Figure VIII Box plot of infant feeding by SES at birth in NTFS study members.
Table of values (rounded) for lower quartile (p25), median (p50) and upper quartile (p75).
There were no significant differences in infant feeding by occupational social class (Kruskal-Wallis equality-of-populations rank test (duration of exclusive/breastfeeding) and ANOVA (solid food introduction) p values>0.1).

Weight categories and qualifications in NTFS (age 50)

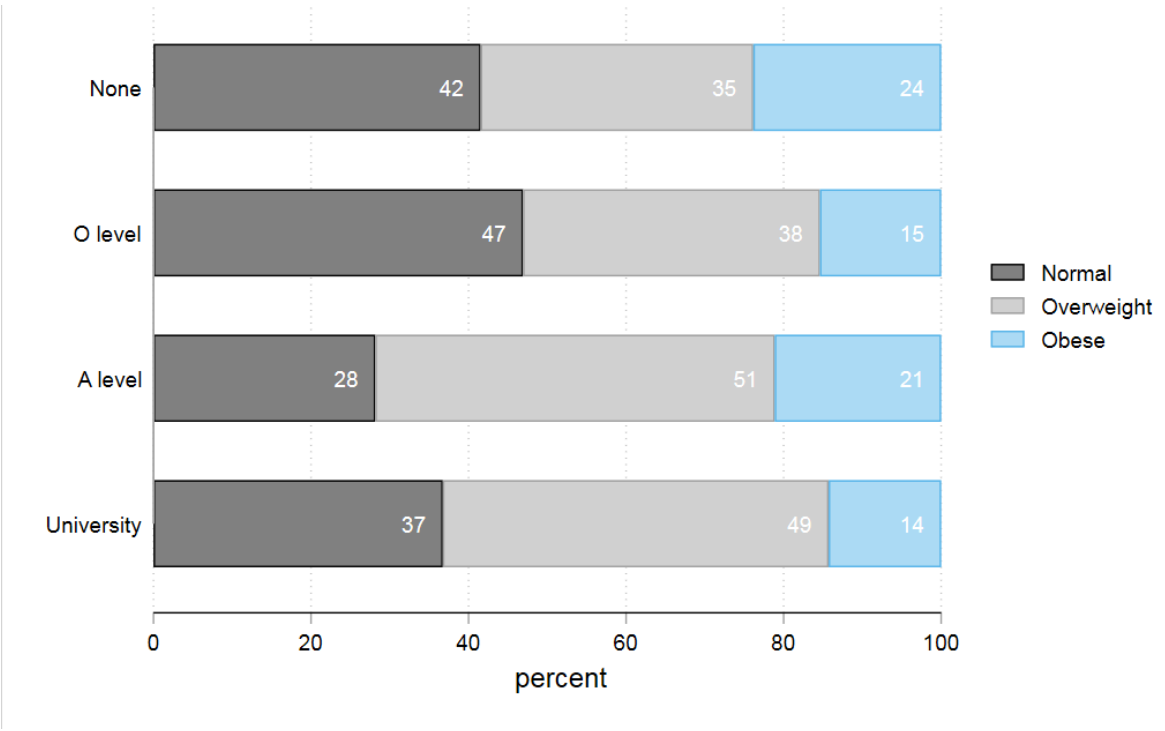


Figure IX Proportion of weight categories by education level achieved at age 50 in NTFS study members.
A large proportion of those without qualifications had obesity, whereas those educated to degree level had the lowest proportion with obesity. However around half of those with A level qualifications and degrees were overweight (Figure IX).
Chi^2 test for differences; Pearson chi2(6) = 11.4717 Pr = 0.075)

Models for bivariate associations in NTFS for weight outcomes (age 50)

Table IV Bivariate linear associations between exposures and BF%, stratified by sex, in NTFS study members (age 50)

		Males			Females		
		coef	CI	p	coef	CI	p
Sex	Female						
Early life	Birthweight (z-score)	-0.52	[-1.67,0.63]	0.37	-0.23	[-1.31,0.86]	0.68
	Gestation (weeks)	0.48	[-0.84,1.80]	0.48	0.08	[-1.02,1.18]	0.89
Maternal age	Continuous	-0.09	[-0.28,0.10]	0.34	-0.12	[-0.32,0.07]	0.22
	<25	-0.88	[-3.39,1.63]	0.49	0.57	[-2.21,3.35]	0.69
	25-34	Ref		.	Ref		.
	35+	-2.23	[-5.12,0.66]	0.13	-2.83	[-6.08,0.42]	0.09
Infant feeding	Never breastfed	Ref		.	Ref		.
	<4 weeks	0.88	[-4.03,5.78]	0.73	-1.39	[-6.77,4.00]	0.61
	4 wk – 6 months	1.05	[-3.25,5.36]	0.63	-0.09	[-4.30,4.12]	0.97
	6 Months +	0.38	[-3.69,4.45]	0.85	2.78	[-1.24,6.80]	0.18
	Breastfed (days)	0.01	[-0.00,0.02]	0.27	0.01	[0.00,0.03]	0.03
	Exclusive (days)	0.00	[-0.02,0.02]	0.96	0.03	[0.00,0.05]	0.02
	Weaning age (days)	0.00	[-0.02,0.03]	0.78	0.00	[-0.03,0.03]	0.92
Early life	Rapid weight gain	-0.34	[-3.99,3.31]	0.85	1.99	[1.90,5.88]	0.317
	Any infection	1.26	[-0.86,3.39]	0.24	1.84	[-0.59,4.27]	0.14
	Number of infections	0.40	[-0.48,1.27]	0.37	0.22	[-0.93,1.36]	0.71
	Bacterial infection	0.29	[-2.67,3.26]	0.85	4.48	[0.70,8.27]	0.02
	Viral infection	-1.53	[-4.10,1.03]	0.24	1.47	[-1.56,4.49]	0.34
SES (childhood)	Least advantaged	Ref		.	Ref		.
	Mid	-2.56	[-4.97,-0.14]	0.04	-1.09	[-3.89,1.71]	0.45
	Most advantaged	-4.24	[-7.78,-0.70]	0.02	-1.61	[-6.26,3.04]	0.50
	Housing score	0.53	[-0.46,1.53]	0.30	0.43	[-0.68,1.54]	0.45
	Overcrowding	1.04	[-1.30,3.37]	0.39	2.25	[-0.40,4.91]	0.10
Adversity	Any adverse event	1.04	[-3.93,6.01]	0.68	-0.45	[-5.85,4.95]	0.87
SES (later life)	Social class	Ref		.	Ref		.
	Mid	1.78	[-1.80,5.36]	0.33	-0.61	[-4.47,3.26]	0.76
	Most advantaged	3.82	[0.33,7.31]	0.03	-2.28	[-5.88,1.31]	0.21
	Educated	3.25	[1.10,5.39]	<0.001	-0.87	[-3.70,1.95]	0.55
	No qualifications	Ref		.	Ref		.
	GCSE/O-level	-0.78	[-3.55,1.98]	0.58	-1.93	[-4.72,0.86]	0.18
	A level	2.82	[-0.14,5.77]	0.06	-0.25	[-4.03,3.52]	0.90
	Degree	2.81	[-0.55,6.18]	0.10	-3.81	[-7.98,0.37]	0.07
	Income	1.17	[0.12,2.23]	0.03	-1.12	[-2.33,0.10]	0.07
	Married	3.60	[0.85,6.35]	0.01	1.78	[-1.10,4.66]	0.23
Lifestyle	Inactive	Ref		.	Ref		.
	Light activity	-0.91	[-5.00,3.18]	0.66	-3.13	[-6.82,0.57]	0.10
	Moderate activity	-1.48	[-5.95,2.98]	0.52	-4.36	[-8.40,-0.31]	0.04
	Heavy activity	-1.72	[-6.38,2.94]	0.47	-6.53	[-10.91,-2.15]	<0.001
	Smoker	-4.02	[-6.40,-1.65]	<0.001	0.05	[-2.62,2.71]	0.97

Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p); Ref, reference group. Bold indicates significant at p<0.05.

Table V Bivariate linear associations between risk factors and WHR in NTFS males and females (age 50)

		Males			Females		
		coef	CI	p	coef	CI	p
Early life	Birthweight (z-score)	-0.01	[-0.02,-0.00]	0.045	0.00	[-0.01,0.01]	0.91
	Gestation (weeks)	0.00	[-0.01,0.01]	0.93	0.00	[-0.01,0.01]	0.93
Maternal age	Continuous	0.00	[-0.00,0.00]	0.14	0.00	[-0.00,0.00]	0.48
	<25	0.02	[-0.00,0.04]	0.12	0.01	[-0.01,0.02]	0.52
	25-34	Ref		.	Ref		.
	35+	0.00	[-0.02,0.02]	0.99	0.00	[-0.02,0.02]	0.95
Infant feeding	Never breastfed	Ref		.	Ref		.
	<4 weeks	0.01	[-0.03,0.04]	0.67	0.03	[-0.01,0.06]	0.14
	4 wk – 6 months	0.02	[-0.01,0.05]	0.26	-0.01	[-0.04,0.02]	0.61
	6 Months +	0.00	[-0.03,0.03]	0.87	0.00	[-0.02,0.03]	0.84
	Breastfed (days)	0.00	[-0.00,0.00]	0.96	0.00	[-0.00,0.00]	0.80
	Exclusive (days)	0.00	[-0.00,0.00]	0.11	0.00	[-0.00,0.00]	0.51
	Weaning age (days)	0.00	[-0.00,0.00]	0.22	-0.00	[-0.00,0.00]	0.09
Early life	Rapid weight gain	-0.02	[-0.05,0.01]	0.14	0.00	[-0.02,0.02]	0.96
	Any infection	0.00	[-0.01,0.02]	0.73	0.01	[-0.00,0.03]	0.07
	Number of infections	0.00	[-0.01,0.01]	0.63	0.00	[-0.00,0.01]	0.20
	Bacterial infection	0.01	[-0.02,0.03]	0.53	0.02	[-0.01,0.04]	0.13
	Viral infection	-0.02	[-0.04,0.00]	0.09	0.02	[-0.00,0.04]	0.06
SES (childhood)	Least advantaged	Ref		.	Ref		.
	Mid	-0.02	[-0.04,-0.00]	0.02	-0.01	[-0.03,0.01]	0.46
	Most advantaged	-0.05	[-0.08,-0.02]	<0.001	-0.01	[-0.05,0.02]	0.34
	Housing score	0.00	[-0.00,0.01]	0.25	0.00	[-0.01,0.01]	0.88
	Overcrowding	0.01	[-0.01,0.03]	0.22	0.01	[-0.01,0.03]	0.29
Adversity	Any adverse event	0.01	[-0.02,0.05]	0.50	0.00	[-0.04,0.04]	0.91
SES (later life)	Social class	Ref		.	Ref		.
	Mid	-0.02	[-0.05,0.01]	0.15	-0.03	[-0.05,-0.00]	0.03
	Most advantaged	-0.03	[-0.06,-0.00]	0.047	-0.01	[-0.04,0.01]	0.24
	Educated	-0.01	[-0.03,0.00]	0.10	-0.01	[-0.02,0.01]	0.47
	No qualifications	Ref		.	Ref		.
	GCSE/O-level	-0.02	[-0.04,0.00]	0.06	-0.01	[-0.03,0.00]	0.14
	A level	-0.02	[-0.04,0.01]	0.12	-0.02	[-0.04,0.01]	0.16
	Degree	-0.04	[-0.07,-0.02]	0.002	-0.01	[-0.03,0.02]	0.60
	Income	-0.01	[-0.02,-0.00]	0.03	-0.01	[-0.02,-0.00]	0.02
	Married	0.00	[-0.02,0.02]	0.96	0.00	[-0.02,0.02]	0.71
Lifestyle	Inactive	Ref		.	Ref		.
	Light activity	0.00	[-0.03,0.04]	0.79	0.00	[-0.03,0.02]	0.78
	Moderate activity	-0.03	[-0.06,0.01]	0.14	0.00	[-0.03,0.02]	0.72
	Heavy activity	-0.02	[-0.06,0.01]	0.19	-0.02	[-0.05,0.01]	0.14
	Smoker	0.00	[-0.02,0.02]	0.72	0.02	[0.00,0.04]	0.02

Reference category for SES was least advantaged. Coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p); Ref, reference group. Bold indicates significant at p<0.05.

Smoking sensitivity analyses

Table VI Regression sensitivity analyses for categories of smoking and pack years and body composition outcomes in NTFS study members (age 50)

Unadjusted		All			Males			Females		
	Smoking status	Coef/ OR	CI	p	Coef/ OR	CI	p	Coef/ OR	CI	p
BMI	Never	Ref		.						
	Ex	0.71	[-0.33,1.75]	0.18						
	Current	-1.02	[-2.10,0.07]	0.066						
	Pack years	0	[-0.03,0.03]	0.90						
Obese	Never	Ref		.						
	Ex	1.02	[0.58,1.79]	0.94						
	Current	0.56	[0.29,1.09]	0.088						
	Pack years	1	[0.98,1.01]	0.76						
WHR	Never	Ref		.	Ref		.	Ref		.
	Ex	0.05	[0.03,0.07]	<0.001	0.03	[0.01,0.05]	0.003	0.01	[-0.01,0.03]	0.38
	Current	0.03	[0.01,0.05]	0.012	0.02	[-0.00,0.04]	0.094	0.02	[0.01,0.04]	0.008
	Pack years	0.00	[0.00,0.00]	<0.001	0.00	[0.00,0.00]	0.006	0.00	[0.00,0.00]	0.004
BF%	Never	Ref		.	Ref		.	Ref		.
	Ex	0.43	[-1.58,2.43]	0.68	0.67	[-1.67,3.01]	0.58	1.98	[-0.97,4.93]	0.19
	Current	-1.25	[-3.35,0.84]	0.24	-3.89	[-6.53,-1.25]	0.004	0.89	[-1.92,3.70]	0.54
	Pack years	-0.04	[-0.09,0.01]	0.15	-0.04	[-0.09,0.02]	0.19	0.07	[-0.03,0.17]	0.17
Adjusted model		All			Males			Females		
	Smoking status	Coef/ OR	CI	p	Coef/ OR	CI	p	Coef/ OR	CI	p
BMI	Never	0	Ref	.						
	Ex	0.81	[-0.25,1.87]	0.14						
	Current	-1.19	[-2.32,-0.05]	0.041						
	Pack years	0	[-0.03,0.03]	0.86						
Obese	Never	1	[1.00,1.00]	.						
	Ex	0.94	[0.53,1.69]	0.84						
	Current	0.49	[0.24,0.98]	0.042						
	Pack years	0.99	[0.98,1.01]	0.39						
WHR	Never	Ref		.	Ref		.	Ref		.
	Ex	0.05	[0.03,0.07]	<0.001	0.03	[0.01,0.05]	0.008	0.01	[-0.01,0.03]	0.28
	Current	0.02	[-0.00,0.05]	0.052	0.01	[-0.01,0.04]	0.30	0.02	[-0.00,0.03]	0.086
	Pack years	0.00	[0.00,0.00]	<0.001	0.00	[0.00,0.00]	0.023	0.00	[0.00,0.00]	0.02
BF%	Never	Ref		.	Ref		.	Ref		.
	Ex	0.46	[-1.62,2.53]	0.67	0.76	[-1.63,3.15]	0.53	2.18	[-0.90,5.25]	0.17
	Current	-1.62	[-3.85,0.62]	0.16	-3.22	[-6.03,-0.40]	0.025	0.25	[-2.76,3.25]	0.87
	Pack years	-0.04	[-0.10,0.02]	0.17	0	[-0.07,0.06]	0.88	0.06	[-0.05,0.17]	0.28

Unadjusted models (top) and adjusted (for adult SES, bottom) presented for each outcome. Coefficients (coef) for linear outcomes or odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Bold indicates p<0.05.

Appendix C

Correlations between explanatory variables in GMS

	Female	Maternal age	First-born	Caesarean	Gestation	Bwt-z	RWG	RT	Breastfeeding categories	Exclusive breastfeeding	Formula-fed	Wean age	Any infection	Adversity	Sleep issues	Townsend	Maternal education	Deprived	SES (age 8)	Upward mobility	% MVPA	Season activity
Female	1.00																					
Maternal age	-0.06	1.00																				
	0.083																					
First-born	0.01	-0.36	1.00																			
	0.659	<0.001																				
Caesarean	0.00	0.19	0.06	1.00																		
	0.968	<0.001	0.053																			
Gestation	0.03	-0.04	0.08	-0.18	1.00																	
	0.412	0.236	0.012	<0.001																		
Bwt-z	0.00	0.13	-0.12	0.08	-0.10	1.00																
	0.983	<0.001	<0.001	0.011	0.001																	
RWG	-0.02	-0.07	0.15	-0.02	0.12	-0.38	1.00															
	0.664	0.040	<0.001	0.510	0.001	<0.001																
RT	0.01	-0.01	0.14	0.03	0.07	-0.03	0.66	1.00														
	0.860	0.690	<0.001	0.465	0.053	0.333	<0.001															
Breastfeeding categories	-0.03	0.31	0.03	0.04	0.05	0.10	-0.07	-0.02	1.00													
	0.351	<0.001	0.432	0.193	0.144	0.001	0.056	0.618														
Exclusive breastfeeding	-0.01	0.21	-0.13	-0.01	0.05	0.10	-0.09	-0.09	1.00	1.00												
	0.821	<0.001	0.004	0.793	0.262	0.022	0.061	0.068	<0.001													
Formula-fed	0.03	-0.26	-0.10	-0.04	-0.05	-0.08	0.02	-0.02	-0.82	-1.00	1.00											
	0.321	<0.001	0.002	0.247	0.114	0.015	0.586	0.601	<0.001	1.000												
Wean age	0.08	0.13	-0.07	0.01	-0.04	-0.02	-0.06	-0.04	0.20	0.34	-0.14	1.00										
	0.036	<0.001	0.066	0.778	0.267	0.678	0.098	0.260	<0.001	<0.001	<0.001											
Any infection	-0.06	0.04	-0.08	0.02	-0.04	-0.01	0.07	0.05	0.05	0.00	0.00	-0.04	1.00									
	0.068	0.235	0.009	0.525	0.231	0.695	0.059	0.148	0.096	0.983	0.948	0.299										
Adversity	-0.03	0.02	0.07	0.05	0.03	0.00	0.04	0.03	0.03	-0.03	-0.04	-0.13	0.07	1.00								
	0.440	0.451	0.040	0.156	0.423	0.901	0.222	0.413	0.436	0.582	0.192	0.001	0.041									
Sleep issues	-0.01	-0.03	0.02	0.00	0.02	0.02	0.00	-0.03	-0.07	-0.10	0.04	-0.10	-0.01	-0.02	1.00							
	0.815	0.526	0.651	0.911	0.603	0.684	0.924	0.466	0.074	0.074	0.281	0.016	0.734	0.660								
Townsend	-0.06	0.18	-0.06	0.02	0.04	0.07	-0.06	-0.03	0.23	0.14	-0.20	0.08	-0.01	0.00	-0.07	1.00						
	0.078	<0.001	0.074	0.556	0.171	0.020	0.083	0.472	<0.001	0.001	<0.001	0.028	0.672	0.881	0.084							
Maternal education	-0.04	0.29	0.09	0.10	0.08	0.16	-0.02	0.04	0.42	0.35	-0.36	0.05	0.01	0.13	0.01	0.22	1.00					
	0.218	<0.001	0.010	0.004	0.016	<0.001	0.642	0.332	<0.001	<0.001	<0.001	0.231	0.743	<0.001	0.901	<0.001						
Deprived	0.05	-0.38	-0.08	-0.04	-0.05	-0.15	0.07	-0.01	-0.30	-0.13	0.30	-0.10	-0.01	-0.05	0.02	-0.23	-0.38	1.00				
	0.090	<0.001	0.008	0.257	0.118	<0.001	0.045	0.724	<0.001	0.003	<0.001	0.008	0.711	0.160	0.574	<0.001	<0.001					
SES (age 8)	-0.09	0.34	-0.03	0.08	-0.05	0.12	-0.02	0.05	0.27	0.12	-0.20	0.10	0.08	-0.02	-0.01	0.15	0.37	-0.42	1.00			
	0.088	<0.001	0.592	0.128	0.357	0.023	0.742	0.313	<0.001	0.124	<0.001	0.082	0.104	0.687	0.891	0.004	<0.001	<0.001				
Upward mobility	0.06	-0.04	-0.03	0.05	-0.01	-0.01	0.03	-0.01	0.01	0.00	-0.02	-0.06	0.03	0.00	-0.05	-0.04	-0.03	0.22	-0.04	1.00		
	0.042	0.236	0.396	0.135	0.643	0.745	0.427	0.720	0.687	0.961	0.534	0.084	0.300	0.883	0.250	0.203	0.308	<0.001	0.489			
%MVPA	-0.13	-0.07	0.10	-0.04	0.04	-0.10	0.12	0.11	-0.09	-0.11	0.08	0.05	0.00	0.03	0.01	-0.06	-0.04	0.05	0.00	0.02	1.00	
	0.003	0.113	0.031	0.418	0.353	0.024	0.013	0.014	0.059	0.103	0.070	0.350	0.967	0.501	0.903	0.160	0.348	0.229	0.948	0.616		
Season	0.04	0.01	-0.04	-0.02	0.06	0.02	-0.06	-0.07	0.05	0.00	-0.01	-0.07	0.03	0.06	-0.06	0.11	0.03	-0.07	-0.01	0.05	-0.21	1.00
	0.421	0.840	0.424	0.711	0.192	0.706	0.220	0.148	0.271	0.983	0.809	0.151	0.514	0.172	0.255	0.014	0.464	0.122	0.877	0.308	<0.001	

Figure X Correlations between all explanatory variables in GMS and p values

SES and early life factors

Table VII Categorical early life factors that differed by Townsend quintile in the age 6-8 sub-sample.

			Townsend score quintile for North East region										
	Total		Least advantaged		2		3		4		Most advantaged		p
	n	%	n	%	n	%	n	%	n	%	n	%	
Maternal age	987		188		201		221		223		154		<0.001
Less than 25	324	32.8	102	54.3	38	18.9	72	32.6	97	43.5	15	9.7	
25-34	543	55	77	41	133	66.2	124	56.1	101	45.3	108	70.1	
35+	120	12.2	9	4.8	30	14.9	25	11.3	25	11.2	31	20.1	
Formula-fed	947		180		191		212		215		149		<0.001
No	481	50.8	48	26.7	120	62.8	101	47.6	101	47	111	74.5	
Yes	466	49.2	132	73.3	71	37.2	111	52.4	114	53	38	25.5	
Total													
Breastfeeding	947		180		191		212		215		149		<0.001
Never	466	49.2	132	73.3	71	37.2	111	52.4	114	53	38	25.5	
<6wk	235	24.8	34	18.9	58	30.4	54	25.5	56	26	33	22.1	
>6wk	88	9.3	3	1.7	25	13.1	18	8.5	16	7.4	26	17.4	
>4m	158	16.7	11	6.1	37	19.4	29	13.7	29	13.5	52	34.9	
Adversity	929		174		193		206		204		152		0.761
No	715	77	135	77.6	151	78.2	152	73.8	161	78.9	116	76.3	
Yes	214	23	39	22.4	42	21.8	54	26.2	43	21.1	36	23.7	

Sample sizes (n), column percentages (%) and Chi-square test statistic presented (p)

Models for bivariate associations for outcomes (age 6-8) in GMS

Table VIII Bivariate (unadjusted) associations between explanatory variables and all body composition outcomes in GMS (age 6-8)

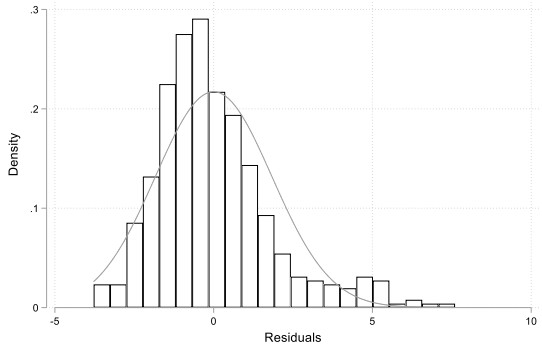
Early life factors	OWOB			BMIz			FMI			Waist OB		
	OR	CI	p	coef	CI	p	coef	CI	p	OR	CI	p
Female	0.89	[0.59,1.34]	0.56	0.02	[-0.16,0.20]	0.85	-0.11	[-0.43,0.20]	0.47	1.01	[0.62,1.64]	0.98
First-born	0.85	[0.57,1.29]	0.45	-0.08	[-0.26,0.11]	0.412	-0.07	[-0.38,0.25]	0.68	0.87	[0.53,1.42]	0.58
Gestation (wk)	1.13	[0.98,1.30]	0.10	0.03	[-0.03,0.09]	0.31	0.02	[-0.08,0.12]	0.66	1.03	[0.89,1.19]	0.71
Preterm	0.64	[0.28,1.47]	0.29	-0.12	[-0.44,0.20]	0.46	-0.14	[-0.69,0.42]	0.63	0.92	[0.41,2.08]	0.85
Normal	Ref		.	Ref		.	Ref		.	Ref		.
Post term	1.67	[0.77,3.63]	0.19	0.07	[-0.33,0.46]	0.74	0.32	[-0.36,1.00]	0.36	0.89	[0.31,2.55]	0.83
Bwtz	1.53	[1.25,1.88]	<0.001	0.26	[0.17,0.34]	<0.001	0.22	[0.07,0.37]	0.004	1.37	[1.09,1.73]	0.007
Categories of birthweight												
SGA	0.63	[0.27,1.44]	0.27	-0.30	[-0.62,0.01]	0.060	-0.06	[-0.62,0.49]	0.82	1.24	[0.55,2.83]	0.60
Normal	Ref		.	Ref		.	Ref		.	Ref		.
LGA	1.67	[0.91,3.08]	0.10	0.25	[-0.06,0.55]	0.11	0.2	[-0.33,0.73]	0.45	1.74	[0.85,3.54]	0.13
Caesarean	1.07	[0.61,1.86]	0.81	0.03	[-0.22,0.28]	0.80	0.25	[-0.18,0.69]	0.26	1.64	[0.90,2.99]	0.11
Maternal age	1.02	[0.98,1.06]	0.27	0	[-0.01,0.02]	0.63	0.01	[-0.02,0.03]	0.64	1.02	[0.98,1.06]	0.30
Less than 25	0.59	[0.36,0.97]	0.038	-0.15	[-0.35,0.06]	0.16	-0.28	[-0.65,0.08]	0.13	0.63	[0.35,1.12]	0.16
25-34	Ref		.	Ref		.	Ref		.	Ref		.
35+	0.67	[0.35,1.28]	0.23	-0.23	[-0.50,0.04]	0.091	-0.39	[-0.87,0.10]	0.12	0.82	[0.37,1.78]	0.61
Categories of breastfeeding												
Never	Ref		.	Ref		.	Ref		.	Ref		.
<6wk	1.2	[0.72,2.00]	0.47	0.11	[-0.12,0.34]	0.35	0.33	[-0.08,0.74]	0.12	1.53	[0.85,2.75]	0.16
>6wk	0.65	[0.30,1.40]	0.27	-0.19	[-0.49,0.12]	0.23	-0.22	[-0.76,0.31]	0.41	0.74	[0.25,2.16]	0.58
>4m	1.19	[0.69,2.04]	0.54	-0.09	[-0.34,0.16]	0.48	-0.18	[-0.62,0.25]	0.41	1.52	[0.78,2.96]	0.22
Exclusive	0.92	[0.33,2.57]	0.88	-0.18	[-0.59,0.23]	0.38	-0.24	[-0.92,0.44]	0.49	1.09	[0.31,3.92]	0.89
Formula fed	0.93	[0.61,1.41]	0.74	0.02	[-0.16,0.21]	0.82	-0.03	[-0.36,0.30]	0.85	0.73	[0.44,1.20]	0.22
Wean age	0.97	[0.91,1.05]	0.47	-0.03	[0.06,0.00]	0.086	-0.02	[0.07,0.04]	0.54	0.97	[0.90,1.06]	0.56
RWG	1.52	[0.98,2.36]	0.060	0.47	[0.27,0.67]	<0.001	0.23	[-0.12,0.58]	0.20	1.46	[0.86,2.49]	0.16
Adversity	2.01	[1.28,3.13]	0.002	0.39	[0.18,0.61]	<0.001	0.62	[0.25,0.98]	0.001	1.49	[0.85,2.61]	0.17
Sleep issues	1.37	[0.77,2.46]	0.28	0.25	[-0.03,0.53]	0.076	0.22	[-0.25,0.69]	0.36	1.30	[0.65,2.62]	0.46
Infection	1.3	[0.71,2.37]	0.40	0.06	[-0.22,0.34]	0.67	-0.05	[-0.53,0.44]	0.85	1.59	[0.79,3.22]	0.20
MVPA (%)	0.82	[0.73,0.91]	<0.001	-0.05	[-0.09,-0.01]	0.016	-0.19	[-0.25,-0.12]	<0.001	0.82	[0.70,0.95]	0.008

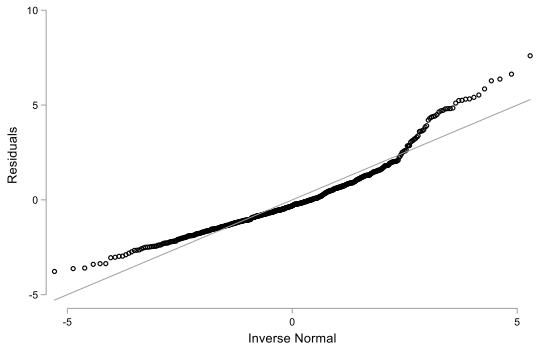
Socioeconomic and lifestyle variables	OWOB			BMIz			FMI			Waist OB		
	OR	CI	p	coef	CI	p	coef	CI	p	OR	CI	p
T score quintile for N region												
Least advantaged	Ref		.	Ref		.	Ref		.	Ref		.
2nd to least advantaged	1.24	[0.62,2.47]	0.54	-0.02	[-0.31,0.27]	0.89	0.04	[-0.47,0.54]	0.88	0.7	[0.30,1.65]	0.42
Mid	1.67	[0.86,3.23]	0.13	0.21	[-0.08,0.49]	0.16	0.33	[-0.17,0.83]	0.20	1.49	[0.73,3.04]	0.28
2nd to most advantaged	1.46	[0.73,2.91]	0.28	0.22	[-0.08,0.52]	0.15	0.35	[-0.17,0.86]	0.19	1.04	[0.49,2.22]	0.92
Most advantaged	1.28	[0.63,2.61]	0.50	-0.07	[-0.37,0.23]	0.65	-0.18	[-0.71,0.34]	0.49	0.74	[0.30,1.83]	0.51
Maternal education (birth)												
None	Ref		.	Ref		.	Ref		.	Ref		.
GCSE	0.99	[0.51,1.94]	0.98	-0.01	[-0.31,0.29]	0.95	-0.13	[-0.64,0.38]	0.61	1.04	[0.51,2.14]	0.92
A level	0.98	[0.42,2.28]	0.96	0.02	[-0.36,0.40]	0.93	-0.19	[-0.83,0.45]	0.56	1.03	[0.39,2.69]	0.96
Degree	1.01	[0.46,2.24]	0.97	0.02	[-0.34,0.38]	0.91	-0.42	[-1.03,0.18]	0.17	0.76	[0.28,2.07]	0.60
Parental occupational social class (childhood)												
Least advantaged	Ref		.	Ref		.	Ref		.	Ref		.
Mid	0.83	[0.44,1.55]	0.56	-0.1	[-0.38,0.17]	0.46	-0.38	[-0.88,0.12]	0.14	1.02	[0.47,2.23]	0.96
Most advantaged	0.76	[0.40,1.44]	0.41	-0.04	[-0.32,0.24]	0.78	-0.43	[-0.94,0.07]	0.093	0.83	[0.37,1.87]	0.65
Upward mobility 0-8	1.01	[0.49,2.10]	0.97	0.11	[-0.22,0.43]	0.53	-0.1	[-0.66,0.47]	0.74	1.95	[0.80,4.76]	0.14

For outcomes of overweight/obesity (OWOB) and waist obesity (OB) (logistic regression), BMIz, and FMI (linear regression). Coefficients (coef) and odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Ref indicates reference category for factor variables. Bold indicates significant at $p<0.05$.

FMI regression model outlier analysis

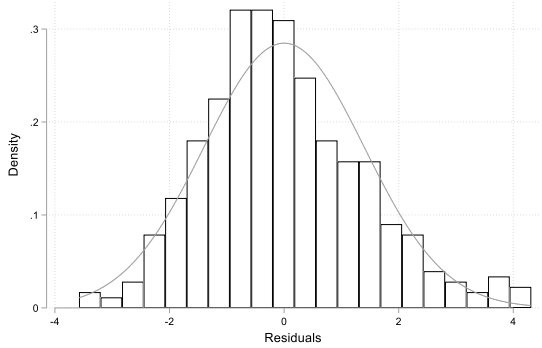
Before removing outliers

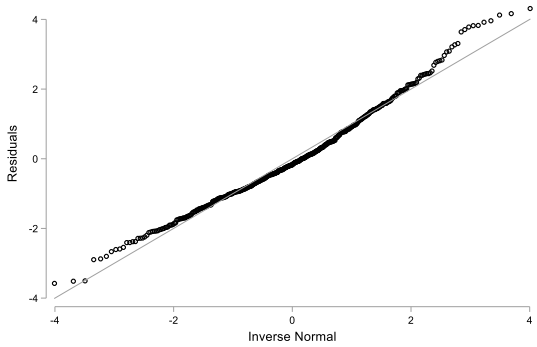




Shapiro-Wilk test for normal data
p value <0.0001

After removal of outliers





Shapiro-Wilk test for normal data
p value=0.00003

Figure XI Regression diagnostic plots for FMI outlier sensitivity analysis.
Plots are for models before (a) and after (b) removal of outliers. The histogram represents the distribution of residuals, whilst the normal probability (pnorm) plot shows the distribution relative to the specified distribution (straight line). Using the Shapiro-Wilk W test for normal data, a p-value less than 0.05, means the null hypothesis that the data are normally distributed is rejected.

Investigating adversity

In order to determine which factors could be driving the associations between adversity and adiposity outcomes, each of the components of adversity was investigated individually.

Parental separation and debt were associated with increased odds of OWOB and increased BMI, and debt was also associated with increased FMI (Table IX). Adversity was not associated with waist OB in any capacity. There were relatively low numbers for some exposures, in particular for death in the family (*n*=12, 2.2%).

Table IX The associations between individual components of adversity and adiposity outcomes (age 6-8).

	n	%	OWOB			BMI		
			OR	CI	p	coef	CI	p
Adversity (all)	144	26.9	2.37	[1.29,4.33]	0.01	0.33	[0.06,0.61]	0.02
Parental separation	33	6.2	3.19	[1.16,8.79]	0.02	0.55	[0.04,1.07]	0.04
Police involvement	28	5.2	1.09	[0.34,3.50]	0.89	0.00	[-0.52,0.52]	0.99
Debt	93	17.4	2.68	[1.37,5.25]	<0.001	0.45	[0.13,0.78]	0.01
Death in the family	12	2.2	1.80	[0.15,21.28]	0.64	0.35	[-0.79,1.50]	0.54
	n	%	FMI			Waist OB		
			coef	CI	p	OR	CI	p
Adversity (all)			0.68	[0.19,1.16]	0.01	1.23	[0.56,2.69]	0.61
Parental separation			0.85	[-0.07,1.77]	0.07	2.86	[0.88,9.25]	0.08
Police involvement			0.25	[-0.67,1.18]	0.59	1.74	[0.46,6.59]	0.41
Debt			0.75	[0.18,1.32]	0.01	1.20	[0.48,2.98]	0.70
Death in the family			1.21	[-0.82,3.24]	0.24	2.20	[0.17,29.40]	0.55

Models are adjusted for sex, gestation, maternal age, Townsend score and maternal education at birth, and parental occupational social class in childhood. Coefficients (coef) or odds ratios (OR) are presented with confidence intervals (CI) and the corresponding level of significance (p). Ref indicates reference category for factor variables. N, number of participants who experienced this adversity and corresponding % of the cohort.

Appendix D

Inverse probability weighting (IPW) for the cohort comparison models

Table X IPW weighing on adjusted models for BMIz for NTFS and GMS

	NTFS - original		NTFS - IPW		GMS - original		GMS -IPW	
	coef	CI	coef	CI	coef	CI	coef	CI
Female	-0.22*	[-0.42,-0.03]	-0.20*	[-0.40,-0.01]	-0.16	[-0.39,0.06]	-0.18	[-0.40,0.05]
Rapid thrive	0.24*	[0.04,0.45]	0.24*	[0.04,0.44]	0.50***	[0.23,0.76]	0.49***	[0.23,0.75]
Bwtz	0.02	[-0.07,0.12]	0.01	[-0.08,0.11]	0.17**	[0.05,0.29]	0.17**	[0.06,0.29]
Height (cm)	0	[-0.02,0.01]	-0.01	[-0.02,0.01]	0.04***	[0.02,0.06]	0.04***	[0.02,0.06]
SES (at birth)								
1 Least advantaged	Ref	.	Ref	.	Ref	.	Ref	.
2	0.04	[-0.32,0.39]	0.04	[-0.30,0.39]	-0.24	[-0.62,0.14]	-0.23	[-0.59,0.14]
3	0.09	[-0.22,0.40]	0.1	[-0.23,0.42]	-0.34	[-0.73,0.05]	-0.37	[-0.77,0.03]
4	0.03	[-0.46,0.52]	0.12	[-0.46,0.70]	-0.19	[-0.58,0.19]	-0.19	[-0.57,0.19]
5 Most advantaged	-0.93	[-1.87,0.00]	-0.91*	[-1.62,-0.20]	-0.18	[-0.57,0.22]	-0.18	[-0.57,0.21]
Adversity					0.30*	[0.05,0.55]	0.30*	[0.04,0.56]
Adjusted R^2	0.026		0.033		0.212		0.217	
n	313		313		269		269	

NTFS model was weighted using SES at birth and maternal age, whilst the GMS model was additionally weighted for adversity as well. Models were adjusted for gestation, maternal age and SES (age 9). Those in most advantaged group had a lower BMIz in NTFS after weighting was applied. However, this may be a spurious finding as this was a small group size (*n*=5, 2% of the 313 in the sample). * indicates p<0.05, ** p<0.01 and *** p<0.001.

Appendix E

Correlations between exposure variables in GMS

	Sex	Gestation length	BWT-z	Firstborn	Antibiotics (6 months)	Maternal age	Old maternal age	Smoked during pregnancy	Pre-pregnancy BMI	Breastfeeding categories	Adversity (pre)	Adversity (post)	Adversity (pre + post)	RWG	RT	SES (birth)	SES (age 8)
Sex	1.00																
Gestation length	0.04	1.00															
	<0.001																
BWT-z	0.00	-0.16	1.00														
	0.593	<0.001															
Firstborn	0.01	0.00	-0.17	1.00													
	0.203	0.651	<0.001														
Antibiotics (6 months)	-0.08	-0.04	0.04	-0.13	1.00												
	<0.001	<0.001	<0.001	<0.001													
Maternal age	-0.02	0.00	0.11	-0.27	0.00	1.00											
	0.007	0.606	<0.001	<0.001	0.737												
Old maternal age	-0.02	-0.02	0.02	-0.11	-0.01	0.72	1.00										
	0.020	0.026	0.051	0.000	0.220	<0.001											
Smoked during pregnancy	-0.02	-0.04	-0.16	-0.01	0.02	-0.24	-0.03	1.00									
	0.010	<0.001	<0.001	0.581	0.035	<0.001	0.001										
Pre-pregnancy BMI	-0.01	0.02	0.16	-0.06	0.03	0.02	0.03	-0.01	1.00								
	0.282	0.046	<0.001	<0.001	0.013	0.094	0.013	0.141									
Breastfeeding categories	0.01	0.07	0.05	0.03	-0.04	0.27	0.08	-0.20	-0.10	1.00							
	0.112	<0.001	<0.001	0.004	<0.001	<0.001	<0.001	<0.001	<0.001								
Adversity (pre)	-0.01	0.14	0.01	-0.03	0.02	0.00	0.01	0.11	0.03	-0.03	1.00						
	0.530	<0.001	0.214	<0.001	0.055	0.981	0.416	<0.001	0.003	0.006							
Adversity (post)	0.00	0.12	0.01	-0.02	0.02	0.03	0.01	0.06	0.02	0.00	0.31	1.00					
	0.912	<0.001	0.374	0.011	0.032	<0.001	0.167	<0.001	0.014	0.647	<0.001						
Adversity (pre + post)	-0.01	0.08	0.01	-0.03	0.02	0.02	0.01	0.09	0.03	0.00	0.59	0.71	1.00				
	0.334	<0.001	0.375	0.001	0.044	0.065	0.276	<0.001	0.005	0.621	<0.001	<0.001					
RWG	-0.03	0.15	-0.38	0.18	0.04	-0.05	-0.02	0.06	-0.04	-0.05	-0.04	-0.04	-0.04	1.00			
	0.238	<0.001	<0.001	<0.001	0.238	0.088	0.603	0.053	0.217	0.088	0.178	0.150	0.207				
RT	-0.0455	0.0322	-0.0198	0.138	0.0758	-0.0517	-0.0475	0.0333	0.0248	-0.0539	0.0057	0.0181	-0.0033	0.6623	1		
	0.118	0.268	0.496	<0.001	0.011	0.076	0.132	0.262	0.420	0.068	0.843	0.534	0.910	<0.001			
SES (birth)	0.00	0.01	0.04	0.07	0.00	0.27	0.08	-0.19	-0.09	0.26	-0.07	-0.04	-0.04	0.02	-0.02	1.00	
	0.642	0.479	<0.001	<0.001	0.675	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.480	0.515		
SES (age 8)	0.01	0.00	0.03	0.06	-0.01	0.19	0.07	-0.13	-0.11	0.20	-0.03	-0.02	-0.03	0.02	0.03	0.43	1.00
	0.509	0.908	0.061	<0.001	0.769	<0.001	<0.001	<0.001	<0.001	<0.001	0.065	0.216	0.038	0.680	0.496	<0.001	

Figure XII Pairwise correlations between exposure variables in GMS and corresponding significance

Models for bivariate associations in ALSPAC

Table XI Bivariate associations between early life factors and adiposity outcomes (BMIz, OWOB) at ages 7 and 17 in ALSPAC study members

Exposures		Age 7						Age 17					
		BMIz			OWOB			BMIz			OWOB		
		Coef	CI	p	OR	CI	p	Coef	CI	p	OR	CI	p
Female		0	[-0.05,0.04]	0.83	0.95	[0.83,1.08]	0.42	0.03	[-0.03,0.10]	0.31	0.98	[0.85,1.12]	0.75
Gestation (weeks)		0.02	[0.01,0.03]	0.005	1.02	[0.98,1.06]	0.37	0	[-0.02,0.02]	0.86	1.01	[0.96,1.05]	0.80
Pre-term		-0.03	[-0.10,0.05]	0.489	1.12	[0.89,1.40]	0.33	0.08	[-0.03,0.19]	0.14	1.2	[0.95,1.51]	0.12
Post-term		0.06	[-0.03,0.14]	0.21	1.08	[0.84,1.40]	0.54	-0.08	[-0.20,0.05]	0.25	0.91	[0.68,1.22]	0.53
Bwt z-score		0.18	[0.16,0.21]	<0.001	1.30	[1.21,1.39]	<0.001	0.12	[0.08,0.15]	<0.001	1.20	[1.12,1.30]	<0.001
Bwt categories	SGA	-0.23	[-0.33,-0.14]	<0.001	0.86	[0.64,1.17]	0.35	-0.04	[-0.17,0.09]	0.56	0.93	[0.69,1.24]	0.62
	LGA	0.35	[0.28,0.42]	<0.001	1.73	[1.43,2.09]	<0.001	0.30	[0.20,0.40]	<0.001	1.64	[1.33,2.01]	<0.001
Maternal age													
Less than 25		0.04	[-0.03,0.10]	0.24	1.09	[0.90,1.31]	0.39	0.24	[0.15,0.33]	<0.001	1.50	[1.24,1.80]	<0.001
35+		0.01	[-0.06,0.08]	0.83	1.04	[0.84,1.28]	0.72	-0.11	[-0.21,-0.01]	0.033	0.83	[0.66,1.05]	0.12
First-born		0.02	[-0.02,0.07]	0.31	1.08	[0.94,1.25]	0.27	0.01	[-0.06,0.08]	0.73	0.95	[0.82,1.10]	0.51
RWG		0.52	[0.37,0.66]	<0.001	2.04	[1.36,3.06]	0.001	0.25	[0.04,0.45]	0.020	1.25	[0.80,1.96]	0.32
RT		0.77	[0.63,0.90]	<0.001	3.82	[2.53,5.76]	<0.001	0.26	[0.05,0.46]	0.014	1.55	[1.00,2.39]	0.051
Adversity	Pre-natal	0.05	[0.00,0.09]	0.045	1.06	[0.92,1.21]	0.42	0.12	[0.05,0.19]	<0.001	1.28	[1.11,1.47]	0.001
	Post-natal	0.02	[-0.03,0.07]	0.34	1.02	[0.88,1.18]	0.79	0.03	[-0.04,0.10]	0.42	1.03	[0.88,1.20]	0.75
	Pre and post	0.08	[0.02,0.14]	0.012	1.08	[0.91,1.29]	0.39	0.14	[0.06,0.23]	0.001	1.38	[1.15,1.65]	0.001
Antibiotics		0.02	[-0.03,0.07]	0.39	0.12	[0.05,0.19]	0.002	0.98	[0.84,1.14]	0.80	1.19	[1.02,1.40]	0.028
SES at birth													
Mid		0.01	[-0.12,0.13]	0.90	0.92	[0.64,1.32]	0.66	-0.1	[-0.29,0.08]	0.26	0.76	[0.53,1.10]	0.14
Most advantaged		-0.03	[-0.15,0.10]	0.68	0.78	[0.55,1.11]	0.17	-0.25	[-0.43,-0.08]	0.005	0.52	[0.37,0.74]	<0.001
SES in childhood													
Mid		-0.01	[-0.13,0.11]	0.85	0.78	[0.54,1.11]	0.17	-0.11	[-0.28,0.05]	0.18	0.71	[0.49,1.03]	0.070
Most advantaged		-0.05	[-0.16,0.07]	0.42	0.65	[0.46,0.91]	0.012	-0.20	[-0.36,-0.05]	0.010	0.58	[0.41,0.81]	0.001

Coefficients (coef) or odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Bold indicates significant at p<0.05. Reference categories were normal term for gestational age categories; the least advantaged group for SES; and age 25-34 for maternal age. SES, socioeconomic status; Bwt, birthweight; RWG, rapid weight gain.

Table XII Multivariable regression models for adiposity outcomes in ALSPAC using RWG and adjusted for SES

Age 7	BMIz									OWOB								
	Adjusted for early life			Adjusted for SES (birth)			Adjusted for SES (childhood)			Adjusted for early life			Adjusted for SES (birth)			Adjusted for SES (childhood)		
	Coef	CI	p	Coef	CI	p	Coef	CI	p	OR	CI	p	OR	CI	p	OR	CI	p
Female	-0.07	[-0.20,0.06]	0.27	-0.07	[-0.21,0.06]	0.27	-0.08	[-0.28,0.12]	0.43	0.63	[0.41,0.98]	0.039	0.63	[0.40,0.99]	0.047	0.71	[0.36,1.40]	0.32
Gestation (weeks)	0	[-0.04,0.05]	0.87	0	[-0.04,0.05]	0.97	0	[-0.07,0.07]	0.99	1	[0.87,1.15]	0.98	1.01	[0.87,1.17]	0.91	1.02	[0.81,1.28]	0.87
Bwtz	0.35	[0.27,0.42]	<0.001	0.35	[0.27,0.42]	<0.001	0.30	[0.18,0.42]	<0.001	1.86	[1.43,2.42]	<0.001	1.72	[1.31,2.26]	<0.001	1.09	[0.73,1.62]	0.67
Maternal age																		
<25	-0.01	[-0.21,0.20]	0.94	0.09	[-0.14,0.32]	0.45	0.40	[-0.03,0.82]	0.066	0.9	[0.46,1.76]	0.76	1.18	[0.58,2.43]	0.64	2.06	[0.65,6.46]	0.22
35+	-0.18	[-0.37,0.01]	0.065	-0.17	[-0.37,0.02]	0.086	-0.04	[-0.32,0.24]	0.77	0.56	[0.27,1.17]	0.13	0.51	[0.23,1.11]	0.089	0.26	[0.06,1.15]	0.076
First-born	0.09	[-0.04,0.23]	0.19	0.06	[-0.08,0.21]	0.38	0.12	[-0.10,0.34]	0.28	1.24	[0.79,1.96]	0.35	1.15	[0.72,1.85]	0.55	1.13	[0.54,2.38]	0.74
RWG	0.79	[0.63,0.94]	<0.001	0.81	[0.65,0.96]	<0.001	0.62	[0.38,0.86]	<0.001	3.36	[2.05,5.52]	<0.001	3.15	[1.90,5.23]	<0.001	1.36	[0.63,2.96]	0.43
Adversity																		
Pre and post-natal	0.09	[-0.09,0.26]	0.34	0.03	[-0.16,0.22]	0.76	0.1	[-0.19,0.39]	0.50	1.1	[0.62,1.96]	0.74	1.03	[0.55,1.92]	0.93	1.7	[0.71,4.04]	0.23
SES at birth																		
Mid				0.23	[-0.11,0.58]	0.19	-0.23	[-0.88,0.42]	0.48				1.25	[0.34,4.53]	0.74	0.38	[0.06,2.40]	0.31
Most advantaged				0.25	[-0.09,0.60]	0.15	-0.09	[-0.74,0.55]	0.77				1.57	[0.44,5.64]	0.49	0.71	[0.12,4.36]	0.71
SES in childhood																		
Mid							0.11	[-0.27,0.50]	0.56							1.03	[0.32,3.30]	0.96
Most advantaged							0.02	[-0.36,0.39]	0.93							0.62	[0.20,1.93]	0.41
N		825			772			382			825			772			382	
Adjusted/ pseudo R-sq		0.142			0.143			0.079			0.066			0.061			0.049	

Age 17	BMIz									OWOB								
	Adjusted for early life			Adjusted for SES (birth)			Adjusted for SES (childhood)			Adjusted for early life			Adjusted for SES (birth)			Adjusted for SES (childhood)		
	Coef	CI	p	Coef	CI	p	Coef	CI	p	OR	CI	p	OR	CI	p	OR	CI	p
Female	-0.08	[-0.27,0.11]	0.42	-0.1	[-0.31,0.10]	0.32	-0.09	[-0.36,0.18]	0.53	0.98	[0.63,1.51]	0.92	0.94	[0.60,1.46]	0.78	0.93	[0.49,1.77]	0.82
Gestation (weeks)	0	[-0.07,0.06]	0.88	-0.01	[-0.08,0.06]	0.74	0	[-0.10,0.09]	0.92	1.01	[0.87,1.17]	0.88	1.01	[0.86,1.17]	0.94	0.99	[0.79,1.23]	0.93
Bwtz	0.12	[0.01,0.24]	0.040	0.12	[-0.01,0.24]	0.063	0.15	[-0.02,0.31]	0.084	1.31	[1.00,1.71]	0.051	1.28	[0.98,1.69]	0.075	1.19	[0.80,1.78]	0.39
Maternal age																		
<25	0.19	[-0.13,0.51]	0.24	0.26	[-0.10,0.62]	0.15	0.41	[-0.28,1.11]	0.24	1.12	[0.56,2.26]	0.75	1.19	[0.55,2.56]	0.66	1.33	[0.30,5.92]	0.71
35+	-0.12	[-0.41,0.17]	0.41	-0.09	[-0.39,0.21]	0.54	0.03	[-0.37,0.42]	0.89	1.35	[0.73,2.47]	0.34	1.41	[0.75,2.62]	0.28	1.91	[0.81,4.48]	0.14
First-born	0.04	[-0.18,0.25]	0.74	0.01	[-0.21,0.23]	0.93	0.11	[-0.19,0.41]	0.47	0.99	[0.62,1.60]	0.98	0.99	[0.61,1.62]	0.97	1.14	[0.56,2.33]	0.72
RWG	0.34	[0.10,0.57]	0.006	0.36	[0.11,0.60]	0.005	0.31	[-0.03,0.65]	0.072	1.65	[0.97,2.81]	0.066	1.64	[0.95,2.81]	0.073	1.17	[0.53,2.61]	0.69
Adversity																		
Pre and post-natal	0.33	[0.05,0.61]	0.021	0.28	[-0.02,0.58]	0.068	0.21	[-0.20,0.63]	0.31	1.84	[1.05,3.25]	0.034	1.66	[0.91,3.03]	0.10	1.43	[0.58,3.50]	0.44
SES at birth																		
Mid				0.42	[-0.13,0.98]	0.16	-0.18	[-1.16,0.80]	0.72				1.59	[0.42,5.96]	0.49	1.42	[0.14,14.36]	0.77
Most advantaged				0.28	[-0.26,0.83]	0.31	-0.51	[-1.48,0.47]	0.31				1.41	[0.38,5.21]	0.60	0.78	[0.08,8.05]	0.84
SES in childhood																		
Mid							-0.22	[-0.78,0.35]	0.45							1.04	[0.28,3.83]	0.95
Most advantaged							-0.06	[-0.60,0.47]	0.82							1.15	[0.33,3.99]	0.83
N		527			498			270			527			498			270	
Adjusted/ pseudo R-sq		0.018			0.017			0.008			0.02			0.018			0.024	

Coefficients (coef) or odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Reference categories were; normal term for gestational age categories; the least advantaged group for SES; and age 25-34 for maternal age.

SES, socioeconomic status; RWG, rapid weight gain

Multivariable parsimonious models in ALSPAC

Table XIII Multivariable linear regression parsimonious models for BMIz at ages 7 and 17

	BMIz7			BMIz17		
	Coef	CI	p	Coef	CI	p
Female	-0.07	[-0.20,0.05]	0.26	-0.09	[-0.29,0.10]	0.35
Gestation (weeks)	0	[-0.04,0.05]	0.90	-0.01	[-0.07,0.06]	0.80
First-born	0.09	[-0.05,0.23]	0.20	0.03	[-0.18,0.24]	0.81
Maternal age						
Less than 25	-0.01	[-0.21,0.20]	0.95	0.19	[-0.12,0.51]	0.23
35+	-0.18	[-0.37,0.01]	0.066	-0.11	[-0.40,0.17]	0.44
Birthweight z-score	0.35	[0.27,0.42]	<0.001	0.12	[0.00,0.24]	0.043
RWG	0.78	[0.63,0.94]	<0.001	0.35	[0.12,0.59]	0.004
Adversity - prenatal				0.34	[0.14,0.55]	0.001
n	825			527		
Adjusted R^2	0.142			0.028		

Coefficients (coef) presented with 95% confidence intervals (CI) and the corresponding level of significance (p). The reference category was age 25-34 for maternal age. n, sample size; SES, socioeconomic status; RWG, rapid weight gain.

Table XIV Multivariable logistic regression parsimonious models for OWOB at ages 7 and 17

	OWOB7			OWOB17		
	OR	CI	p	OR	CI	p
Female	0.63	[0.40,0.97]	0.037	0.81	[0.64,1.03]	0.081
Gestation (weeks)	1.00	[0.87,1.15]	0.99	1.03	[0.95,1.11]	0.45
First-born	1.24	[0.79,1.95]	0.36	1.04	[0.81,1.33]	0.75
Maternal age						
Less than 25	0.90	[0.46,1.77]	0.76	1.42	[0.97,2.09]	0.073
35+	0.56	[0.27,1.17]	0.13	0.98	[0.70,1.38]	0.93
Birthweight z-score	1.86	[1.43,2.42]	<0.001	1.26	[1.10,1.43]	0.001
RWG	3.35	[2.04,5.50]	<0.001			
Adversity - prenatal				1.43	[1.12,1.82]	0.004
SES at birth						
Mid				0.50	[0.25,0.97]	0.041
Most advantaged				0.39	[0.20,0.76]	0.006
SES in childhood						
Mid				0.74	[0.50,1.09]	0.12
Most advantaged				0.67	[0.45,0.98]	0.039
n	825			1952		
pseudo R-sq	0.066			0.026		

Odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Reference categories were; the least advantaged group for SES; and age 25-34 for maternal age. n, sample size; SES, socioeconomic status; RWG, rapid weight gain.

Table XV Multivariable regression models for adiposity outcomes in ALSPAC using RWG and adjusted for SES

Age 7	BMIz									OWOB								
	Adjusted for early life			Adjusted for SES (birth)			Adjusted for SES (childhood)			Adjusted for early life			Adjusted for SES (birth)			Adjusted for SES (childhood)		
	Coef	CI	p	Coef	CI	p	Coef	CI	p	OR	CI	p	OR	CI	p	OR	CI	p
Female	-0.07	[-0.20,0.06]	0.27	-0.07	[-0.21,0.06]	0.27	-0.08	[-0.28,0.12]	0.43	0.63	[0.41,0.98]	0.039	0.63	[0.40,0.99]	0.047	0.71	[0.36,1.40]	0.32
Gestation (weeks)	0	[-0.04,0.05]	0.87	0	[-0.04,0.05]	0.97	0	[-0.07,0.07]	0.99	1	[0.87,1.15]	0.98	1.01	[0.87,1.17]	0.91	1.02	[0.81,1.28]	0.87
Bwtz	0.35	[0.27,0.42]	<0.001	0.35	[0.27,0.42]	<0.001	0.30	[0.18,0.42]	<0.001	1.86	[1.43,2.42]	<0.001	1.72	[1.31,2.26]	<0.001	1.09	[0.73,1.62]	0.67
Maternal age																		
<25	-0.01	[-0.21,0.20]	0.94	0.09	[-0.14,0.32]	0.45	0.40	[-0.03,0.82]	0.066	0.9	[0.46,1.76]	0.76	1.18	[0.58,2.43]	0.64	2.06	[0.65,6.46]	0.22
35+	-0.18	[-0.37,0.01]	0.065	-0.17	[-0.37,0.02]	0.086	-0.04	[-0.32,0.24]	0.77	0.56	[0.27,1.17]	0.13	0.51	[0.23,1.11]	0.089	0.26	[0.06,1.15]	0.076
First-born	0.09	[-0.04,0.23]	0.19	0.06	[-0.08,0.21]	0.38	0.12	[-0.10,0.34]	0.28	1.24	[0.79,1.96]	0.35	1.15	[0.72,1.85]	0.55	1.13	[0.54,2.38]	0.74
RWG	0.79	[0.63,0.94]	<0.001	0.81	[0.65,0.96]	<0.001	0.62	[0.38,0.86]	<0.001	3.36	[2.05,5.52]	<0.001	3.15	[1.90,5.23]	<0.001	1.36	[0.63,2.96]	0.43
Adversity																		
Pre and post-natal	0.09	[-0.09,0.26]	0.34	0.03	[-0.16,0.22]	0.76	0.1	[-0.19,0.39]	0.50	1.1	[0.62,1.96]	0.74	1.03	[0.55,1.92]	0.93	1.7	[0.71,4.04]	0.23
SES at birth																		
Mid				0.23	[-0.11,0.58]	0.19	-0.23	[-0.88,0.42]	0.48				1.25	[0.34,4.53]	0.74	0.38	[0.06,2.40]	0.31
Most advantaged				0.25	[-0.09,0.60]	0.15	-0.09	[-0.74,0.55]	0.77				1.57	[0.44,5.64]	0.49	0.71	[0.12,4.36]	0.71
SES in childhood																		
Mid							0.11	[-0.27,0.50]	0.56							1.03	[0.32,3.30]	0.96
Most advantaged							0.02	[-0.36,0.39]	0.93							0.62	[0.20,1.93]	0.41
N		825			772			382			825			772			382	
Adjusted/ pseudo R-sq		0.142			0.143			0.079			0.066			0.061			0.049	

Age 17	BMIz									OWOB								
	Adjusted for early life			Adjusted for SES (birth)			Adjusted for SES (childhood)			Adjusted for early life			Adjusted for SES (birth)			Adjusted for SES (childhood)		
	Coef	CI	p	Coef	CI	p	Coef	CI	p	OR	CI	p	OR	CI	p	OR	CI	p
Female	-0.08	[-0.27,0.11]	0.42	-0.1	[-0.31,0.10]	0.32	-0.09	[-0.36,0.18]	0.53	0.98	[0.63,1.51]	0.92	0.94	[0.60,1.46]	0.78	0.93	[0.49,1.77]	0.82
Gestation (weeks)	0	[-0.07,0.06]	0.88	-0.01	[-0.08,0.06]	0.74	0	[-0.10,0.09]	0.92	1.01	[0.87,1.17]	0.88	1.01	[0.86,1.17]	0.94	0.99	[0.79,1.23]	0.93
Bwtz	0.12	[0.01,0.24]	0.040	0.12	[-0.01,0.24]	0.063	0.15	[-0.02,0.31]	0.084	1.31	[1.00,1.71]	0.051	1.28	[0.98,1.69]	0.075	1.19	[0.80,1.78]	0.39
Maternal age																		
<25	0.19	[-0.13,0.51]	0.24	0.26	[-0.10,0.62]	0.15	0.41	[-0.28,1.11]	0.24	1.12	[0.56,2.26]	0.75	1.19	[0.55,2.56]	0.66	1.33	[0.30,5.92]	0.71
35+	-0.12	[-0.41,0.17]	0.41	-0.09	[-0.39,0.21]	0.54	0.03	[-0.37,0.42]	0.89	1.35	[0.73,2.47]	0.34	1.41	[0.75,2.62]	0.28	1.91	[0.81,4.48]	0.14
First-born	0.04	[-0.18,0.25]	0.74	0.01	[-0.21,0.23]	0.93	0.11	[-0.19,0.41]	0.47	0.99	[0.62,1.60]	0.98	0.99	[0.61,1.62]	0.97	1.14	[0.56,2.33]	0.72
RWG	0.34	[0.10,0.57]	0.006	0.36	[0.11,0.60]	0.005	0.31	[-0.03,0.65]	0.072	1.65	[0.97,2.81]	0.066	1.64	[0.95,2.81]	0.073	1.17	[0.53,2.61]	0.69
Adversity																		
Pre and post-natal	0.33	[0.05,0.61]	0.021	0.28	[-0.02,0.58]	0.068	0.21	[-0.20,0.63]	0.31	1.84	[1.05,3.25]	0.034	1.66	[0.91,3.03]	0.10	1.43	[0.58,3.50]	0.44
SES at birth																		
Mid				0.42	[-0.13,0.98]	0.16	-0.18	[-1.16,0.80]	0.72				1.59	[0.42,5.96]	0.49	1.42	[0.14,14.36]	0.77
Most advantaged				0.28	[-0.26,0.83]	0.31	-0.51	[-1.48,0.47]	0.31				1.41	[0.38,5.21]	0.60	0.78	[0.08,8.05]	0.84
SES in childhood																		
Mid							-0.22	[-0.78,0.35]	0.45							1.04	[0.28,3.83]	0.95
Most advantaged							-0.06	[-0.60,0.47]	0.82							1.15	[0.33,3.99]	0.83
N		527			498			270			527			498			270	
Adjusted/ pseudo R-sq		0.018			0.017			0.008			0.02			0.018			0.024	

Coefficients (coef) or odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Reference categories were; normal term for gestational age categories; the least advantaged group for SES; and age 25-34 for maternal age.

SES, socioeconomic status; RWG, rapid weight gai

Table XVI Multivariable linear regression parsimonious models for BMIz at ages 7 and 17

	BMIz7			BMIz17		
	Coef	CI	p	Coef	CI	p
Female	-0.07	[-0.20,0.05]	0.26	-0.09	[-0.29,0.10]	0.35
Gestation (weeks)	0	[-0.04,0.05]	0.90	-0.01	[-0.07,0.06]	0.80
First-born	0.09	[-0.05,0.23]	0.20	0.03	[-0.18,0.24]	0.81
Maternal age						
Less than 25	-0.01	[-0.21,0.20]	0.95	0.19	[-0.12,0.51]	0.23
35+	-0.18	[-0.37,0.01]	0.066	-0.11	[-0.40,0.17]	0.44
Birthweight z-score	0.35	[0.27,0.42]	<0.001	0.12	[0.00,0.24]	0.043
RWG	0.78	[0.63,0.94]	<0.001	0.35	[0.12,0.59]	0.004
Adversity - prenatal				0.34	[0.14,0.55]	0.001
n	825			527		
Adjusted R^2	0.142			0.028		

Coefficients (coef) presented with 95% confidence intervals (CI) and the corresponding level of significance (p). The reference category was age 25-34 for maternal age.
n, sample size; SES, socioeconomic status; RWG, rapid weight gain.

Table XVII Multivariable logistic regression parsimonious models for OWOB at ages 7 and 17

	OWOB7			OWOB17		
	OR	CI	p	OR	CI	p
Female	0.63	[0.40,0.97]	0.037	0.81	[0.64,1.03]	0.081
Gestation (weeks)	1.00	[0.87,1.15]	0.99	1.03	[0.95,1.11]	0.45
First-born	1.24	[0.79,1.95]	0.36	1.04	[0.81,1.33]	0.75
Maternal age						
Less than 25	0.90	[0.46,1.77]	0.76	1.42	[0.97,2.09]	0.073
35+	0.56	[0.27,1.17]	0.13	0.98	[0.70,1.38]	0.93
Birthweight z-score	1.86	[1.43,2.42]	<0.001	1.26	[1.10,1.43]	0.001
RWG	3.35	[2.04,5.50]	<0.001			
Adversity - prenatal				1.43	[1.12,1.82]	0.004
SES at birth						
Mid				0.50	[0.25,0.97]	0.041
Most advantaged				0.39	[0.20,0.76]	0.006
SES in childhood						
Mid				0.74	[0.50,1.09]	0.12
Most advantaged				0.67	[0.45,0.98]	0.039
n	825			1952		
pseudo R-sq	0.066			0.026		

Odds ratios (OR) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p). Reference categories were; the least advantaged group for SES; and age 25-34 for maternal age. n, sample size; SES, socioeconomic status; RWG, rapid weight gain.

RWG regression diagnostic plots

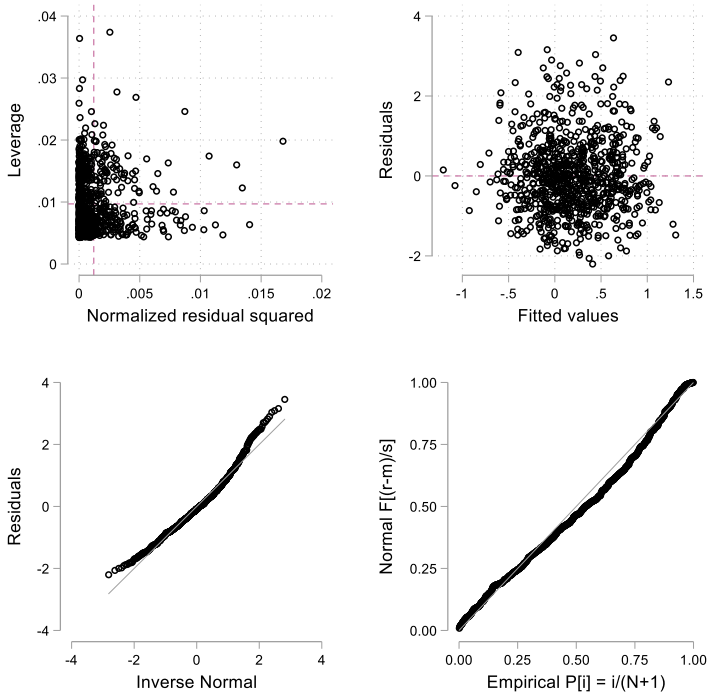


Figure XIII Regression diagnostic plots for the parsimonious model for RWG and BMIz7 in ALSPAC.
From left-right and top-bottom: Leverage vs. squared residual plot; residuals vs. predicted values (Yhat) plot, Quintile-normal plot and Standardize normal probability plot.

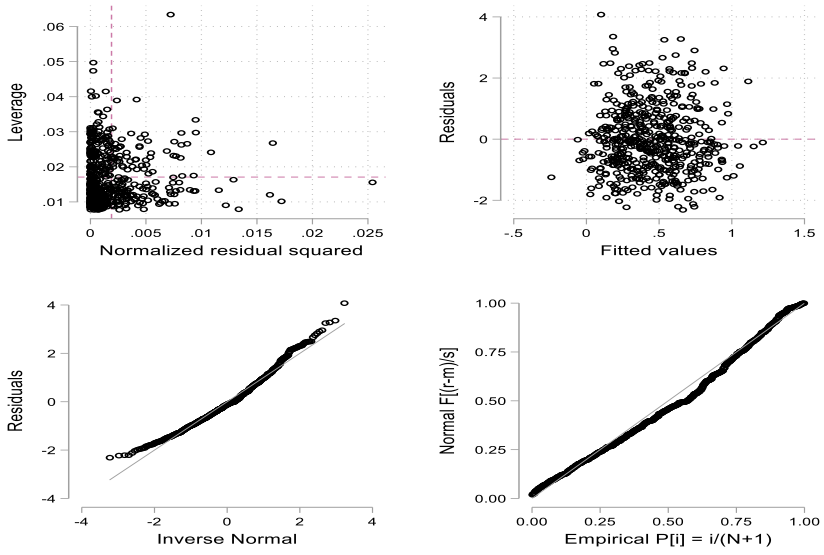
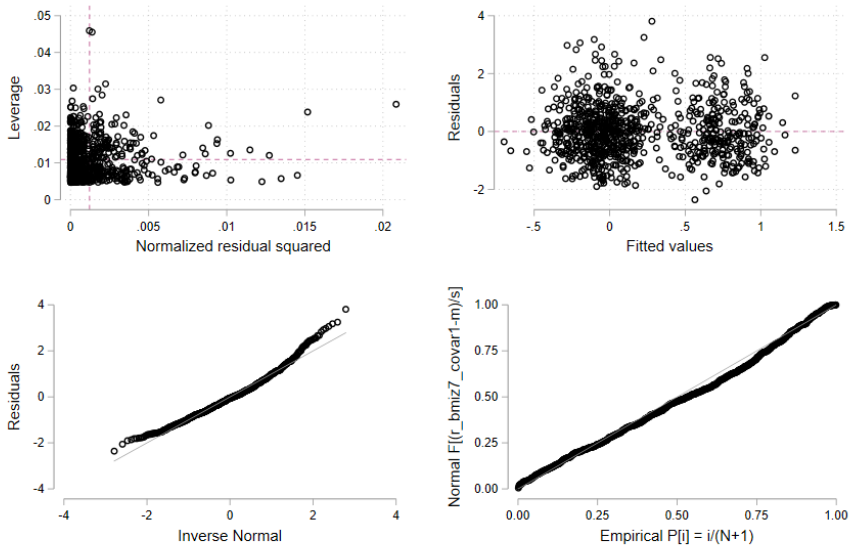
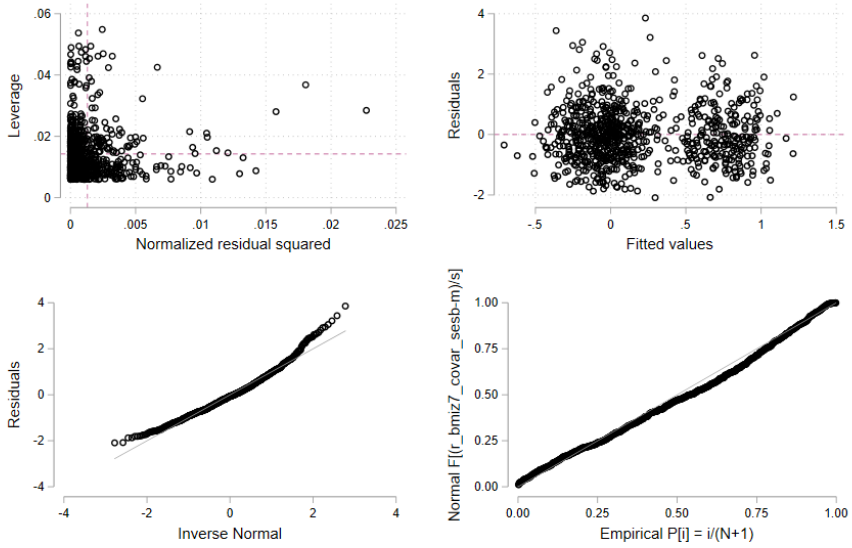


Figure XIV Regression diagnostic plot for the parsimonious model for RWG and BMIz17 in ALSPAC.
From left-right and top-bottom: Leverage vs. squared residual plot; residuals vs. predicted values (Yhat) plot, Quintile-normal plot and Standardize normal probability plot.

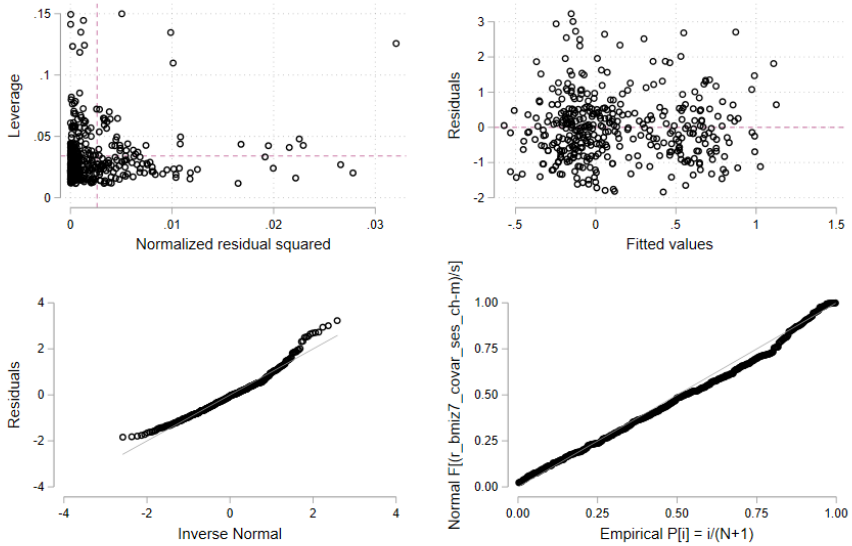
A. BMIz7 – basic model



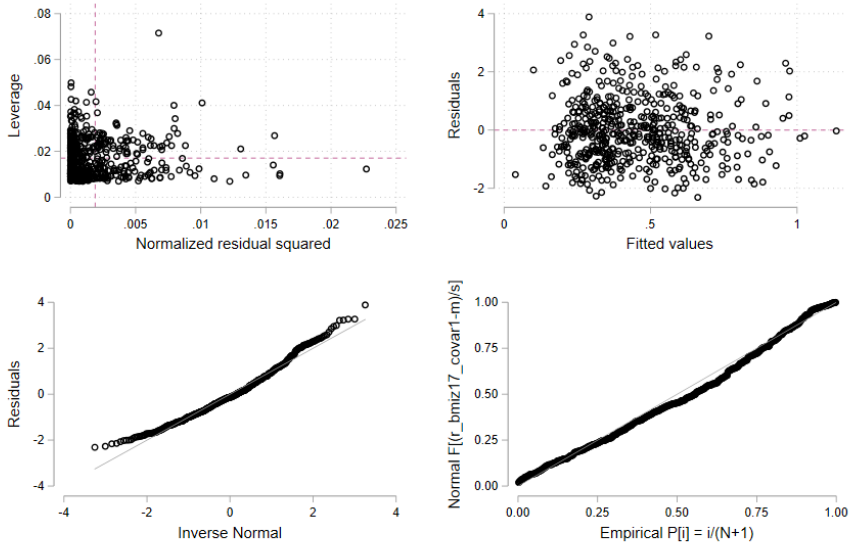
B. BMIz7 – basic model + SES (birth)



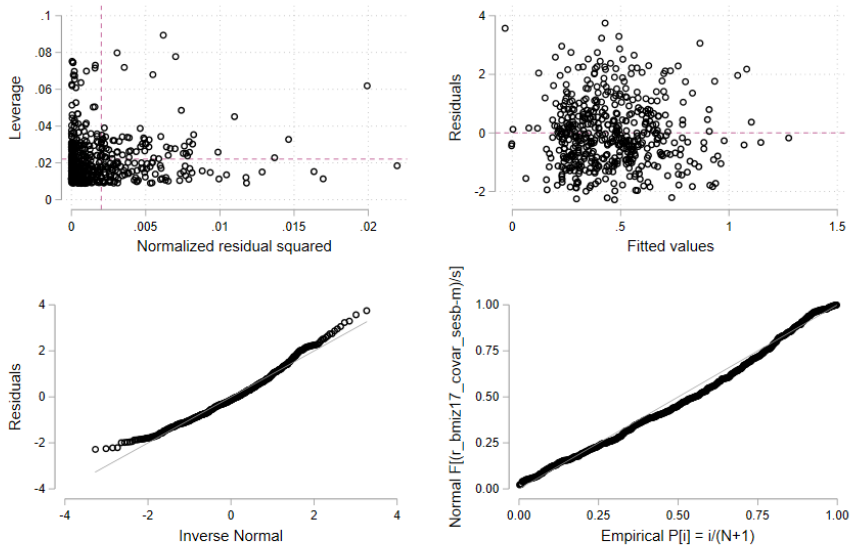
C. BMIz7 – basic model + SES (birth) + SES (childhood)



D. BMIz17 – basic model



E. BMIz17 – basic model + SES (birth)



F. BMIz17 – basic model + SES (birth) + SES (childhood)

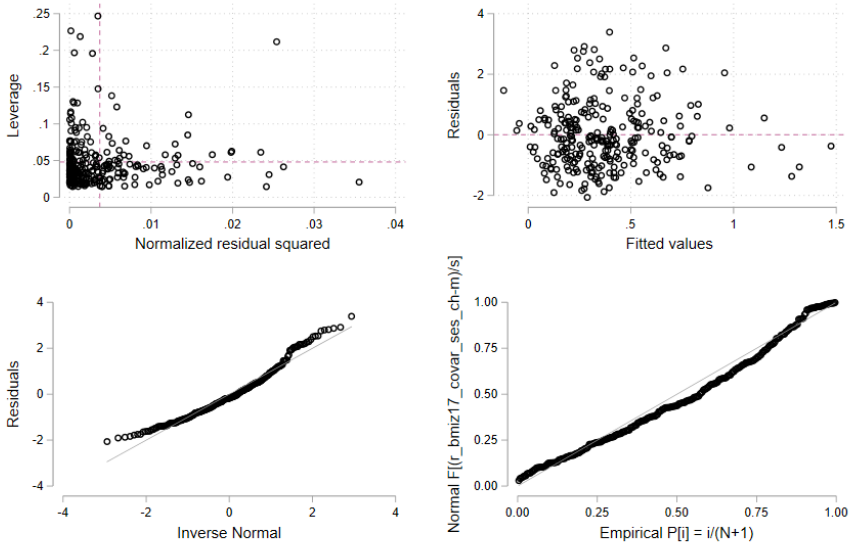
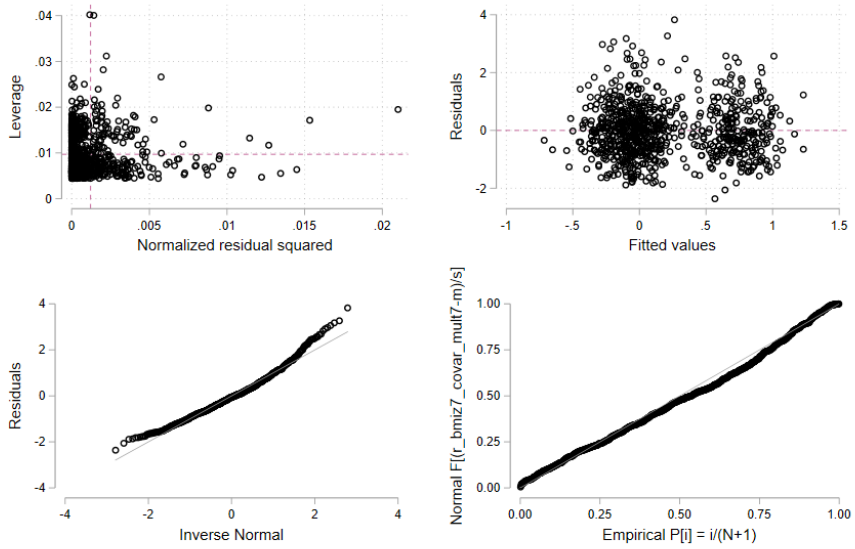


Figure XV Regression diagnostic plots for linear regression models (basic, adjusted for SES (birth) and for SES (childhood)) for BMIz at age 7 and 17.
From left-right and top-bottom: Leverage vs. squared residual plot; residuals vs. predicted values (Yhat) plot, Quintile-normal plot and Standardize normal probability plot.

Rapid thrive regression diagnostic plots

BMIz7



BMIz17

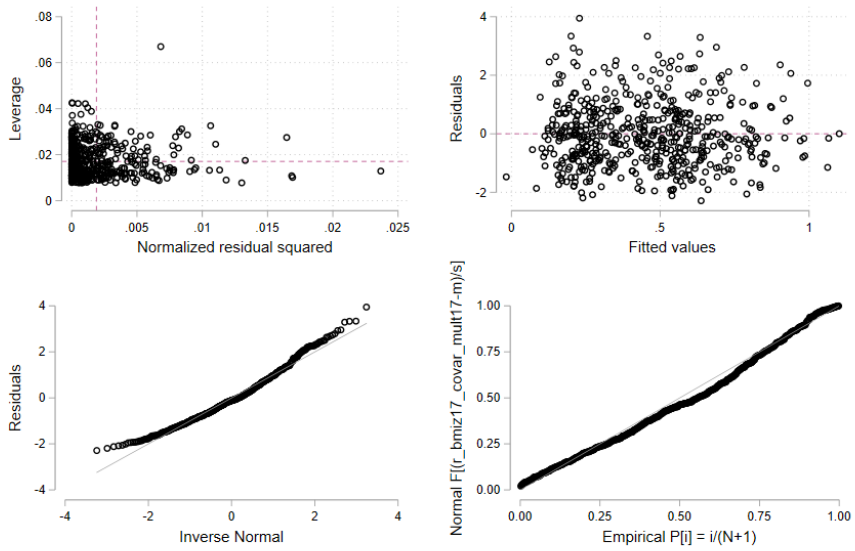
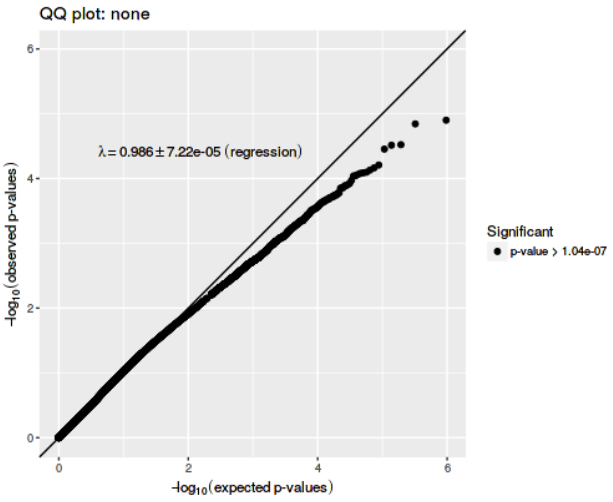


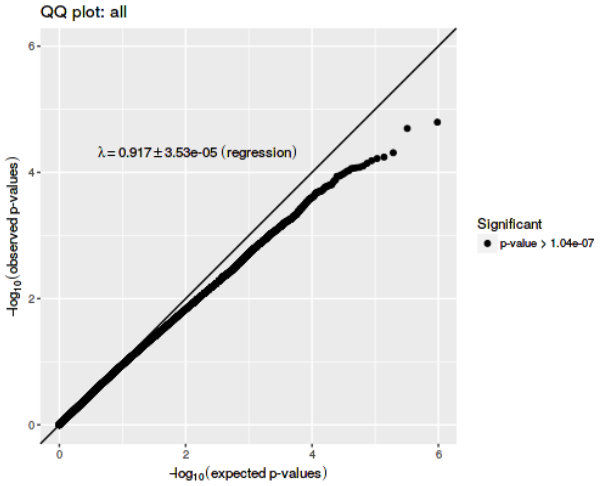
Figure XVI Regression diagnostic plots for the RT multivariable linear regression models for BMI at age 7 and age 17. From left-right and top-bottom: Leverage vs. squared residual plot; residuals vs. predicted values (Yhat) plot, Quintile-normal plot and Standardized normal probability plot.

EWAS regression diagnostic plots

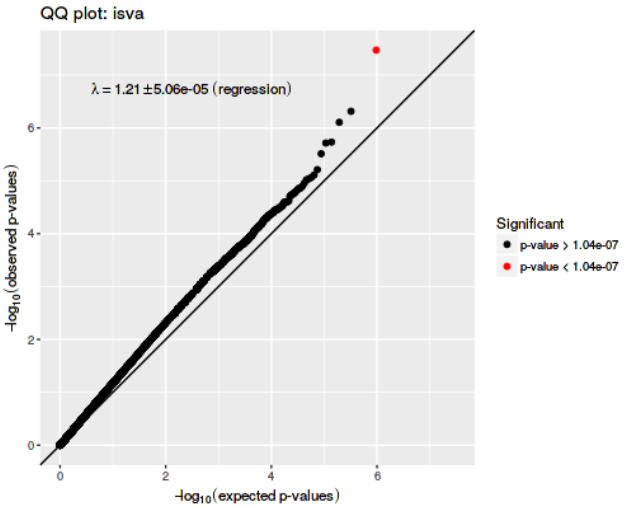
1. Not adjusted for covariates



2. Adjusted for covariates



3. ISVA adjusted model



4. SVA adjusted model

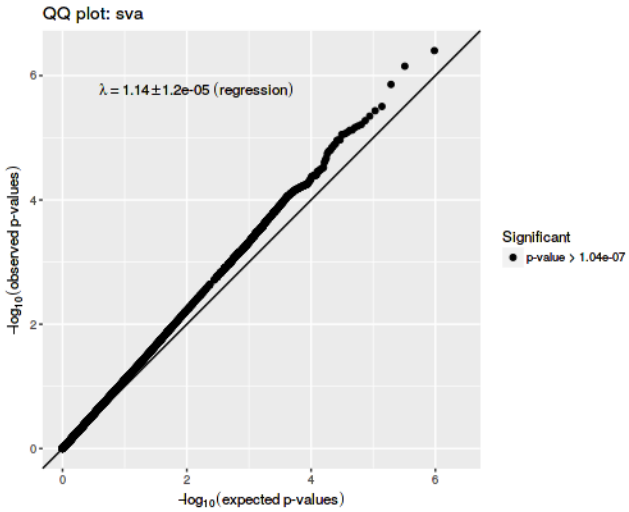
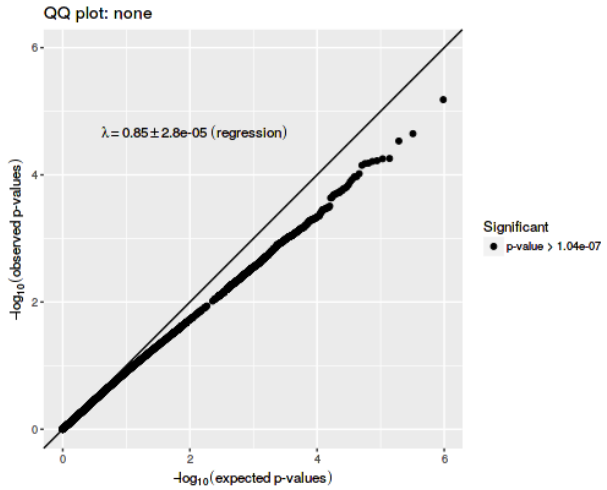
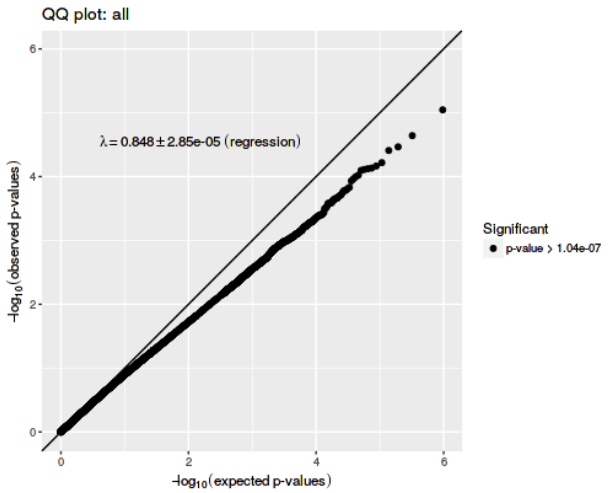


Figure XVII *EWAS regression diagnostic plots.*
An example of meffil q-q- plots for each adjustment model (none (1), all (2), ISVA (3) and SVA (4)). The Q-Q plots present the distribution of the p value for the association between CpG site methylation and RWG. The straight line is the expected distribution under the null hypothesis. For example, in the ISVA model the locus highlighted in red deviated from the expected distribution and was significantly associated with the exposure (RWG at the Bonferonni cut-off (Figure XVII)).

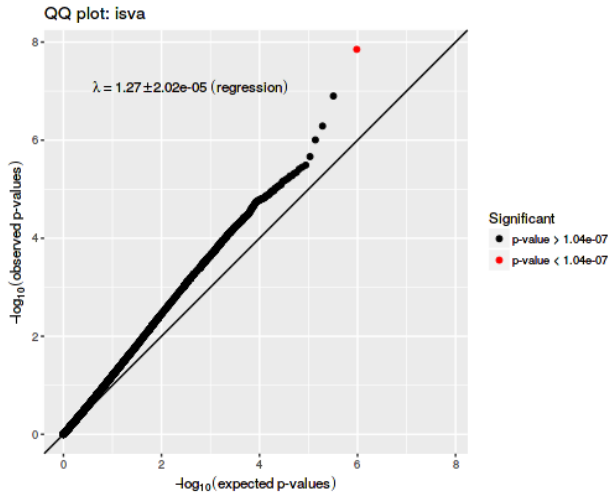
1. Not adjusted for covariates



2. Adjusted for covariates



3. ISVA adjusted model



4. SVA adjusted model

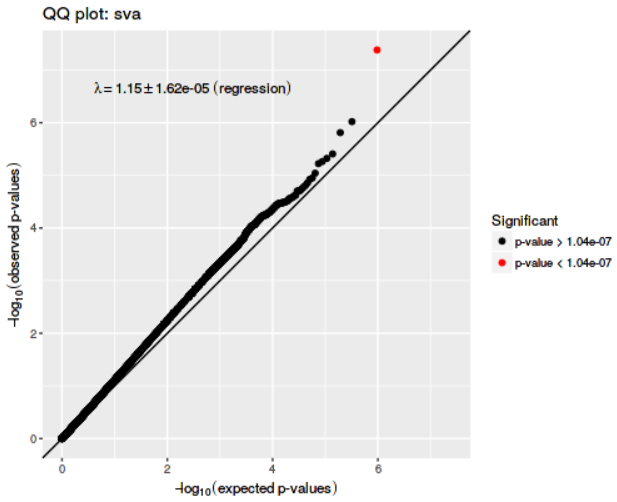


Figure XVIII EWAS regression diagnostic plots for RWG cell count sensitivity analysis.
Q-Q plots of cell counts sensitivity analysis EWAS (age 7) for RWG for each adjustment model (none, all, ISVA and SVA). The Q-Q plots present the distribution of the p value for the association between CpG site methylation and RWG. The straight line is the expected distribution under the null hypothesis.

CpG island and gene associations

Table XVIII Linear associations between RWG and CpG loci in nearby genes and/or CpG islands in ALSPAC participants (age 7)

	CpG loci	RWG Coef	p	R^2
CpGs in island	cg24802244	0.00	0.255	0.014
chr17:17206527-17207306	cg07527330	0.00	0.075	0.004
	cg21614420	0.00	0.021	0.062
	cg01379158	0.01	0.001	0.247
	cg04495270	0.01	0.329	0.035
	cg08693337	-0.00	0.932	0.131
	cg09810313	-0.00	0.578	-0.014
	cg18850434	-0.00	0.777	0.263
	cg19118161	-0.00	0.771	0.021
	cg27410828	0.01	0.479	-0.004
CpGs in island	cg02012576	0.00	0.618	-0.012
chr12:133484658-133485739	cg20307302	0.00	0.431	0.080
	cg25928819	-0.01	0.235	0.142
	cg24427504	-0.00	0.612	0.022
	cg03279164	0.01	0.459	-0.001
	cg07388347	0.00	0.102	0.170
	cg08572734	0.00	0.418	0.082
	cg09831026	0.00	0.192	0.042
	cg11531579	0.01	<0.0001	0.110
	cg15607538	0.00	0.182	0.130
	cg16562275	-0.00	0.288	0.063
	cg24459893	0.01	0.510	0.217
CpGs in NT5M	cg24802244	0.00	0.255	0.014
NT5M	cg07527330	0.00	0.075	0.004
	cg21614420	0.00	0.021	0.062
	cg01379158	0.01	0.001	0.247
	cg01979266	-0.01	0.203	0.014
	cg04495270	0.01	0.329	0.035
	cg08693337	-0.00	0.932	0.131
	cg09810313	-0.00	0.578	-0.014
	cg15761954	0.01	0.193	0.128
	cg18850434	-0.00	0.777	0.263
	cg19118161	-0.00	0.771	0.021
	cg27410828	0.01	0.479	-0.004

Linear regression models with methylation age 7 as the outcome, and RWG as the main exposure. All models are adjusted for WBCs, age and sex. Underlined are the significant CpG loci from the EWAS. n=116. R^2 is adjusted r-squared.

Table XIX EWAS linear associations between significant CpG loci (age 7) and RT in ALSPAC participants

Exposure	CpG name	n	Age	Model	No. of surrogate variables	Estimate	SE	P value
With cell counts								
RT	cg01379158	116	7	SVA	12	0.0078	0.0025	0.0022
Without cell counts								
RT	cg01379158	125	7	SVA	10	0.0065	0.0025	0.0107
RT	cg11531579	125	7	SVA	10	0.0056	0.0028	0.0493

Models are adjusted for age, sex, SVAs and with or without adjustment for cell counts.
Estimates represent beta coefficients. n, sample size; SE, standard error; SVA, surrogate variable analysis.

Consortium CpG loci featured in sub-analyses

Table XX Consortium CpG loci and associated genes from (Wahl et al., 2016) featured in the candidate gene analysis

1	cg00094412	GABBR1	95	cg11614585	ANGPT4
2	cg00108715	NT5DC2	96	cg11650298	SMIM2-AS1
3	cg00138407	KLHL18	97	cg11832534	WRAP73
4	cg00144180	HDAC4	98	cg11927233	NPM1
5	cg00238353	PTPRE	99	cg11969813	P4HB
6	cg00244001	FAM53B	100	cg12484113	AHDC1
7	cg00431050	ELOVL3	101	cg12593793	LMNA
8	cg00574958	CPT1A	102	cg12992827	ZPLD1
9	cg00634542	SLC11A1	103	cg13097800	RPL10L
10	cg00673344	LINC00880	104	cg13123009	LY6G6F
11	cg00711896	ZNF48	105	cg13274938	RARA
12	cg00863378	BBS2	106	cg13591783	ANXA1
13	cg00973118	AXIN1	107	cg13781414	NACC2
14	cg01101459	LINC00184	108	cg13922488	PKN1
15	cg01243823	NOD2	109	cg14020176	SLC9A3R1
16	cg01511901	UBE2L5P	110	cg14264316	PRRC2B
17	cg01798813	ZZEF1	111	cg14476101	PHGDH
18	cg02119938	ACSBG1	112	cg15323828	TMEM63A
19	cg02286155	SLC34A1	113	cg15357118	UGGT1
20	cg02560388	LPIN1	114	cg15681239	DLEC1
21	cg02650017	PHOSPHO1	115	cg15721584	SOX2-OT
22	cg02711608	SLC1A5	116	cg16163382	CDC42EP3
23	cg02716826	AQP3	117	cg16578636	PCGF5
24	cg03050965	S1PR1	118	cg16594806	PHBP3
25	cg03159676	GSE1	119	cg16611584	AKAP10
26	cg03318904	TAB1	120	cg16815882	KIAA0319L
27	cg03327570	ZEB2	121	cg16846518	EEFSEC
28	cg03433986	BSCL2	122	cg17178175	NFE2L2
29	cg03523676	CPNE6	123	cg17260706	BCL9L
30	cg03725309	SARS	124	cg17501210	RPS6KA2
31	cg03885055	SPATA21	125	cg17901584	DHCR24
32	cg03940776	SYNJ2	126	cg17971578	STK40
33	cg03957124	COX6A1P2	127	cg18098839	GOLIM4
34	cg04011474	RNA5SP89	128	cg18120259	C6orf223
35	cg04126866	C10orf99	129	cg18181703	SOCS3
36	cg04232128	TMEM173	130	cg18217136	PPIAP3
37	cg04524040	CREB3L3	131	cg18219562	MEOX1
38	cg04577162	RFC2	132	cg18513344	MUC4
39	cg05063895	SLC9A3R2	133	cg18608055	SBNO2
40	cg05095590	MAD1L1	134	cg19217955	ACADVL

41	cg05648472	PRDM11	135	cg19373099	CRYGFP
42	cg05720226	ST7	136	cg19566658	TRIP6
43	cg05845030	DCN	137	cg19589396	RPL5P24
44	cg06012428	ARID1B	138	cg19695507	BEND7
45	cg06164260	BCL6	139	cg19750657	UFM1
46	cg06192883	MYO5C	140	cg19881557	RNASE10
47	cg06500161	ABCG1	141	cg19998073	ZC3H14
48	cg06559575	IGFBP6	142	cg21108085	CD82
49	cg06603309	KCNQ1	143	cg21429551	GARS
50	cg06690548	SLC7A11	144	cg21486834	RHBDF2
51	cg06898549	CNTN1	145	cg22012981	ACOX2
52	cg06946797	RMI2	146	cg22103219	SH2B2
53	cg07021906	SLC7A5	147	cg22488164	PLBD1
54	cg07037944	DAPK2	148	cg22534374	RPS10P7
55	cg07136133	PRR5L	149	cg22590032	FLT4
56	cg07202479	DARC	150	cg22695339	CHD3
57	cg07471614	LINC00964	151	cg22700686	S100A2
58	cg07504977	LINC00263	152	cg23032421	IL5RA
59	cg07682160	UPF1	153	cg23232188	EAF2
60	cg07728579	FSD2	154	cg24174557	VMP1
61	cg07769588	ATG4D	155	cg24403644	TOX2
62	cg08305942	MAF	156	cg24469729	HOXA-AS2
63	cg08309687	LINC00649	157	cg24531955	LOXL2
64	cg08443038	CBFA2T3	158	cg24679890	MYO9B
65	cg08548559	PIK3IP1	159	cg25001190	NFIA
66	cg08648047	C1orf127	160	cg25096107	IGHA2
67	cg08726900	ANKRD11	161	cg25197194	EFCC1
68	cg08813944	CPSF4L	162	cg25217710	BCAN
69	cg08857797	VPS25	163	cg25435714	RN7SL142P
70	cg09152259	MAP3K2	164	cg25570328	SULT1C2
71	cg09222732	EXOC2	165	cg25649826	USP22
72	cg09315878	SDF4	166	cg26033520	ANAPC16
73	cg09349128	CRELD2	167	cg26253134	TGFA
74	cg09554443	CD247	168	cg26357885	HSPA2
75	cg09613192	FTH1P20	169	cg26361535	ZC3H3
76	cg09664445	CLUH	170	cg26403843	RNF145
77	cg09777883	BCO2	171	cg26542660	CEP135
78	cg10179300	TRIO	172	cg26663590	NFATC2IP
79	cg10438589	LINC00504	173	cg26687842	LINC00598
80	cg10505902	PDE4DIP	174	cg26804423	ICA1
81	cg10513161	ABCC5	175	cg26836479	DEDD2
82	cg10549088	PRICKLE2	176	cg26878209	SMC3
83	cg10717869	SLC41A1	177	cg26894079	CLMP
84	cg10734665	ATP10A	178	cg26952928	SLC45A4
85	cg10814005	GPR68	179	cg27050612	NFE2L1

86	cg10919522	ELMSAN1	180	cg27087650	BCL3
87	cg10922280	DUS2L	181	cg27115863	CARD10
88	cg10927968	CTSD	182	cg27117792	DRAM1
89	cg10975897	JARID2	183	cg27184903	APBA2
90	cg11024682	SREBF1	184	cg27269962	SND1
91	cg11080651	ROPN1L	185	cg27547344	TIE1
92	cg11183227	MAN2A2	186	cg27614723	SLCO3A1
93	cg11202345	LGALS3BP	187	ch.2.30415474F	LBH
94	cg11376147	SLC43A1			

Relationship between methylation and adiposity outcomes

For cg01379158 methylation at age 7 and adiposity outcomes at age 7, there was evidence from the results of the ANOVA that at least one pair of means were not equal. There was a statistically significant difference between groups as determined by one-way ANOVA ($F(3,121) = 3.35$, $p = .017$). After all possible pairwise comparisons were performed, a Bonferroni post-hoc test revealed that DNAm was statistically significantly higher in those who had RWG and were OWOB (age 7) (2.2% increase, $p = .049$) compared to those who did not have RWG and were a healthy weight (age 7) (Figure *XIX*, left). There was no statistically significant difference between the other groups or for outcomes at age 17 (Figure *XIX*, right).

As samples sizes for some categories were small, there is reduced likelihood of observing significant differences, and therefore whilst these results give an indication of methylation patterns, they are not conclusive.

Table XXI Descriptive characteristics of CpG (cg01379158) methylation (age 7) by phenotype (at ages 7 and 17) in ALSPAC participants

Groups		n	Mean	SD	Median	Min	Max
Age 7							
Healthy weight	No RWG	74	0.069	0.017	0.068	0.033	0.125
	RWG	35	0.078	0.025	0.076	0.033	0.149
OWOB	No RWG	10	0.075	0.021	0.071	0.042	0.106
	RWG	6	0.091	0.019	0.090	0.064	0.114
	Total	125	0.073	0.021	0.073	0.033	0.149
Age 17							
Healthy weight	No RWG	39	0.071	0.017	0.071	0.039	0.103
	RWG	26	0.078	0.026	0.077	0.033	0.149
OWOB	No RWG	13	0.067	0.016	0.062	0.042	0.106
	RWG	6	0.088	0.024	0.083	0.056	0.118
	Total	84	0.074	0.021	0.073	0.033	0.149

n, total in each group; SD, standard deviation; min, minimum; max, maximum.

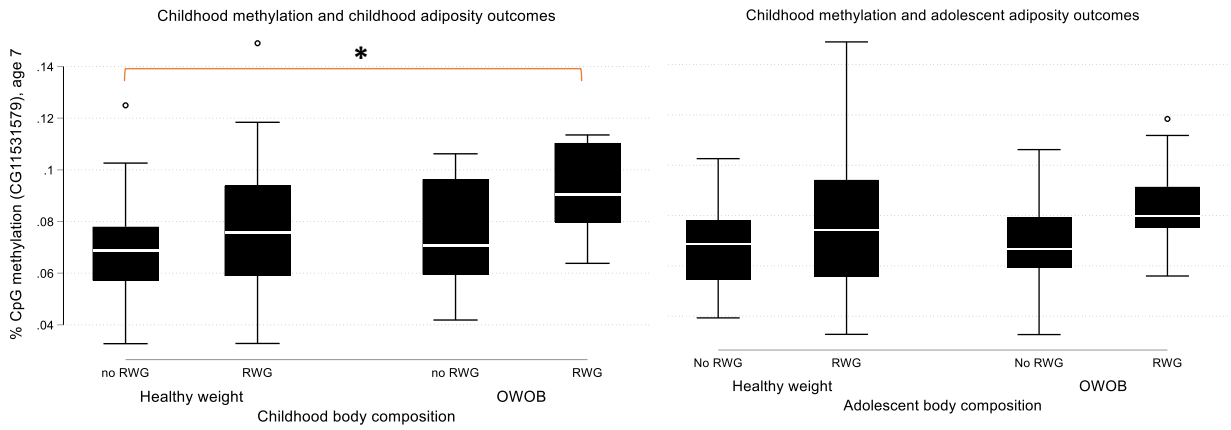


Figure XIX Box plots of methylation level (age 7), RWG and OWOB (age 7 (left), age 17 (right)).
* indicates P_{Bonf}<0.05. *Differences between groups were non-significant at age 17.*

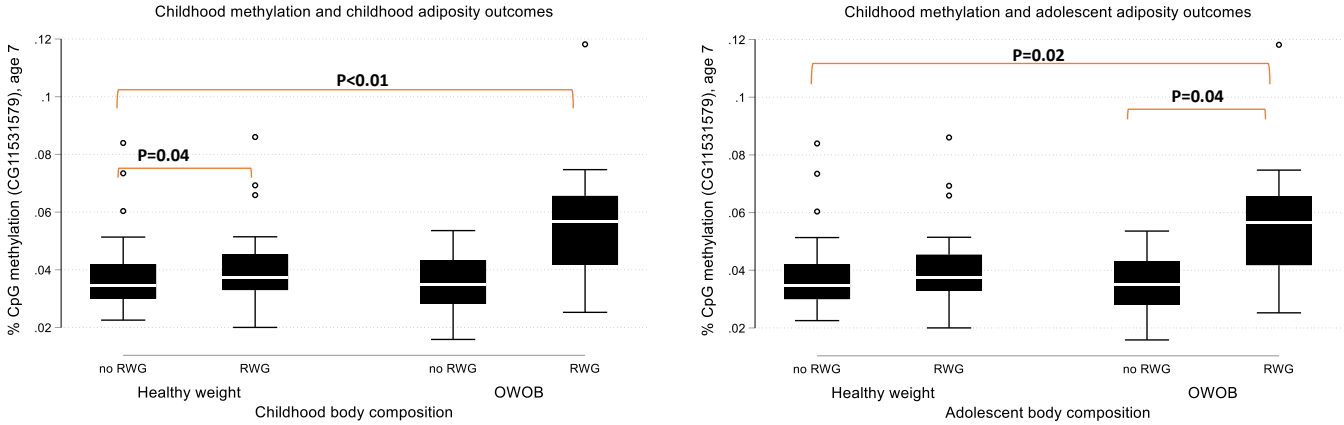
For cg11531579 methylation at age 7 and adiposity outcomes at age 7, there was evidence of a statistically significant difference between methylation and the phenotype groups (χ2 (3)=14.556, p 0.002). Correcting for multiple comparisons revealed that the differences were between no RWG and healthy weight and a comparative small increase in methylation in those who had RWG and were healthy weight (p=0.04)(Figure XX, left). Also, compared to those with no RWG and healthy weight, those with RWG and OWOB had a notably higher methylation (+2%, p=0.003), however there were few in this group (n=6).

Table XXII Descriptive characters of CpG (cg11531579) methylation (age 7) by phenotype (at ages 7 and 17) in the ALSPAC cohort

Groups		n	Mean	SD	Median	Min	Max
Age 7							
Healthy weight	No RWG	65	0.036	0.011	0.034	0.016	0.084
	RWG	35	0.045	0.020	0.039	0.020	0.118
OWOB	No RWG	10	0.039	0.011	0.041	0.024	0.054
	RWG	6	0.057	0.013	0.059	0.042	0.075
		116	0.040	0.016	0.036	0.016	0.118
Age 17							
Healthy weight	No RWG	47	0.037	0.012	0.035	0.023	0.084
	RWG	28	0.041	0.014	0.037	0.020	0.086
OWOB	No RWG	18	0.036	0.010	0.035	0.016	0.054
	RWG	9	0.059	0.027	0.057	0.025	0.118
	Total	102	0.040	0.015	0.036	0.016	0.118

n, total in each group; SD, standard deviation; min, minimum; max, maximum. Testing the assumptions of the ANOVA indicated that there was a slight positive skew in the residuals. Furthermore, Levene's test statistic (to test for homogeneity of variances) was significant; therefore, variances were not equal. This was true when analysing adiposity outcomes in both childhood and adolescence. Instead, the non-parametric KW test was used to assess group differences, plus Dunn's test with Bonferroni adjustment for multiple testing.

For methylation at age 7 and adiposity outcomes at age 17, there was evidence from the results from the KW test that there was a statistically significant difference between methylation and the phenotype groups (χ 2 (3)=8.735, p 0.033). Correcting for multiple testing revealed that there was a statistically significant difference between RWG and OWOB with both: no RWG and healthy weight (p=0.02) and no RWG and OWOB (p=0.04) (Figure XX, right).



Levene's test statistic; F(3, 112) = 3.55, p= 0.017

Levene's test statistic; F(3, 98)=3.04, p= 0.033

Figure XX Boxplots of childhood methylation and adiposity outcomes in childhood and adolescence.
The tests for significance (p values) from the KW test with Bonferroni correction for multiple testing.

Overall, methylation was consistently higher in those who were OWOB and experienced RWG, compared to those who did not experience RWG and were a healthy weight, at both time points. In those that had RWG, there is the potential that methylation at this CpG site may be able to act as a predictive biomarker of subsequent OWOB (Figure XXI); those who had RWG and were OWOB at age 7 or age 17, had higher methylation. Furthermore, those who were healthy weight at age 7, but then became OWOB at age 17 had higher methylation at age 7. Although this included few study members (n=6), methylation at age 7 could have indicated future risk of OWOB in this group.

Whereas, those who had RWG but were a healthy weight at age 7 and 17, had consistently lower levels of methylation. On average, methylation was lower in those who did not have

RWG regardless of weight status. Sample sizes are for those with complete data at that time point. Group sizes were small for some phenotypes.

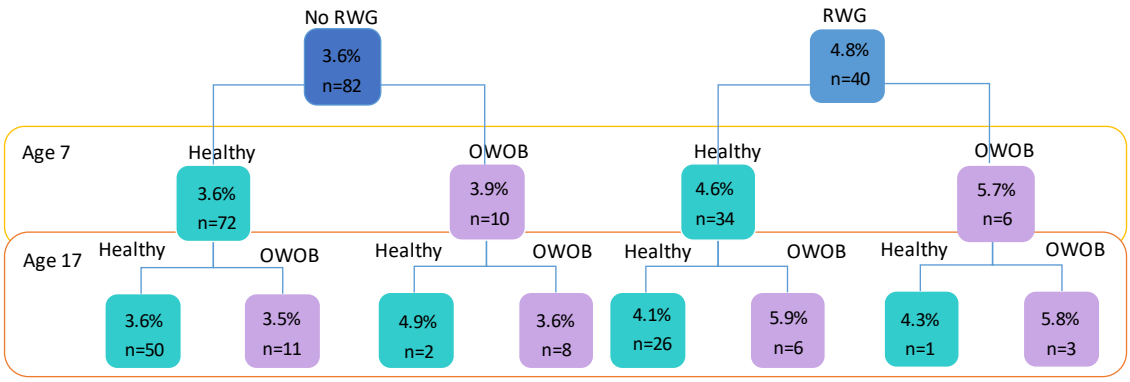


Figure XXI Pathways of mean methylation levels (%, age 7) and body composition (at ages 7 and 17).
Sample sizes are for those with complete data at that time point. Group sizes were small for some phenotypes.

Appendix F

Additional laboratory methods

Table XXIII Primers designed for examination of the 2 Significant CpG loci in the NTFS samples

CpG site	Direction	Size (bp)	Tm	Sequence
cg11531579	Forward	25	49.5	AGTTTTTGTGGAAATTTAGAAGTAA
	Forward (IUPAC)	25	49.5	RGTTTTTGTGGAAATTTARAAGTAA
	Reverse (bio)	20	52.3	AACCAACCCCATCCTAAATC
	Sequencing	15	38.2	TTTTTTGGGAATGAA
cg01379158	Forward	17	42.7	GGAGGAGAAGTTTTTAA
	Forward (IUPAC)	17	41.5	GRAGGAGAAGTTTTTAA
	Forward (IUPAC 3' edit)	17	41.9	GRAGGAGAAGTTTTTAR
	Reverse (bio)	27	46.7	CCATAATAATCRATACAATAAAAAAAA
	Sequencing	15	38.0	YGTTTTTAGAAGGTT

All primers are in the 5'→3' direction. If the original primers failed, they were redesigned with IUPAC 'wobble' base pairing to account for potential SNPs (underlined). bp, base pairs; Tm, melting temperature.

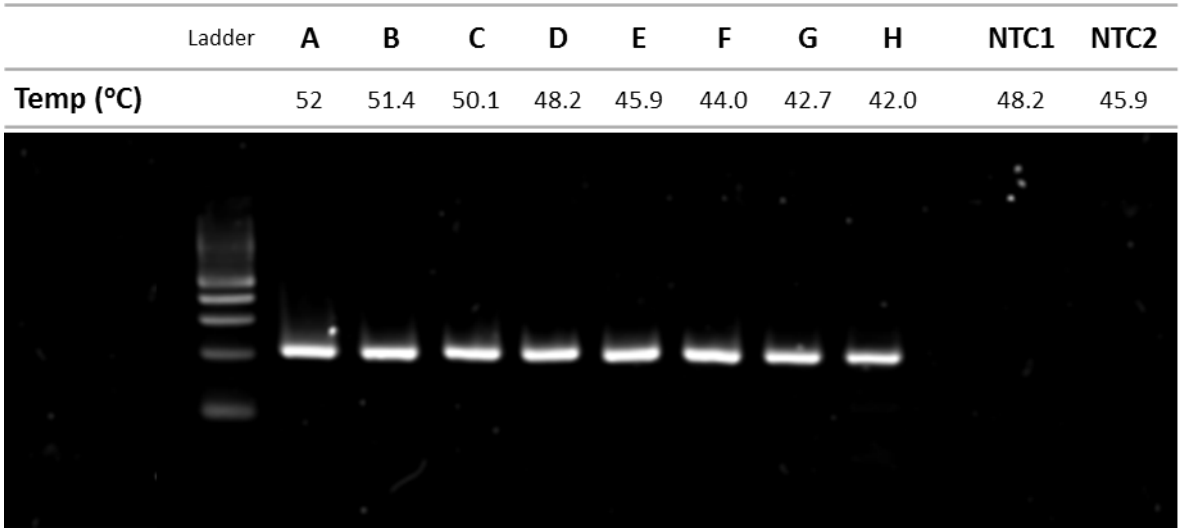
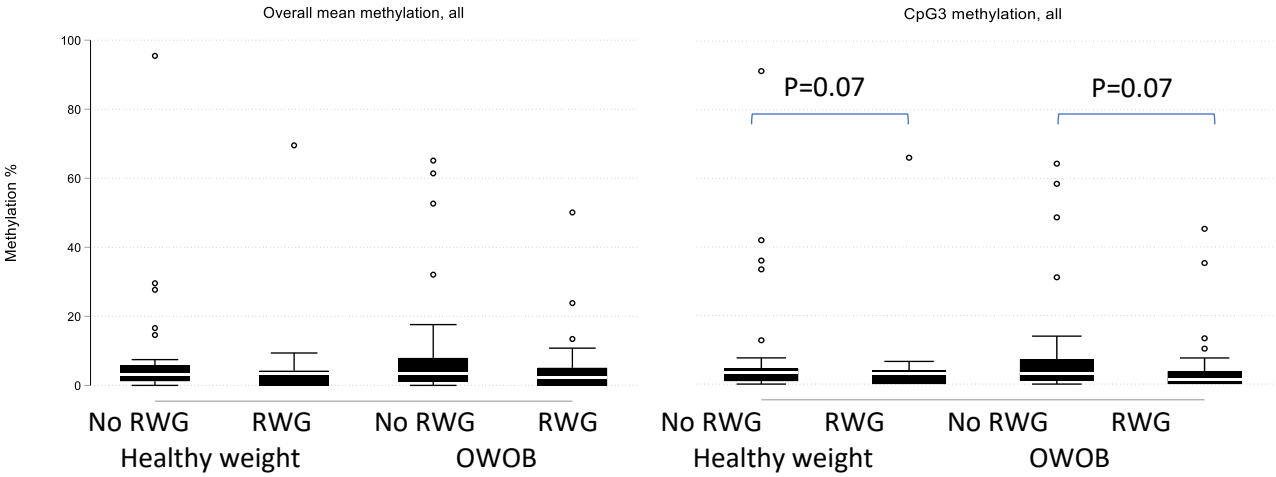


Figure XXII Gel electrophoresis image for the cg11531579 assay.
Gel image for the temperature gradient PCR for cg11531579 performed using HEK cell-line bsDNA.
NTC, no template control.

Examining the relationship between RWG, DNAm and body composition in NTFS

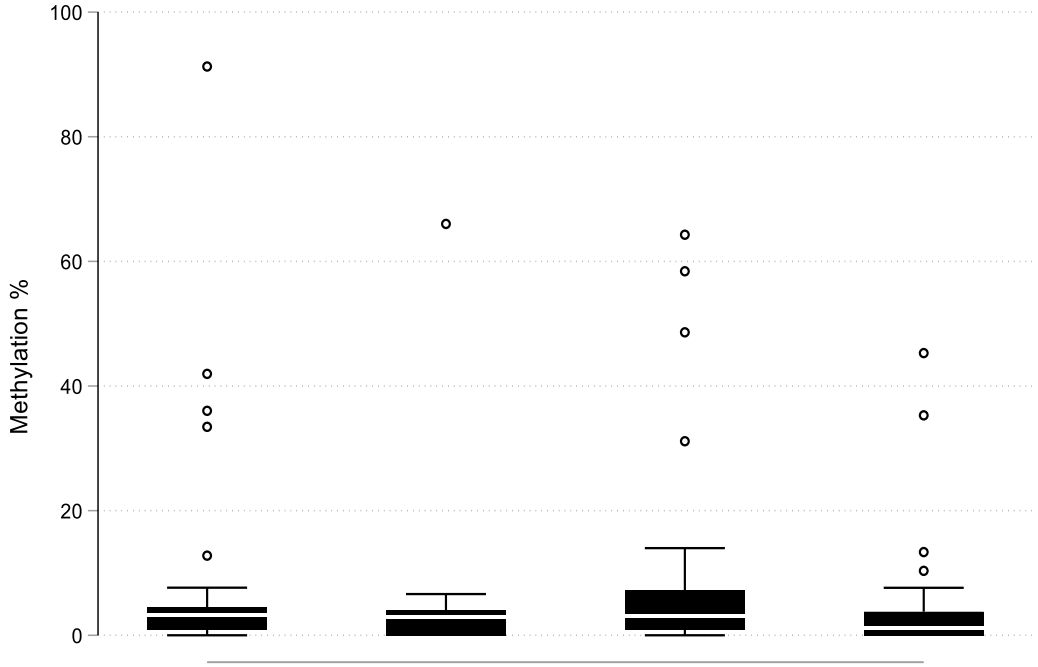
Within either the OWOB or healthy weight categories, there were differences in CpG3 methylation levels between those who had/did not have RWG (p<0.1) (Figure XXIII). There were no significant differences observed for CpG 1 or 2.



	Mean methylation					Cpg3 methylation				
	Healthy weight		OWOB			Healthy weight		OWOB		
	No RWG	RWG	No RWG	RWG	Total	No RWG	RWG	No RWG	RWG	Total
N	39	19	38	36	132	39	19	38	36	132
p50	3.05	3.4	3.46	2.19	3.08	3.24	2.98	3.16	1.22	2.53
min	0	0	0	0	0	0	0	0	0	0
max	95.46	69.53	65.12	50.09	95.46	91.27	66.01	64.28	45.28	91.27
p25	1.22	0	1.18	0	0.97	0.85	0	0.91	0	0
p75	5.81	4.28	7.96	5.1	5.18	4.5	4.02	7.25	3.78	4.42
p	0.22					0.13				

Figure XXIII Box plots of DNA methylation by phenotype groups.
Kruskal-Wallis p value shows differences across groups. There were no significant differences at p<0.05 between groups (Bonferroni adjusted using Dunn's Pairwise Comparison), however differences were observed at p<0.1.

Similar to results for body composition at age 50, those who had RWG and were subsequent OWOB (age 60) had the lowest levels of methylation (Figure XXIV). However, overall there were no significant differences between groups (Kruskal-Wallis, $\chi 2(3) = 5.507$, $p = 0.14$). There were no significant differences for CpG 1 or 2.



Age 60	Cpg3				
	Healthy weight		OWOB		
	No RWG	RWG	No RWG	RWG	Total
N	16	10	38	27	91
p50	2.81	1.80	3.16	1.41	2.42
min	0	0	0	0	0
max	5.32	4.02	91.27	45.28	91.27
p25	1.47	0	0.98	0	0
p75	4.36	3.53	7.64	4.01	4.45
p value	0.14				

Figure XXIV Box plot for CpG3 methylation (age 50) by RWG and subsequent adiposity outcomes (age 60). The tests for significance (p values) from the Kruskal-Wallis test with Bonferonni correction for multiple testing. The Kruskal-Wallis test for differences between groups was non-significant overall, $\chi2(3) = 5.507$, $p = 0.14$.

Examining the relationship between DNAm and body composition in those who had RWG

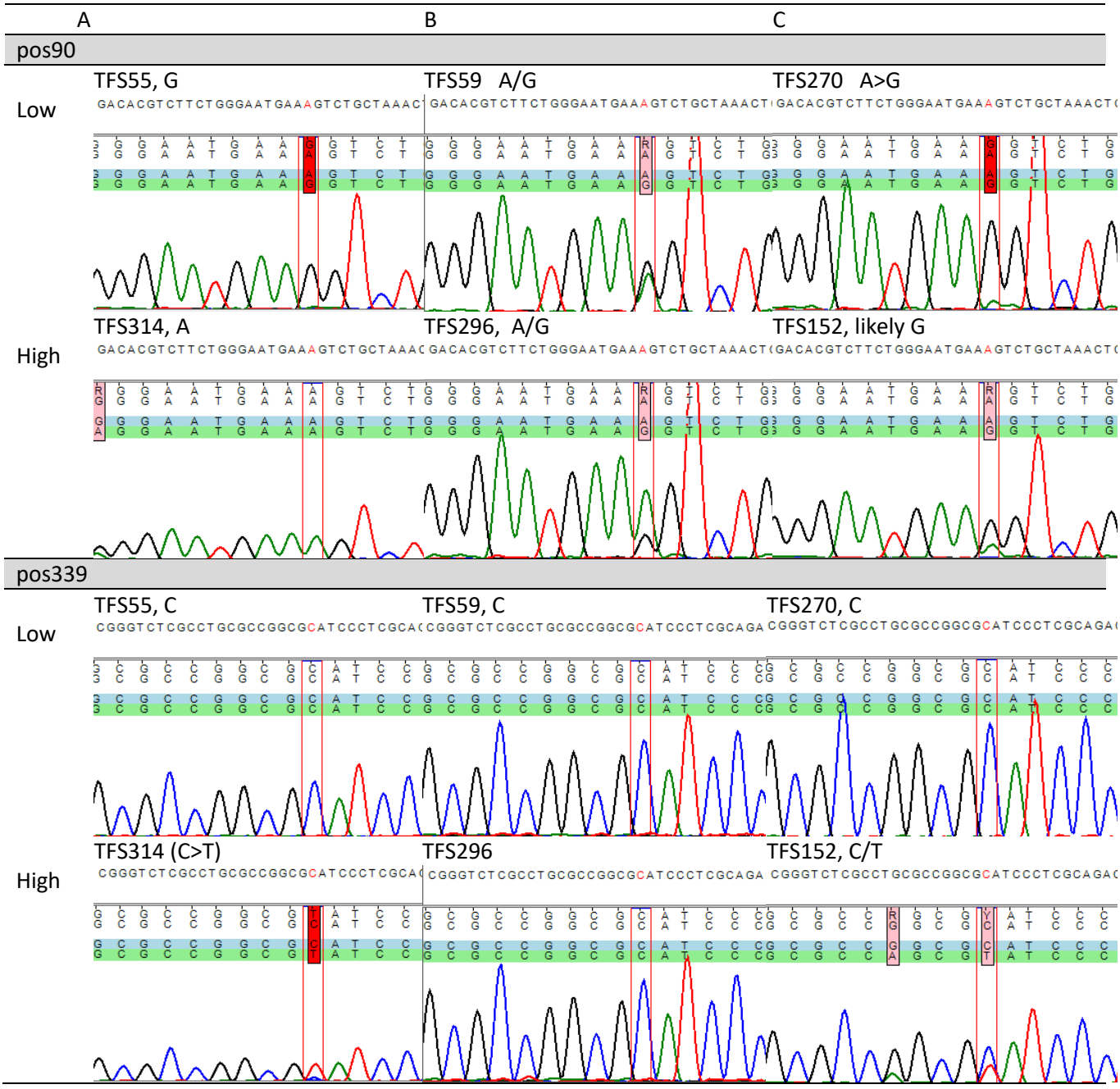
Methylation was investigated further with respect to body composition only in those who experienced RWG in infancy, as this was the 'exposed' group in which differences in methylation were previously observed (in the ALSPAC cohort). There were no significant associations between DNAm and body composition in those who had RWG (Table XXIV).

Table XXIV Logistic/linear regression models for weight outcomes (age 50) and DNAm (age 50), investigated only in those who had infancy RWG

	OB			OWOB			BMI		
	coef	CI	p	coef	CI	p	coef	CI	p
Mean methylation	0.97	[0.88,1.07]	0.57	0.99	[0.95,1.04]	0.72	0	[-0.10,0.10]	0.94
Cpg1	0.93	[0.77,1.13]	0.48	0.99	[0.94,1.04]	0.63	-0.02	[-0.12,0.08]	0.72
Cpg2	0.98	[0.92,1.06]	0.68	1	[0.95,1.04]	0.86	0	[-0.09,0.10]	0.96
Cpg3	0.98	[0.90,1.07]	0.61	0.99	[0.94,1.04]	0.69	0.01	[-0.10,0.11]	0.92

Weight outcome was the dependent variable and DNAm the independent variable. Healthy weight was the reference category for the OWOB and OB logistic regression models. Only in those who had infancy RWG. Adjusted for sex. *n*=55. Odds ratios (OR) and coefficients (coef) are presented with 95% confidence intervals (CI) and the corresponding level of significance (p)

NTFS samples (age 50) sequence traces



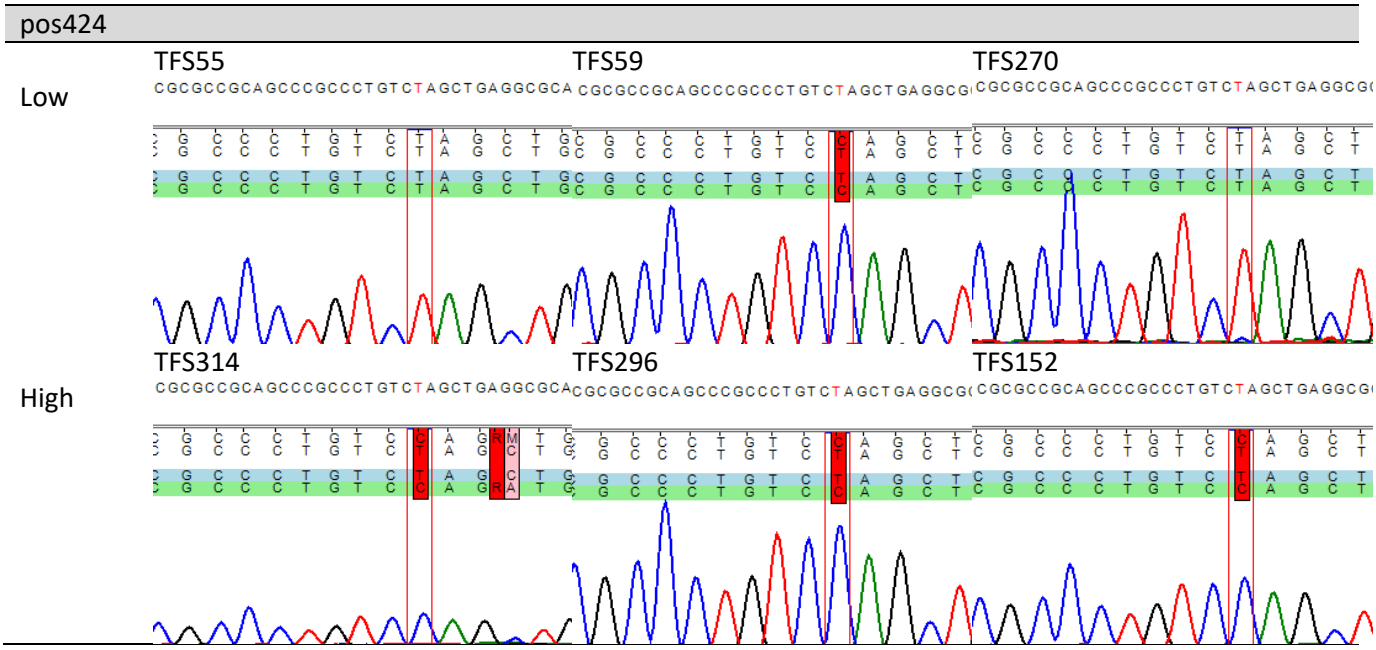


Figure XXV SNP patterns in sequenced matched samples.
There were differences observed at positions (pos) 90, 339 and 424 for matched pairs of high and low methylation. In these samples, there were SNP differences between high and low methylation in matched pairs A and C. These pairs demonstrated similarities in that they had RWG and a higher BMI, contrary to pair B which did not exhibit differences between high and low methylation.

Linkage disequilibrium

D prime (D') values for the query SNPs

RS_number	*rs2873193*	*rs4758916*
rs2873193	1.0	0.095
rs4758916	0.095	1.0

```
Query SNPs:
rs2873193 (chr12:133484722)
rs4758916 (chr12:133485056)

GBR Haplotypes:
                rs4758916
                C       T
              ------------------
           A | 123    | 2      | 125    (0.687)
rs2873193     ------------------
           G | 54     | 3      | 57     (0.313)
              ------------------
                177      5        182
               (0.973)  (0.027)

          A_C: 123 (0.676)
          G_C: 54 (0.297)
          G_T: 3 (0.016)
          A_T: 2 (0.011)

          D': 0.4176
          R2: 0.0108
       Chi-sq: 1.9662
     p-value: 0.1609

rs2873193 and rs4758916 are in linkage equilibrium
```

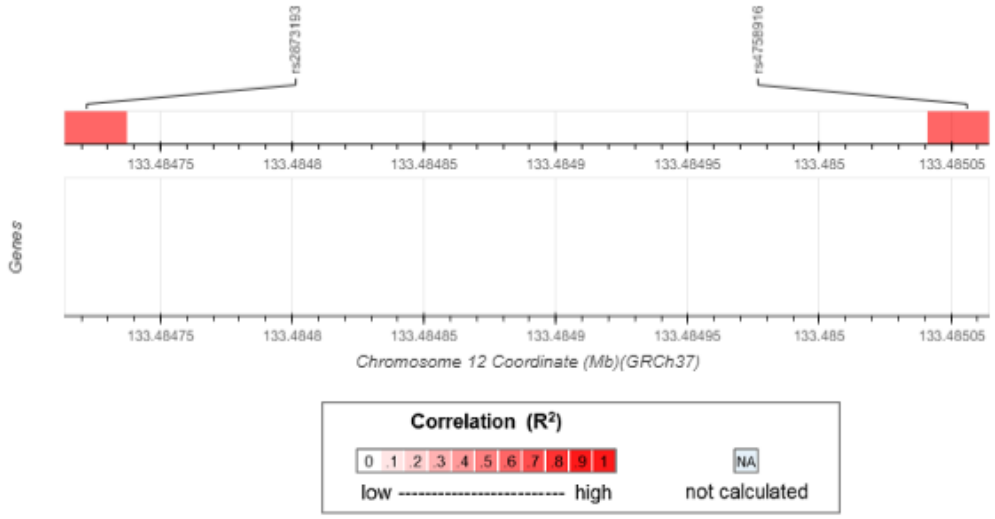


Figure XXVI LDlink output for the SNPs in the region of interest for cg11531579 for European populations.
The image output shows the chromosomal location and proximity of the SNPs, and the text output indicates there is no correlation between them (R^2=0.01). The X^2 and p-value (=0.16) indicate that the haplotypes do not deviate from the expected values and therefore there was no evidence of linkage disequilibrium.