The development of a dietary assessment tool to predict future obesity risk in young people

Catherine Elizabeth Rycroft

Submitted in accordance with the requirements for the degree of Doctor of Philosophy.

The University of Leeds

School of Food Science and Nutrition

May 2020

Declaration

The candidate confirms that the work submitted is her own and that appropriate credit has been given within the thesis where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Catherine Elizabeth Rycroft to be identified as Author of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

Firstly, thank you to the University of Leeds for awarding me an Emma and Leslie Reid scholarship in 2014. Without that PhD funding I could not have started my research.

Thanks are also given to the University of Bristol for granting access to data from the Avon Longitudinal Study of Parents and Children (ALSPAC) for model development and testing.

I want to offer a huge thank you to my supervisors, Charlotte and Janet (Dr Charlotte Evans and Professor Janet Cade), for supporting my application and providing guidance throughout my PhD, especially when things did not go to plan. I have really appreciated your kindness and encouragement and working with you both.

I also want to acknowledge many other people at the University of Leeds and elsewhere who helped me along the way:

- Christine Bosch (advice at transfer and annual reviews), Victoria Burley and Lizzie Caperon (literature search strategy ideas), Neil Hancock (IT support), Jane Hetherington (MS Office training), Darren Greenwood (STATA in general and meta-analysis in particular) and George Ellison (advice on causal inference and directed acyclic graphs).
- Marion Héry, Ben Clapinson and Alice Kininmonth for help with systematic review screening and quality assessment in duplicate.
- Fellow post graduate researchers in the Nutritional Epidemiology Research Group, especially Ziyi Li and Ayyoub Taher for teamwork on the ALSPAC research proposal.
- ALSPAC data buddy Louise Jones.
- Corresponding authors Gina Ambrosini, Ruth Weissman, Lars Libuda, Eduardo Villamor and Vasanti Malik.
- Examiners Professor Jacob Seidell and Professor Marion Hetherington for a lively debate in my viva voce and their constructive criticism.

On a personal level I want to thank medical staff at St. James's University Hospital, Leeds, including Mr. Navin, Dr Choudhry, Dr Trainor and Macmillan Nurse Belinda Archer. Without your expertise and care in 2018/19 I would not have been able to complete this PhD.

Thank you too to my friends and family (especially Marie and my daughters Lindsey, Alice and Isobel) for your unquestioning support and keeping faith.

Most of all, I want to thank my husband Phil, for your love, countless cups of Yorkshire Tea and help with final edits and proofreading. No more "apidosity" – let's get our life back on track.

Abstract

Worldwide prevalence of childhood and adolescent obesity continues to rise. It warrants prevention, but finite resources dictate targeted interventions. This research developed and evaluated an obesity risk algorithm, translated into a questionnaire and risk score to identify childhood communities at higher risk of obesity by early adolescence.

A systematic review of children's diet and adiposity outcomes found evidence for 24 potential predictors of future obesity. 20 predictors, including food and drink intakes and other factors at 10+ years, were matched to variables in a dataset from a UK birth cohort (Avon Longitudinal Study of Parents and Children). The data (n = 5,486) was randomly split, 75% for derivation of the algorithm and 25% for internal validation. Purposeful selection of covariates determined a predictive logistic regression model for adolescent obesity at 13+ years. Predictive metrics were run. Risk scores were based on β coefficients of the final model in the combined dataset.

Evidence from 14 longitudinal childhood cohorts showed that foods and drinks which contributed to energy dense dietary patterns, plus some eating habits, health behaviours and familial factors, were associated with adverse adiposity outcomes.
The final model had 9 predictive variables: Intake of vegetables, milk, dairy foods and snacks/treats, sugar sweetened beverage frequency, early puberty, mother's overweight, child's body satisfaction and active travel to school.
In the derivation sample the model had good overall predictive ability (Brier score = 0.04), acceptable discrimination (AUROC = 0.76) and showed potential usefulness (PPV = 10%). Metrics were similar in the validation sample, showing reproducibility.

The Children's Obesity Risk Assessment (CORA) is the first predictive model of childhood obesity known to include detailed measures of diet. The model and risk score require external validation to demonstrate transportability to different populations. A discriminating and well calibrated model could help target obesity prevention interventions more effectively.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

24-HDR	24-hour dietary recall
ADP	Air displacement plethysmography
ALSPAC	Avon Longitudinal Study of Parents and Children (UK cohort study)
AUROC	Area under the Receiver Operating Characteristic curve
BDPP	Bienestar Diabetes Prevention Programme (USA cohort study)
BHS	Bogalusa Heart Study (USA cohort study)
BIA	Bioelectrical-impedance analysis
BMI	Body Mass Index
BMI-SDS / BMIz	BMI standard deviation score / BMI z score
BSCC	Bogotá School Children Cohort Study (Colombian cohort study)
CA	Cluster Analysis
CASP	Critical appraisal skills programme
CDC	Centers for Disease Control and Prevention
CDC 2000	USA National growth reference
CENTRAL	Cochrane central register of controlled trials
CI	Confidence interval
CHD	Coronary heart disease
CORA	Childhood Obesity Risk Assessment
CORE	Childhood Obesity Risk Evaluation
CT	Computer tomography scan
CVD	Cardiovascular disease
CVDPoRT	Cardiovascular Disease Population Risk Tool
DAG	Directed Acyclic Graph
DAPA	Diet, anthropometry and physical activity measurement toolkit
DARE	Cochrane Database of Reviews of Effect
DASH	Dietary Approaches to Stop Hypertension
DAT	Dietary assessment tool
DLW	Doubly labelled water
DMP	Data management plan
DONALD	Dortmund Nutritional and Anthropometric Longitudinally Designed Study (German cohort study)
DP	Dietary Pattern
DXA	Dual energy X-ray absorptiometry
ED	Energy Density

EER	Estimated Energy requirement
EI	Energy Intake
EQUATOR	Enhancing the QUality And Transparency Of health Research
EYHS	European Youth Heart Study (Danish part, cohort study)
F10	ALSPAC Focus 10+ clinic (at 10+ years)
FA	Factor Analysis
FAMS	Female Adolescent Maturation Study (USA girls cohort study)
FCS	Framingham Children's Study (USA cohort study)
FAO	Food and Agriculture Organization (agency of the United Nations)
FFA	Fried food away from home
FD	Food diary
FFQ	Food frequency questionnaire
g-o-f	Goodness of fit
GUTS	Growing Up Today Study (USA cohort study)
GUTS II	Growing Up Today Study II (USA cohort study)
HSE	Health Survey for England
IDEA & ECHO	Identifying Determinants of Eating and Activity and Etiology of Childhood Obesity (USA, combined cohort study)
IMD	Index of Multiple Deprivation
IOTF	International Obesity Task Force
IQR	Interquartile range
IROC	Infant Risk of Overweight Checklist
LOWESS	Locally weighted scatterplot smoothing
MA	Meta-analysis
MAR	Missing at random
MCAR	Missing completely at random
MCS	Millennium Cohort Study
MDA	Mean decrease in (predictive) accuracy
MeSH	Medical subject headings
METS	Metabolic equivalents – a measure of Energy expenditure
MNAR	Missing not at random
MRC	Medical Research Council
MRI	Magnetic resonance imaging
NC	Neck circumference
NCD	Non-communicable disease
NCMP	National Child Measurement Programme (England)

NDNS	National Diet and Nutrition Survey (UK)
NHANES	National Health and Nutrition Examination Survey (USA)
NHES	National Health Examination Survey (USA)
NGHS	National Heart, Lung and Blood Institute's Growth and Health Study (USA girls cohort study)
NHLBI	National Heart, Lung and Blood Institute (USA)
NHS	National Health Service (UK)
NICE	National Institute for Clinical Excellence (UK)
NOS	Newcastle-Ottawa Scale
NPV	Negative predictive value
OR	Odds ratio
ORT	Obesity Risk Tool
o/w	overweight
PAL	Physical Activity Level
PCA	Principal component analysis
PHE	Public Health England
PICO-S	Participants, Interventions, Comparators, Outcomes, Study type (a framework)
PPV	Positive predictive value
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis
PROGRESS	Prognosis Research Strategy group
PROSPERO	International prospective register of systematic reviews
Project EAT	Eating and Activity in Teens and young adults (USA cohort study)
RAINE	Western Australia Pregnancy Cohort (Raine) Study (Australian cohort study)
RCT	Randomised Controlled Trial
RRR	Reduced rank regression
RSD	Regular soft drink (i.e. sugar sweetened, not diet)
SACN	Scientific advisory committee on nutrition
SD	Standard deviation
SDIL	Soft drinks industry levy (UK)
s.e.	Standard error
SES	Socio-economic status
SFT	Skin fold thickness
SIGN	Scottish Intercollegiate Guidelines Network
SMD	Standardised Mean Difference
SNP	Single Nucleotide Polymorphism

SSB	Sugar sweetened beverage
STROBE	Strengthening the reporting of observational studies in epidemiology
STROBE-nut	Strengthening the reporting of observational studies in epidemiology – nutritional epidemiology
TEI	Total Energy Intake
TF2	ALSPAC Teen Focus 2 clinic (at 13+ years)
TF3	ALSPAC Teen Focus 3 clinic (at 15+ years)
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
UK	United Kingdom
UK90	UK National growth reference, 1990
USA	United States of America
WC	Waist circumference
WHtR	Waist-to-height ratio
WHO	World Health Organization
YAQ	Youth adolescent questionnaire (an FFQ)

Useful terms

Overweight and obesity	The World Health Organization defines overweight and obesity as: "abnormal or excessive fat accumulation that presents a risk to health".
Body Mass Index (also known as the Quetelet index)	Body Mass Index (BMI) is a proxy measure of body fat in adults, based on anthropometric measures of height and weight, using the formula: • BMI = Weight or body mass (in kilograms) ÷ by the square of Height (in metres). In Caucasian populations adult overweight is defined as a BMI ≥ 25 kg/m^2 and < 30 kg/m^2. Adult obesity is defined as a BMI ≥ 30 kg/m^2.

Chapter 1 Introduction

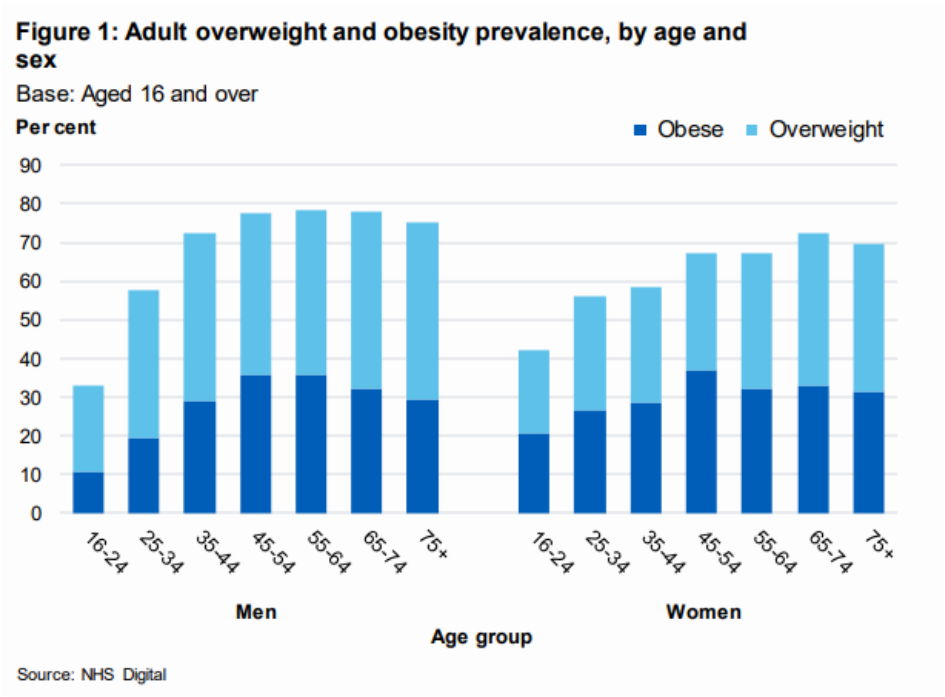
1.1 Obesity prevalence and health risks for adults and children

The World Health Organization (WHO) defines overweight and obesity as "abnormal or excessive fat accumulation that presents a risk to health" (WHO, 2015). The rising prevalence of obesity and overweight is a public health concern in the United Kingdom (UK) and many other countries.

Worldwide, obesity has almost tripled since 1975. In most regions of the world, other than parts of sub-Saharan Africa and Asia, more people have obesity than have underweight. In 2008 over 500 million adults had obesity (WHO, 2015), rising to 650 million or 13% of all adults by 2016. A further 39% of the World's adults aged 18 years and above had overweight (WHO, 2020).

In some countries, the prevalence of adult overweight and obesity is even higher. The Health Survey for England found that in 2017 approximately 29% of adults (27% of men and 30% of women) were classified as obese, with a body mass index (BMI) of 30kg/m^2 or above, with another 35% (40% of men and 31% of women) classified as overweight, with a BMI of 25kg/m^2 or above but below 30kg/m^2 (NHS Digital, 2018). The proportion of adults in England with overweight or obesity increased with age in both sexes, as illustrated in Figure 1-1 below.

Figure 1-1 Adult overweight and obesity prevalence in England in 2017, by age and sex



In adulthood there are well established links between overweight and obesity and risk of ill health, but for individuals there are varying degrees of cardiometabolic risk even in people with the same BMI. The highest risk of cardiovascular disease (CVD) is seen in adults with excess visceral adiposity, which is linked with greater accumulations of lipids in the heart and the liver (Neeland et al., 2019). Adult obesity is also associated with a higher risk of type 2 diabetes and some cancers (Kopelman, 2007). Such non-communicable diseases (NCD) reduce longevity and the potential for individuals to be socio-economically active and diminish the quality of life. Public Health England (PHE) estimated that the cost to the UK economy of overweight and obesity was £15.8 billion per year in 2007 (PHE, 2015a), rising to £27 billion per year by 2014/15, including a cost of £6.1 billion to the National Health Service (PHE, 2017).

In 2004 an estimated 10% of the world's school-aged children had overweight or obesity, with the highest prevalence (approximately 20 to 30%) of children with overweight in North America, Europe and parts of the Western Pacific (Lobstein et al., 2004). Since then childhood overweight and obesity has risen in low and middle income countries, especially in urban environments. Between 1975 and 2016 it is estimated that the global prevalence of overweight and obesity among children and adolescents aged 5 to 19 years more than quadrupled, from 4% to 18%. Of the 340 million school-aged children and adolescents with overweight or obesity in 2016, more than 124 million were classified as obese. (WHO, 2020).

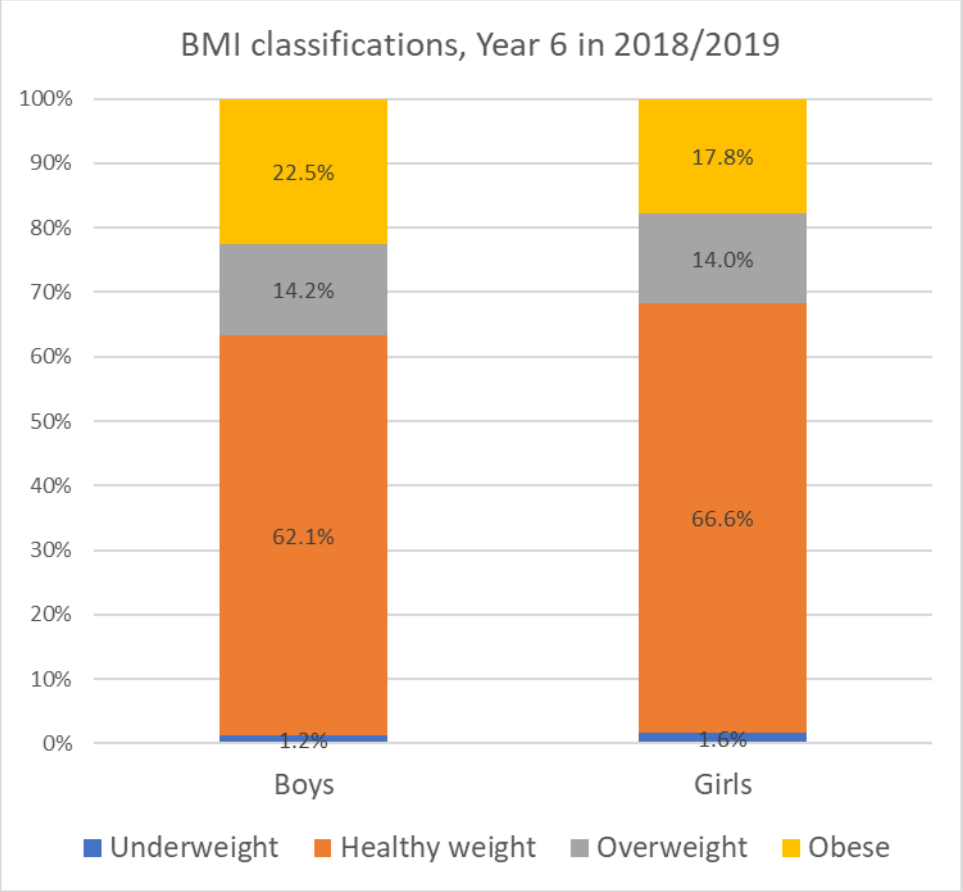
Since 2006/2007 in England, each year the National Child Measurement Programme (NCMP) has measured Reception class children (aged 4 to 5 years old) and Year 6 primary school children (aged 10 to 11 years old) who are attending mainstream state-maintained schools. For this annual population monitoring exercise, each child's Body Mass Index (BMI) calculated from measured height and weight is classified by centile distribution, using the British 1990 growth reference (UK90) (Cole et al., 1995). See Table 1-1 below.

Table 1-1 Children's BMI classification based on UK90, as used by the NCMP

BMI classification	Centile of UK90 BMI distribution
Underweight	Less than or equal to 2nd
Healthy weight	Greater than 2nd and less than 85th
Overweight	Greater than or equal to 85th and less than 95th
Obese	Greater than or equal to 95th (obese plus severely obese)
Excess weight	Greater than or equal to 85th (overweight plus obese)
Severely obese	Greater than or equal to 99.6th

Data from the latest NCMP in 2018/19 shows that on average over a fifth (22.6%) of children in Reception class were classified as either overweight (12.9%) or obese (9.7%). (NHSDigital, 2019). Over a third (34.3%) of children in Year 6 were classified as either overweight (14.1%) or obese (20.2%). A greater proportion of Year 6 boys were classified as obese, compared with girls, as shown in Figure 1-2.

Figure 1-2 NCMP 2018/2019 Year 6 BMI classifications by sex



When trends in childhood obesity were examined over time by the NCMP, the picture was mixed. Between 2006/2007 and 2018/2019 overall obesity prevalence in Reception class children decreased from 9.9% to 9.7%, with downward trends in boys' excess weight (obesity and overweight together) but upward tends in girls' excess weight. In a shorter time frame between 2009/2010 and 2018/2019 overall obesity prevalence in Year 6 children increased from 18.7% to 20.2%. The only downward trend in Year 6, when children are on the cusp of adolescence, was for boys' overweight prevalence.

Obesity/overweight and obesity-related behaviours have been shown to track from childhood to adulthood (Craigie et al., 2011). Consequently, adolescents with overweight and obesity have an increased risk of becoming adults with overweight (Singh et al., 2008). It is thought that some adult cardiovascular morbidity may originate from childhood obesity, which itself is

associated with cardiovascular risk factors such as high blood pressure, dyslipidaemia and insulin resistance (Reilly and Kelly, 2011). Obesity during childhood is also associated with an increased risk of asthma, musculoskeletal disorders and low self-esteem (Reilly et al., 2003).

Overweight and obesity are physiological consequences of surplus energy stored as fat. Once established, both childhood and adolescent obesity are resistant to treatment (Mead et al., 2017) (Al-Khudairy et al., 2017) but overweight and obesity are not inevitable. Faced with the adverse effects of obesity on children's health and development and their future well-being in adulthood, the prevention (rather than the treatment) of childhood obesity has become an international public health priority. In 2004 the World Health Assembly adopted the WHO Global Strategy on Diet, Physical Activity and Health. This strategy asks stakeholders to act locally, regionally and globally to improve diets and patterns of physical activity at the population level (WHO, 2015).

1.2 Evidence framework for childhood obesity

Childhood obesity can be considered using the evidence framework developed by the International Obesity Taskforce (IOTF) (Swinburn et al., 2005, Swinburn, 2010) which asks five sets of questions:

- Should we do something?
- What should we target?
- Who, how and where should we intervene?
- What could we do?
- What should we do?

To answer each question demands different types of evidence.

The first question is answered by childhood obesity prevalence and upward trends and its negative impact upon short and long term health: the burden of childhood obesity is clear, and action is warranted.

The second question is answered by evidence about the determinants of obesity. Obesity interventions often focus on diet and/or physical activity because they are modifiable, although the contribution of diet to obesity risk is not fully understood. Dietary patterns describing the whole diet may be better at explaining or predicting obesity and disease risk than individual nutrients or foods (Hu, 2002). The interactions between environmental, behavioural and genetic determinants of excess weight gain are also uncertain (Rennie et al., 2005) and more socio-cultural research is needed to understand beliefs, attitudes and practices that may explain the differences in obesity prevalence between countries or sections

of society. For example, why do parents have a limited ability to perceive when their child is overweight (Jones et al., 2011) and why are they unwilling to acknowledge that it may affect a child's adult health (Syrad et al., 2015)?

The third question is about the framework for action. Strategies to reduce obesity prevalence and its associated diseases range from education and communication to environmental, infrastructure and policy changes, applied in a variety of settings such as schools, communities, workplaces, health and commercial sectors. In the widest context "who?" may include policy makers, company shareholders, health services, schools or any player that influences the environment or information that shapes behaviour, but in practice "who?" is often a whole population or a targeted group that is regarded as vulnerable, such as children.

The fourth question is about the efficacy and cost-effectiveness of potential strategies. Much of the evidence about childhood obesity prevention is from intervention studies, often in school settings. A Cochrane systematic review of interventions for preventing childhood obesity that use a controlled study design found evidence of beneficial effects (Waters et al., 2011). Many of the most promising strategies were applied in school settings, but it was hard to establish which components were most effective as included studies were often small or of short duration, with a lot of heterogeneity between studies.

This Cochrane review was later updated (Brown et al., 2019), investigating three times as many studies (153 vs 55), conducted in child care centres, schools, homes and health care centres or in community or recreation centres. The updated review found with "moderate certainty" that combined diet and physical activity interventions reduced the risk of obesity in young children aged 0 to 5 years while physical activity interventions reduced the risk of obesity in children aged 6 to 12 years, compared with control groups. In adolescents aged 13 to 18 years old, physical activity interventions to reduce the risk of obesity also seemed to be effective. There was weak evidence that dietary interventions alone were beneficial for children age 0 to 5 years. In older children and adolescents there was no evidence that dietary interventions on their own were helpful for reducing the risk of obesity, but combined diet and physical activity interventions may be effective in this age group. Brown et al noted that, although the population-wide clinical significance of modest reductions in childhood obesity risk is hard to judge, diet and physical behaviours established during childhood are known to track into adult life. If sustained, even small changes towards healthier diets and increased physical activity in childhood may yield rewards in terms of healthier weights and other health benefits for adult populations. The authors concluded that interventions to prevent childhood obesity do not seem to cause adverse effects (such as weight concerns or eating disorders) and there was no

evidence that they added to health inequalities, but very few studies looked at costs or cost-effectiveness. Evidence about cost effectiveness remains scarce.

The fifth question is about setting priorities and selecting appropriate interventions to reduce the burden of obesity. The IOTF's recommendation is that "a balanced portfolio of specific, promising interventions" should be agreed, while acknowledging that this is a challenge, especially with limited funding.

The IOTF framework shows where more evidence would help in addressing the issue of childhood obesity and its prevention. Such evidence includes:

- A better understanding of the determinants of obesity, including diet.
- Ways to identify at-risk groups who might benefit from timely obesity prevention measures.
- Ways to evaluate the effectiveness of obesity prevention interventions.

1.3 Research aims and objectives

1.3.1 Aim

The aim of this research is to develop and internally validate a dietary assessment tool to identify, at a population level, children who are at risk of obesity during early adolescence.

1.3.2 Objectives

- Identify longitudinal studies that have quantified dietary intake and measured adiposity of children and adolescents at more than one time and assess their methodological quality.
- Investigate longitudinal associations between childhood diet (food and drink intakes, dietary patterns and eating habits) and outcomes of overweight and obesity in adolescence.
- Use the reported evidence as the basis of a dietary assessment tool.
- Develop and internally validate a predictive model using data from a high quality childhood cohort or cohorts.

1.3.3 Hypothesis

- Diet during childhood predicts overweight and obesity outcomes during adolescence.

1.3.4 Methods

- Systematically review longitudinal research examining dietary influences on childhood and adolescent overweight and obesity.
- Examine reported associations between childhood diet and children's growth/adiposity over time to determine the key dietary factors that increase or decrease the likelihood of obesity in adolescence.
- Design a prototype risk questionnaire based on the evidence from the systematic review.
- Using data from an identified high quality childhood cohort or combined cohorts, build a logistic regression model (with candidate variables matched to the questionnaire) to predict 3 year risk of obesity in young people.
- Test the reliability and internal validity of the predictive model.
- Translate the final model coefficients into a questionnaire with risk scores.
- Externally validate and pilot test the predictive model (beyond the scope of this research).

1.3.5 Significance

Preventing obesity is a major public health challenge, but resources are finite. This study sets out to develop and internally validate a dietary assessment tool which includes dietary risk factors for future obesity, adding evidence to the IOTF framework.

A better understanding of the determinants of obesity, including diet.
This study will synthesise evidence about diet and childhood obesity, which may influence policy and dietary recommendations and raise public awareness of dietary patterns that promote healthy weight gain in childhood and adolescence.

A way to identify at risk groups who might benefit from obesity prevention measures.
After external validation and pilot testing of the predictive model, risk scores from the questionnaire could be used by health professionals to identify childhood populations at high risk of adolescent obesity, in order to target public health messages and interventions designed to reduce the risk.

A way to evaluate the effectiveness of obesity prevention interventions.
Risk scores from the questionnaire could be used to measure immediate and longer-term change in risk to evaluate the effectiveness of an intervention.

1.4 Childhood obesity risk tools

Before developing a new predictive tool for the risk of childhood obesity it is helpful to consider what has already been done. Ten years ago, while working as a professor of child health at the University of Leeds, Mary Rudolf was asked by the Department of Health (UK) for her views on childhood obesity prevention (Rudolf, 2011). Professor Rudolf was aware of infant risk factors that increased the risk of childhood obesity, as identified by epidemiological studies. One such study originated from the Avon Longitudinal Study of Parents and Children (ALSPAC) and investigated early life risk factors for obesity (Reilly et al., 2005). Reilly et al ascertained that parental obesity, more than eight hours watching television per week at 3 years and short sleep duration at 3 years, as well as aspects of infant weight gain, were associated with a higher risk of obesity at 7 years old.

This prompted the idea of an evidence-based obesity risk tool (ORT) to predict a baby's obesity risk using routinely collected perinatal information. Instead of relying solely on growth charts and alerting parents only when an infant's growth trajectory crossed two centiles or lay above the 98th percentile, guidance could be given to parents *before* their child developed obesity. Rudolf and colleagues reviewed the evidence around early-life risk factors, looked at how main risk factors could be built into a simple tool for use in primary care settings, considered further research and development and discussed the practical, ethical, legal and policy aspects of using such a tool (Levine et al., 2012).

The feasibility of using perinatal risk factors to predict childhood overweight and obesity was explored using data from the ALSPAC cohort and from the Millennium Cohort Study (MCS). Factors considered as putative predictors included parental BMI, maternal age, ethnicity, education, smoking, sleeping patterns, birth weight and infant weight gain. Maternal pre-pregnancy BMI was one of the strongest risk factors. Predictors included in the final model were ethnicity, household obesity (maternal and/or paternal), child's early weight gain, birth weight (large for gestational age) and mother's education level (degree). A paper-based version of the tool, with a simple scoring system, was created. However, the sensitivity (ability to correctly identify infants who will develop obesity) and specificity (ability to correctly ignore infants who will not develop obesity) of the prototype ORT were not acceptable for use with individual cases in primary care. Further development work was proposed.

This influential paper raised important ethical issues, pertinent to the use of any risk tool. Concerns were voiced about the potential of the ORT to do more harm than good, either by stigmatising or antagonising parents of "at risk" infants to the point of disregarding advice or being uncooperative, or by alarming parents so much that they might adopt inappropriate feeding regimes for their child (Levine and Rudolf, 2011). A further ethical consideration was

that of follow-up. What level of intervention or remedial help would be made available to the infants and families deemed at risk?

Since the early 2000s, more childhood obesity risk tools have been developed and published. Several papers included in systematic reviews of models/tools to predict overweight or obesity cited Levine, Dahly and Rudolf's earlier work.

A recent systematic review of prediction models for childhood overweight/obesity looked for studies that used maternal and early life risk factors for the individual estimation of future risk of childhood overweight and obesity, finding eight studies (Ziauddeen et al., 2018). One included study used data from the ALSPAC cohort and 9 other cohorts to develop and validate an infant risk score for obesity at 7 and 11 years, based primarily on infant weight gain and maternal BMI (Druet et al., 2012). Druet et al acknowledged that additional factors might improve the predictive ability of their infant risk score but stated that data was either not available or was not generalisable across all 10 cohorts.

Across all the studies included in Ziauddeen's review, the most selected predictors were maternal BMI, birthweight and infant's sex, with breastfeeding and/or the introduction of solids included in two models. Six studies defined risk predictors a priori, with four giving a rationale for choosing predictors. Papers reported a median 23 of the 37 items recommended by the TRIPOD statement for Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (Collins et al., 2015). Titles clearly identified each study as developing/validating a multivariable prediction model and all papers reported their objectives, study designs, sources of data, participants, sample size, predictors, outcomes, statistical analysis methods and funding sources. Most papers described the flow of participants through the study and participant characteristics of those with and without the outcome, giving numbers of those missing data for predictors and outcomes. The majority presented a full prediction model with all regression coefficients/O.R.s and the intercept, with at least some performance measures and an explanation of how to use the model, an interpretation of results and a discussion of study limitations, potential use of the model and implications for future research. Reporting items most often omitted included actions for blind assessment of the outcome and predictors for the outcome, how missing data was handled (complete cases or imputation method), comparisons of development and validation settings (highlighting differences in demographics or distribution of predictors or outcomes), the number of participants and events in each analysis, and discussion of results in the validation data with reference to performance in the development data.

A subsequent systematic review of tools to predict infant, childhood and adult obesity identified 12 papers, describing 12 tools (Canfell et al., 2018). As most of the included papers

aimed to predict overweight/obesity between the ages of 2 to 12 years; there is overlap with the eight papers included in the review by Ziauddeen et al.

All but three of the thirteen tools in the two systematic reviews were internally validated. One study was internally and externally validated by the same authors (Redsell et al., 2016; Weng, S. F. et al., 2013) which may have introduced bias. The same authors later examined the feasibility of using the externally validated Infant Risk of Overweight Checklist (IROC) (Redsell et al., 2016) in clinical practice, to differentiate between infants at or above the population risk of overweight (Redsell et al., 2017). Two studies were externally validated, and both used data from the ALSPAC cohort to validate models to predict obesity at two years (Santorelli et al., 2013) and overweight at five years (Redsell et al., 2016).

Only two studies reported all three performance measures (calibration, discrimination and decision curve analysis) that are recommended for the validation of clinical prediction models (Steyerberg and Vergouwe, 2014). Discrimination (the ability to correctly differentiate between individuals with and without the outcome of interest) of validated models or tools, as measured by the concordance statistic, ranged from 0.64 (poor) to 0.89 (excellent). To date, none of these obesity risk tools appear to have been used in a real-life setting.

Putative predictors of future obesity considered for the different tools in both systematic reviews included anthropometric, socio-economic/demographic and clinical variables plus infant/maternal diet history, with two tools adding genetic variables associated with obesity in their attempts to enhance predictive accuracy. See Table 1-2. Not all predictors were included in the final models.

The Childhood Obesity Risk Evaluation (CORE) featured in both reviews, and used a score based on mother's pre-pregnancy BMI, education level and smoking plus infant weight gain and child's sex, to predict obesity among Greek children or teenagers (Manios et al., 2013). Unusually, this study utilised cross-sectional data with retrospective data collection, rather than data from a prospective cohort. In their examination of the utility of CORE, the authors adjusted for children's breakfast, fruit and vegetable frequency and leisure time activity level (Manios et al., 2016), although these factors were not used specifically as predictors.
No other study in either review considered dietary intake, physical activity or sleep in childhood as putative predictors of future obesity. Consequently, these exposures did not feature in any of the final models. This research gap presents an opportunity to investigate children's diet as a potential predictor of future obesity, possibly alongside other factors related to energy balance, which were shown to predict obesity risk in early life (Reilly et al., 2005).

Table 1-2 Examples of putative predictors considered by studies in two Systematic Reviews of tools to predict obesity in infants, children and adults

Type of predictor	Examples
Anthropometric	Maternal weight/BMI, pre-pregnancy or current
	Paternal BMI
	Child's birthweight
	Child's weight gain in infancy or early childhood
	Child's BMI z score
Socio-economic and socio-demographic	Maternal age
	Maternal education level
	Paternal education level
	Paternal employment
	Full time work vs at-home mother
	Family income
	Mother's marital status
	Number in household
	Child's sex/gender
	Ethnicity
	Living in highly urban environment
Clinical	Smoking, during pregnancy, parental or household
	Birth order/ number of siblings
	Hospital delivery
	Delivery type
	Gestational age/premature birth
	Infant fussiness
	Infant developmental stages
	Maternal health
	Maternal alcohol consumption
Diet history	Breastfeeding/formula feeding
	Solids at < or > 6 months
	Maternal vegetable consumption in pregnancy
Genetic	Single nucleotide polymorphisms associated with obesity

1.5 Thesis structure

The research for this PhD was carried out in three phases:

- Phase 1: Systematic Review and meta-analysis
- Phase 2: Preparation of cohort data and model development methods
- Phase 3: Dietary risk tool development – model fitting and internal validation

The thesis is structured broadly in line with these phases. See Figure 1-3. Numbered chapters relating to each phase are also shown in the figure. In brief, the thesis chapters present the following:

Chapter 1: Introduction

- Obesity prevalence in adults and children, worldwide and in England, and risks to health. An evidence framework for childhood obesity. Research aims and objectives. Brief literature review of published obesity risk tools.

Chapter 2: A systematic review of childhood and adolescent cohorts measuring whole diet and subsequent adiposity: Methods

- Systematic review protocol and registration. Methods used to develop and execute a search strategy, screen records in duplicate, assess the quality of included studies and extract data.

Chapter 3: Systematic Review: PRISMA results with characteristics of included cohorts

- Results of literature searches and screening on title, abstract and full text, with record numbers in a PRISMA flow chart. Level of agreement between reviewers. Quality assessment of included papers using the Newcastle-Ottawa scale for cohort studies. Overview of extracted information including cohort characteristics.

Chapter 4: Narrative Review: Dietary assessment and measures of adiposity

- Dietary assessment tools (food diaries, 24 hour dietary recalls and food frequency questionnaires) and adiposity measures used by included studies, their pros and cons and the steps taken to minimise measurement error. Baseline mean energy intakes in each childhood cohort are compared by age, DAT and country. Macronutrients share of energy intake are compared with WHO nutrient intake goals. Children's mean Body Mass Index values are compared with the UK 1990 growth reference.

Chapter 5: Narrative Review: Reported foods and drink intakes and adiposity outcomes

- Narrative synthesis of quantified intakes of whole grains, dairy foods including milk, fruit and vegetables including juice, fish, convenience foods, snack foods, sugar sweetened and diet beverages and reported longitudinal associations with adiposity.

Chapter 6: Narrative Review: Dietary patterns, eating habits and multiple predictors, and subsequent adiposity

- Narrative synthesis of dietary patterns, diet quality scores and eating habits and reported longitudinal associations with adiposity. Narrative synthesis of multiple predictors (dietary and non-dietary) of future overweight.

Chapter 7: Meta-analysis of sugar sweetened beverage intakes and adiposity outcomes

- Exploratory quantitative synthesis of reported SSB intakes and various adiposity outcomes. Meta-analysis of SSB intake and change in BMI, with forest plots.

Chapter 8: Tool development: Pre-specification of candidate predictor variables and preparation of the ALSPAC data

- Risk algorithms in different settings. Principles of predictive model development. Pre-specification of childhood predictors of adolescent obesity. Check of assumptions using a directed acyclic graph. Evidence based questionnaire, matched to candidate variables (potential predictors) in the Avon Longitudinal Study of Parents and Children (ALSPAC). Preparation of ALSPAC variables, including imputation. Random 3:1 split of the dataset into derivation and validation samples for model development and internal validation.

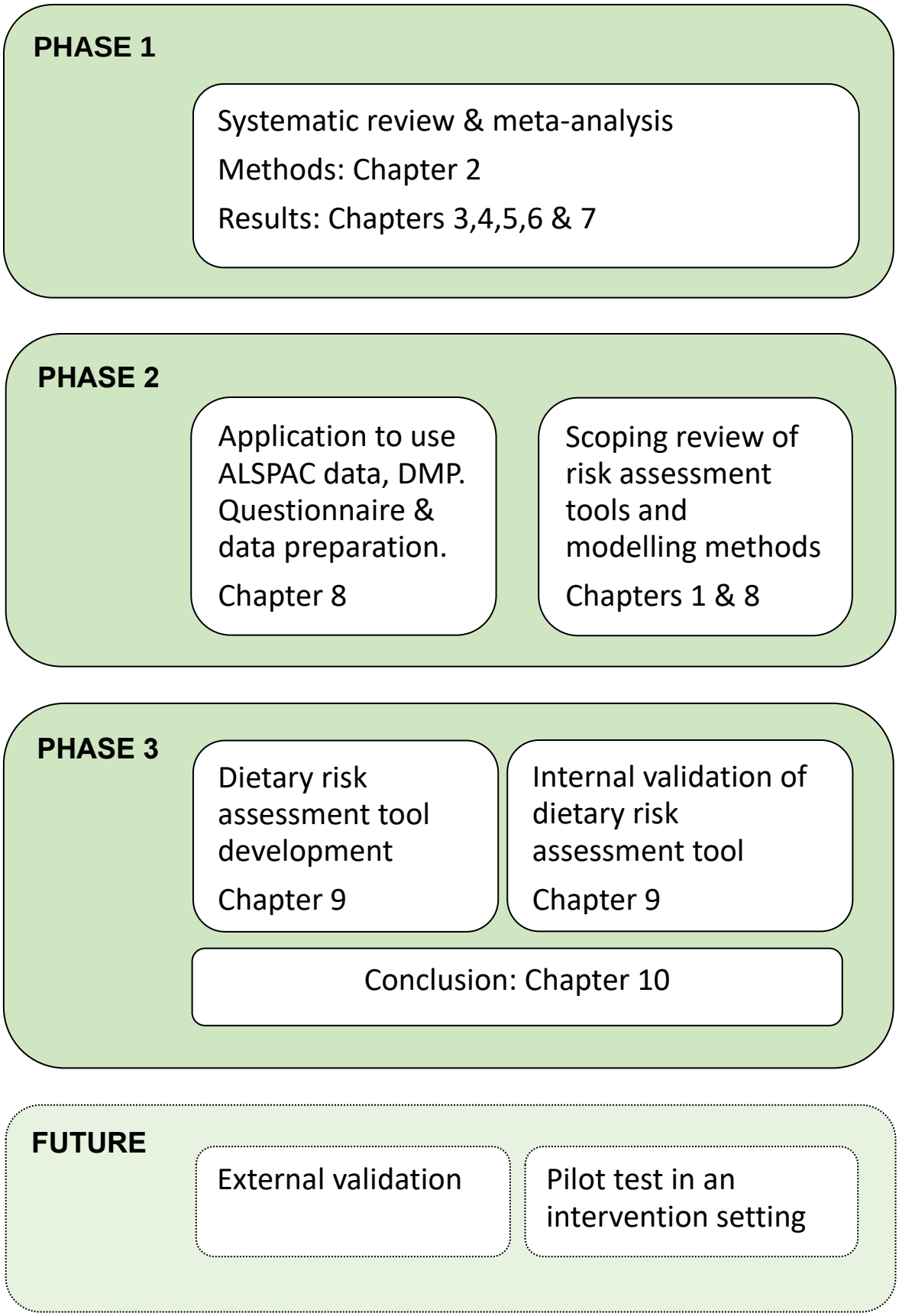
Chapter 9: Tool development: Model fitting, internal validation and risk score allocation

- Methods used to develop and internally validate a logistic regression model in the ALSPAC data to predict obesity at 13+ years, based on potential predictors measured at or close to 10+ years old. Presentation of the reduced, interim and final models. Results of internal validation (discrimination, calibrations and clinical usefulness).

Chapter 10: Conclusions and recommendations for future research

- Conclusions, limitations and challenges, strengths, future research and public health implications.

Figure 1-3 Overview of research and thesis structure



Chapter 2 A systematic review of childhood and adolescent cohorts measuring whole diet and subsequent adiposity: Methods

2.1 Aims of the Systematic Review

The systematic review aims:

- To identify cohorts with quantitative data on diets (the exposure) including total energy intake during childhood and/or adolescence, that have also taken anthropometric measurements of some or all participants at a follow-up during later adolescence, enabling an assessment of participants' adiposity, overweight or obesity status (the outcome) to be made with defined dietary exposures.
- To critically appraise identified cohorts and their chosen methods of dietary assessment and measures of adiposity (validated or not).
- To synthesise the reported evidence about diet in childhood or adolescence and longitudinal adiposity, overweight or obesity outcomes.

The review sets out to address the following research question:

> *To what extent does diet during childhood or adolescence influence future indicators of overweight or obesity?*

2.2 Summary

A systematic review protocol was developed in accordance with PRISMA-P guidelines, PROSPERO registration: CRD42015030081.

The research question was broken down into concepts using a PICO-S (participants, intervention/exposure, comparators, outcomes, study type) framework to specify study characteristics and inclusion criteria.

Five databases, Ovid Medline, Embase, Cochrane central register of controlled trials (CENTRAL), Scopus and Web of Science, were searched using the following search terms:

- (Child OR Adolescent) AND (Diet OR Diet record OR Energy intake OR Nutrition assessment) AND (Anthropometry OR Body composition OR Body Mass Index OR Waist circumference OR (Body height AND Body weight)) AND (Clinical trial AND Follow-up) *or* AND (Cohort study)

Relevant reviews were checked for additional articles.

Records were managed using Endnote X7. De-duplicated records were screened on title and remaining records were screened against the inclusion/exclusion criteria by two independent reviewers, first on title and abstract, then full text articles. A third reviewer resolved differences. Numbers at each screening stage were noted for a PRISMA flow diagram.

The quality of included studies was assessed using the Newcastle-Ottawa scale for cohort studies. Pre-specified data was extracted from included papers, to Excel.

Included studies were grouped by common cohort and by the dietary exposures they reported for the narrative synthesis. Where statistical synthesis was feasible, extracted data was imported to STATA for random effects meta-analysis.

2.3 Protocol and Registration

During Autumn 2014 a protocol was written with reference to the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement (Moher et al., 2015). This statement was downloaded from the EQUATOR (Enhancing the QUality And Transparency Of health Research) website at http://www.equator-network.org/.

Protocols prepared in accordance with PRISMA-P are designed to help authors register the protocol and to write up a full text systematic review that complies with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) reporting guidelines. The PRISMA-P statement has a checklist of 17 numbered items to include in protocols of systematic reviews and meta-analyses; the wording of the checklist is harmonized with the longer 27 item PRISMA checklist for the full systematic review report (Moher et al., 2009). The methods and results are reported broadly in-line with this PRISMA checklist. See http://prisma-statement.org/

The protocol, entitled "A protocol for a systematic review of childhood and adolescent cohorts which measure whole diet and subsequent adiposity", includes a description of the planned methods for the systematic review.

In November 2015, when the systematic review was underway but not completed, an updated summary version of the protocol was prepared (See Appendix A) and registered with PROSPERO, an international database of prospectively registered systematic reviews. The registered protocol can also be found via the PROSPERO website: http://www.crd.york.ac.uk/PROSPERO, registration number CRD42015030081.

By providing a list of systematic reviews registered at the protocol stage the PROSPERO register aims to avoid unplanned duplication, promote transparency and minimise the risk of bias.

2.4 Eligibility criteria

As recommended by PRISMA a PICO-S (participants, interventions, comparators, outcomes, study type) framework was used to breakdown the research question into main concepts. As the studies required were observational, "exposure" was substituted for "interventions". Concepts are shown in Table 2-1.

Table 2-1 Concepts identified using a PICO-S framework

Participants	Intervention or Exposure	Comparators or Settings	Outcome	Study Type
CHILDREN **ADOLESCENTS**	DIET NUTRITION FOOD INTAKE ENERGY INTAKE DIET QUALITY	ANY	BODY COMPOSITION HEIGHT and WEIGHT	COHORT STUDY Or CLINICAL TRIAL with FOLLOW UP

2.4.1 Study Characteristics

Studies meeting the following criteria were considered eligible for inclusion:

- Mixed or single sex studies of healthy children or adolescents where half or more of the participants were aged from 8 to 19 years old at baseline. Participants may be classified as underweight, normal weight, overweight or obese at baseline, reflecting the range of weights that exist within populations.
- Studies which used an objective measure of *whole* dietary intake derived from a weighed or un-weighed diet diary, 24 hour diet recall or a quantitative or semi-quantitative food frequency questionnaire, from which total or partial intake of foods and drink (the exposure) was quantified and reported.
- Studies which used anthropometric measurements (as a minimum self-reported height and weight) taken at least two years after measuring diet, from which adiposity or overweight/obesity status (the primary outcome) was assessed and reported.
- Observational, longitudinal or cohort studies or clinical trials/intervention studies with an untreated control group that had followed up participants for at least 2 years.

2.4.2 Report characteristics

Reports meeting the following criteria were considered eligible for inclusion:

- Reports of observational (prospective) cohort studies published since 1990. The relatively few papers published before this date are about earlier cohorts whose diet may be less typical of the present day.
- Reports of clinical trials with follow-up where there is an untreated or placebo control group, published since 2010. This date overlaps with the searches reported in a Cochrane Systematic Review of interventions for preventing obesity in children (Waters et al., 2011). The overlap helped to avoid replicating their work while allowing for some variation in the timings of database updates (Yoshii et al., 2009).

Literature published in languages other than English was logged but not included in the review.

2.5 Information sources

Existing reviews were searched for in the Cochrane Database of systematic reviews and in "other reviews" held in the Cochrane Database of Reviews of Effect (DARE).

Five bibliographic databases were searched to find eligible reports: Ovid Medline, Embase, Cochrane central register of controlled trials (CENTRAL) for trials only, Scopus, Web of Science.

Reference lists of relevant reviews were examined for articles not found by database searches.

2.6 Search strategies

Concepts identified using the PICO-S framework helped to develop search strategies for the bibliographic databases based on key words and related medical subject headings (MeSH) terms. MeSH terms used in papers about known cohorts and in relevant reviews provided helpful ideas. Scope notes in the Medline database provided clear definitions of MeSH terms. For example

- CHILD, PRESCHOOL = an individual 2 to 5 years of age
- CHILD = a person 6 to 12 years of age
- ADOLESCENT = a person 13 to 18 years of age
- ADULT, YOUNG = a person 19 to 24 years of age

Some reviews featured their search strategies (as recommended by PRISMA) which provided additional guidance. Developed search strategies were tried out several times, with and without ready-made search filters. After consultation with the Mathematics and Physical

Sciences Faculty librarian, Lizzie Caperon, a simpler, third set of strategies using fewer search terms was devised. Truncation and wildcards were used, and search terms were tailored to fit the specific requirements of each bibliographic database.

Two search strategies were deployed in April 2015:

- All five databases (Ovid Medline, Embase, CENTRAL, Scopus, and Web of Science) were searched for reports of clinical trials with follow-up, published since 2010 and limited to human studies with no restriction on language.
  Search strategy: (Child OR Adolescent) AND (Diet OR Diet record OR Energy intake OR Nutrition assessment) AND (Anthropometry OR Body composition OR Body Mass Index OR Waist circumference OR (Body height AND Body weight)) AND (Clinical trial AND Follow-up).
- Four of the chosen databases (Ovid Medline, Embase, Scopus, and Web of Science) were searched for reports of cohort studies published since 1990 and limited to human studies with no restriction on language.
  Search strategy: (Child OR Adolescent) AND (Diet OR Diet record OR Energy intake OR Nutrition assessment) AND (Anthropometry OR Body composition OR Body Mass Index OR Waist circumference OR (Body height AND Body weight)) AND (Cohort study).

Full details of the nine searches, with dates, are included in the Appendices, see Appendix B.

A simple search strategy, (Child OR Adolescent) AND Diet, run in April 2015, was used to find relevant systematic reviews in the Cochrane Database of Systematic Reviews and to find other reviews in the Cochrane Database of Reviews of Effect (DARE). A search for relevant systematic reviews in the NHS Evidence database was employed in June 2015. The reference lists of relevant reviews were hand-searched for additional potentially relevant articles that had not been found by database searches.

2.7 Record Management

Endnote X7 was used throughout to manage the results of the searches, de-duplication and screening against the inclusion/exclusion criteria, creating a new Endnote library (or libraries) with back-up copies for each stage in the study selection process. See Appendix C.

Records found by the nine searches were exported from each bibliographic database into two *Database searches* Endnote libraries (One for Clinical trials and follow-up and one for Cohort studies) in the following priority order: Ovid Medline, Embase, Cochrane central register of controlled trials (CENTRAL) for trials only, Scopus, Web of Science.

In Ovid Medline and Embase the complete reference was exported, but records from Cochrane CENTRAL, Scopus and Web of Science currently provide less bibliographic information. Endnote software automatically assigns a unique record number (#n) to each record as it is imported to the Endnote library, displaying the unique record number at the top left corner of the record page. The pre-specified database export order helped ensure that the first record imported into the Endnote library (with the lowest unique record number) was likely to be the most complete record available (Rader et al., 2014).

Records were grouped by their origin (named bibliographic database search or additional search) and the number of references from each source was noted. Master copies of the two *Database searches* Endnote libraries were saved.

Bibliographic records of potentially relevant articles cited in reviews, but not found by the database searches, were sought and imported into a *Combined All searches plus additional sources* Endnote library alongside the database search results, after de-duplication.

2.8 De-duplication

De-duplication was carried out in May 2015. All references were sent to two *De-duplication* Endnote libraries, created specifically for the de-duplication process, with records sorted by Author. Using the "Preferences>Display Fields" option in the drop-down list from the Edit tab, the library window was set up to show the following fields for each record: Name of Database, Author, Year, Title, Journal, Pages, Accession number (a unique record number or alphanumeric ID assigned to a reference by the online bibliographic database provider) and URL (Uniform Resource Locator or address for documents accessible over the internet).

References, grouped alphabetically by first author, were searched in batches for duplicate records using the "Find duplicates" command in the drop-down list from the References tab. Fields for matching were pre-specified using the "Preferences > Duplicates" option in the drop-down list from the Edit tab. Several combinations of fields were used to search for duplicates. E.g. Author, Year and Title; Author and Year; Author and Title; Title; Author, Year and Journal; Year, Journal and pages.

The option to "automatically discard duplicates" was *not* selected. Instead retrieved records were visually compared to ensure that all display fields matched exactly.

For each set of duplicate records, the first imported record is displayed on the left, with potential duplicate records shown on the right for comparison one-by-one. The check box option to "keep this record" automatically sends all other duplicates in the set to trash. This step was cancelled so that a summary list of retrieved duplicates was displayed. The first

imported record is listed first with potential duplicate records highlighted beneath. After a visual comparison of the display fields, records identified as duplicates were selected and manually moved to trash. Duplicates were immediately restored from trash into unfiled records and filed in a duplicates group.

Duplicates were also sought "by eye". Due to different presentations of authors' names and journal names between databases, mismatching abbreviations, capitalizations and misspelling not all duplicate records could be retrieved using "Find duplicates".

Duplicate records were copied to *Duplicates kept* Endnote libraries before deletion, so that no records were deleted altogether.

Only the first listed record with the lowest unique record number (first imported, so likely the most complete record available) was kept in the de-duplicated libraries (Clinical trials and follow-up and Cohort studies). A backup copy of each de-duplicated Endnote library was made, and the numbers of records were noted. The two libraries were combined and re-checked for duplicates. Then, as previously described, additional records found from the reference lists of relevant reviews were added to the *Combined All searches plus additional sources* Endnote library ready for screening.

2.9 Screening process and study selection

The screening process had three stages:

- First stage screening on publication date and title, May – July 2015
- Second stage screening on title and abstract, July - October 2015
- Third stage screening on full text, October 2015 - March 2016

Endnote libraries were created for each stage of systematic review screening and for each independent reviewer, with back-up copies. Record numbers were noted after each stage of screening and recorded in a PRISMA flow diagram. The flow diagram numbers were updated each time inclusion/exclusion decisions were revised – third stage screening was an iterative process.

2.9.1 First stage screening on publication date and title

The publication dates of all records in a de-duplicated *First pass #1 screening* Endnote library were checked for agreement with the dates specified in the search strategies. Titles were checked for clearly irrelevant studies from which to generate a list of keywords. Searches in the title field using these keywords were used to find other potentially irrelevant papers. The keywords used for searches in the title field are listed in the Appendices, see Appendix D.1. If clearly ineligible from the title the record was excluded. Remaining records were copied to a *Second pass #2 screening* Endnote library for the next screening stage.

2.9.2 Second stage screening on title and abstract

Using duplicate copies of the *Second pass #2 screening* Endnote library two reviewers independently checked the titles and abstracts of the remaining records against the selection criteria using a simple screening questionnaire, based on the study characteristics identified using the PICO-S framework. The #2 screening questionnaire is shown in Appendix D.2

During second stage screening the independent reviewers were research postgraduate student, Catherine Rycroft (CR) and visiting French undergraduate student, Marion Héry, (MH).

Each reviewer entered their second stage screening decisions into their *Second pass #2 screening* Endnote library as a simple code in the Research notes field of each record, with the option to add explanatory comments. Based on the decision code, reviewers then placed each record in one of four groups:

- Non-English Abstract
- Exclude (with subgroups for study type, participants, exposure or outcome)
- Potentially Include
- Unclear

The simple #2 screening questionnaire was piloted twice with 20, then 30 records, to help achieve a high level of inter-rater reliability. Some studies were specifically about breast-feeding, a dietary factor which was not intended to be within the scope of this review; the decision was made to exclude such studies. Once all records had been independently screened and placed in a group the level of agreement between Reviewer 1 (CR) and Reviewer 2 (MH) was assessed using Cohen's Kappa coefficient (Altman, 1991), which considers the agreement in excess of what would be expected by chance. (A value of 0 is no better than chance, 1 is perfect agreement.)

All records placed in "Potentially Include" and "Unclear" groups by either reviewer were copied to the *Resolving #2 differences* Endnote library for comparison of the specific

differences between Reviewer 1 (CR) and Reviewer 2 (MH) decisions. A check for duplicates revealed those records where both reviewers had decided to place the record in either the "Potentially include" or "Unclear" group. The remaining single records were those where there was a difference of screening decision (i.e. Only one reviewer had decided to place the record in the "Potentially include" or "Unclear" group, whereas the other reviewer had decided to exclude it.) These differences were resolved by a third reviewer, Dr Charlotte Evans (CE). All records deemed unclear as well as those to potentially include were copied to a *Third pass #3 screening* Endnote library, going forward to the third and final screening stage.

2.9.3 Third stage screening on full text

Full-text articles were sought, firstly by using the "find full text" function in Endnote X7, then by manually searching Find at Leeds (University of Leeds on-line library) and Google Scholar and finally papers in author profiles on Research Gate or by direct e-mail request. A pdf copy of the full-text article, if available, was attached to the bibliographic record in the *Third pass #3 screening* Endnote library. At this stage papers were only printed if no pdf version was available.

One reviewer (CR) scanned titles, author names, abstracts and the introduction and method sections of full-text articles in order to identify named studies. Records were grouped in Endnote by their common study, where multiple publications may be based on the same dataset.

We used a Cochrane template of a data collection form for intervention reviews as the basis of a two page data extraction/screening form. The template, which has sections that can be added to or removed if not required, is available as ""Good practice data extraction form" at *http://epoc.cochrane.org/epoc-specific-resources-review-authors*. In parallel with the data extraction/screening form we created a #3 screening questionnaire based on the inclusion and exclusion criteria. Copies of the #3 screening questionnaire and the #3 screening form are shown in Appendix D.3 and Appendix D.4. The questionnaire and form were piloted with 8 records and found to be helpful.

Using duplicate copies of the *Third pass #3 screening* Endnote library with pdf attachments, two sets of independent reviewers checked each full-text paper against the #3 screening questionnaire and each reviewer separately completed a copy of the #3 data extraction/screening form. Two forms, one per reviewer, were completed for each full-text paper.

During third stage screening CR was again Reviewer 1. Three individuals acted as Reviewer 2, each screening a section of the Endnote library. University of Leeds undergraduate student

Ben Clapinson (BC) screened pdfs of full texts, first authors A to C. CE screened pdfs of full texts, first authors D to Z and Professor Janet Cade (JC) screened printed papers with no pdfs available.

On the screening form full text articles were identified using first author surname and year of publication, together with the *Third pass #3 screening* Endnote library record number. If given in the text, the study or cohort name was recorded and Endnote library record numbers of other articles from the same study or cohort were noted. When extracting data each reviewer entered their name and the date on the #3 screening form and noted the article type and whether it was available as full-text in English. If applicable, reviewers went on to consider study eligibility using the #3 screening questionnaire as a prompt, answering each question in turn. If the answer to a 3# screening question was YES or UNCLEAR, the next question was asked. If the answer to a question was NO that record was excluded and subsequent questions were not considered. Extracted information about the study type, participants, exposure and outcome measures was recorded on the #3 screening form up to the point where a record was excluded, or all questions were answered. Each reviewer marked their overall decision (INCLUDE, UNCLEAR or EXCLUDE with the reason for deciding to exclude a record) at the end of the form.

Reviewer 1 entered details of all third stage screening decisions into a Reviewer 1 *Third pass #3 screening* Endnote library. Decisions were entered as a code in the Research notes field of each record, with the option to add explanatory comments. Each record, according to the decision code, was placed in one of three main groups:

- Exclude (with subgroups for the specific exclusion reason – Abstract only, Non-English full text, Wrong study type, Short study under 2 years, Unrepresentative participants, Participants too young, Participants too old, Diet not measured, Used diet history, Whole diet not quantified, Adiposity not measured, Adiposity measured only at baseline, Adiposity measured less than 2 years after measuring diet, Association between diet and adiposity not reported)
- Unclear
- Include

This process was not repeated in full detail for all the third stage screening decisions of Reviewer(s) 2. Only reviewers' inclusion decisions were entered in the Reviewer 2 *Third pass #3 screening* Endnote library.

Summaries of Reviewer 1 (CR) and Reviewer 2 (BC, CE and JC) decisions were copied into an Excel spreadsheet to compare and identify differences of inclusion or exclusion. Again, the level of agreement between the two sets of reviewers was assessed using Cohen's Kappa coefficient.

All records placed in the "Include" groups of the *Third pass #3 screening* Endnote library by either/any reviewer were copied to the *Resolving #3 differences* Endnote library for comparison of the specific differences between Reviewer 1 and Reviewer 2 decisions. A check for duplicates revealed where both reviewers had agreed to include the record. The remaining single records were those where there was a difference of screening decision. i.e. Only one reviewer had decided to place the record in the "include" group, whereas the other reviewer had decided to exclude it or else was unclear whether to include it or not. These differences were resolved by a third reviewer (CE or JC) who each only considered disputed records that they had not previously screened.

Third reviewer decisions were recorded on the Excel spreadsheet and in the *Resolving 3# differences* Endnote Library. All records included by two reviewers, or included after third review, were copied to the *Included studies for data extraction* Endnote library.

2.10 Quality Assessment

2.10.1 Selection of a quality assessment tool

Quality assessment is a vital part of the systematic review process as the quality of evidence generated by a systematic review depends upon the quality of the studies that have been included within it. (In this context quality means the methodological quality or internal validity of the primary studies, rather than the quality of reporting.)

Many methodological quality assessment tools have been developed. Almost 200 different tools were mentioned in recent reviews and evaluations (Deeks et al., 2003) (Sanderson et al., 2007) (Zeng et al., 2015).

Methodological quality assessment tools try to answer the question, "What was done in each individual study to promote internal validity and reduce the risk of bias?". Internal validity can be affected by all sorts of bias E.g. selection bias, performance bias, detection bias, attrition rates, confounding. Some of these biases are related to the study design, so the review team must evaluate each study based on the risks of bias specific to the study design. A quality assessment tool designed for randomised controlled trials (RCTs) will not work well with a cohort study and vice-versa.

A range of tools for the different types of study design was reported in a systematic review of methodological quality assessment tools. This review found an abundance of tools for randomised controlled studies but fewer tools for assessing observational cohort studies. (Zeng et al., 2015). Three main types of tool were described:

- Item – individual components relevant to clinical research methodology, revealing whether results might be biased. E.g. was there concealment and blinding?
- Checklist – many items for assessment of study quality and risk of bias, no scoring.
- Scale – many items for assessment of study quality and risk of bias, with every item scored and combined to give a summary score.

The review by Zeng et al recommended the Cochrane Collaboration's tool for assessing risk of bias in randomised controlled trials (Higgins et al., 2011) and featured both the Critical appraisal skills programme (CASP) checklist for RCTs and the National Institute for Clinical Excellence (NICE) methodology checklist for RCTs. For cohort studies the same review featured the CASP checklist for cohort studies, the Scottish Intercollegiate Guidelines Network (SIGN) methodology tools and the Newcastle-Ottawa Quality Assessment scale.

Another systematic review of quality assessment tools suited specifically to observational studies in epidemiology (Sanderson et al., 2007) assessed 86 tools: 41 simple checklists, 12 checklists with summary judgements and 33 scales. Most tools addressed one study design type (case-studies, cohorts or cross-sectional studies) or a combination of case-control and cohort studies. A third of the tools reviewed were intended for the critical appraisal of single studies, another third of the studies were created for single use in a specific systematic review. One sixth of the tools reviewed were generically designed for use in any systematic review of observational studies while the purpose of remainder was ambiguous.

The content of each tool was evaluated against the following domains:

- selection of participants *.
- appropriate measurement of both exposure and outcome variables *.
- appropriate methods to address design-specific sources of bias, such as recall bias, interviewer bias, blinding, loss to follow-up.
- control of confounding (using the appropriate design and/or analytical methods) *.
- appropriate use of statistics
- declaration of conflict of interest/funding sources

No tool included all six domains. Half of the tools included all three of the starred * domains that the reviewers regarded as fundamental while most tools included the first three domains

on the list, three-quarters included the next two. Only three tools considered conflict of interest.

In their summary, Sanderson et al made no recommendations but advised that tools should include:

- a small number of key domains (including three fundamental * domains)
- be specific to the study design
- use a simple checklist rather than a scale. (The authors cautioned that scales with summary scores, while appearing simple, inherently include weighting of items and some of these items may not be related to validity. This caution was given, albeit to a lesser extent, for checklists with summary judgements.)
- show evidence of careful development, validity and reliability

An earlier evaluation of 194 checklists and scales that had been used for assessing the quality of non-randomised studies (Deeks et al., 2003) reported that many were poorly developed while some tools for RCTs had been inappropriately applied to another study design. The top fourteen tools, assessed for internal and external validity and quality of reporting, featured at least five of six domains for internal validity (creation of groups, blinding, sound information about the intervention or exposure and outcomes, adequacy of follow-up, comparability of groups, appropriateness of analysis). Two tools, CASP and Newcastle-Ottawa (Wells et al., 2009), were designed for the quality assessment of cohort studies. The Downs & Black tool (Downs and Black, 1998) designed to assess the methodological quality of both randomised and non-randomised studies was also in the top fourteen. In their evaluation Deeks et al found that the Downs & Black tool, although long, was clear and easy to use (Deeks et al., 2003).

This brief survey of the literature demonstrated that there is no one outstanding methodological quality assessment tool; all have their pros and cons. To help decide which quality assessment tool (or tools) to use, key factors about the strongest contenders as described in the reviews, were compared. Seven quality assessment tools were considered: CASP for RCT, CASP for cohorts, SIGN 50 for RCT, SIGN 50 for cohorts, NICE for quantitative intervention studies, Downs & Black and the Newcastle-Ottawa scale.

The chosen instrument needed to be suitable for use in a systematic review where the aim is to compare quality across many studies, rather than a tool intended to help critically appraise single studies, ideally to be a checklist and be easy to use.

Evidence from the reviews and from each quality assessment tool was summarised in a comparison table (See Table 2-2) from which two potential options emerged:

- NICE Quality Appraisal checklist.
- Newcastle-Ottawa quality assessment scale (so-called, but a checklist).

In the protocol we stipulated the use of the NICE Quality Appraisal checklist which, with modifications, can be used for the quality appraisal of both RCT and cohort studies. However, as the systematic review sought evidence from cohort studies or the untreated control arm of RCTs, it was decided that a dual approach or flexible tool was not necessary – a quality assessment tool designed for cohort studies would be suitable. The Newcastle-Ottawa quality assessment tool for Cohort studies was used instead of the NICE Quality Appraisal checklist. The Newcastle-Ottawa tool was simpler to customize, easier to understand and much shorter (9 questions rather than 27) and therefore quicker to use, but still valid.

Table 2-2 Choosing a quality assessment tool

Tool Name and attributes	CASP	CASP	SIGN 50	SIGN 50	NICE for quantitative intervention studies	Downs & Black	Newcastle Ottawa Scale
Which study type? RCT or cohort?	RCT	Cohort	RCT	Cohort	Both	Both	Cohort
Suitable for systematic review?	No - X	No - X	No? -X	No? - X	Yes?	Yes?	Yes
Simple checklist or Scale?	Ch.list	Ch.list	Ch.list	Ch.list	Ch.list	Scale- X	Ch.list
Validated?	N	N	Yes?	Yes?	Yes?	Yes?	No
Number of questions?	11	12	17	19	27	27	9
Reporting Quality?	N	N	N	N	N	Y	N
Internal validity – participants	Y	Y	Y	Y	Y	Y	Y
Internal validity – exposure	Y	Y	Y	Y	Y	Y	Y
Internal validity – outcome(s)	Y	Y	Y	Y	Y	Y	Y
Internal validity –design biases	Y	Y	Y	Y	Y	Y	Y
Internal validity – confounding	Y	Y	N	Y	Y	Y	N
Internal validity – statistics	Y	Y	N	Y	Y	Y	N
Internal validity – conflict of interest or funding?	N	N	Y	Y	N	N	N
External validity?	Y	Y	N	N	Y	Y	Y
Decision	NO	NO	NO	NO	**YES**	NO	**YES**

2.10.2 The Newcastle-Ottawa Scale

The Newcastle-Ottawa Scale (NOS) is a collaboration between the University of Newcastle, Australia and the University of Ottawa, Canada. NOS was developed to assess the quality of non-randomised studies and features a version for case-control studies and a version for cohort studies. Both are available through a web-based link to the Ottawa Hospital Research Institute (Wells et al., 2009). Each version is a simple checklist with questions across three domains: selection, comparability and either exposure (case-control studies) or outcome (cohort studies).

On the website the authors claim that the face/content validity and inter-rater reliability of NOS have been established while its validity in comparison with other quality assessment tools and its intra-rater reliability are being examined. The NOS has seemingly not been published in peer-reviewed journals (Stang, 2010). Nevertheless it has been widely used in recent Systematic Reviews and meta-analyses (O'Sullivan et al., 2013; Threapleton et al., 2013; Zhang et al., 2013; Chowdhury et al., 2014; Liu et al., 2014).

The Newcastle-Ottawa quality assessment tool for Cohort studies has eight questions in three domains. Some questions are general while others can be customised to suit the specific systematic review. Stars are awarded depending on the answer to each question, to a maximum of nine stars, as follows:

Selection (and exposure) ****

- Representativeness of the exposed cohort?
- Selection of the non-exposed cohort?
- Ascertainment of exposure?
- Demonstration that the outcome of interest was not present at the start?

Comparability **

- Comparability of cohorts based on design or analysis? (Stars for up to two control factors)

Outcome ***

- Assessment of outcome?
- Was follow-up long enough for outcomes to occur?
- Adequacy of follow-up of cohort?

2.10.3 Customising the Newcastle-Ottawa tool

A customized version of the Newcastle-Ottawa tool was created, specifying the characteristics of the exposed cohort (children or adolescents in the community), the best quality measures of the exposure (whole diet measured using a food record, food diary or multiple 24 hour recalls), the analysis (controlled for age and Total energy intake, TEI), the best quality measures of the outcome (adiposity measured by trained staff), length of follow-up (2 years or longer) and loss to follow-up (less than 25%).

Controlling for age in analyses was important as cohort participants were not all the same age at baseline and children's growth and development varies with age. Total energy intake is a factor in energy balance so is often considered to be on the causal pathway between dietary exposures and adiposity outcomes. It is usual practice to adjust for energy intake in analyses of dietary factors and weight change or adiposity outcomes, as a way of adjusting for bias in food and nutrient intake reporting by weight status (Laska et al., 2012).

Loss to follow up is common in longitudinal studies and may result in an unrepresentative sample if the characteristics of participants who dropout differ markedly from those of participants who remain in the study. Only one aforementioned systematic review paper which used NOS gave the cut-off that they applied for adequacy of follow-up, awarding prospective cohort studies one star if follow-up was ≥ 75% (Liu et al., 2014). Guided by this, an equivalent loss to follow-up of < 25% was used in the customised version of NOS.

2.10.4 Quality Assessment using the customised Newcastle-Ottawa tool

The internal validity of each study/paper that met the systematic review inclusion criteria was assessed by two independent reviewers (Recent graduate, Alice Kininmonth, AK, and CR) using the customised Newcastle-Ottawa quality assessment scale for cohort studies (See Appendix D.5). Each reviewer wrote their decisions on a quality appraisal form (See Appendix D.6). Awarded stars were entered onto a summary Excel spreadsheet to compare differences between reviewers. Differences were resolved by a third reviewer, CE.

2.11 Data Extraction

In preparation for data extraction, each potentially included paper was given a unique identification, based on the record number automatically generated by Endnote in the *Third pass #3 screening* Endnote library, plus the first author name and the publication year. E.g. #5 Affenito 2005, #390 Zheng 2015.

The unique ID was added to each record in the *Included studies for data extraction* Endnote library in the Label field. Using the "Preferences>Display Fields" option in the drop-down list from the Edit tab, the library window was set up to show the label field containing the unique ID. A paper copy of each potentially relevant article was also printed and marked up with the unique ID.

For every included paper, where available the following data was extracted directly to an Excel spreadsheet and checked:

- Paper unique ID, First author, Publication year, Title, Journal.
- Study/cohort name, country, study type, study aims, length (year established and times) length of follow up, setting of study, sampling frame, sample size and attrition rates.
- Sample population age/age range at baseline, gender, nationality/ethnicity, SES.
- Dietary assessment method with timings/ages at assessment and exposures reported (Total energy intake, macro and micronutrient intakes, foods and beverages, dietary patterns/habits, diet quality score).
- Adiposity assessment method (measured or self-report) with timings/ages at assessment and outcomes reported (Continuous: Body Mass Index, BMI centile, BMI z score, waist circumference, hip circumference, waist to height ratio, waist to hip ratio, skinfold thickness, body fat percentage or mass, lean body mass. Categorical: underweight, normal weight, overweight, overweight or obese, obese).

If applicable papers were grouped by their common study/cohort. The titles of included papers and the extracted data about reported exposures were surveyed to identify papers that had reported the same or similar dietary exposures.

Papers from several cohorts reported sugar sweetened beverage (SSB) intake and adiposity outcomes. In order to perform meta-analyses we sought additional information about SSB intake at baseline, follow-up and change, SSB serving size and data from energy adjusted and/or the most adjusted models, aiming to extract either odds ratio or mean difference, comparing the most exposed group with the least exposed group, or alternatively the β

coefficient (the degree of change in the adiposity outcome variable for every one unit change in sugar sweetened beverage intake).

During data extraction it became clear that some potentially included papers were not eligible for inclusion. Several papers measured whole diet and subsequent adiposity but did not report their association. An additional exclusion code and group was created for this reason.

As a screening quality measure, checks of whether included papers had reported TEI at baseline and/or had adjusted for TEI in their analysis were made. Those papers which did neither were double-checked to ensure that whole diet had been measured. If whole diet had not been measured the paper was excluded.

2.12 Data synthesis

Included papers were grouped by their common cohorts, by the type of dietary factors that they reported and by whether the dietary exposures under investigation showed positive, negative or no associations with future adiposity. Studies which reported similar dietary exposures were grouped according to how they reported exposure variables and adiposity outcomes. The opportunity to carry out a statistical synthesis was explored.

Based on previous systematic reviews of diet and anthropometric variables in prospective cohort studies (Te Morenga et al., 2013) (Schwingshackl et al., 2015) we anticipated four main methods of reporting:

- β coefficient for the continuous association between baseline intake (the dietary exposure) and adiposity outcome.
- β coefficient for the continuous association between change in intake over time and adiposity outcome.
- Odds ratio for the risk of overweight or obesity comparing participants with the highest intakes with those who had the lowest intakes.
- Mean difference in change in measures of adiposity over time between participants with the highest intakes and those who had the lowest intakes.

Looked for adiposity outcomes (at follow-up or change) were:

- Body Mass Index or BMI z score.
- Waist circumference.
- Skin fold thickness.
- Body fat percentage.
- Classification of overweight/obesity.

Where similar dietary factors were presented in a format unsuitable for meta-analysis, or were reported in only one or two cohorts, results are presented as part of a narrative review. However, sugar sweetened beverage intake and adiposity outcomes were reported in several cohorts. Individual papers in the sugar sweetened beverage group reported odds ratios (for adiposity outcomes that were binary - overweight/obese versus not overweight/obese) or β coefficients (for categorical or continuous exposure variables with adiposity outcomes modelled as continuous variables) or compared change in adiposity measures between different quantiles of intake. Where feasible pooled estimates were generated for the sugar sweetened beverage subgroups using metan commands with random effects in STATA. Heterogeneity between studies was evaluated using the I^2 statistic. Statistical analyses undertaken will be described more fully in the relevant results chapter.

2.13 Method differences from the protocol

2.13.1 Aims

The main aims in the protocol were to identify and critically appraise cohorts with quantitative data on diets during childhood and adolescence that had also taken anthropometric measurements at a follow-up. Ambitiously the systematic review protocol also set out to determine if included studies had measured other exposures such as physical activity, sedentary behaviour and other risk behaviours or if they had assessed other outcome measures of cardiovascular disease risk besides obesity. Due to time constraints these exposures and outcomes were not investigated.

2.13.2 Study characteristics

The protocol defined studies eligible for inclusion as those which had used measures of whole dietary intake among children or adolescents and, at least two years afterwards, anthropometric measurements from which to assess their adiposity or overweight status.

During screening on title and abstract many studies specifically about infant breast-feeding were encountered. This dietary factor was not intended to be within the scope of this review as neither breast milk nor formula milk form part of the usual diet during later childhood or adolescence. Inclusion/exclusion criteria were amended so that breast-feeding studies were excluded.

During screening on full-text and data extraction it became clear that measurements alone (with no reported relationship or association between diet and subsequent adiposity) were insufficient to include a study. In practice this Systematic Review only includes studies which

reported aspects of dietary intake as an exposure *and reported* adiposity as an outcome. Inclusion criteria were re-worded to reflect this change and an additional exclusion code group was created.

2.13.3 Information sources

Studies were found by searching bibliographic databases and examining the reference lists of relevant reviews as set out in the protocol. A formal examination of the reference lists of eligible studies and a citation search of databases to find additional studies were not done.

2.13.4 Data extraction

The protocol stated that data would be extracted to Review Manger 5. Instead data was extracted to Excel and STATA.

Data extraction was not carried out in duplicate, due to a lack of a second reviewer for this stage in the review. Data was extracted once (by CR) and carefully checked.

In order to run meta-analyses in STATA using β coefficients, the standard error of the β coefficient was required. Some papers did not give this value so it could not be extracted directly. Instead the SE variable was derived from either the extracted low and high 95% confidence intervals or from the extracted P value (assuming standard normal distribution) using formulas in Excel.

2.13.5 Quality assessment

The protocol proposed the use of the NICE Quality Appraisal checklist, which can be modified to appraise the quality of both RCT and cohort studies. However, a dual approach was not required as the systematic review sought evidence from cohort studies or from longitudinal follow-ups of the untreated control arm of randomised controlled trials. Instead a customised version of the Newcastle-Ottawa quality assessment tool for Cohort studies was used to assess the quality of all included studies.

Chapter 3 Systematic review: PRISMA results with characteristics of included cohorts.

3.1 Summary

This chapter presents the results of literature searches and the screening process, with numbers of records included or excluded at each stage, as recommended by the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2009). The number of records at each stage of the systematic review process are shown in the PRISMA flow diagram in Figure 3-1.

11,604 records were retrieved by searches. After de-duplication and screening on title, abstract and full text, 35 full text records were included in the review. Screening was done in duplicate and agreement between reviewers was good to moderate based on Cohen's kappa coefficient.

Included papers came from 14 observational cohorts of children and adolescents that had measured diet using 3 day food records, 24 hour recalls or semi-quantitative food frequency questionnaires. Most cohorts took objective measures of children's height and weight as well as other measures such as waist circumference, skinfold thicknesses or body fat percentage. A quality assessment using a customised Newcastle-Ottawa quality assessment scale for cohort studies showed that two-thirds (24) of included papers were of higher quality (8 or 7 stars). The rest were of moderate quality (6, 5 or 4 stars).

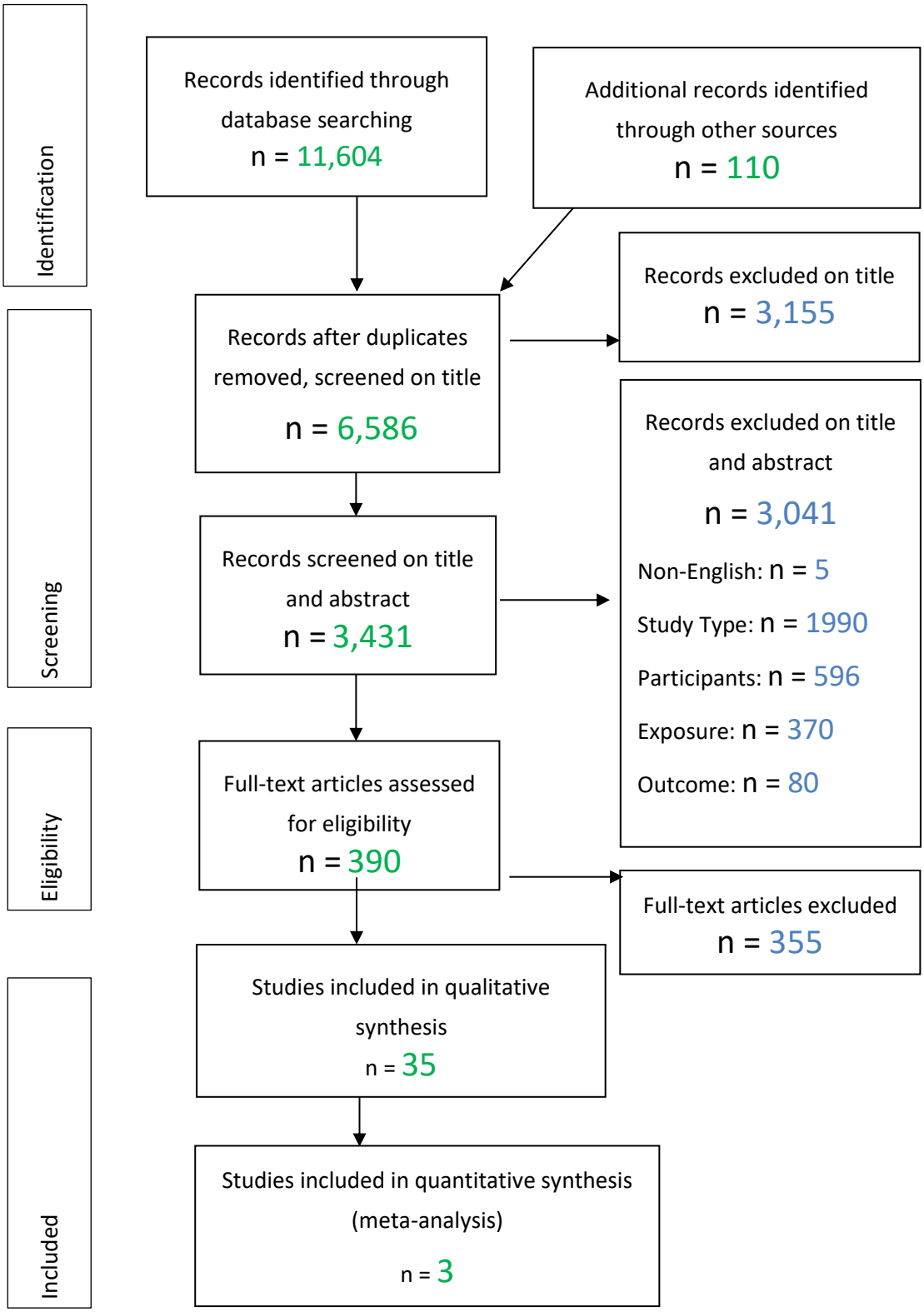
Cohort characteristics are summarised. Included studies were established between the mid-1970s to the mid-2000s and most are from the USA or Northern Europe. Cohort sizes ranged from small (100s) to medium (1,000s) to large (10,000+), but sample sizes in analyses are smaller. Loss to follow up (attrition) ranged from < 2% to almost 50%.

Included papers reported a range of dietary exposures including quantified intakes of specific food and drinks, dietary patterns and eating habits, as well as non-dietary predictors of overweight and obesity. Opportunities for quantitative synthesis were limited.

Dietary assessment tools and adiposity measures are reviewed in Chapter 4. Dietary exposures and associations with adiposity outcomes are narratively synthesised in Chapters 5 and 6. A meta-analysis of sugar sweetened beverages and adiposity outcomes is attempted in Chapter 7

.

Figure 3-1 PRISMA flow diagram



3.2 Literature searches

The chosen search strategy, conducted as nine separate searches in the five bibliographic databases, yielded a total of 11,604 records, which were imported into an Endnote library and used for this systematic review. Search results by database are summarised in Table 3-1 below. Most records were found using the cohort search.

Table 3-1 Numbers of records imported from each bibliographic database

Bibliographic database	Cohort search, published since 1990	Clinical trial with follow-up search, pub. since 2010	Both search strategies	Percentage of total records imported
Medline	3,772	261	4,033	34.8%
Embase	2,526	213	2,739	23.6%
Cochrane CENTRAL	n/a	266	266	2.3%
Scopus	2,557	135	2,692	23.2%
Web of Science	1,851	23	1,874	16.1%
Total imported from databases	10,706	898	11,604	100.0%
Percentage split	92.3%	7.7%	100.0%	

The simple search in the Cochrane Database of Systematic Reviews found 62 Cochrane systematic reviews, of which one, “Interventions for preventing obesity in children” was directly relevant (Waters et al., 2011). The search in the Cochrane Database of Reviews of Effect found 105 other reviews, of which 19 were potentially relevant based on title alone. A search of the NHS evidence database additionally found an independent update of the 2011 Cochrane systematic review (Peirson et al., 2015). The examination of the reference lists of all 21 relevant reviews found 110 potentially relevant cited articles which had not been found by the database searches. These additional articles were also added to the Endnote library. Revised numbers of records are summarised in Table 3-2.

Table 3-2 Total number of records including additional records found via Cochrane systematic reviews and other reviews in DARE

	Cohorts	Trials with follow-up	All
Total imported from databases	10,706	898	11,604
Additional records from reference lists of reviews	10	100	110
Revised total	10,716	998	11,714
Revised percentage split	91.5%	8.5%	100%

3.3 De-duplication

The review began with 11,714 records. After removing (not deleting) 5,128 records identified as duplicates, 6,586 unique records remained for screening.

3.4 First stage screening on publication date and title.

All 6,586 records had been published since 1990, as specified in the search strategy. Almost half (48%) of the records were clearly irrelevant based on their titles alone. Some of these were animal studies that were retrieved despite applying limits to the search strategies to restrict the results to human studies only. 3,431 potentially eligible records went forward to the next screening stage. The numbers of records after the first stage screening are in Table 3-3.

Table 3-3 Results from first stage screening on publication date and title

	Number of records
Records for 1# screening	6,586
Removed, pre 1990	0
Removed, irrelevant title	3,155
Records remaining for 2# screening	3,431

3.5 Second stage screening on title and abstract

Two reviewers independently screened 3,431 records on title and abstract. Reviewers agreed that 2,874 (84%) records should be excluded and that 281 (8%) records were potentially eligible or else unclear based on the title and abstract. Reviewers had a difference of opinion about 276 (8%) records.

The level of agreement between the two independent reviewers, calculated using Cohen's Kappa coefficient, was 0.47 (moderate) after the first pilot of the simple #2 screening questionnaire (20 records). This improved to 0.75 (good) after the second pilot (30 records). The level of agreement for the whole second stage screening on title and abstract was 0.63 (good) based on the actual and calculated numbers shown in Table 3-4.

Table 3-4 Second stage screening, level of agreement between two reviewers

		First reviewer, CR		
		Include/Unclear	Exclude	Total
Second reviewer, MH	Include/Unclear	**281**	**186**	467
	Exclude	**90**	**2,874**	2,964
	Total	371	3,060	3,431
Agreement		281	2,874	155
By chance		50.5	2,643.5	2,694
Cohen's Kappa coefficient		**0.63**	Good Agreement	

Once differences of opinion were resolved by the third reviewer, almost 89% of the records were excluded. The remaining 390 records were either potentially eligible or not clearly ineligible based on the title and abstract alone.

The numbers of records after the second stage screening by each reviewer, the differences between the two reviewers and the numbers of records after differences were resolved by a third reviewer are displayed in Table 3-5. The most common reason for exclusion was wrong study type. Many of these studies were of less than 2 years duration.

Table 3-5 Results from second stage screening on title and abstract, done in duplicate

Decisions	First reviewer, CR	Second reviewer, MH	Difference between reviewers	After differences resolved by third reviewer, CE	Final percentage split
Total screened	3,431	3,431	0	*	100%
Non English abstract	5	2	3	5	<1%
Exclude on study type	1,992	1,160	832	1,990	58%
Exclude on participants	593	1,185	-592	596	18%
Exclude on exposure	391	441	-50	370	11%
Exclude on outcome	79	176	-97	80	2%
Total to Exclude	**3,060**	**2,964**	**96**	**3,041**	**89%**
Total to screen on full text	**371**	**467**	**-96**	**390**	**11%**

3.6 Third stage screening on full text

3.6.1 Finding full texts

In total 390 records went forward for screening on full text. Using the "find full text" function in Endnote X7 retrieved 294 pdfs of abstracts or full texts which were automatically attached to the corresponding record in the Endnote library. After manually searching other sources for full texts for the remaining 96 records, 85 corresponding pdfs were found and attached to the record in Endnote. Eight full texts were not in English and 38 records were meeting or conference abstracts with no full report available. Seven papers were printed as paper copies as no pdf was available. The following four papers could not be obtained:

- Barba 2001 (Barba et al., 2001) e-mail request sent to co-authors (as Barba now deceased) but no reply.
- Burkhard-Jagodzinska 2001 (Burkhard-Jagodzinska et al., 2001) Polish journal, full text unlikely to be in English.
- Kimm 1993 (Kimm et al., 1993) Excluded on abstract alone.
- Siqueira 2006 (Siqueira, 2006) Journal not found.

3.6.2 Identifying common studies

After scanning abstracts and full-text articles available in English, 82 named studies were identified from 265 of the 390 records. Identified named studies are presented in Table 3-6. No specific study name was identified for the remainder of the records.

Once records were grouped by their common study, multiple publications (including meeting or conference abstracts as well as full-text articles) from the same cohort or dataset were evident. Most cohorts yielded less than a handful of papers at this stage of the screening process. However ten studies had between five and nine publications still included and there were five studies with publication numbers in double figures: Amsterdam Growth And Health Longitudinal Study (19), Avon Longitudinal Study of Parents and Children (16), Dortmund Nutritional And Longitudinally Designed study (19), Growing Up Today Study (13) and the National Heart Lung and Blood Institute's Growth and Health Study (19)

Table 3-6 Named studies identified from records to be screened on full text

Study	No. of records	Country
Add Health National Longitudinal Study of Adolescent Health	2	USA
Adelaide Nutrition Study Cohort	3	Australia
AGAHLS Amsterdam Growth and Health Longitudinal Study	19	Netherlands
ALSPAC Avon Longitudinal Study of Parents and Children	16	UK
ASH 30	2	UK
ATLS Arab Teens Lifestyle Study	1	Kuwait
BAEW Be Active Eat Well	2	Australia
BDPP Bienestar Diabetes Prevention Project	1	USA
BHS Bogalusa Heart Study	1	USA
BSCC Bogotá School Children Cohort	1	Colombia
Bratteby et al Swedish Adolescents	2	Sweden
Bt20 Birth to Twenty cohort	1	South Africa
Burke et al Perth	2	Australia
CARDIA Coronary Artery Risk Development in Young Adults	2	USA
CATCH Child and Adolescent Trial for Cardiovascular Health	4	USA
CFPS China Family Panel Studies	1	China
Challenge! Obesity Prevention Programme	1	USA
ChiBS Children's Body Composition and Stress study	5	Belgium
CHNS China Nutrition and Health Surveys	2	China
CLHNS Cebu Longitudinal Health and Nutrition Survey	4	Philippines
Cretan Health and Nutrition Education Programme	4	Greece
DISC Dietary Intervention Study in Children	5	USA
DONALD Dortmund Nutritional and Longitudinally Designed study	19	Germany
ECHO Etiology of Childhood Obesity	1	USA
ECLS-K Early Childhood Longitudinal Survey– Kindergarten cohort	4	USA
EPITeen	2	Portugal
European Youth Heart Study	5	Europe
FAT ain't PHAT intervention	1	Netherlands
Female Adolescent Maturation Study	3	Hawaii, USA
FinnTwin16	1	Finland
Follow up of NVS (German nation-wide nutrition survey)	1	Germany
Framingham Children's Study	3	USA
Gateshead Millennium Study GMS	2	UK
GEMS Girls' health Enrichment Multi-site Studies)	2	USA
Gimme 5	2	USA
GUTS Growing Up Today Study	13	USA
GUTS II Growing Up Today Study II	1	USA
HBSC Health Behaviour of Children Study	1	UK
HEAPS Healthy Eating and Play Study	1	Australia
HeLP Healthy Lifestyles Programme	1	UK
Ho Chi Minh City	3	Vietnam
Hong Kong "Children of 1997"	1	Hong Kong
IDEA Identifying Determinants of Eating and Activity	1	USA

Study	No. of records	Country
IDEFICS Identification and prevention of dietary- and lifestyle-induced health effects in children and infants study	7	Europe
INTCS Institute of Nutrition of Central America and Panama	1	Guatemala
IYM It's Your Move	2	Australia
KOPS Kiel Obesity Prevention Study	1	Germany
Lifeways Cross-generation Cohort Study	1	Ireland
Longitudinal Eating and Activity Study	1	Belgium
LOOK Lifestyle of our Kids project	1	Australia
LSAC Longitudinal Study of Australian Children	5	Australia
Massachusetts Institute of Technology (MIT) Growth & Development Study	6	USA
MUSP Mater University Study of Pregnancy and its outcomes	1	Australia
Mysore Parthenon Birth Cohort	1	Greece
NCDS National Child Development Study 1958	3	UK
NEAT Girls Nutrition and Enjoyable Activity for Teen Girls study	1	Australia
Nepean Study	1	Australia
New Delhi Birth Cohort study	1	India
NFBC 1966 Northern Finland Birth Cohort 1966	1	Finland
NFBC 1986 Northern Finland Birth Cohort 1986	1	Finland
NHLBI / NGHS National Heart Lung and Blood Institute Growth and Health Study	19	USA
Northern Ireland Young Hearts Project	5	N. Ireland
Nurses' Health Study II	2	USA
Oslo Youth Study	2	Norway
PBMAS Pediatric Bone Mineral Accrual Study	2	Canada
Pelotas (Brazil) 1982 birth cohort study	1	Brazil
Pelotas (Brazil) 1993 birth cohort study	2	Brazil
Penn State Young Women's Health Study	3	USA
Project EAT I, II, III, Eating Among Teens/ Eating and Activity in Teens and Young Adults	9	USA
Project Heartbeat	3	USA
Quebec Family study	2	Canada
Quebec Heart Health Demonstration project	1	Canada
READI Resilience for Eating and Activity Despite Inequality study	2	Australia
San Diego Family Health project	1	USA
SNPI School Nutrition Policy Initiative Philadelphia	1	USA
STRIP Special Turku Coronary Risk Factor Intervention Project	4	Finland
Sydney Childhood Eye Study	2	Australia
SWS Southampton Women's Survey	1	UK
Tehran Lipid and Glucose Study	3	Iran
Ventura et al, girls in Pennsylvania	6	USA
Western Australian Pregnancy Cohort (RAINE) study	3	Australia
Young Finns study (Cardiovascular risk in young Finns)	7	Finland

Collectively the 82 named studies represent many regions of the world. There were 28 named studies from North America (U.S.A. and Canada), 28 from Europe (Multi-country studies plus Belgium, Finland, Germany, Greece, Ireland, Netherlands, Norway, Portugal, Sweden and United Kingdom), 13 from Oceania (all Australia), six from Asia (China, Hong Kong, India, Philippines and Vietnam), four from South and Central America (Brazil, Colombia and Guatemala), two from the Middle East (Iran and Kuwait) and one named study from Africa (South Africa).

3.6.3 Level of agreement

After independently screening 390 records on full text reviewers agreed that 272 (70%) of full text records should be excluded and that 40 (10%) records were potentially eligible. Reviewers had a difference of opinion about 78 (20%) records.

The overall level of agreement between the two independent sets of reviewers for the whole third stage screening on full text, calculated using Cohen's Kappa coefficient, was 0.40 (moderate) based on the actual and calculated numbers shown below in Table 3-7.

Table 3-7 Third stage screening, level of agreement between two sets of reviewers

		First reviewer, CR		
		Include/Unclear	Exclude	Total
Second reviewers, BC, CE and JC	Include/Unclear	**40**	**70**	110
	Exclude	**8**	**272**	280
	Total	48	342	390
Agreement		40	272	312
By chance		13.54	245.54	259.08
Cohen's Kappa coefficient		**0.40**	Moderate Agreement	

The level of agreement between reviewers for the third stage screening on full text varied depending upon who was Reviewer 2. Agreement between CR and BC (104 pdf records screened, first authors A to C) was 0.30 (fair). Agreement between CR and CE (279 pdf records

screened, first authors D to Z) was 0.45 (moderate). Agreement between CR and JC (7 printed records screened) was 0.36 (fair).

3.6.4 Results of screening on full text

Once the 78 differences of opinion were resolved by third reviewers (47 by JC, 31 by CE) 61 of 390 full text records were potentially included and each was assigned a unique identification label. During quality assessment and data extraction 26 more records were found to be ineligible, leaving 35 (9%) included and 355 (91%) excluded records. A summary of the number of included and excluded records after overall screening on full text is displayed in Table 3-8. The number of records by reason for exclusion are also given if available.

A full text paper in English was not found for 13.1% of records. Many of these were abstracts from conference proceedings. (One such abstract appeared to be a modified version of an earlier abstract by different authors. The title and wording of both abstracts was the same, but location details, timing, study length and results were slightly different. The first published abstract had a study length of 19 months and was excluded. The later published conference abstract described a study length of over 2 years, so went forward to third stage screening and was excluded then.)

Examination of available full texts revealed that 13.5% of papers used an ineligible study design and 17.4% of papers did not have predominantly healthy study participants in the specified age range at baseline. The dietary assessment method used to measure exposure was the most common reason for exclusion, ruling out one third (33.6%) of full text records screened. A further 10.5% of papers did not have a 2 year follow-up of anthropometry measures from which to assess adiposity outcomes. The last 11 excluded papers (2.8%), described studies of children or adolescents that measured both diet and adiposity, but did not report an association between them. Two of these eleven studies described dietary intake, one paper described risk profiles for metabolic syndrome. Four intervention studies focussed on the outcome of their interventions. Two papers considered the association of diet with bone mass, one studied the association of physical activity and fat mass development, adjusting for energy intake, and one paper looked at children's stress and adiposity outcomes.

Table 3-8 Results from third stage screening on full text, done in duplicate

Decisions	First reviewer, CR	Second reviewers, BC,CE, JC	Difference between reviewers	After third reviewer resolved differences	Final percentage split
Total screened	**390**	**390**	**0**	**.**	**100%**
Abstract only	43	n/a	n/a	43	11.0%
Non-English full text	8	n/a	n/a	8	2.1%
Exclude on study type	29	n/a	n/a	31	7.9%
Study too short	23	n/a	n/a	22	5.6%
Participants not representative	15	n/a	n/a	16	4.1%
Participants too young	44	n/a	n/a	39	10.0%
Participants too old	14	n/a	n/a	13	3.3%
Dietary intake not measured	22	n/a	n/a	27	6.9%
Used diet history	28	n/a	n/a	27	6.9%
Whole diet not quantified	71	n/a	n/a	77	19.7%
Anthropometry not measured	6	n/a	n/a	6	1.5%
Anthropometry at baseline only	19	n/a	n/a	20	5.1%
Anthropometry follow-up too short	20	n/a	n/a	15	3.8%
Association not reported	0	n/a	n/a	11	2.8%
Total to exclude	**342**	**280**	**62**	**355**	**91%**
Total to include	**48**	**110**	**-62**	**35**	**9%**

3.7 Quality Assessment

Two reviewers (CR and AK) independently assessed the internal validity of 48 potentially included papers, using the customised Newcastle-Ottawa quality assessment scale for cohort studies. During the quality assessment process 13 papers were deemed ineligible. These exclusions were checked and confirmed by a third reviewer (CE), leaving 35 included records to be quality assessed.

Both reviewers were in complete agreement about the methodological quality of 17 of the 35 included records (48%), awarding the same stars for each question. For the other 18 records (52%), reviewers did not agree. In two instances, although both reviewers gave the same total number of stars, the stars were awarded for different domains or questions. For 14 records the total differed by one star and for two records the total differed by two stars. Most disagreements arose due to differing assessments in the Comparability domain (did the study control for Age or Total energy Intake?) or the Outcome domain (loss to follow-up). Once these differences were resolved by a third reviewer (CE) all the quality assessments ranged between four and eight stars.

Resolved quality assessment results are summarised in Table 3-9, with results grouped by common studies or cohorts.

Apart from one paper (Rehkopf et al., 2011) which was given no stars in the Comparability domain, all included papers were awarded the following four stars:

- two stars in the Selection domain for representativeness of the exposed cohort and for selection of the non-exposed persons from the same population as the exposed persons.
- one star in the Comparability domain for controlling for one of either age or Total energy intake in their analysis.
- one star in the Outcome domain for a follow-up period 2 years or longer - this was one of the inclusion eligibility criteria.

No paper received nine stars as a star could not be given in the Selection domain for a demonstration that the outcome of interest (body fatness or adiposity) was not present at the start. Some children/ adolescents in every cohort had overweight or obesity at baseline, representing the range of weights that exist within populations.

Table 3-9 Newcastle-Ottawa Quality Assessment stars

Study	Author and Year	Selection	Comparability	Outcome	Star total
ALSPAC	Ambrosini 2012	***	*	***	7
	Bigornia 2014	***	**	***	8
	Noel 2011	***	**	**	7
	Noel 2013	***	**	***	8
Bienestar Diabetes Prevention Programme	Balvin Frantzen 2013	***	**	**	7
Bogalusa Heart Study	O'Neil 2015	***	**	**	7
Bogotá Schoolchildren Study	Shroff 2014	**	**	***	7
DONALD	Alexy 2011	***	**	**	7
	Cheng 2009	***	**	**	7
	Libuda 2008	***	**	**	7
European Youth Heart Study (Danish part)	Zheng 2014	**	**	***	7
	Zheng 2015	**	**	***	7
Female Adolescent Maturation Study	St-Jules 2014	***	**	**	7
Framingham Children's Study	Hasnain 2014	***	**	**	7
Growing Up Today Study	Field 2004	**	**	*	5
	Field 2003	**	**	*	5
	Field 2003	**	**	*	5
	Taveras 2005	**	**	**	6
Growing Up Today Study II	Field 2014	**	*	*	4
IDEA and ECHO	Laska 2012	***	**	***	8
NHLBI's Growth and Health Study	Affenito 2005	***	**	***	8
	Albertson 2007	***	**	***	8
	Albertson 2009	***	**	***	8
	Barton 2005	***	**	**	7
	Berz 2011	***	**	***	8
	Franko 2008	***	**	***	8
	Rehkopf 2011	***		***	6
	Ritchie 2012	***	*	***	7
	Ritchie 2007	***	*	***	7
	Striegel-Moore 2006	***	**	***	8
Project EAT	Berge 2015	**	*	**	5
	Cutler 2012	**	*	*	4
	Fulkerson 2008	**	**	*	5
	Quick 2013	**	**	*	5
RAINE study	Ambrosini 2013	**	*	***	6

Nine papers (26% of the 35 included) were assessed as having the highest achievable quality (eight stars). They originated from three studies, ALSPAC, IDEA and ECHO, and NHLBI's Growth & Health Study. Each of these studies measured diet with a 3 day food record/diary, single or multiple 24 hour recalls and assessed outcomes of body fatness or adiposity from measures taken by trained research staff. The individual highest quality papers also described controlling for both age and Total energy intake in their analyses and reported the loss to follow-up.

Fifteen papers (43%) were awarded seven stars. Most described studies that had used 3 day food records/ diaries or multiple 24 hour recalls but three papers were from two studies (Bogotá Schoolchildren and European Youth Heart Study) that had used food frequency questionnaires or a single 24 hour recall to measure diet. Three papers controlled for only one of age or Total energy intake in their analyses. Nine papers did not clearly report the loss to follow-up.

Three papers (9%) were given six stars. Two of them described studies that had employed food frequency questionnaires. One paper was from the Growing Up Today Study I study that used self-reported height and weight to assess adiposity outcomes – the other two described studies that used assess adiposity outcomes from measures taken by trained research staff. All reported the loss to follow-up.

Six papers (17%) were given five stars and two papers (5%) were given four stars. These eight papers assessed as having the lowest quality (four to five stars) were from Project EAT and the Growing Up Today Study, GUTS I and GUTS II. Both studies employed semi-quantitative food frequency questionnaires to measure diet exposures and were the only included cohorts that relied upon self-reported height and weight to assess adiposity outcomes.

3.8 Data Extraction

Data from still included records was extracted by one reviewer (CR) into Excel. During this process more papers were deemed ineligible. Exclusion decisions were checked and confirmed by another reviewer (CE), leaving 35 included records from which to extract data.

3.8.1 Grouping by common study

The 35 included records originated from 14 common studies. The first author, publication year and title of the included papers, grouped by their common study or cohort, are listed in Table 3-10. Eight studies have only one included paper. The other six studies (ALSPAC, DONALD, EYHS, GUTS, NGHS and Project EAT) account for 27 papers, over three quarters of the total.

Table 3-10 Included papers, grouped by cohort or study

Cohort/study and *country*	First author and publication year	Title and reference
ALSPAC, Avon Longitudinal Study of Parents and Children, *U.K.*	Ambrosini 2012	Identification of a dietary pattern prospectively associated with increased adiposity during childhood and adolescence (Ambrosini et al., 2012)
	Bigornia 2014	Dairy intakes at age 10 years do not adversely affect risk of excess adiposity at 13 years (Bigornia et al., 2014)
	Noel 2011	Milk intakes are not associated with percent body fat in children from ages 10 to 13 years (Noel et al., 2011)
	Noel 2013	Associations between flavoured milk consumption and changes in weight and body composition over time: differences among normal and overweight children (Noel et al., 2013)
BDPP, Bienestar Diabetes Prevention Programme, *U.S.A.*	Balvin Frantzen 2013	Association between frequency of ready-to-eat cereal consumption, nutrient intakes, and body mass index in fourth- to sixth-grade low-income minority children (Balvin Frantzen et al., 2013)
BHS, Bogalusa Heart Study, *U.S.A.*	O'Neil 2015	Candy consumption in childhood is not predictive of weight, adiposity measures or cardiovascular risk factors in young adults (O'Neil et al., 2015)
BSCC, Bogotá School Children Cohort Study, *Colombia*	Shroff 2014	Adherence to a snacking dietary pattern and soda intake are related to the development of adiposity: a prospective study in school-age children (Shroff et al., 2014)
DONALD, Dortmund Nutritional and Anthropometric Longitudinally Designed Study, *Germany*	Alexy 2011	Convenience foods in children's diet and association with dietary quality and body weight status (Alexy et al., 2011)
	Cheng 2009	Relation of dietary glycemic index, glycemic load, and fiber and whole-grain intakes during puberty to the concurrent development of % body fat and body mass index (Cheng et al., 2009)
	Libuda 2008	Pattern of beverage consumption and long-term association with body-weight status in German adolescents (Libuda et al., 2008)
EYHS, European Youth Heart Study (Danish part), *Denmark*	Zheng 2014	Sugar-sweetened beverages consumption in relation to changes in body fatness over 6 and 12 years among 9-year-old children (Zheng et al., 2014)
	Zheng 2015	Substituting sugar-sweetened beverages with water or milk is inversely associated with body fatness development from childhood to adolescence (Zheng et al., 2015)

Cohort/study and *country*	First author and publication year	Title and reference
FAMS, Female Adolescent Maturation Study, *Hawaii U.S.A.*	St-Jules 2014	Estimation of fish intake in Asian and white female adolescents, and association with 2-year changes in body fatness and body fat distribution (St-Jules et al., 2014)
FCS, Framingham Children's Study, U.S.A.	Hasnain 2014	Beverage intake in early childhood and change in body fat from preschool to adolescence (Hasnain et al., 2014)
GUTS, Growing Up Today Study, *U.S.A*.	Field 2004	Snack food intake does not predict weight change among children and adolescents (Field et al., 2004)
	Field 2003	Relation between dieting and weight change among preadolescents and adolescents (Field et al., 2003a)
	Field 2003	Association between fruit and vegetable intake and change in body mass index among a large sample of children and adolescents in the United States (Field et al., 2003b)
	Taveras 2005	Association of consumption of fried food away from home with body mass index and diet quality in older children and adolescents (Taveras et al., 2005)
GUTS II, Growing Up Today Study II, *U.S.A*.	Field 2014	Association of sports drinks with weight gain among adolescents and young adults (Field et al., 2014)
IDEA, Identifying Determinants of Eating and Activity and ECHO, Etiology of Childhood Obesity, *U.S.A.*	Laska 2012	Longitudinal associations between key dietary behaviors and weight gain over time: transitions through the adolescent years (Laska et al., 2012)
NGHS, National Heart, Lung and Blood Institute's Growth and Health Study, *U.S.A*	Affenito 2005	Breakfast consumption by African-American and white adolescent girls correlates positively with calcium and fiber intake and negatively with body mass index (Affenito et al., 2005)
	Albertson 2007	Longitudinal patterns of breakfast eating in black and white adolescent girls (Albertson et al., 2007)
	Albertson 2009	Prospective associations among cereal intake in childhood and adiposity, lipid levels, and physical activity during late adolescence (Albertson et al., 2009)
	Barton 2005	The relationship of breakfast and cereal consumption to nutrient intake and body mass index(Barton et al., 2005)
	Berz 2011	Use of a DASH food group score to predict excess weight gain in adolescent girls (Berz et al., 2011)

Cohort/study and *country*	First author and publication year	Title and reference
	Franko 2008	The relationship between meal frequency and body mass index in black and white adolescent girls: more is less (Franko et al., 2008)
	Rehkopf 2011	The relative importance of predictors of body mass index change, overweight and obesity in adolescent girls (Rehkopf et al., 2011)
	Ritchie 2012	Less frequent eating predicts greater BMI and waist circumference in female adolescents (Ritchie, 2012)
	Ritchie 2007	Dietary patterns in adolescence are related to adiposity in young adulthood in black and white females (Ritchie et al., 2007)
	Striegel-Moore 2006	Correlates of beverage intake in adolescent girls (Striegel-Moore et al., 2006)
Project EAT, Eating and Activity in Teens, *U.S.A.*	Berge 2015	The protective role of family meals for youth obesity: 10-year longitudinal associations (Berge et al., 2015)
	Cutler 2012	Association between major patterns of dietary intake and weight status in adolescents (Cutler et al., 2012)
	Fulkerson 2008	Family meal frequency and weight status among adolescents: cross-sectional and 5-year longitudinal associations (Fulkerson et al., 2008)
	Quick 2013	Personal, behavioral and socio-environmental predictors of overweight incidence in young adults (Quick et al., 2013)
RAINE, Western Australia Pregnancy Cohort Study, *Australia.*	Ambrosini 2013	Prospective associations between sugar-sweetened beverage intakes and cardiometabolic risk factors in adolescents (Ambrosini et al., 2013)

3.8.2 Characteristics of included cohorts

The main characteristics of the 14 included cohort studies are summarised in Table 3-11 and the follow on Table 3-12. More detailed descriptions are provided in Appendix E.

All the included studies had a longitudinal cohort study design. Only one study (Bienestar Diabetes Prevention Programme) was a follow-up of the control arm of a randomised controlled trial. Nine cohorts were from North America (all USA), three were from Europe, one from South America and one from Australia. Included studies commenced from the mid-1970s, with most studies established in the 1990s or in the first decade of the 2000s. At the time of writing several cohorts, including two pregnancy/birth cohorts, are ongoing. Reported study lengths were between two and twenty plus years.

Two cohorts (the Female Adolescent Maturation Study and the NHLBI's Growth and Health Study, or NGHS) were of girls only. The other twelve cohorts included both boys and girls. Although some cohorts had recruited as many as 14,000+ participants, the included papers reported sample sizes from 100 to 6,500+ children, with an age range between 3 and 15 years old at baseline. Loss to follow-up was not reported in every paper, so is not available for all cohorts. The Bogotá Schoolchildren Study reported the lowest attrition rate, 1.5% loss after 2.5 years. The RAINE Study reported the highest attrition rate, 49.5% loss after 3 years.

Five cohorts employed semi-quantitative food frequency questionnaires to assess diet, four had either a single or repeated 24 hour recalls, and five cohorts used 3 day diet records / food diaries.

Three of the USA cohorts assessed adiposity outcomes by calculating Body Mass Index (BMI) based on self-reported height and weight, whereas all the other cohorts used measured height and weight to calculate BMI. Ten cohorts reported at least one other adiposity assessment, such as measured waist circumference, skinfold thickness or body fat percentage based either on bioelectrical impedance or dual energy X-ray absorptiometry (DXA scan).

.

Table 3-11 Characteristics of 14 included cohorts, country, study design and length

Cohort	No. incl. papers	Country	Study design	Study began	Study ended	Length, years
ALSPAC	4	UK	Birth cohort	1990 - 1992	ongoing	3 to 8
BDPP	1	USA	Longitudinal cohort (Control arm of RCT)	2001	2004	3
Bogalusa Heart Study	1	USA	Longitudinal cohort	1973	1996	20+
Bogotá School Children	1	Colombia	Longitudinal cohort	2006	2008	2.5
DONALD Study	3	Germany	Open cohort	1985	ongoing	4 to 5
EYHS (Danish only)	2	Denmark	Longitudinal cohort	1997	2009	12
Female Adolescent Maturation Study	1	USA	Cohort study, girls only	2000-2001	.	2
Framingham Children's Study	1	USA	Longitudinal cohort	1987	.	12
GUTS I	4	USA	Cohort study	1996	1999	3
GUTS II	1	USA	Cohort study	2004	2011	7
IDEA and ECHO	1	USA	2 Longitudinal cohorts, combined	2006/7 and 2007/8	.	2
NGHS	10	USA	Longitudinal biracial cohort, girls only	1987	1997	10
Project EAT,	4	USA	Longitudinal cohort	1998/99	2008/09	10
RAINE Study	1	Australia	Pregnancy cohort - offspring	1989-1991	ongoing	3

Table 3-12 Characteristics of 14 included cohorts, size, attrition rate, age, sex, DAT and adiposity measures

Cohort	Cohort Size	Sample sizes in analyses	Reported attrition rate	First measure, age (years)	Sex	Dietary Assessment Tool	Adiposity measures reported
ALSPAC	14,701	2,245 to 6,772	16%	7	Girls & Boys	3 day food diary/record	Measured Height and Weight, DXA scan
BDPP	706	625	n/a	9	Girls & Boys	3 x 24 hour recall	Measured Height and Weight
Bogalusa Heart Study	4,000 of school age	355	37%	10	Girls & Boys	1 x 24 hour recall	Measured Height and Weight, Waist circumference, Skinfold thickness
Bogotá School children	3,202	975	1.5%	5 to 12	Girls & Boys	Food frequency questionnaire	Measured Height and Weight, Waist circumference, Skinfold thickness
DONALD Study	1,400 by 2010	215 to 585	n/a	2 (measured annually)	Girls & Boys	3 day food diary/record	Measured Height and Weight, Skinfold thickness
EYHS (Danish only)	590	590	34%	9	Girls & Boys	1 x 24 hour recall	Measured Height and Weight, Waist circumference, Skinfold thickness
Female Adolescent Maturation Study	349	200	43%	9 to 14	Girls only	3 day food diary/record	Measured Height and Weight, Waist circumference, Skinfold thickness, DXA scan

Cohort	Cohort Size	Sample sizes in analyses	Reported attrition rate	First measure, age (years)	Sex	Dietary Assessment Tool	Adiposity measures reported
Framingham Children's Study	103	98	n/a	3 to 9	Girls & Boys	3 day food diary/record	Measured Height and Weight, Waist circ., Skinfold thickness, DXA scan
GUTS I	16,882	14,918 to 14,977	32%	9 to 14	Girls & Boys	Food frequency questionnaire	Self-reported Height and Weight
GUTS II	10,919	7,559	n/a	9 to 15	Girls & Boys	Food frequency questionnaire	Self-reported Height and Weight
IDEA and ECHO	723	693	15.6%	9 to 15	Girls & Boys	Up to 3 x 24 hour recalls	Measured Height and Weight, Bioelectrical impedance
NGHS	2,379	2,117 to 2,379	11%	9 to 10	Girls only	3 day food diary/record	Measured Height and Weight, Waist circumference, Skinfold thickness, Bioelectrical impedance
Project EAT,	4,746	2,117 to 2,516	52%	15 (mean age)	Girls & Boys	Food frequency questionnaire	Self-reported Height and Weight
RAINE Study	2,868	1,433	49.5%	14	Girls & Boys	Food frequency questionnaire	Measured Height and Weight, Waist circumference

3.8.3 Reported energy and nutrient intakes

All 14 included studies had measured and quantified whole diet. Hence there was the potential to adjust for energy intake in analyses, although not every paper did so. Energy intake was reported in 21 of the 35 included papers, encompassing all studies apart from the GUTS II cohort. Most of these papers reported energy intake at baseline; papers from the NGHS gave energy intakes averaged from 3 day diet records collected during the 9 years of the study. Enough data was provided to explore mean energy intakes by age of cohort, as set out in Chapter 4.

Selected macronutrient intakes, pertinent to the dietary exposure under investigation, were reported by 13 papers from eight of the studies (ALSPAC, BDPP, DONALD, EYHS, Female Adolescent Maturation Study, Framingham Children's Study, GUTS, and NGHS). Three papers from two studies (BDPP and NGHS) reported selected vitamins, while six papers from three studies (ALSPAC, BDPP and NGHS) reported selected minerals. Most papers gave baseline intakes but some NGHS papers gave averaged intakes. Few papers gave the equivalent information at follow-up.

A summary of which nutrients were reported for each cohort is set out in Table 3-13.

Most papers reported a final model which adjusted for Energy. The majority adjusted for total energy intake (TEI), but papers from the DONALD study adjusted for residual energy intake (the energy intake from everything other than the specific food or drink under investigation).

Seven papers did not report an analytical model which adjusted for energy intake. Arguably, the three papers which investigated dietary patterns (Ambrosini et al., 2012, ALSPAC, Ritchie et al., 2007, NGHS, Cutler et al., 2012, Project EAT), one which ranked the relative importance of predictors of BMI change, including total calories, (Rehkopf et al., 2011, NGHS) and one which focussed on the protective role of family meals (Berge et al., 2015, Project EAT), had no need to adjust for energy intake. Ambrosini et al., 2013 (RAINE study) investigated prospective associations between sugar sweetened beverages and cardio-metabolic risk factors; the authors stated that associations were unchanged after additional adjustment for TEI, so this analysis and underlying information was not shown.

Field et al., 2014 (GUTS II) investigated sports drinks and soda consumption and weight gain. Despite using the same validated semi-quantitative food frequency questionnaire (Youth Adolescent Questionnaire or YAQ) as the earlier GUTS cohort, baseline mean energy was not given and there was no adjustment for energy intake in analyses.

Table 3-13 Papers and cohorts which reported baseline or averaged * Energy, macro and micronutrients

Paper	Study Name	Reports Energy	Adjusts for Energy	Reported Macronutrients	Reported Vitamins	Reported Minerals
Noel 2011	ALSPAC	No	Yes			
Noel 2013	ALSPAC	Yes	Yes	Carbohydrates, Sugars, NMES, Fat, Saturated fat, Protein, Fibre		Ca
Bigornia 2014	ALSPAC	Yes	Yes	Fat, Protein, Fibre		Ca
Ambrosini 2012	ALSPAC	No	No			
Balvin Frantzen 2013	BDPP	Yes	Yes	Carbohydrates, Fat, Saturated fat, Cholesterol, Fibre	D C Bs E	Ca Fe Zn K Na
O'Neil 2015	Bogalusa Heart Study	Yes	Yes			
Shroff 2014	Bogotá School-children cohort	Yes	Yes			
Cheng 2009	DONALD	Yes	Yes	Carbohydrates, Sugars, Fat, Saturated fat, Protein, Fibre		
Libuda 2008	DONALD	Yes	Yes			
Alexy 2011	DONALD	Yes	Yes	Carbohydrates, Fat, Saturated fat, Protein		
Zheng 2015	EYHS (Danish part)	Yes	Yes	Carbohydrates, Fat, Protein, Fibre		
Zheng 2014	EYHS (Danish part)	Yes	Yes	Carbohydrates, Sugars, Fat, Protein		
St-Jules 2014	Female Adolescent Maturation Study	Yes	Yes	Carbohydrates, Fat, Protein		

Paper	Study Name	Reports Energy	Adjusts for Energy	Reported Macronutrients	Reported Vitamins	Reported Minerals
Hasnain 2014	Framingham Children's Study	Yes	Yes	Added sugars, Fat, Protein		
Field 2014	GUTS II	No	No			
Field 2004	GUTS	No	Yes			
Field 2003	GUTS	No	Yes			
Field 2003a	GUTS	Yes	Yes			
Taveras 2005	GUTS	Yes	Yes	Fat, Saturated fat, Trans fat, Fibre		
Laska 2012	IDEA and ECHO	Yes	Yes			
Berz 2011	NGHS	Yes *	Yes			
Albertson 2007	NGHS	Yes *	Yes			
Affenito 2005	NGHS	No	Yes	Fibre		Ca
Barton 2005	NGHS	No	Yes	Fat, Cholesterol, Fibre *	C Folic acid *	Ca Fe Zn *
Albertson 2009	NGHS	No	Yes			
Franko 2008	NGHS	No	Yes			
Rehkopf 2011	NGHS	No	No			
Ritchie 2012	NGHS	No	Yes			

Paper	Study Name	Reports Energy	Adjusts for Energy	Reported Macronutrients	Reported Vitamins	Reported Minerals
Ritchie 2007	NGHS	Yes *	No	Fat, Saturated fat, Cholesterol, Fibre *	A Bs C D Folic Acid *	Ca Fe Mg P Zn Na *
Striegel-Moore 2006	NGHS	No	Yes			
Berge 2015	Project EAT	No	No			
Cutler 2012	Project EAT	No	No			
Fulkerson 2008	Project EAT	Yes	Yes			
Quick 2013	Project EAT	Yes	Yes			
Ambrosini 2013	RAINE study	Yes	No			

3.8.4 Grouping by reported dietary exposure.

A survey of the titles of the included papers indicated that a wide range of dietary exposures and their associations with adiposity had been investigated. From the titles alone it was not always clear exactly which dietary exposures had been investigated, but close inspection of abstracts, full texts and extracted data revealed what was reported.

Included papers had several themes:

- **Intake of specific foods and drinks**. E.g. Whole grains, dairy foods, candy, convenience food, fish, snack foods, fruit, vegetables, juice, milk, soda, sports drinks and sugar-sweetened beverages.
- **Diet quality scores or dietary patterns based on intake.** E.g. Adherence of observed dietary intakes to a healthy eating pattern established á priori, or patterns calculated á posteriori from observed dietary intakes, using reduced rank regression, cluster analysis or principal components analysis.
- **Eating habits or behaviours** such as family meals, breakfast/cereal eating, meal or eating frequency, fast food purchases, snacking and dieting.
- **Multiple predictors** of overweight and obesity, including dietary and non-dietary variables.

These broad themes were used to group included papers for qualitative synthesis, set out in Chapters 5 and 6 as a narrative review. There is overlap between groups; some specific foods and drinks contributed to a dietary pattern or score and certain eating habits and behaviours were also considered as obesity/ overweight risk predictors.

3.8.5 Opportunities for quantitative synthesis

Papers which reported similar or related dietary exposures were initially grouped together as follows:

- Breakfast, cereal and whole grain (11 papers from 5 studies)
- Candy, snack foods, snacking (7 papers from 5 studies)
- Convenience food, fried food away from home, fast food (6 papers from 5 studies)
- Dairy foods including milk (7 papers from 3 studies)
- Energy dense food, starchy food (4 papers from 4 studies)
- Family meals, breakfast and other meal frequency, eating frequency and dieting (11 papers from 4 studies)
- Fish (2 papers from 2 studies)

- Fruit and vegetables including fruit juice (8 papers from 5 studies)
- Sugar sweetened beverages, diet and regular soda, sports drinks (13 papers from 10 studies)

The papers in each group are listed alphabetically by cohort in Table 3-14. Again, there is overlap, with several papers appearing in more than one group, whilst beverages (milk, juice, soda etc.) feature across several groups. Quantitative intakes were needed for meta-analysis, but most groups contained too few papers with quantitative intakes for this to be feasible.

The single most reported dietary exposure was sugar sweetened beverages (SSB), which was investigated in 10 cohorts and reported by 13 included papers. Three papers included SSBs in a dietary pattern, diet score or predictive risk model. The other ten papers reported the association between quantified SSB intake and future obesity risk in various ways. The opportunity for meta-analysis, as set out in Chapter 7, was limited by methodological heterogeneity between the studies.

The next most reported dietary exposure was breakfast eating, cereal and whole grain which was investigated in five studies and reported by 11 papers, seven of them from the NGHS. Three studies investigated whole grain intakes, one of them in the context of a dietary pattern or predictive risk model. Insufficient data was presented for quantitative synthesis.
Three studies, including the NGHS, considered breakfast eating frequency and adiposity, using different adiposity outcome measures. Two studies, again including the NGHS, reported associations between cereal/ready to eat cereal frequency and adiposity outcomes. As frequencies were reported meaningful quantitative synthesis was not feasible.

Fruit and vegetables and/or juice were investigated in eight papers from five different cohorts, using different adiposity outcome measures. Apart from one paper, dairy foods were only considered in the wider context of a dietary pattern or diet score. Milk intake as a beverage and its association with adiposity was reported by three papers from three studies, but again different adiposity outcome measures were used. Associations of other foods or drinks and adiposity outcomes were reported by one or at most two included papers, or were reported as part of a dietary pattern, diet score or predictive risk model. Insufficient data was presented for quantitative synthesis.

Table 3-14 Reported dietary exposures

Reported dietary exposure	Cohort/study	First author and publication year	Paper's main exposure focus
Breakfast eating, cereal and whole grain	BDPP	Balvin Frantzen 2013	Ready to eat cereal frequency
	DONALD	Cheng 2009	Whole grain
	IDEA & ECHO	Laska 2012	*Key dietary behaviours*
	NGHS	Affenito 2005	Breakfast consumption
		Albertson 2007	Breakfast eating
		Albertson 2009	Cereal frequency
		Barton 2005	Breakfast and cereal frequency
		Berz 2011	*DASH food group score*
		Rehkopf 2011	*Obesity predictors*
		Ritchie 2007	*Dietary Patterns*
	Project EAT	Quick 2013	*Overweight predictors*
Candy, snack foods, snacking	Bogalusa Heart Study	O'Neil 2015	Candy
	Bogotá Schoolchildren Study	Shroff 2014	Snacking and soda
	GUTS	Field 2004	Snack foods
		Field 2003	Dieting (and snacking)
	NGHS	Rehkopf 2011	*Obesity predictors*
		Ritchie 2007	*Dietary Patterns*

Reported dietary exposure	Cohort/study	First author and publication year	Paper's main exposure focus
Candy, snack foods, snacking, continued	Project EAT	Cutler 2012	*Dietary Patterns*
Convenience foods, fried food away from home, fast food	DONALD	Alexy 2011	Convenience foods
	GUTS I	Taveras 2005	Fried food away from home
	IDEA & ECHO	Laska 2012	*Key dietary behaviours*
	NGHS	Rehkopf 2011	*Obesity predictors*
		Ritchie 2007	*Dietary Patterns*
	Project EAT	Quick 2013	*Overweight predictors*
Dairy foods, milk	ALSPAC	Bigornia 2014	Dairy intake
		Noel 2011	Milk intake
		Noel 2013	Flavoured milk consumption
	Framingham Children's Study	Hasnain 2014	Beverage intake
	NGHS	Berz 2011	*DASH food group score*
		Ritchie 2007	*Dietary Patterns*
		Striegel-Moore 2006	Beverage intake (milk)
Energy dense food, starchy food	ALSPAC	Ambrosini 2012	*Dietary Pattern*
	Bogotá Schoolchildren Study	Shroff 2014	Snacking (Energy dense) *dietary pattern*

Reported dietary exposure	Cohort/study	First author and publication year	Paper's main exposure focus
Energy dense food, starchy food continued	NGHS	Ritchie 2007	*Dietary Patterns*
	Project EAT	Cutler 2012	*Dietary Patterns*
Family meals, breakfast and other meal frequency, eating frequency, dieting	IDEA & ECHO	Laska 2012	*Key dietary behaviours*
	NGHS	Affenito 2005	Breakfast consumption
		Albertson 2007	Breakfast eating
		Barton 2005	Breakfast and cereal
		Franko 2008	Meal frequency
		Rehkopf 2011	*Obesity predictors*
		Ritchie 2012	Eating frequency
	GUTS	Field 2003	Dieting
	Project EAT	Berge 2015	Family meal frequency
		Fulkerson 2008	Family meal frequency
		Quick 2013	*Overweight predictors*
Fish	Female Adolescent Maturation Study	St. Jules 2014	Fish
Fruit and vegetables, fruit juice	DONALD	Libuda 2008	Beverage consumption – fruit juice
	Framingham Children's Study	Hasnain 2014	Beverage intakes – fruit *and* vegetable juices
	GUTS	Field 2003	Fruit and Vegetables
	NGHS	Berz 2011	*DASH food group score*
		Ritchie 2007	*Dietary Patterns*

Reported dietary exposure	Cohort/study	First author and publication year	Paper's main exposure focus
Fruit and vegetables, fruit juice continued		Striegel-Moore 2006	Beverage intake (juice)
	Project EAT	Cutler 2012	*Dietary Patterns*
		Quick 2013	*Overweight predictors*
Sugar sweetened beverages, diet and regular soda, sports drinks	Bogotá Schoolchildren Study	Shroff 2014	Snacking and soda
	DONALD study	Libuda 2008	Beverage consumption
	EYHS	Zheng 2014	SSBs
		Zheng 2015	SSBs
	Framingham Children's Study	Hasnain 2014	Beverage intake
	GUTS II	Field 2014	Sports drinks
	GUTS I	Taveras 2005	Fried food away from home
	IDEA & ECHO	Laska 2012	*Key dietary behaviours*
	NGHS	Rehkopf 2011	*Obesity predictors*
		Ritchie 2007	*Dietary Patterns*
		Striegel-Moore 2006	Beverage intake
	Project EAT	Quick 2013	*Overweight predictors*
	RAINE	Ambrosini 2013	SSBs

3.9 Discussion

It was soon apparent that the diet and health of children and adolescents is an enormous research field, about which many research papers have been written. Initial exploratory literature searches yielded more than 19,000 "hits" from the five bibliographic databases searched, with many duplicates. It was a challenge to develop a search strategy with the right balance of sensitivity and precision, that would:

- find all potentially relevant reports (sensitivity) so that they could be included, thereby reducing bias in the systematic review and reducing random error in any meta-analysis.
- minimise the number of irrelevant reports retrieved (precision) so that the burden on researchers' time was reduced.

In Medline ready-made search filters for study types and to exclude non-human studies were tried, but they made a marginal difference. Instead results retrieved by individual search terms were examined to find the search terms that were most useful and to eliminate search terms that brought back irrelevant studies. The revised strategy yielded 11,604 "hits" of which 45% were duplicates. The Ovid Medline and Embase bibliographic databases were chosen because both allowed the export of complete references; searches yielded similar results in both, contributing to the high proportion of duplicates. Using two search strategies, one for cohort studies and another for clinical trials with follow-up, also added more work for little gain. Only one included study was the control arm of a randomised controlled trial, followed up as an observational cohort study. All other included papers were from observational cohorts.

Screening the 6,536 de-duplicated records took a long time and almost half (48%) of the de-duplicated records were irrelevant on title alone, suggesting that a more precise strategy would have been helpful. Ultimately 29 included papers were originally retrieved by Ovid Medline, the other 6 included papers came from Scopus or Web of Science. In hindsight, searching for cohort studies only in these three databases may have sufficed, with little impact on sensitivity.

Meaningful titles and well written abstracts were useful for the second stage of the screening process. However, many abstracts had incomplete descriptions of participants' ages or state of health, failed to state the dietary assessment method used or did not indicate the timings of dietary and anthropometric measures so decisions could not be made on abstracts alone. Guidelines such as the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) checklist (Vandenbroucke et al., 2014) try to counter such poor quality

reporting, advocating ways to "improve the completeness and transparency" of reports. Guidelines for epidemiological studies of nutrition (STROBE-nut) have subsequently been published (Lachat et al., 2016) and guideline number one is that nutritional studies should "state the dietary/nutritional assessment method(s) used in the title, abstract, or keywords". After screening on title and abstract 390 records remained to screen on full text. Finding full texts was easier than anticipated as many papers were available in pdf format, which also eliminated the need to print hundreds of documents.

The level of agreement between reviewers at second stage screening on title and abstract was "good" based on Cohen's kappa. Running a pilot at the start and having only two reviewers throughout undoubtedly helped. Screening on full text was not piloted as fully, even though three different individuals acted as the second reviewer. This may partly explain why third stage screening proved more challenging and why agreement on inclusion/exclusion was only "moderate" to "fair".

Using Endnote as a reference management system, with a series of libraries for each screening stage, worked well. However, with such a large review, the libraries and their back-up copies used a lot of computer storage space. For example, the initial Endnote library for studies retrieved by the cohort search strategy with duplicates was over 56 GB. A research data management plan specifying the exact series of libraries at the outset with the use of zip files as back-up copies would have helped to minimise storage requirements.

Initial thoughts on quality assessment were that, as this systematic review was expected to include both observational cohort studies and clinical trials, a dual approach might be appropriate, using either two specific quality assessment tools or a flexible tool that could be adapted to suit both study types. (A dual approach was taken in another systematic review (Gorber et al., 2007) using the Downs & Black tool to assess the quality of non-randomised study designs and the Jadad's scale (Jadad et al., 1996) to assess the quality of randomised controlled trials. The Cochrane handbook (Higgins and Wells, 2011) discourages use of Jadad's scale as its rating of study attrition is subjective, it does not account for allocation concealment and it overemphasises reporting quality.) In the end only observational studies were included, so only one type of quality assessment tool was required.

Although the Newcastle-Ottawa tool allowed an assessment of the quality of the included cohort studies, it is important to recognise that this was a measure of methodological quality *in relation to the other assessed studies,* not an absolute measure. As subjective dietary assessment methods used by the included studies are prone to measurement error (bias) (Lachat et al., 2016) none of the studies provide the highest quality evidence about dietary intake. This issue will be further explored and discussed in Chapter 4.

The Newcastle-Ottawa quality assessment tool for Cohort studies was straightforward to customise, but one of the stars in the Selection domain (demonstration that the outcome of interest i.e. adiposity, was not present at the start) could not be awarded to any included study, so was redundant. In the Comparability domain, the number of control factors was limited to two (Age and Total Energy Intake). Most included papers received at least one star in this domain. Although age and TEI were chosen as the most important control factors, sex is also important when considering children's growth and development. The ability of the NOS to differentiate between studies methodological quality may have been improved by allocating a third star in the Comparability domain, for adjusting analyses by sex, or for separate analyses of girls and boys.

With nine questions to consider, the NOS was relatively quick for reviewers to use. Much of the sought after information was in the method section of each paper, but some details were not reported or were very hard to spot. As a result, reviewers awarded different numbers of stars across the three domains. Most disagreements were in the Comparability domain, or for loss to follow-up. Elsewhere, reviewers' assessments of study quality of cohorts using NOS have been compared with the study authors' assessments using the same tool (Lo et al., 2014). Reviewers gave significantly higher NOS scores than authors. Inter-rater reliability (compared using kappa statistics) by each item in the NOS ranged from "slight" to "poor". Inter-rater reliability for the overall NOS score was also "poor". Lo et al. argued that inherent subjectivity in the NOS negatively affects inter-rater reliability. They concluded that when applying the NOS in systematic reviews, reviewers should contact the authors for missing information, rather than making assumptions which introduce bias in the quality assessment. This finding, and our own experience with NOS, also points to the need for better reporting of cohorts studies, as put forward by STROBE (Vandenbroucke et al., 2014).

Data extraction to Excel was time-consuming. Characterising cohorts was easier if the study had a web page, as was the case with ALSPAC, DONALD (in German and English), GUTs and Project EAT. Published cohort profiles and study protocols, such as those for ALSPAC (Boyd et al., 2013; Fraser et al., 2013), the Bogalusa Heart Study (Berenson and Bogalusa Heart Study, 2001), DONALD (Kroke et al., 2004), the IDEA study (Lytle, 2009) and the EYHS (Riddoch et al., 2005) were also useful. For other cohorts, characteristics had to be pieced together from included papers. Article length restrictions imposed by journals may have been an impediment to completeness of reporting; some authors cited earlier papers from the same study for more details. A few papers gave information about supplementary material, available online.

Due to the breadth of dietary exposures and adiposity outcomes included in the systematic

review, measurement methods, units of measurement and analytical approaches showed much heterogeneity. Extracting comparable information about energy and nutrient intakes, specific dietary exposures and adiposity outcomes was a challenge. Surprisingly, some key items of information, such as portion sizes and loss to follow-up, were missing from some papers.

As well as assessing relative methodological quality, it would have been useful to formally assess the quality of reporting of included papers against the STROBE and STROBE-nut reporting guidelines. This approach has been taken elsewhere. For example Ziauddeen et al assessed the reporting of predictive models included in their systematic review (Ziauddeen et al., 2018) against the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement (Collins et al., 2015).

Chapter 4 Systematic Review: Dietary assessment and adiposity measures.

4.1 Summary

In this chapter (Chapter 4) the dietary assessment methods and measures of adiposity used by included studies are narratively reviewed. All cohorts used subjective measures of diet, either food diaries, 24 hour recalls or semi-quantitative food frequency questionnaires. A range of adiposity measures were employed, including Body Mass index (BMI) based on height and weight, BMI percentiles, BMI z scores or overweight/obesity derived from growth references, waist circumference, or body fat percentage based on skin fold thicknesses, bio-electrical impedance or dual energy x-ray absorptiometry. Most papers reported steps taken to minimise measurement error and biases in the cohort study. Some studies also carried out validation studies to gauge the accuracy of their chosen measurement methods.

Baseline mean energy intakes reported by each cohort were compared by age, dietary assessment tool and geography, and against UK age specific estimated average requirements and were found to be plausible. Under and over reporting was apparent. Macronutrient contributions to energy compared with World Health Organization nutrient intake goals showed that many children and adolescents had diets that were high in fat and low in carbohydrates and dietary fibre.

Reported mean BMI values increased with age and were plausible when compared with the UK 1990 growth reference. The prevalence of overweight/obesity increased with the age of cohort participants, although the use of different growth references impeded direct comparisons between cohorts.

4.2 Introduction

This systematic review seeks evidence about childhood and adolescent diet and future obesity risk. The first criterion was that included cohort studies must have measured and quantified the whole diet, so that estimated energy intake could be adjusted for in longitudinal analyses of specific dietary exposures and adiposity outcomes. The rationale is that energy intake is on the casual pathway between dietary intake and weight status, as total energy intake is a factor in energy balance and surplus energy promotes adiposity.

The second criterion was that studies must have measures from which an assessment of adiposity at follow-up could be made, whilst recognising that it can be hard to classify overweight or obesity status with certainty in growing children, as will be explained later in this chapter.

When attempting to establish diet and nutrition-related risks for health outcomes such as overweight or obesity, it is important that diet and the outcome are measured as accurately as possible. However, measured values contain random errors and systematic errors (bias) so are estimates of the true values. Well-reported measurement methods and an understanding of their limitations are vital for the correct interpretation of results (Lachat et al., 2016). The methods used by each study to measure diet and adiposity and steps taken to improve accuracy, or assess the extent of inaccuracies, are examined, before investigating the plausibility of reported mean energy and macronutrient intakes and selected adiposity measures.

Reported mean energy intakes by age between cohorts and against UK age-specific estimated average requirements (SACN, 2011) are compared. As a gauge of diet quality within cohorts, mean macronutrient contributions to energy intake are compared with World Health Organization population nutrient intake goals (WHO/FAO, 2003) and mean dietary fibre intakes are compared with current UK recommendations. Reported mean Body Mass Index (BMI) values are compared by age across the included cohorts and against the UK 1990 age and sex specific growth reference.

4.3 Dietary assessment

Diet is a complex exposure that is challenging to measure, as diet differs between and within individuals, and varies daily according to the combinations and quantities of foods and drinks that are consumed. Depending upon the research question, diet can be investigated at several levels (Total energy intake, macro and micronutrients, individual foods or food groups, dietary patterns, and/or eating habits/behaviours), as described by the Medical Research Council's Diet, Anthropometry and Physical Activity (DAPA) measurement toolkit (MRC, 2014).

Dietary assessment has two main objectives:

i. To extract information about usual diet (the recent or long-term average of either an individual or a population) as accurately as possible.
ii. To convert reported intakes into estimated nutrient data.

4.3.1 Dietary assessment methods

Diet can be measured objectively or subjectively. Objective dietary assessment methods aim to capture an individual's intake of foods and drinks as it occurs, using technology or direct observation by trained and independent observers. Subjective dietary assessment methods rely on verbal or written reports given by an individual, or sometimes their proxy (MRC, 2014).

Subjective dietary assessment methods include prospective methods such as food diaries (FD) or food records, with details recorded by an individual as and when they consume food or drink, and retrospective methods such as diet histories, 24 hour dietary recall (24-HDR) and food frequency questionnaires (FFQ), with food and drink intake reported afterwards. Traditionally such dietary assessment tools (DAT) have been paper based. Manually coding reported intakes for conversion into nutrient data, based on a food composition database suitable for the population being studied, is a costly and lengthy process. To reduce error (coding bias) it is usually done by qualified nutritionists or trained researchers.

It is important to choose a DAT that suits the study aims and resources, but all established methods have inherent limitations. For example, objective methods are expensive and may be intrusive – their use may be inappropriate for large scale studies. Prospective methods place a high burden on respondents who may alter their intake, while retrospective methods reliant on memory contain recall bias. Respondent bias in studies using subjective methods of dietary assessment includes both under-reporting and over-reporting. (MRC, 2014)

4.4 Dietary assessment tools used by included cohorts

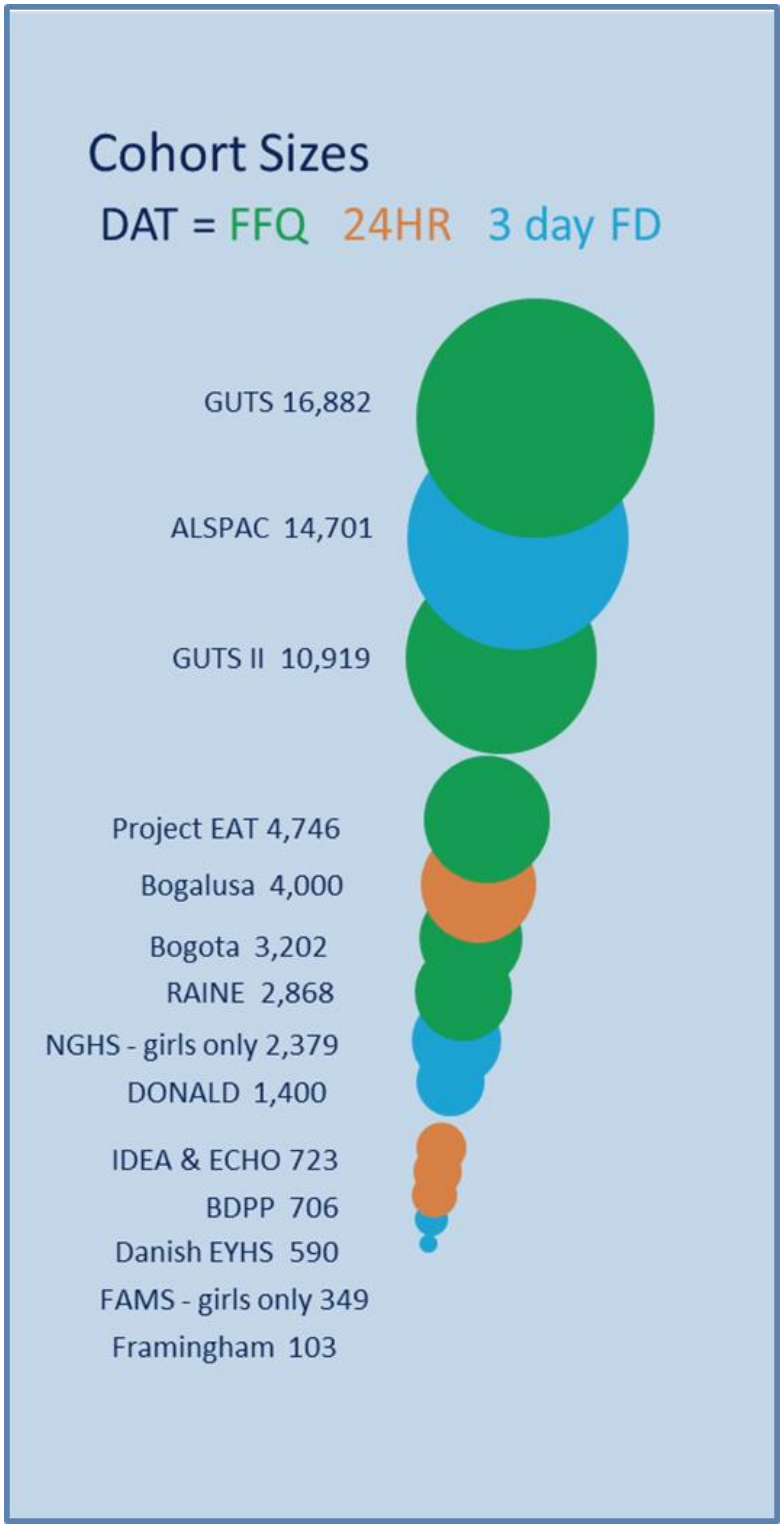
The dietary assessment tools employed by each cohort were previously summarised in Chapter 3, Table 12. All 14 included cohorts used a subjective dietary assessment method (3 day FD, 24-HDR or FFQ) to measure the usual diet. Food frequency questionnaires are often favoured for dietary assessment in large cohorts as they are less costly to administer, but some of the larger cohorts included in this review used a 3 day food diary. See Figure 4-1.

Studies that used diet histories were excluded. Respondents often find it hard to estimate their usual portion sizes, so a diet history is rarely a reliable measure of energy intake. Detailed diet histories are time-consuming and best suited to measuring the dietary habits of an individual. They are not an efficient way to collect diet data about populations and diet histories are seldom used in large scale studies today.

Only one included cohort, NGHS, used observation as an objective measure of diet at the start of their study. Covert observers recorded the types and amounts of food eaten by 58 girls (9 to 10 years old) at school lunch, to compare with the girls' own reports from a 3 day food record, a 24 hour dietary recall and five food frequency questionnaires (Crawford et al., 1994).

The 3 day FD had the least missing and the fewest phantom foods (25% & 10% respectively) compared with the 24-HDR (30% & 33%) and 5 x FFQ (46% & 40%). Based on agreement between observed and reported intakes, a 3 day food record was chosen as the best method of dietary assessment for the NGHS cohort.

Figure 4-1 Dietary assessment tools and size of cohort



4.4.1 Food diaries

Five included cohorts (ALSPAC, DONALD, FAMS, Framingham Children's study and NGHS) repeatedly used 3 day food diaries, as summarised in Table 4-1.

Table 4-1 Comparison of 3 day food diaries by cohort

Cohort & Country	3 day FD	Quality measures and validation	Food composition database, software for analysis
ALSPAC UK	3 day FD 2 weekdays 1 w/e day (non-consecutive) At age 7, 10 & 13 yrs.	Parent completed FD if child aged 7 yrs. Parental input thereafter. Household measures FD checked at clinic visit. Misreporting via EI: EER	McCance & Widdowson 5th edition DIDO (Diet in, data out) BRIGADE nutrient analysis programme
DONALD Germany	3 day FD Weighed 75% weekdays 25% w/e days (consecutive) Annually, 0 to 18 yrs.	Parents helped younger subjects to complete FD Food scales for weighing	In-house LEBTAB (German food tables, supplemented by food tables from UK, USA and the Netherlands)
Female Adolescent Maturation Study Hawaii, USA	3 day FD Thurs, Fri & Sat. At 9 to 14 yrs. & at 2 year follow-up	Girls assisted by parent/guardian Measuring cup, spoon and ruler given	Shared Nutrition Food Composition database version 1999 (USDA database plus foods eaten in Hawaii)
Framingham Children's Study USA	3 day FD Up to 4 annually from 3 to 5 yrs. for 12 years	Parent completed FD if child under 10 yrs. Child over 10 yrs. assisted by parent. Instruction on estimating portion sizes.	Nutrition Data System (NDS), University of Minnesota USDA Continuing Survey of Food Intakes by Individuals (CSFII)
NGHS USA	3 day FD 2 weekdays 1 w/e day (consecutive) Annual visits 1 to 5, 7, 8 & 10 (Ages 9 to 19 yrs.)	Girls given instruction. Dietitians reviewed FDs with each girl. FD estimates compared with observed intakes at school lunch. Top and bottom 1% reviewed.	Nutrition Data System (NDS), University of Minnesota. Dietary Data Entry Center

The food diary (FD) is a prospective dietary assessment method. An individual (or proxy, such as a child's parent or carer) is asked to record food and drink at the time of consumption, over 3 or more days, often including a weekend day and a weekday to capture variation between those days. Portion sizes may be quite accurate if foods are recorded as they are eaten, but if recording is delayed food diaries may contain error due to recall bias and mistakes in assessing portion sizes. Individuals may amend habitual intake or be unwilling to report certain items (social desirability bias) (Patterson and Pietinen, 2004). E.g. A person (or their proxy) who is aware of nutrition guidelines might exaggerate vegetable intake or downplay the true intake of fried food.

Cohorts enlisted parental help to complete the child's food diary, especially for children under 10 years, which probably improved reporting accuracy but risked the introduction of social desirability bias from the child's parent or carer. The exception was the NGHS, which prioritised confidentiality over additional information from parents, even at the ages of 9 and 10 years (Affenito et al., 2005).

Most cohorts asked for 3 consecutive days to be recorded and only the Framingham Children's study did not state that weekend and weekdays were surveyed (Hasnain et al., 2014). With each additional day of a food diary there is a tendency to record fewer food and drink items, a form of reporting bias. As a result, food intake can be under-reported in food diaries (measurement error), especially those that last longer than 4 days (Patterson and Pietinen, 2004). Unusually, ALSPAC requested non-consecutive days (Ambrosini et al., 2012), perhaps to counteract study fatigue and under-reporting.

Food diaries are onerous, more so if respondents are asked to weigh foods and log the time of eating. Only DONALD study children kept a weighed diary using electronic food scales (±1g) with parents help. They were allowed semi-quantitative recording with household measures if weighing was not possible (Alexy et al., 2011).

Other cohorts gave instruction on estimating portion sizes from household measures and FAMS provided girls and their families with measuring tools (St-Jules et al., 2014), all of which will have improved the quantitative accuracy of reported intakes. Girls from NGHS were asked to log the time of eating/drinking. Despite this extra task, NGHS retention rates were high. Over 90% of girls attended annual visits 1 to 4 and over 80% attended later visits (Affenito et al., 2005).

Food diaries require high levels of literacy and motivation from the individual or their proxy, with a tendency for more dropouts if more days are needed. These factors can lead to non-response, giving a less representative sample (selection bias). In ALSPAC a nutritionist checked

the food diary with the child and parent at the clinic visit and in the NGHS dietetics professionals reviewed the food record with each girl, to ensure completeness of each diary.

The ALSPAC and NGHS cohorts also examined accuracy and misreporting. The 3 day food record used in the NGHS was validated by an observation study at baseline (Crawford et al., 1994) finding that 25% of actual foods consumed was unrecorded and 10% of reported foods had not been consumed. 10% of all returned food records were reviewed and outlying nutritional values and food records in the top and bottom 1% were checked (Albertson et al., 2009).

In ALSPAC dietary misreporting was based on the ratio of energy intake to estimated energy requirement (EI: EER) (Noel et al., 2010). Physical activity, measured with an accelerometer, and body composition were used to calculate EER which was compared with diet data at the same age. At 13 years old 35% were under-reporters, 44% were plausible reporters and 21% were over-reporters (Noel et al., 2013). At 10 years old 42% had plausible dietary intakes (Noel et al., 2011).

Translating the reported intakes from a food diary into nutrient data requires an extensive nutritional database, suitable for the population being studied. All five cohorts used a nutritional database matched to the country/population, with the University of Minnesota's Nutrition Data System used by two USA cohorts. FAMS supplemented their chosen USA database with typically Hawaiian foods. In the NGHS all nutrition coding was done professionally by the Dietary Data Entry Center at Cincinnati Children's Hospital and Medical Center.

4.4.2 Twenty four hour dietary recalls

Four included cohorts (BDPP, Bogalusa Heart Study, EYHS and IDEA & ECHO) used 24 hour dietary recalls at baseline and follow-up, as summarised in Table 4-2

A 24 hour dietary recall (24-HDR) is a subjective short-term dietary assessment method, traditionally conducted as a structured interview, asking about foods and drinks consumed the previous day. Unlike a food diary, a retrospective interview does not influence the respondents eating behaviour. The respondent burden is lower, and an interviewer administered 24-HDR does not need high levels of literacy, which can boost response rates (Patterson and Pietinen, 2004). However, a 24-HDR relies on the individual's short-term memory, ability and willingness to report their diet, so may contain respondent biases (recall bias, reporting bias, social desirability bias). Consequently, 24-HDR tends to under report individual intakes. To counter this, interviewers often use a "multiple-pass" to probe for details about portion size, how foods were prepared or time of eating, or to prompt respondents about easily forgotten items (MacIntyre, 2009).

Table 4-2 Comparison of 24 hour dietary recalls by cohort

Cohort	24-HDR	Quality measures and validation	Food composition database & software for analysis
BDPP USA	24-HDR x 3 Sun, Mon & Tues. Face-to-face interviews on 3 consecutive days, in school At 9 yrs. & at 2 & 3 year follow-up	20 trained & certified (NDS-R) interviewers Multiple pass, scripted	Nutrition Data System for Research (NDS-R) version 2006
Bogalusa Heart Study USA	24-HDR x 1 Face-to-face interview, in school At age 10 yrs. in 1973/74, 1978/77 or 1978/79 At 1st follow-up in 1981/91	Duplicate recalls in 10% sub-sample to assess interviewer variability	The Moore Extended Nutrients database (MENu) formerly the Extended Table of Nutrient values USDA data
EYHS Denmark	24-HDR x 1 Face-to-face interview, in school conducted Mon to Fri. At 9, 15 & 21 yrs.	9 year olds kept a FD the day before, with help from parent. Utensils and food pictures Under reporting EI: BMR	Danish Food Composition Tables 2006 Dankost 3000 software
IDEA & ECHO USA	24-HDR x 3 (sometimes 2) Telephone administered 1 w/e day 2 weekdays At 14+ & 16+ yrs.	Trained & certified interviewers (NDS-R)	Direct data entry linked to Nutrition Data System for Research (NDS-R)

In cohorts with younger children (10 years old or less) baseline 24-HDR were carried out face-to-face, in school settings. 24-HDR were only conducted by telephone in the IDEA & ECHO cohorts, where children were already teenagers at baseline (Laska et al., 2012). In the IDEA & ECHO cohorts up to three interviews were done by trained staff from the University of Minnesota Nutrition Coordination Center, which will have reduced observer bias but added to the research costs. Interviewers used direct data entry linked to a nutrient database, which is time efficient and cost effective, but may rule out later quality checks.

A single 24-HDR for many representative individuals can be enough to assess dietary exposure in a population but does not capture the day-to-day variation in the diet of an individual (MacIntyre, 2009). Repeated 24-HDRs are needed to assess individual intake and capture foods which are only eaten episodically (Patterson and Pietinen, 2004). The Bogalusa Heart Study and the Danish European Youth Heart study used a single 24-HDR at baseline and follow-up, whereas the BDPP children were surveyed each time with 3 recalls on consecutive days.

Dietary recalls in the BDPP were conducted by trained interviewers who used a "multiple-pass" technique, with a standardised script for dialogue, prompting and recording, thereby reducing recall bias and observer bias. Children in BDPP were asked on Monday about what they ate on Sunday, on Tuesday about Monday and on Wednesday about Tuesday, so that the child only had to recall one day at a time (Balvin Frantzen et al., 2013). This captured variation in children's diets between weekdays and weekend days, but as the latter recalls were not "unannounced" some children may have altered their habitual intake, or experienced study fatigue – disadvantages usually associated with 3 day Food diaries.

An adapted 24-HDR method was validated for use in children from the Bogalusa Heart Study. Quality control protocols improved data reliability (Frank et al., 1984) and reduced variability between interviewers (observer bias). Another 24-HDR was used for the first follow-up, but for the second follow-up the youth adolescent questionnaire (YAQ) was applied. The YAQ had been used with young adults elsewhere and gave similar results to the 24-HDR (Nicklas, 1995).

As an aide memoire to reduce recall bias, the EYHS asked children (with parental help at age 9 years) to keep a qualitative food record for the day that was to be recalled, which may have introduced other forms of bias. During the recall interview, different sized cups, dishes and spoons were used alongside pictures of food, to improve estimations of portion size. The EYHS also identified misreporting, by checking the ratio of Energy intake: Basal metabolic rate (EI: BMR). The Goldberg cut-off value of 0.9 was applied (Goldberg et al., 1991). 26 children (out of 283) were identified as under reporters and excluded from analyses.

All four cohorts used a nutritional database matched to the population being studied. Again, the Nutrition Data System was used by two USA cohorts. The computerised Moore Extended Nutrients database used by the Bogalusa Heart Study is updated periodically, to show nutrient changes in food products. As the number of foods that can be included in a database is limited (although it may be a large number) and new food products are continually introduced, no food database can remain comprehensive for long (MRC, 2014).

4.4.3 Food frequency questionnaires

Food frequency questionnaires (FFQ) are mainly used to assess long-term average intakes in individuals and populations (Patterson and Pietinen, 2004). Six included cohorts (Bogalusa Heart Study, Bogotá School Children, GUTS, GUTS II, Project EAT and the RAINE study) used FFQs, as summarised in Table 4-3. The Bogotá School Children cohort used an FFQ only at baseline, while four cohorts used FFQs at baseline and follow-up. The Bogalusa Heart Study switched from using a 24-HDR in childhood to a FFQ at follow-up in adulthood, although only childhood candy consumption at baseline was investigated as an exposure (O'Neil et al., 2015).

Table 4-3 Comparison of food frequency questionnaires by cohort

Cohort	FFQ	Quality measures and validation	Food composition and nutrition data
Bogalusa Heart Study USA	131 item YAQ (past year) *At 2nd follow up only in 1995/96 as young adults*	YAQ validated by comparison with 3 non-consecutive 24-HDR	Channing Laboratory, Boston, Massachusetts
Bogotá School children Colombia	38 item FFQ (usual intake in past month) Mothers asked when child 5 to 12 yrs. old	Based on FFQ validated in Costa Rica. Trained research dietitians. Excluded if > 3 missing	USDA Standard reference food composition database & the Composition table of Colombian Foods
GUTS USA	131/132 item YAQ (past year) Mailed At 9 to 14 yrs. & at annual follow up, over 3 years.	YAQ validated by comparison with 3 non-consecutive 24-HDR, corr = 0.54	Harvard's FFQ database NFCS Foods commonly eaten by individuals. USDA Handbook No.8 McCance & Widdowson 4th & 5th edition
GUTS II USA	131 item YAQ (past year) Mailed At 9 to 16 yrs. in 2004 and follow ups in 2006, 2008 & 2011	As above	As above
Project EAT USA	149 or 152 item YAQ (past year) Classroom survey at 12 or 15 yrs. Project EAT I Follow up by mail after 5 years. Project EAT II	As above	As above
	Semi-quantitative Willett's FFQ On-line or by mail Follow-up after 10 years. Project EAT III	Validated Willett's FFQ Estimates compared with YAQ in a sub-sample	n/a
RAINE Australia	212 item FFQ (past year) At 14 yrs. & 17 yrs.	14 year olds had help from parent. FFQ validated by comparison with 3 day FD in cohort Plausible reporting EI: EER	Australian food composition database NUTTAB95. Data input verified by Commonwealth Scientific and Research Organisation

The FFQ is applied retrospectively, so does not influence eating habits. Like 24-HDR it relies on memory, ability and willingness to report their diet, so can contain respondent bias.
FFQs are finite lists of foods with a response option to indicate how often each food is/was usually consumed in a given time frame (day, week, month etc.). It has been recommended that recall periods should not exceed 1 month or that the FFQ should be repeated in different seasons throughout the year (MacIntyre, 2009). Mothers of children in the Bogotá School children cohort were asked about their child's diet only in the past month. Other included studies asked about intakes in the past year, at baseline and follow-up, but with no apparent consideration of seasonal variations in diet.

FFQs about the whole diet usually contain between 80 to 120 generic food items that are typically consumed in the population being studied (Patterson and Pietinen, 2004). FFQs about specific nutrients are shorter, and may be qualitative, asking only about frequency. Quantitative FFQs ask the respondent to indicate the quantity they usually consume, whereas semi-quantitative FFQs provide standard portion sizes. Quantitative data can be used to estimate average nutrient intakes in the sample population. Most FFQs employed by included cohorts were semi-quantitative, but in the RAINE study (Ambrosini et al., 2013) adolescents were asked to record their usual serving size if it differed from the example serving size in the FFQ.

For a modest cost, FFQs can be sent to many participants, so they are often chosen for large scale studies. However, response rates can be low, which adds selection or attrition bias. In GUTS barely two-thirds of children responded to the mailed-out survey at baseline (Field et al., 2003a). In Project EAT, among those who could be contacted, the mail response rate at follow-up was 66%, representing less than half the original cohort (Quick et al., 2013).

Although the researcher burden is relatively low, completing an FFQ is cognitively challenging. It may be easy for respondents to decide if they ever eat any of the listed foods, but it is harder to recollect how often or to estimate the quantity usually consumed (Patterson and Pietinen, 2004). Hence food intakes can be misreported, leading to significant measurement error. The Bogotá School children study asked mothers to complete the FFQ for their children and the RAINE study asked younger teenagers to complete the FFQ with help from a parent. GUTS surveys were sent to children whose mothers had agreed to their enrolment – possibly they helped their children too. Parental assistance may have improved the accuracy of responses. Plausible reporting of the chosen FFQ was only examined in the RAINE study, by checking the ratio of Energy Intake: Estimated Energy requirement (EI: EER) and applying the Goldberg equation (Goldberg et al., 1991). At 14 years old (when teenagers had parental help) 25% were under-reporters, 63% were plausible and 12% were over-reporters. At 17 years old,

(when teenagers completed the FFQ independently) 37% were under-reporters, 54% were plausible and 9% were over-reporters (Ambrosini et al., 2013).

Four USA cohorts (Bogalusa Heart Study, GUTS, GUTS II and Project EAT) used versions of the youth adolescent questionnaire (YAQ) developed by the Channing laboratory (Rockett et al., 1995). In their validation study (Rockett et al., 1997), comparing the YAQ with three 24-HDR in a sample of children of women from the Nurses' Health Study II cohort (NHS II), the average correlation coefficient was 0.54, which suggests that the YAQ yields valid nutritional information about older children and adolescents.

Before using any FFQ it is advisable to calibrate it against a FD or 24-HDR in a sub-sample of the same population (Lachat et al., 2016). When this was done in the Bogalusa Heart Study the energy intakes measured by the YAQ and by 24-HDR were similar (Nicklas, 1995). GUTS researchers relied on the original validation of YAQ, as children in the cohort were also children of women from the NHS II cohort of the same era (Field et al., 2004). In GUTS II, children of women from the NHS II cohort were recruited almost a decade later, but there is no reference to any re-validation of the YAQ in the new cohort (Field et al., 2014). In Project Eat the YAQ was not validated but was tested for comprehension, which was found to be acceptable for middle school children (Cutler et al., 2012).

Project EAT also used the adult form of the YAQ, Willett's FFQ, (Feskanich et al., 1993) at follow-up. When the YAQ and Willett's FFQ were compared (Larson et al., 2012) they gave similar rankings by quartiles of intake. However, absolute intakes were only moderately correlated (r= 0.4 to 0.6) and it was advised that the two questionnaires should not be used to describe change in intake over time. Despite this, one paper from Project EAT (Quick et al., 2013) did investigate change in fruit, vegetables, whole grains and soft drinks using the two instruments. No other validation studies in Project EAT were cited.

The short semi-quantitative FFQ employed in the Colombian Bogotá School children cohort was based on an adult FFQ validated in Costa Rica (Isanaka et al., 2007). The FFQ contained the most frequently consumed foods in the Colombia National Nutrition Survey 2005 but was not calibrated against another form of DAT in the cohort. TEI was estimated, although this 38-item FFQ may not have captured the whole diet of some children.

The longest FFQ, with 212 items, was developed for use in the RAINE study. It was validated (Ambrosini et al., 2009) against a 3 day Food diary. Agreement was "less than ideal", but the FFQ was able to correctly rank most nutrient intakes by tertile.

All FFQs were based on nutritional data appropriate for the country/population being studied. The RAINE study entered FFQ data twice, with independent verification of input to minimise coding errors.

4.5 Adiposity measures

Overweight and obesity are defined by the World Health Organisation as "abnormal or excessive fat accumulation that may impair health" (WHO, 2015).

Approaches for assessing the extent of an individual's fat accumulation (adiposity) include:

- Measure the increased risk of health impairment.
- Measure body fat percentage directly.
- Estimate body fat by anthropometry.
- Estimate body fat from anthropometric indices.

4.5.1 Assessing adiposity in adults and children

Increased health risks due to excess adiposity depend upon the degree of the excess and how long it persists. Although measuring the increased risk of health impairment is an appropriate way to assess obesity in adults, many of the metabolic disorders and other ill effects linked with obesity manifest gradually. It is a less clear measurement method for children, as such ill effects have not had enough time to become apparent (Neovius et al., 2004).

Body fat percentage can be assessed by air displacement plethysmography (ADP) or hydrostatic underwater weighing, bioelectrical-impedance analysis (BIA), whole body dual energy x-ray absorptiometry (DXA or alternatively, DEXA), magnetic resonance imaging (MRI) or computer tomography (CT) scan (Flodmark et al., 2004). DXA scans measure 3 components of body composition including bone mineral and bone-free lean mass as well as fat mass (Simmonds et al., 2015). Body composition can also be assessed using doubly labelled water (DLW) (Beato et al., 2019), whereby study participants ingest a small dose of deuterium enriched water ($^{2}H_2O^{18}$). The difference in markers present in urine or saliva before and after ingestion is then used in a formula to calculate total body water and fat-free mass, using the difference between total body mass and fat-free mass to calculate fat mass (Simmonds et al., 2015). These accurate methods are time-consuming and costly, requiring specialist equipment and trained operators. Under water weighing and scans are not feasible ways of measuring body fat in infants and younger children, but infant ADP is now possible (Simmonds et al., 2015). MRI and CT scans are generally limited to clinical settings and small studies although BIA and DXA have been used in adult and children's cohorts.

Body fat can be estimated by anthropometric measures such as skin fold thicknesses (SFT), measured with spring-tensioned callipers at different sites on the body (E.g. sub-scapular, supra-iliac, abdominal, biceps, triceps, thigh, calf). The SFT measures are totalled (sum of skinfolds) to rank individuals in terms of relative amounts of subcutaneous fat, or the measures are used in population based equations to calculate an estimate of body fat (Slaughter et al., 1988). SFT measurements often become less precise as more are made (tiring to do) and there can be high levels of inter-observer variability. For adults, waist circumference (WC) or hip circumference serve as simple measures of the central or abdominal fat associated with cardio-vascular disease and Type II diabetes.

Skinfold methods are widely used for assessing body fat in children (although callipers are impractical for measuring the youngest) and children's waist circumference is sometimes used as a comparative measure. However, body fat percentage thresholds for overweight/obesity in children are not established, as the assumptions needed may not be true while children are growing and yet to reach maturity (Freedman et al., 2004). A small study of Japanese youth found that the sensitivity and specificity of waist circumference measurements for detecting obesity (confirmed by abnormal values of serum triglyceride and insulin) were >70%, which is high enough for WC to be used as a clinical diagnostic measure (Asayama et al., 2005), but the identified WC threshold for childhood obesity (82cm) may not apply in other populations or age groups. Measures of neck circumference (NC) may also be reliable (LaBerge et al., 2009) and could prove useful for assessing overweight and obesity in childhood, subject to the establishment of age and gender-adjusted references in a given population (Hatipoglu et al., 2010).

Body fat can also be estimated using anthropometric indices. Various indices exist, such as the fat mass index (fat mass/ height ^{x}) (Ambrosini et al., 2012) and the waist-to-height ratio (WHtR) indicative of central adiposity. Waist-to-height ratio has been shown to be useful in children and adolescents (McCarthy and Ashwell, 2006) and can be used as an alternative to skinfold thicknesses (Brambilla et al., 2013). WHtR is another measure influenced by a child's age, sex and ethnicity that has not been standardised, perhaps due to a lack of suitable reference data (Simmonds et al., 2015).

The most widely used anthropometric index is the Quetelet index or Body Mass Index (BMI), which can be readily calculated using the formula:

BMI = Weight or body mass (in kilograms) ÷ by the square of Height (in metres).

Weight and height measurements are easy, safe and inexpensive to obtain, and BMI is a helpful proxy measure of body fat for adults. It assumes that body mass reflects fat mass, although individual differences in body mass are not solely due to differences in adiposity. However, BMI correlates more closely with alternative measures of adiposity (such as skinfold thickness) than other weight for height indices such as Weight/Height^3 (Freedman et al., 2004) and has been validated by the International Obesity Task Force against measures of body fat taken by DXA scan (Dietz and Bellizzi, 1999). At a population level BMI is a useful measure of overweight and obesity in adults as it is the same for all ages and both sexes (WHO, 2015). In the West, an adult BMI of 25 to 30kg/m^2 is classed as overweight and a BMI above 30kg/m^2 is classed as obesity. These cut-offs are based on known health risks for the different BMI levels, but there are ethnic differences. For Caucasian adults the risk of ill health due to excessive fat accumulation starts at or above 25kg/m^2, whereas for Asian adults the risk starts at or above 23kg/m^2.

In a small study of Italian children and adolescents (n = 198) BMI was found to be strongly associated with body fat measured by DXA, although with wide variations (Pietrobelli et al., 1998). The authors concluded that BMI was a useful measure of body fatness in groups of children/adolescents but less reliable for individuals and cautioned against comparing BMI when children were from different age groups.

As there are natural fluctuations in BMI during childhood, which depend upon age, race and gender and differing maturation rates between children, it has proved difficult to establish cut-offs for childhood overweight and obesity (Neovius et al., 2004). It is hard to know at which childhood BMI increased health risks begin, as there is a time lag between excess fat accumulation and the development of disease.

4.5.2 Growth reference data

Given that few ways of measuring adiposity are ideal for growing children, the accepted method to determine overweight or obesity in children is by classification, using national or international growth reference data. Anthropometric measurements are adjusted by age and sex, creating growth indices that reflect normal childhood growth and possible under or over nutrition. Clinicians are advised to use tried and tested national growth reference data to assess obesity accurately and safely in individual children and adolescents. For comparative research purposes international reference data are recommended (Reilly, 2002). International reference data are *not* intended for national epidemiological or clinical use (Cole et al., 2000) and testing does not support their use in this way (Reilly et al., 2000; Reilly, 2002).

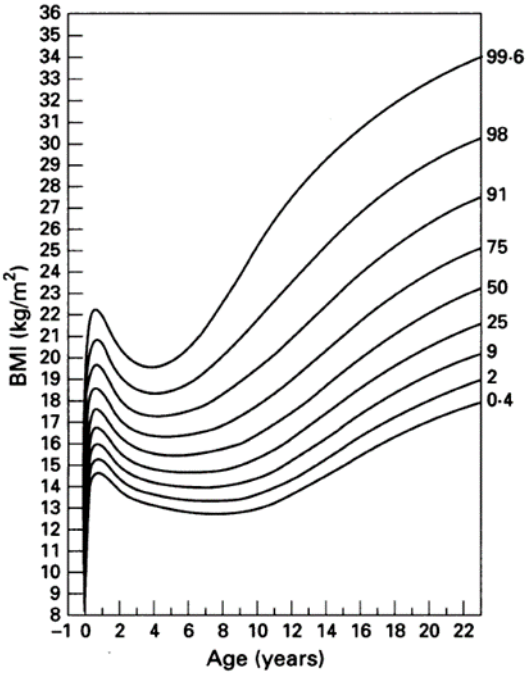
Using national or international references that have different definitions and cut-offs for childhood obesity may give different classifications of obesity for an individual child and will produce different estimates of obesity prevalence in the same population, as shown in the ALSPAC cohort (Reilly et al., 2000) and by the European Childhood Obesity Group (ECOG) (Flodmark et al., 2004). Researchers should consider which growth reference best serves their research aims, or use both (Reilly, 2002). As children's patterns of growth are shifting over time, growth reference data are updated periodically. The most up to date growth reference may not be the best option if it does not match the era of the population being studied.

4.5.2.1 National reference data

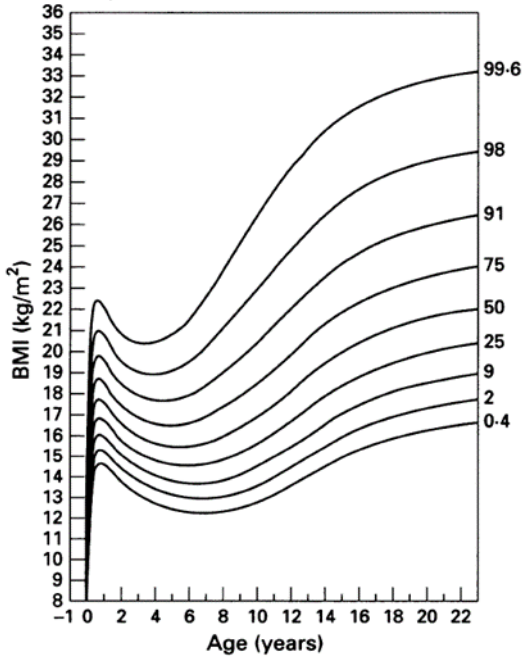
The UK, USA and some European countries use national growth reference curves, based on population samples, to check children's height and weight growth trajectories. Childhood height and height are correlated, so independently they reflect the child's size (small or large) but not whether a child is thin or fat. Weight for height gives a measure of adiposity, but must be adjusted for age and sex, as body fat percentage varies quite naturally during childhood, rising in infancy, then falling through early childhood before the adiposity rebound that heralds puberty, rising through adolescence towards adult maturity. Based on centile growth curves the BMI of an individual child can be represented as a percentile or as a BMI for age z-score (a SD score which shows how much the child deviates from the population mean BMI, without implying if the score is "healthy" or not). To calculate BMI percentiles or z-scores measures of height and weight, the child's sex and age at measurement and growth charts from a reference population that matches the child's background are needed. Examples of National Reference data for children include the UK 1990 growth reference (Cole et al., 1995) (See Figure 4-2) and the USA's Center for Disease Control CDC 2000 growth charts (Kuczmarski et al., 2002).
The UK 1990 reference data (UK90) uses cut-offs between the 85th and 94th percentile (BMI z-score ≥ 1.04 to < 1.64) to define childhood overweight and ≥ 95th percentile (BMI z-score ≥ 1.64) to define childhood obesity for population monitoring purposes (Hughes et al., 2011). (The UK90 cut-offs for clinical purposes are ≥ 91st percentile and ≥ 98th percentile respectively.)
The CDC 2000 was developed using data from 5 national surveys (National Health Examination Survey II 1963 to 1965, NHES III 1966 to 1970, National Health and Nutrition Examination Survey I 1971 to 1974, NHANES II 1976 to 1980, NHANES III 1988 to 1994) and provides growth curves up to the age of 20 years. As recommended by an expert committee (Barlow, 2007), the CDC 2000 growth reference uses cut-offs between the 85th and 94th percentile for children of the same age and sex to define overweight (previously called "at risk of overweight") and at or above the 95th percentile for children of the same age and sex to define obesity (previously called "overweight").

Figure 4-2 Sex specific growth charts for the UK 1990 Growth Reference

The chart shows nine centile curves for BMI in British **boys** in 1990.



The chart shows nine centile curves for BMI in British **girls** in 1990.



Data were collected between 1978 and 1990. Smoothed summary curves were calculated by the least mean squared method and penalized likelihood (Cole and Green, 1992) which adjusts the BMI distribution at different ages, allowing for varying amounts of skewness, to adjust the data to normality.

Age (years) lies on the x axis.

BMI (kg/m^2) lies on the y axis.

Centiles are spaced two thirds of a SD score apart.

The 50th centile curve is the median (M) curve.

BMI increases steeply in infancy, with all centiles peaking at much the same age. The centile curves then dip, flattening at around 5.5 years when median BMI is 15.5 kg/m^2, before the adiposity rebound. For boys, age at adiposity rebound is over three years sooner on the higher rather than the lower centiles.

For girls, age at adiposity rebound is two years sooner on the higher rather than the lower centiles.

After the rebound, BMI increases more rapidly in girls than in boys until the age of 18 years, when median BMI is 21.0 kg/m^2. Thereafter boys have a higher median BMI for age than girls. (Cole et al., 1995).

4.5.2.2 International reference data

At the instigation of the International Obesity Task Force (IOTF) an international definition of childhood overweight and obesity was developed, based on six nationally representative cross sectional growth studies between 1963 and 1993, with varying levels of obesity prevalence. Growth studies came from Brazil, Great Britain, Hong Kong, the Netherlands, Singapore and the USA, totalling over 190,000 males and females from birth to age 25 years (Cole et al., 2000). The IOTF wanted the standards used to identify overweight and obesity in children to agree with the established standards for adult overweight and obesity (Dietz and Bellizzi, 1999). Accordingly, centile curves for boys and girls were drawn for each growth study, with extra curves that passed through the widely accepted adult overweight (25 to 30kg/m^2) and obesity (>30kg/m^2) cut-offs at 18 years. Curves were averaged to give international age and sex specific cut-offs for overweight and obesity between 2 and 18 years. When the international cut-offs were applied to the national datasets, overweight prevalence at 18 years ranged from 5 to 18% and obesity prevalence ranged from 0.1 to 4%.

4.5.2.3 World Health Organisation growth standards and references

It is increasingly recognised that using population samples to create a growth reference, when there is an underlying trend towards increasing prevalence of overweight and obesity, may result in skewness that underestimates overweight and obesity and overestimates under nutrition (de Onis, 2004).

The World Health Organisation (WHO) developed growth standards for children from birth to 5 years, based on approximately 8,500 breast-fed children growing up in optimal conditions (MGRS, 2006). These standards represent an *aspirational* pattern of growth and have been adopted in over 110 countries (de Onis and Lobstein, 2010).

For older children, where a national growth reference was not available, WHO previously recommended the National Centre for Health Statistics/WHO growth reference for children above 5 years old, which included a BMI for age reference based primarily on USA growth data (Must et al., 1991). The National Centre for Health Statistics/WHO growth reference had evident shortcomings and in 2006 a decision was made to redevelop it. The current WHO Growth Reference for children aged 5 to 19 years has BMI centile curves that tie in with the WHO Child Growth Standards for children under 5 years, although there is not an exact match at age 60 months (de Onis and Lobstein, 2010). It also aligns closely with accepted adult cut-offs for overweight and obesity at age 19 years (de Onis et al., 2007). Overweight is defined as a BMI-for-age greater than 1 standard deviation above the WHO Growth Reference median. Obesity is defined as greater than 2 standard deviations above the WHO Growth Reference median (WHO, 2015). At 19 years old, this definition of overweight equates to a BMI of 25.0

kg/m^2 for girls and 25.4 kg/m^2 for boys, and obesity equates to a BMI of 29.7 kg/m^2 for both sexes (de Onis et al., 2007).

4.5.3 Measuring adiposity change in growing children

As children grow, their BMI (kg/m^2) naturally changes. Population based growth references, from which BMI percentiles and BMI z scores for age and sex are generated, are presented as growth curves, but the growth of an individual child does not follow a smooth trajectory. Instead children and adolescents tend to have non-linear peaks in growth (growth spurts); height and weight do not necessarily increase in tandem, so their BMI percentile or BMI z score will fluctuate.

A study of Italian kindergarten children found that the BMI z score was best suited to assessing adiposity on a single occasion but BMI or BMI percentile were better for measuring change (Cole et al., 2005). This finding was reinforced by a simulation study using data from the Growing Up Today study (GUTS) (Berkey and Colditz, 2006) which demonstrated that for adolescents, change in BMI is a better measure than change in BMI z score in longitudinal studies and is more readily interpreted.

4.6 Adiposity measures used by included cohorts

The adiposity measures employed by each cohort were previously summarised in Chapter 3, Table 12. All fourteen included cohorts used height and weight to calculate children's BMI. Using BMI and a growth reference, two-thirds of the included papers calculated BMI z-scores, BMI percentiles or classified children/adolescents as overweight or obese at baseline and/or follow-up. Papers from ten of the included cohorts used at least one other adiposity measure, such as waist circumference, skinfold thickness or body fat percentage.

4.6.1 Body Mass Index

Most included cohorts used *measured* height and weight to calculate BMI. Typically, height was measured in bare or stockinged feet using a stadiometer) and weight was measured without shoes and in light clothing, with digital weighing scales. A counterbalance scale was used for the Framingham Children's study (Hasnain et al., 2014) and electric chair scales were used by the RAINE study (Ambrosini et al., 2013). The IDEA & ECHO study (Laska et al., 2012) and the NGHS (Albertson et al., 2009) employed trained staff to take measurements at clinic visits, as did the Bogotá School Children cohort (Shroff et al., 2014), which will have reduced observer bias. Certified examiners who measured NGHS girls took two measures of height and of weight to calculate an average, taking a third measure if the discrepancy was greater than 0.5cm for height or 0.3kg for weight.

Three of the USA cohorts (Project EAT, GUTS and GUTS II) assessed adiposity outcomes by calculating BMI from *self-reported* height and weight.

In Project EAT height and weight were measured and self-reported by school students at baseline but only self-reports were available at the 5 year (Fulkerson et al., 2008) and 10 year follow-up (Berge et al., 2015). For consistency self-reports were used throughout. An earlier paper from Project EAT I compared self-reported height and weight with measured values for 3,797 children aged 12 to 18 years. Both male and female adolescents tended to overestimate their height and underestimate their weight, such that BMI was underestimated by - 2.2kg/m^2 for males and by - 2.5kg/m^2 for females (r = 0.89 and 0.85) (Himes et al., 2005).

No internal validation studies of self-reported vs. measured height and weight were referenced in the GUTS (Field et al., 2003b) and GUTS II (Field et al., 2014) cohorts. As evidence of the reliability of self-reported height and weight to calculate BMI, adiposity status and weight change over time, authors cited studies in other USA adolescent cohorts. In the NHANES III study, despite under reporting of weight which varied with race and gender (Strauss, 1999), 94% of adolescents received the correct classification of weight status based on self-reported height and weight. Similar under-reporting of weight was observed in both sexes in the National Longitudinal Study of Adolescent Health (NLSAH) cohort (Goodman et al., 2000) but based on BMI calculated from self-reported height and weight 96% of teenagers were correctly classified. In the NLSAH under-reporting of weight was found to be consistent over time, so weight change based on self-report under-estimated true weight change by only 2 or 3 lbs or 1.5kg. (Field et al., 2007)

BMI at follow-up was the adiposity outcome in the Bogalusa Heart Study (O'Neil et al., 2015), Framingham Children's study (Hasnain et al., 2014) and NGHS (Affenito et al., 2005).

Change in BMI was the adiposity outcome in the EYHS (Zheng et al., 2014), GUTS (Taveras et al., 2005), GUTS II (Field et al., 2014), IDEA & ECHO (Laska et al., 2012) and NGHS (Berz et al., 2011) (Ritchie, 2012; Ritchie et al., 2007) (Striegel-Moore et al., 2006).

Adult overweight/obesity at follow-up (based on BMI from self-reported height and weight) was the outcome of interest employed in Project EAT.

4.6.2 Growth references

Several included cohorts used a growth reference only for classifying children's overweight or obesity status at baseline, but papers from approximately half the cohorts used a growth reference to generate BMI z scores or BMI percentiles (at follow-up or change) or to determine overweight/obesity at follow-up as their outcome measure.

Papers from included USA cohorts generally selected the USA national CDC 2000 growth reference (Kuczmarski et al., 2002). The CDC 2000 growth reference data broadly matched the era of included USA cohorts that used it . The timing of baseline measures of these USA cohorts and the purpose for which they used the CDC 2000 growth reference, are listed below:

- BDPP aged ~ 9 years in 2001/2002
  BMI percentiles. (Balvin Frantzen et al., 2013)
- Bogalusa Heart study aged ~ 10 years old in 1973 to 1984
  Children's overweight and obesity classification at baseline. (O'Neil et al., 2015)
- FAMS girls aged 9 to 14 years in 2000 to 2001
  Change in BMI z score. (St-Jules et al., 2014)
- GUTS aged 9 to 14 years in 1996
  Change (increase) in BMI z score. (Field et al., 2003a)
- NGHS girls aged 9 or 10 years in 1987
  BMI z score at each annual visit. (Albertson et al., 2007)
  Change in BMI percentile, onset of overweight or obesity (Rehkopf et al., 2011)
- Project EAT adolescents aged ~ 15 years in 1998 to 1999
  Adolescents' overweight and obesity classification at baseline. (Quick et al., 2013)

In the Bogalusa Heart Study at follow-up as young adults, overweight was defined as a BMI from > 24.9 to 29.9 kg/m^2 and obesity as BMI ≥ 30 kg/m^2.

In Project EAT participants were adults by the 10 year follow-up, so adult cut-offs for overweight (BMI ≥ 25 and < 30kg/m^2) and obesity (BMI ≥ 30 kg/m^2) were applied by some papers (Berge et al., 2015) (Quick et al., 2013). Other Project EAT papers (Cutler et al., 2012; Fulkerson et al., 2008), in an attempt to avoid the "discontinuity" at the 20 year age boundary between the CDC 2000 growth reference and the usual adult classifications for overweight and obesity, instead used an obesity reference for children and adults developed by Must et al. Must's reference is based on data collected between 1971 and 1974 by NHANES I from over 20,000 individuals aged between 6 and 74 years old (Must et al., 1991). Participants were classified as overweight/obese if their BMI was ≥ 85th percentile for age and sex.

The IOTF growth reference (Cole et al., 2000) was used in the Australian RAINE study to define overweight and obesity among teenagers aged 14 years old at baseline in 2003 to 2006 and to assess the risk of overweight and obesity at 3 year follow-up (Ambrosini et al., 2013).
The IOTF growth reference (Cole et al., 2000) was used by one included paper from the USA GUTS II cohort (Field et al., 2014) to classify overweight and obesity, using the usual adult classifications for participants who were 18 years or older.

The IOTF reference was also used by one included paper from the UK ALSPAC cohort (Bigornia et al., 2014) to classify overweight and obesity, although a previously published paper from ALSPAC (Reilly et al., 2000) had established that the obesity cut-off in the national UK 1990 growth reference (Cole et al., 1995) had a greater sensitivity than the IOTF obesity cut-off among 7 year old children in the cohort. Another included paper from the UK based ALSPAC cohort used the USA CDC 2000 growth charts (rather than the UK 1990 or IOTF growth reference) to classify overweight/obesity at baseline for a stratified analysis (Noel et al., 2013).

Elsewhere in Europe the DONALD study used a growth reference based on a population of 17,147 boys and 17,275 girls, aged 0 to 18 years, compiled from 17 regional studies in Germany (Kromeyer-Hauschild et al., 2001). Following guidelines from the Arbeitsgruppe Adipositas im Kindes und Jugendalter (Working Group on obesity in childhood and adolescence) the 90th and 97th percentiles in this reference population were used as cut-offs for overweight and obesity in German children and adolescents in one paper from DONALD (Libuda et al., 2008). Another DONALD paper used the same German reference to calculate BMI but applied the age and sex-specific cut-offs for overweight and obesity proposed by the IOTF (Cheng et al., 2009). BMI calculated from height and weight were converted to standard deviation scores (BMI-SDS or BMI z scores) using the least mean squares method (Cole, 1990). Change in BMI-SDS was used as an outcome variable.

In the Danish arm of the EYHS (Zheng et al., 2015) age and sex specific BMI z scores were generated using the least mean squares method (Cole and Green, 1992). Overweight/obesity classifications were not reported, and no growth reference was cited. Instead change in BMI z score between baseline at 9 years and follow-up at 15 years was calculated for use as an outcome variable.

In the Colombian Bogotá School Children cohort, with children aged ~ 8 years old at recruitment in 2006, BMI z scores were calculated using WHO growth references (de Onis et al., 2007), defining overweight or obesity as BMI-for-age z-score > 1 and obesity as BMI-for-age z-score > 2 (Shroff et al., 2014).

4.6.3 Waist circumference

Seven cohorts measured waist circumference to assess children's central adiposity, at the "level of the umbilicus", the "narrowest part of the torso" or at the "midpoint between the ribs and the iliac crest", usually measuring twice to obtain an average in mm or cm. A third measure was taken if there was a wide discrepancy between the two.

Waist circumference was measured by trained observers following protocols in the Bogalusa Heart Study (O'Neil et al., 2015) and by trained researchers using standardised techniques in the Bogotá School Children cohort (Shroff et al., 2014).

The Framingham Children's study used cloth measuring tapes (Hasnain et al., 2014), which may stretch with repeated use, introducing systematic measurement error. All other cohorts used "inextensible" measuring tapes.

One included paper that measured waist circumference described calculating waist-to-height ratio in their methods (Albertson et al., 2009), which was then used as an outcome measure of adiposity in NGHS girls.

4.6.4 Skin fold thicknesses

Seven cohorts measured skin fold thicknesses (SFT) in children with callipers, measuring twice to obtain an average in mm, thereby reducing measurement error. Cohorts measured two, three or four skinfolds, reported in different ways:

The Bogalusa Heart Study (O'Neil et al., 2015) reported subscapular and triceps SFT as two separate measures, the Bogotá School Children cohort (Shroff et al., 2014) reported a ratio.

Subscapular and triceps SFT in the DONALD study (Alexy et al., 2011) were used to calculate body fat percentage using Slaughter's equations (Slaughter et al., 1988).

Supra-iliac, subscapular and triceps SFT were used to calculate body fat percentage in NGHS girls (Ritchie et al., 2007).

Supra-iliac, subscapular and triceps SFT plus either biceps or abdominal SFT were used to calculate the sum of 4 skinfolds in the Framingham study (Hasnain et al., 2014), the EYHS (Zheng et al., 2015) and in girls from the Female Adolescent Maturation study (St-Jules et al., 2014).

4.6.5 Body fat percentage by BIA or DXA

The IDEA & ECHO cohorts (Laska et al., 2012) and the NGHS (Albertson et al., 2009) measured body fat percentage at the same time as weight, using bioelectric-impedance scales.

Three cohorts assessed fat mass using a Lunar DXA scanner but processed the information in different ways.

In ALSPAC, children's fat mass was measured by DXA during clinics at 11+, 13+ and 15+ years (Ambrosini et al., 2012). Fat Mass index (FMI) was calculated separately for boys and girls, as fat mass divided by height, raised to an optimum power which varied by age and sex. Log transformations were used to translate FMI into a standardised z-score.

In the Female Adolescent Maturation study at follow-up each girl's ratio of trunk-to-peripheral fat was calculated (St-Jules et al., 2014).

In the Framingham Children's Study at follow-up percentage body fat was calculated by dividing total fat mass by total body weight (Hasnain et al., 2014).

4.6.6 Summary of adiposity measures

Measures of height and weight or estimates from self-report were universally employed by included cohorts to calculate BMI, but only half the studies converted this information to comparable BMI z scores, BMI percentiles or classification of overweight/obesity using a growth reference. Five different age and sex specific growth references were used (CDC 2000, Must's , IOTF, German national, WHO).

Waist circumference was a simple and popular measure, used by half the cohorts. Skin fold thickness, requiring more expertise but relatively inexpensive equipment, was an equally popular way of assessing body fat percentage or fat distribution in children, although there were substantial methodological differences in the way these measures were taken and reported.

Assessment of adiposity using more sophisticated equipment was less common. Although BIA scales are now established as a convenient way to estimate body fat percentage at the same time as measuring weight, in the 1980s and 1990s only two included cohorts made use of them. Three cohorts employed DXA scans. Perhaps because DXA machinery was novel, all three studies processed the fat mass data in different ways, which precluded direct comparison of results.

4.7 Comparison of reported energy and macronutrient intakes, BMI and overweight prevalence in each cohort

4.7.1 Methods

Data extraction

As described in Chapter 2, data from every included paper was extracted to an Excel spreadsheet. Extracted data included the Study/cohort name, country, sample size in analyses, sex and dietary assessment method with ages at assessment and, if reported, mean total energy intake, macronutrient intakes (Carbohydrate, protein and fat in g/day or as a percentage of energy intake) at baseline and follow-up, plus adiposity measures with ages at baseline and at follow-up.

Data synthesis: Mean Energy, macronutrient and fibre intakes

Whole diet was measured and quantified in all 14 included studies. Where provided, mean age at dietary assessment (baseline or follow-up) and mean values of energy intake (kcal/day) with s.e. or S.D. measures of variance, macronutrient intakes (as % of energy intake) and fibre intake (g/day) for each cohort were copied directly from included papers.

When papers gave an age range the mid-point was used as an approximation of the mean age. Mean energy intakes given in kJ/day were divided by 4.184 to convert to kcal/day.

Measures of variance from the mean energy intake given as s.e. were converted to SD, based on the number of participants in the cohort (or in a category) using the formula:

$$SD = s.e. \times \sqrt{n}$$

When mean macronutrient intakes were presented as g/day, they were converted to % of mean energy intake using the metabolizable energy conversion factors from McCance & Widdowson's The Composition of Foods seventh summary edition (McCance, 2013) as follows:

% Energy from Carbohydrate =
((Carbohydrate intake, g/day x **3.75** kcal/g) / Energy intake, kcal/day) x 100

% Energy from Fat =
((Fat intake, g/day x **9.0** kcal/g) / Energy intake, kcal/day) x 100

% Energy from Protein =
((Protein intake, g/day x **4.0** kcal/g) / Energy intake, kcal/day) x 100

Some papers did not provide mean energy intakes and macronutrient intakes for the whole cohort. Instead data was presented for each category of participants (E.g. by sex or ethnic group) or by each quantile or category of intake for the dietary exposure under investigation. Using the number of participants in each category the mean energy intake and nutrient intake was calculated for the whole cohort, as follows:

E.g. Energy intake shown for each category of flavoured milk intake.
Mean energy intake for the whole cohort =
((No. in cat 1 x energy intake in cat 1) + (No. in cat 2 x energy intake in cat 2)) / (No. in cat 1 + No. in cat 2)

The extent of variance in the measure of energy intake or macronutrient intake for the whole cohort could not be derived from s.e. or SD for each category in this way. After converting s.e. to SD, SD was calculated as a percentage of the mean energy intake for each category (or for the whole cohort if reported) as a comparative measure of the extent of variance in measures of energy intake between cohorts and by dietary assessment tool. Variance for categories of macronutrient intake were not calculated.

Reported and derived mean energy intakes in each cohort were sorted by mean age at dietary assessment. Sometimes energy intake in the same study at the same age was given by more than one paper, with slightly different values; the mean energy intake from the analysis with the largest number of participants was preferred.

Energy intakes in all cohorts, at baseline or follow-up, were plotted against age, using copied/reported values when available, rather than mean values calculated from categories. Baseline measures of energy intake were also plotted against age as sub-plots to explore differences by sex (mixed, girls, boys), geography (USA vs Europe) and dietary assessment tool (FD, 24-HDR, FFQ). The extent of variance in measured energy intakes were compared. Mean Energy intakes by age were compared with UK estimated average requirements (EAR) (SACN, 2011).

Directly copied and derived values for macronutrient intakes as % of energy were plotted as a stacked bar chart. If only two macronutrient values were provided, the third macronutrient was assumed to contribute the balance of energy intake, to bring the total to 100%. Macronutrient intakes as % of energy were compared with World Health Organization population nutrient intake goals (WHO/FAO, 2003). Mean dietary fibre intakes, where provided, were also compared with UK recommendations.

Data synthesis: BMI and overweight/obesity

Most included studies reported BMI, calculated from measured or self-reported height and weight, at baseline and sometimes also at follow-up, which could be readily compared by age and against the UK 1990 growth reference. Other adiposity measures (BMI z score or BMI percentile, waist circumference, sum of skinfold thicknesses, % body fat) were reported by fewer papers, or were not readily comparable due to methodological differences between studies. Several papers reported overweight/obesity prevalence in the cohort at baseline or follow-up.

Where presented for the whole cohort, mean age at assessment (at baseline and follow-up) and the corresponding mean BMI values were extracted directly.
The mid-point of an age range was used as an approximation of the mean age.
When BMI was presented by categories, with the number of participants in each category, a mean for the whole cohort was derived as before. Mean BMIs reported in all cohorts, at baseline or follow-up, were plotted against age, again using copied rather than derived mean values whenever possible.
Where reported, the percentage of participants with overweight/obesity were plotted, by cohort and age, and across all ages.

4.7.2 Results: Mean energy intakes

Energy intakes were reported by at least one paper from every cohort except GUTS II (Field et al., 2014). Many papers presented energy intakes measured at baseline and some also reported repeat or follow-up measures. Reported and calculated mean energy intakes for

each cohort/paper are presented in Table 4-4, with the DAT, sex and age of participants at the time of measurement, alongside macronutrient and fibre intakes where available. Note that energy and nutrient intakes for girls in the NHLBI Growth and Health study (NGHS) were reported as averages from eight annual measurements taken during the nine year study, between the ages of 9/10 years and 18/19 years, not separately for each year.

The relative extent of variance in the measure of energy intake for each cohort, organised by DAT, is summarised in Table 4-5. This table presents the calculated percentage of SD/mean energy intake for the whole cohort if available, or the minimum and maximum calculated % of SD/mean energy intake reported across categories. Numbers of observations in categories were sometimes small, so percentages are a crude measure of relative variance. Energy intakes measured with a 3 day food diary generally had less variance than those measured by 24-HDR. Energy intakes measured by FFQ had the greatest variance, even in larger cohorts. In FAMS a median value and inter quartile range for energy intake was reported, so SD could not be derived (St-Jules et al., 2014). In Project EAT the energy intake of young adults at follow-up was measured using Willett's FFQ, but a SD could not be derived as only change from baseline using the YAQ was reported (Quick et al., 2013).

Reported mean energy intakes from each cohort (baseline, repeated and follow-up measures) were plotted against age as shown in Figure 4-3. Again, note that in FAMS median energy intakes were reported. Ages at dietary assessment ranged from 6 to 16 years old at baseline and from 15+ to 23+ years at follow-up, so there is an overlap between the baseline and follow-up measures. As expected, the trend was for energy intake to increase with increasing age of children in the cohorts. When baseline mean energy intakes are considered separately for girls and boys or for mixed sex cohorts, increasing energy intake with age is still seen, but boys tended to have higher energy intakes than girls at the same age. See Figure 4-4.

In a plot of reported baseline mean energy intakes against age for USA cohorts and for European cohorts (without the Colombian Bogotá School Children cohort or the Australian RAINE study), see Figure 4-5, the trend lines suggest that children in USA cohorts have slightly higher mean energy intakes than their European counterparts at the same age. The apparently higher mean energy intake of USA cohorts may in part be due to the choice of dietary assessment tool. When baseline mean energy intakes against age are compared by the different DATs, see Figure 4-6 it shows that FFQs (used by large USA cohorts including GUTS, GUTS II and Project EAT) tended to give higher estimates of mean energy intake than 3 day Food diaries at the same age.

Table 4-4 Mean Energy intakes, macronutrient intakes (as % Energy) and mean fibre intake by cohort and paper

Values in italics are calculated from reported values.

Mean macronutrient values in red do not meet WHO recommendations. Mean fibre intakes in red do not meet current UK age recommendations.

Cohort	Paper	D.A.T.	No.	Sex	Mean age, years	Energy intake, kcal/ day	Energy intake, kJ/ day	Carbs. as % Energy	Added Sugars as % Energy	Total fat as % Energy	Sat fat as % Energy	Protein as % Energy	Fibre, g/day
Recommended Nutrient intakes								55% to 75% Energy	5% to 10% Energy	15% to 30 % Energy	< 10% Energy	10%to 15% Energy	
ALSPAC	Noel 2011					No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
ALSPAC	Noel 2013	3 day FD Baseline	2,270	Girls & Boys	10.6	*1,942*	*8,124*	*48.7*	*n/a*	*35.4*	*13.6*	*12.9*	*11.8*
ALSPAC	Bigornia 2014	3 day FD Baseline	2,455	Girls & Boys	10.6	*1,880*	*7,866*	n/a	n/a	36.2	n/a	*13.0*	*11.8*
ALSPAC	Ambrosini 2012					No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
BDPP	Balvin Frantzen 2013	24-HDR x3 Baseline	625	Girls & Boys	9.1	*1,627*	*6,808*	*46.6*	n/a	*35.8*	*13.0*	n/a	*10.7*
Bogalusa	O'Neil 2015	24-HDR x1 Baseline	355	Girls & Boys	10.1	*2,130*	8,912	n/a	n/a	n/a	n/a	n/a	n/a
Bogalusa	O'Neil 2015	24-HDR x1 Follow up	355	Women & Men	23.6	*2,283*	9,553	n/a	n/a	n/a	n/a	n/a	n/a
Bogotá	Shroff 2014	FFQ Baseline	961	Girls & Boys	5 to 12	*1,540*	*6,445*	n/a	n/a	n/a	n/a	n/a	n/a
DONALD	Cheng 2009	3 day FD Baseline	215	Girls & Boys	*9.4*	1,690	*7,071*	51.0	13.6	35.9	15.8	13.1	18.2

Cohort	Paper	D.A.T.	No.	Sex	Mean age, years	Energy intake, kcal/ day	Energy intake, kJ/ day	Carbs. as % Energy	Added Sugars as % Energy	Total fat as % Energy	Sat fat as % Energy	Protein as % Energy	Fibre, g/day
DONALD	Cheng 2009	3 day FD Follow up	215	Girls & Boys	*13.4*	2,061	*8,623*	52.1	n/a	34.6	15.0	13.2	21.3
DONALD	Libuda 2008	3 day FD Baseline	125	Boys	11.9	*2,120*	8,870	n/a	n/a	n/a	n/a	n/a	n/a
DONALD	Libuda 2008	3 day FD Baseline	119	Girls	11.8	*1,825*	7,636	n/a	n/a	n/a	n/a	n/a	n/a
DONALD	Libuda 2008	3 day FD Follow up	125	Boys	16.8	*2,661*	11,135	n/a	n/a	n/a	n/a	n/a	n/a
DONALD	Libuda 2008	3 day FD Follow up	119	Girls	16.8	*1,945*	8,139	n/a	n/a	n/a	n/a	n/a	n/a
DONALD	Alexy 2011	3 day FD Baseline	296	Boys	3 to 18	*1,850*	7,740	n/a	n/a	n/a	n/a	n/a	n/a
DONALD	Alexy 2011	3 day FD Baseline	289	Girls	3 to 18	*1,551*	6,490	n/a	n/a	n/a	n/a	n/a	n/a
DONALD	Alexy 2011	3 day FD Baseline	585	All (girls & boys)	3 to 18	*1,702*	7,122	52.7	n/a	33.9	14.9	13.1	n/a
EYHS	Zheng 2015	24-HDR x1 Baseline	358	Girls & Boys	9.6	*2,175*	9,100	*49.7*	n/a	*33.2*	n/a	12.8	18.6
EYHS	Zheng 2014	24-HDR x1 Baseline	283	Girls & Boys	9.6	*2,226*	9,315	*53.8*	*19.7*	*32.2*	n/a	*13.1*	n/a
EYHS	Zheng 2014	24-HDR x1 Follow-up	187	Girls & Boys	15.7	*2,407*	*10,070*	*58.0*	*22.1*	*27.4*	n/a	*13.6*	n/a
FAMS	St Jules 2014	3 day FD Baseline	200	Girls only	11.5	*1,721*	*7,203*	*53.0*	n/a	*32.5*	n/a	*15.0*	n/a
Framingham	Hasnain 2014	3 day FD Baseline	98	Girls & Boys	3 to 9	*1,724*	*7,214*	n/a	*16.5*	*33.8*	n/a	*13.7*	n/a
GUTS II	Field 2014	FFQ Baseline	3,438	Boys	12.9	No report	No report	n/a	n/a	n/a	n/a	n/a	n/a

Cohort	Paper	D.A.T.	No.	Sex	Mean age, years	Energy intake, kcal/ day	Energy intake, kJ/ day	Carbs. as % Energy	Added Sugars as % Energy	Total fat as % Energy	Sat fat as % Energy	Protein as % Energy	Fibre, g/day
GUTS II	Field 2014	FFQ Baseline	4,121	Girls	13.0	No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
GUTS	Field 2004					No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
GUTS	Field 2003					No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
GUTS	Field 2003a	FFQ Baseline	6,175	Boys	11.8	2,290	*9,581*	n/a	n/a	n/a	n/a	n/a	n/a
GUTS	Field 2003a	FFQ Baseline	8,203	Girls	12.0	2,050	*8,577*	n/a	n/a	n/a	n/a	n/a	n/a
GUTS	Taveras 2005	FFQ Baseline	14,335	All (girls & boys)	9 to 14	2,140	*8,953*	n/a	n/a	*30.6*	*10.9*	n/a	*16.8*
IDEA & ECHO	Laska 2012	24-HDR x3 Baseline	327	Boys	14.6	2,196	*9,189*	n/a	n/a	n/a	n/a	n/a	n/a
IDEA & ECHO	Laska 2012	24-HDR x3 Baseline	339	Girls	14.6	1,777	*7,433*	n/a	n/a	n/a	n/a	n/a	n/a
IDEA & ECHO	Laska 2012	24-HDR x3 Follow-up	276	Boys	16.5	2,230	*9,329*	n/a	n/a	n/a	n/a	n/a	n/a
IDEA & ECHO	Laska 2012	24-HDR x3 Follow-up	286	Girls	166	1,767	*7,394*	n/a	n/a	n/a	n/a	n/a	n/a
NGHS	Berz 2011	3 day FD Visit 1 - 8	2,327	Girls only	9/10 to 16/17	*1,873*	*7,835*	n/a	n/a	n/a	n/a	n/a	n/a
NGHS	Albertson 2007	3 day FD Visit 1 - 10	2,352	Girls only	9/10 to 18/19	1,900	*7,950*	n/a	n/a	n/a	n/a	n/a	n/a
NGHS	Affenito 2005	3 day FD Visit 1 - 10	2,379	Girls only	9/10 to 18/19	No report	No report	n/a	n/a	n/a	n/a	n/a	10.5 to 13.1
NGHS	Barton 2005			Girls only		No report	No report	n/a	n/a	n/a	n/a	n/a	n/a

Cohort	Paper	D.A.T.	No.	Sex	Mean age, years	Energy intake, kcal/ day	Energy intake, kJ/ day	Carbs. as % Energy	Added Sugars as % Energy	Total fat as % Energy	Sat fat as % Energy	Protein as % Energy	Fibre, g/day
NGHS	Albertson 2009			Girls only		No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
NGHS	Franko 2008			Girls only		No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
NGHS	Rehkopf 2011			Girls only		No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
NGHS	Ritchie 2012			Girls only		No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
NGHS	Ritchie 2007	3 day FD Visit 1 - 10	2,371	Girls only	9/10 to 18/19	*1,873*	*7,836*			*34.8*	*12.7*		*12.4g / 2,000 kcal*
NGHS	Striegel-Moore 2006			Girls only		No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Berge 2015					No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Cutler 2012					No report	No report	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Fulkerson 2008	FFQ Baseline	2,516	All	14.9	*2,097*	*8,774*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Fulkerson 2008	FFQ Baseline	367	Male	12.8	2,293	*9,594*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Fulkerson 2008	FFQ Baseline	763	Male	15.9	2,187	*9,150*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Fulkerson 2008	FFQ Baseline	439	Female	12.8	2,203	*9,217*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Fulkerson 2008	FFQ Baseline	947	Female	15.9	1,187	*7,895*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Fulkerson 2008	FFQ Follow-up	2,516	All	*19.9*	1,818	*7,607*	n/a	n/a	n/a	n/a	n/a	n/a

Cohort	Paper	D.A.T.	No.	Sex	Mean age, years	Energy intake, kcal/ day	Energy intake, kJ/ day	Carbs. as % Energy	Added Sugars as % Energy	Total fat as % Energy	Sat fat as % Energy	Protein as % Energy	Fibre, g/day
Project EAT	Fulkerson 2008	FFQ Follow-up	367	Male	*17.8*	2,099	*8,782*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Fulkerson 2008	FFQ Follow-up	763	Male	*20.9*	2,075	*8,682*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Fulkerson 2008	FFQ Follow-up	439	Female	*17.8*	2,050	*8,577*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Fulkerson 2008	FFQ Follow-up	947	Female	*20.9*	1,706	*7,138*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Quick 2013	FFQ Baseline	756	Male not o/w	15.0	1,941	*8,123*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Quick 2013	FFQ Baseline	887	Female not o/w	15.0	2,271	*9,502*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Quick 2013	FFQ Follow-up	756	Male not o/w at start	25.4	2,373	*9,929*	n/a	n/a	n/a	n/a	n/a	n/a
Project EAT	Quick 2013	FFQ Follow-up	887	Female not o/w at start	25.4	2,045	*8,588*	n/a	n/a	n/a	n/a	n/a	n/a
RAINE Study	Ambrosini 2013	FFQ Baseline	1,632	Girls & Boys	14.0	*2,326*	*9,731*	n/a	n/a	n/a	n/a	n/a	n/a

Table 4-5 Comparison of variance in reported mean Energy intakes by DAT

DAT	Cohort	Paper	Percentage of SD/mean Energy intake for the whole cohort or across categories	
			Minimum	Maximum
3 day Food diary	ALSPAC	Noel 2013	23%	25%
	ALSPAC	Bigornia 201	16%	20%
	DONALD	Cheng 2009	21%	24%
	DONALD	Libuda 2008	19%	21%
	DONALD	Alexy 2011	26%	33%
	FAMS girls	St Jules 2014	n/a	n/a
	Framingham	Hasnain 2014	10%	17%
	NGHS girls	Berz 2011	15%	23%
	NGHS girls	Albertson 2007	21%	
	NGHS girls	Ritchie 2007	17%	24%
3 x 24 hour recall	BDPP	Balvin Frantzen 2013	33%	45%
	IDEA & ECHO	Laska 2012	28%	32%
1 x 24 hour recall	Bogalusa	O'Neil 2013	38%	46%
	EYHS	Zheng 2015	25%	
	EYHS	Zheng 2014	22%	33%
212 item FFQ	RAINE	Ambrosini 2013	31%	33%
131/132 item YAQ	GUTS	Field 2003	31%	32%
	GUTS	Taveras 2005	31%	33%
149 item YAQ	Project EAT	Fulkerson 2008	48%	
151 item YAQ	Project EAT	Quick 2013	43%	50%
Willett's FFQ	Project EAT	Quick 2013	n/a	n/a
38 item FFQ	Bogotá	Shroff 2014	30%	54%

Figure 4-3 Reported mean Energy intakes sorted by age

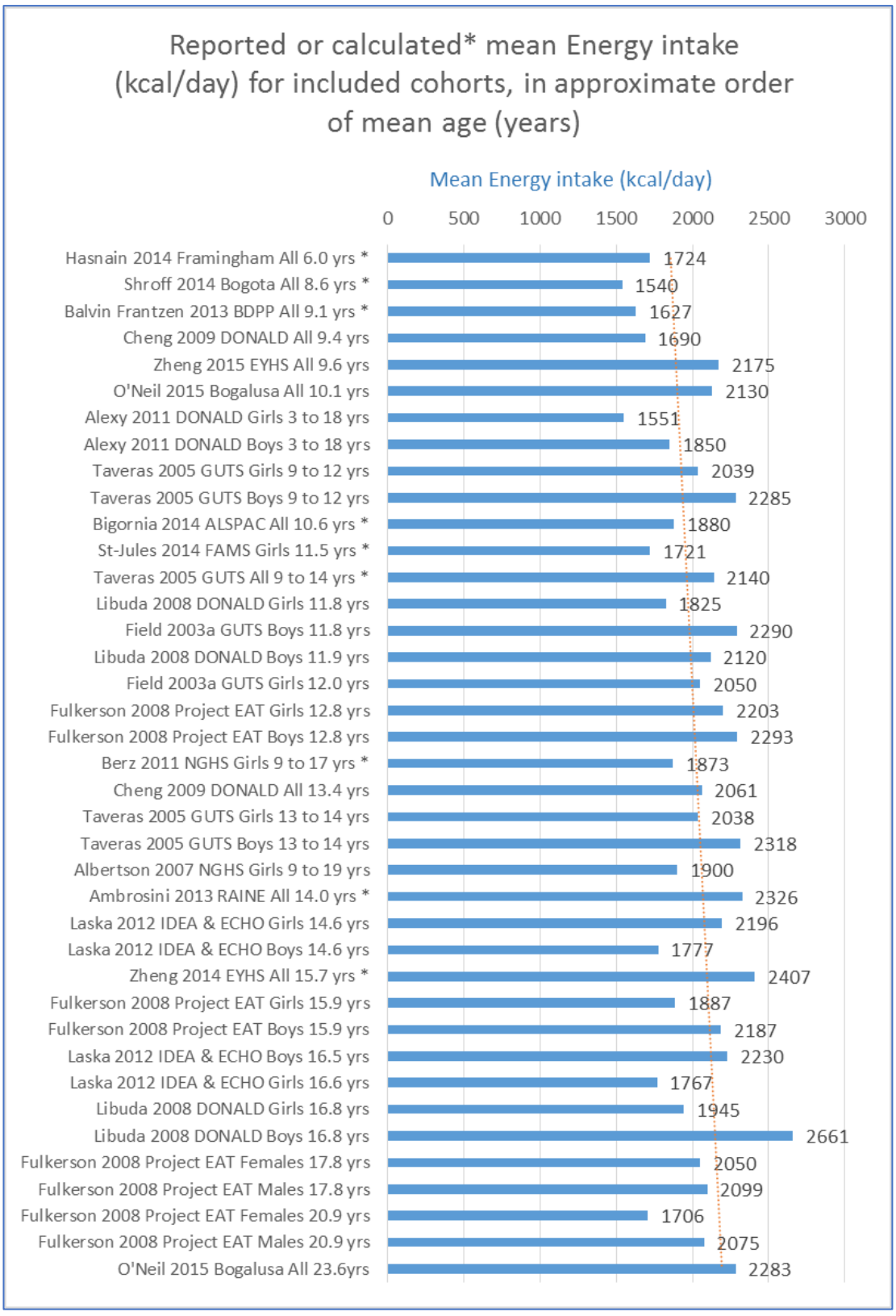


Figure 4-4 Mean Energy intakes at baseline by age and sex of included cohorts

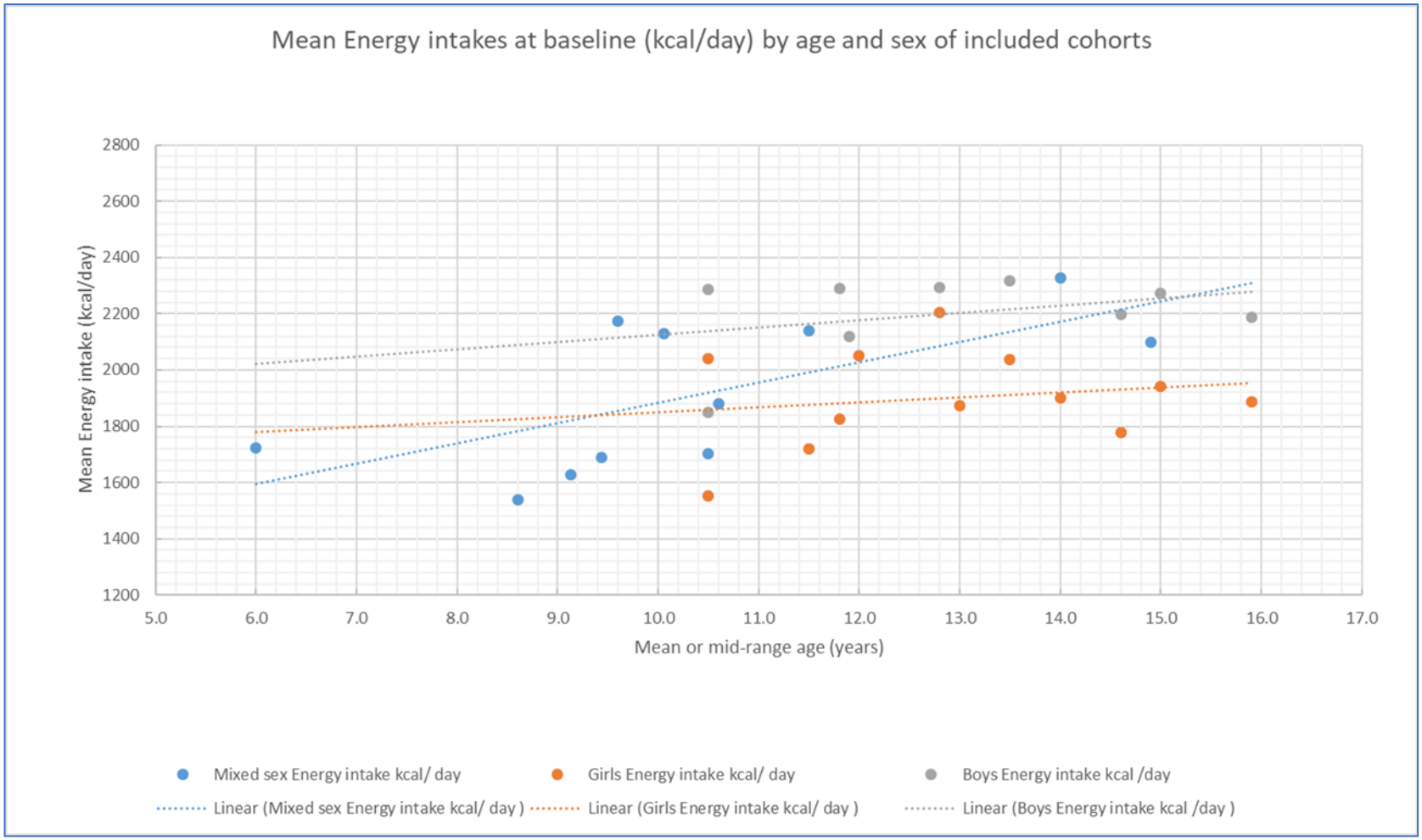


Figure 4-5 Mean Energy intakes at baseline by age and geography of included cohort

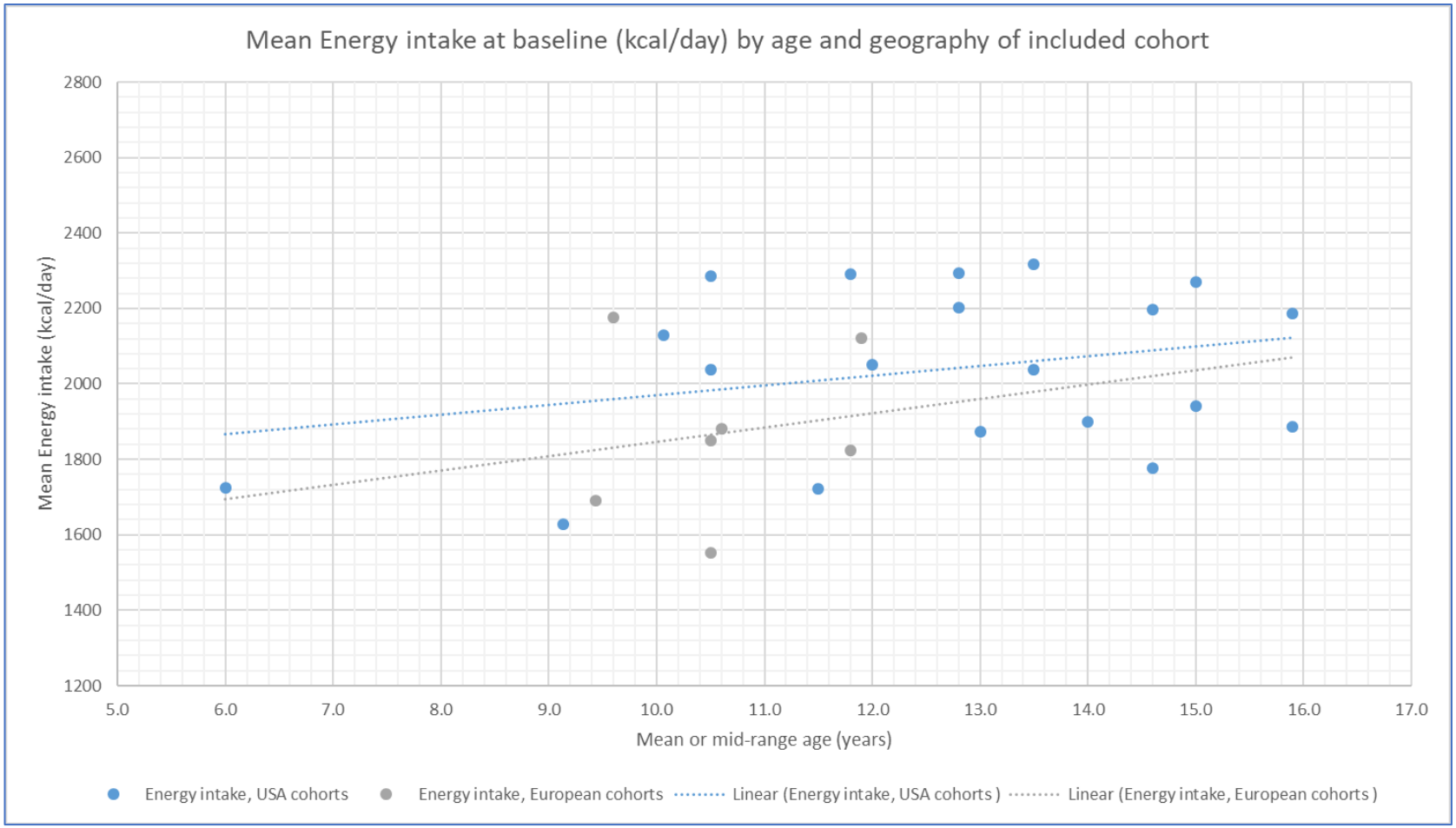
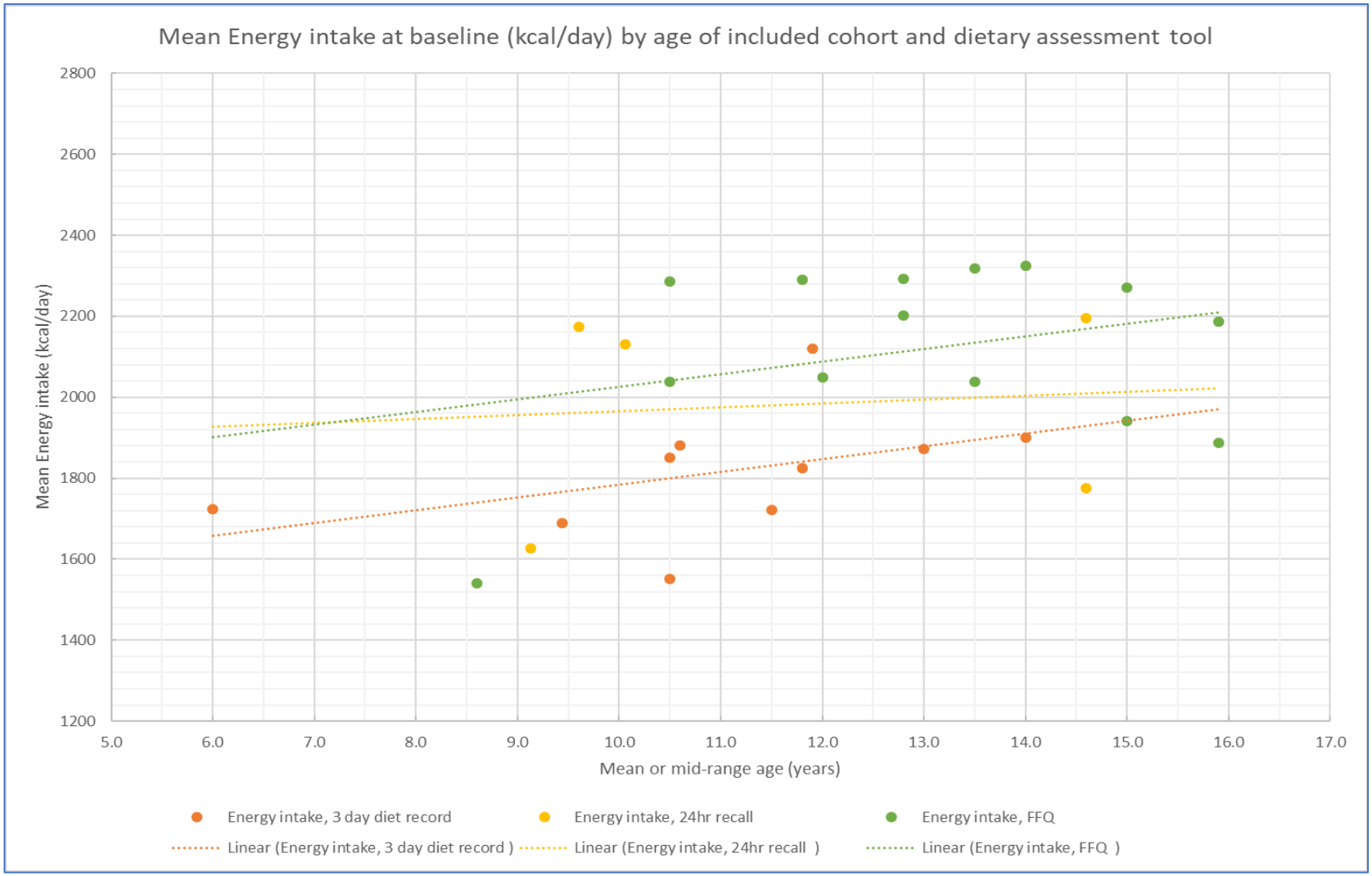


Figure 4-6 Mean Energy intakes at baseline by age of included cohort and dietary assessment tool



Most reported baseline mean energy intakes by age were similar (within ± 200 kcal/day) to UK estimated average requirements (EAR) in pre-teenage children, with the exception of Danish children from EYHS, whose baseline mean energy intake was almost 400 kcal/day more than UK EAR, and Hawaiian girls from FAMS, whose baseline mean energy intake was more than 300 kcal/day less than UK EAR.

In older teenaged children reported baseline mean energy intakes were much less than UK EAR, which may be indicative of under-reporting by older children. See Table 4-6 below.

Table 4-6 UK age specific estimated average requirements (EAR) and baseline mean energy intakes

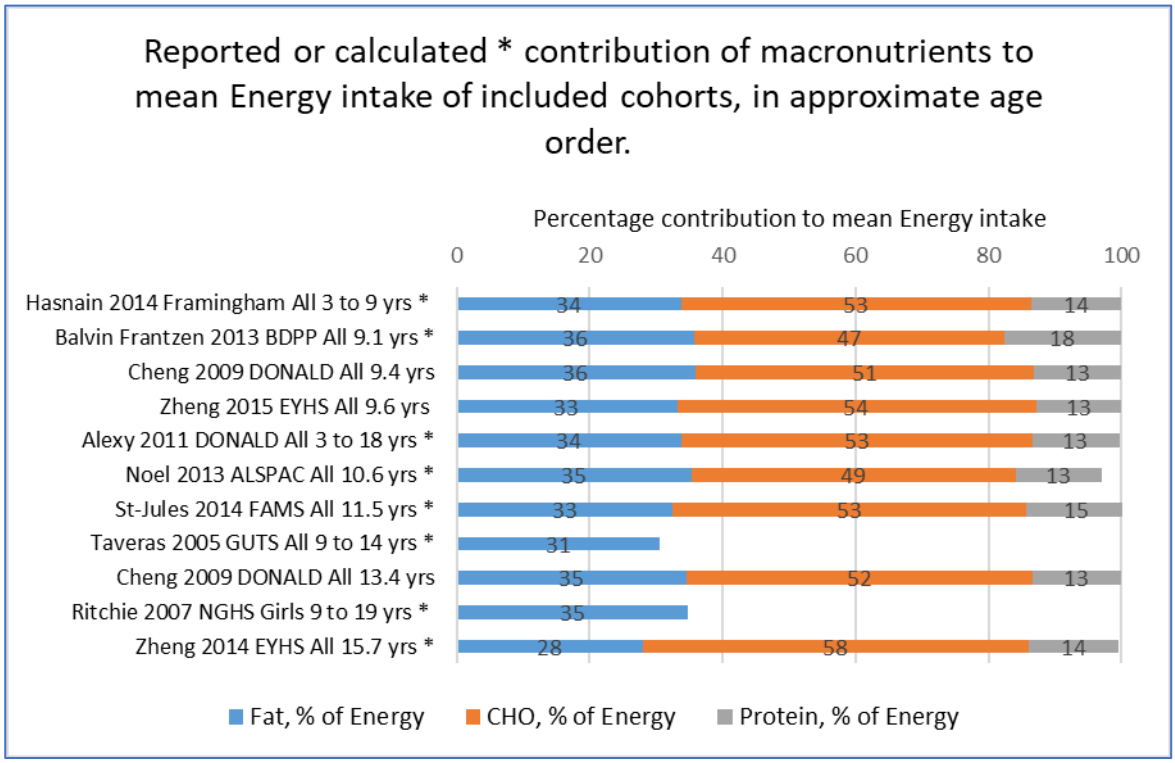
Cohort, paper, sex and age	UK age-specific EAR (kcal/day)	Mean Energy intake (kcal/day)	Difference (kcal/day)
Framingham Children, Hasnain 2014, All 6.0 yrs.	1,530	1,724	195
Bogotá Schoolchildren, Shroff 2014, All 8.6 yrs.	1,685	1,540	-145
BDPP, Balvin Frantzen 2013, All 9.1 yrs.	1,781	1,627	-153
DONALD study, Cheng 2009, All 9.4 yrs.	1,781	1,690	-91
EYHS, Zheng 2015, All 9.6 yrs.	1,781	2,175	394
Bogalusa Heart Study, O'Neil 2015, All 10.1 yrs.	1,984	2,130	146
ALSPAC, Bigornia 2014, All 10.6 yrs.	1,984	1,880	-104
FAMS, St. Jules 2014, Girls only 11.5 yrs.	2,032	1,721	-311
GUTS, Field 2003a, Boys 11.8 yrs.	2,127	2,290	163
GUTS, Field 2003a, Girls 12.0 yrs.	2,103	2,050	-53
Project EAT, Fulkerson 2008, Girls 12.8 yrs.	2,103	2,203	100
Project EAT, Fulkerson 2008, Boys 12.8 yrs.	2,247	2,293	46
RAINE study, Ambrosini 2013, All 14.0 yrs.	2,486	2,326	-160
IDEA & ECHO, Laska 2012, Girls 14.6 yrs.	2,342	2,196	-146
IDEA & ECHO, Laska 2012, Boys 14.6 yrs.	2,629	1,777	-853
Project EAT, Fulkerson 2008, Girls 15.9 yrs.	2,390	1,887	-503
Project EAT, Fulkerson 2008, Boys 15.9 yrs.	2,820	2,187	-633

4.7.3 Results: Macronutrients as a percentage of energy intake

Mean intakes of selected macronutrients were reported in 3 European and 5 USA cohorts, as presented in Table 4-4. Reported values for variance are not replicated in this summary table. Apart from the GUTS cohort, which used a semi-quantitative FFQ (Taveras et al., 2005), all the other cohorts which reported macronutrient intakes measured diet with a 3 day FD or 24-HDR. Papers from GUTS and NGHS reported only fat intakes. Papers from the other six cohorts (ALSPAC, BDPP, DONALD, EYHS, FAMS and the Framingham Children's study) provided information about at least two of three macronutrients (Fat, Carbohydrate, Protein) from which the contribution of each macronutrient to total energy intake (TEI) could be estimated. A stacked bar chart showing each macronutrient's share of TEI at different ages, paints a consistent picture in the eight cohorts. See Figure 4-7. In the cohorts where all three macronutrients could be plotted, on average fat made up a third of TEI. Protein typically accounted for 13% to 15% of TEI and carbohydrate contributed the rest.

The Framingham Children's study reported only fat and protein intakes; their respective contributions to TEI were calculated and carbohydrate was assumed to make up the balance. Similarly, the BDPP only reported fat and carbohydrate. Protein's assumed 18% share of TEI to bring the total to 100% may be over-estimated. Contributions of each macronutrient to TEI in the ALSPAC cohort were derived from categorical dietary intakes, and do not add up to 100% due to rounding.

Figure 4-7 Reported macronutrients as an approximate percentage of energy intake at baseline



The World Health Organization has published population nutrient intake goals of how much fat and carbohydrate (as a percentage range) should contribute to total energy intake (WHO/FAO, 2003):

- Total fat 15% to 30% of TEI
- Saturated fatty acids < 10% of TEI
- Total carbohydrate 55% to 75% of TEI
- Free sugars < 10 % of TEI

The mean macronutrient values highlighted in red in Table 4-4 do not meet WHO recommendations. In the cohorts which reported fat intakes (ALSPAC, BDPP, DONALD, EYHS, FAMS, Framingham Children's study, GUTS and NGHS) mean intakes exceeded the 30% recommended maximum, suggesting that at least half the children consumed too much fat. In some cohorts (ALSPAC, DONALD, Framingham Children's study), based on reported standard deviations in individual papers, as many as 5 out of 6 children were consuming too much fat. The exception was older teenagers from the Danish EYHS cohort, who reported consuming 28% of their TEI as fat on average. Where reported (ALSPAC, BDPP, DONALD, GUTS and NGHS) mean intakes of saturated fat also exceeded the intake goal of < 10% of TEI. Mean intakes of saturated fat were particularly high in the German DONALD cohort, averaging ~15% of TEI at all ages.

Papers from five cohorts (ALSPAC, BDPP, DONALD, EYHS and FAMS) reported carbohydrate intakes. Almost all mean intakes of carbohydrate fell slightly below the recommended minimum of 55% of TEI. Reported SDs suggest that only the top sixth of each cohort did meet the recommended minimum of carbohydrate intake. Again, the exception was older teenagers in the Danish EYHS cohort whose average carbohydrate intake was 58% of TEI, meeting the lower end of WHO recommendations, although their corresponding mean intakes of added sugars (22% of TEI) exceeded the goal for free sugars of < 10% of TEI. Reported mean intakes of added sugars also exceeded this goal in the DONALD (13.6% of TEI) and FAMS (16.5% of TEI) cohorts. Added sugars were not reported separately for the other cohorts.

4.7.4 Results: Fibre intakes

Intakes of fibre were reported in 3 European (ALSPAC, DONALD and EYHS) and 3 USA cohorts (BDPP, GUTS and NGHS). All mean fibre intake values shown in the final column of Table 4-4 are highlighted in red, as they fall below current UK age recommendations.

The World Health Organization's recommended intake of fruit and vegetables and whole grain foods (WHO/FAO, 2003) provides > 20g/day of non-starch polysaccharides (NSP), equivalent to > 25g/day of total dietary fibre, a goal primarily intended for adults. In the UK in 2015 the

Scientific Advisory Committee on Nutrition (SACN) recommended that adult population intakes of dietary fibre (measured by the AOAC method) should be 30g/day (PHE, 2015b).

SACN also recommended intakes of dietary fibre for children, dependent on age:

- 6 months to 2 years old. Gradual introduction of whole grains, fruits and vegetables.
- 2 to 5 year olds ~ 15 g/day.
- 5 to 11 year olds ~ 20g/day.
- 11 to 16 year olds ~ 25g/day.
- 16 to 18 year olds ~ 30g/day (equivalent to adult recommendation).

The highest reported mean intakes of dietary fibre were in the German DONALD study (21.3g/day, SD 7.0g/day at age 13 years), which corresponds with the highest reported mean intakes of whole grains, see Chapter 5, section 5.3.1. The next highest mean intakes of dietary fibre were reported by the Danish EYHS cohort (18.6g/day, SD 7.6g/day at age 9 years) and the USA GUTS cohort (16.8g/day, SE 0.07g/day at ages 9 to 14 years). Reported ranges indicate that about one sixth of children in each of these cohorts had adequate levels of dietary fibre intake. However, reported mean dietary fibre intakes in the other cohorts (ALSPAC, BDPP, NGHS) were considerably lower, between 10g and 13 g/day, with standard deviations which suggest that less than 3% of children/teenagers in these cohorts met the UK recommended intake of dietary fibre for their age group.

4.7.5 Results: BMI and overweight/obesity

Body Mass Index was reported in at least one paper from every cohort except FAMS. Reported mean BMI from included cohorts (including baseline, repeat and follow-up measures) sorted by age are summarised in Figure 4-8. The trendline on the plot shows increasing BMI through childhood and adolescence to young adulthood, as expected. Most mean BMI values lie slightly above median BMI in the UK 1990 reference (See Figure 4-2 again) but are plausible.

It is not possible to gauge the proportion of children in each cohort who were overweight, based on mean BMI alone. However, reported mean BMI for young adults in NGHS, Project EAT, and the Bogalusa Heart study lie close to or slightly above the accepted adult cut-off for overweight (25 kg/m^2). This shows that many participants in these three cohorts were overweight by adulthood.

Childhood growth reference and adult cut-offs were used to classify overweight/obesity in childhood/adolescence and in young adulthood in eight cohorts. Overweight/obesity was reported at baseline in the GUTS and GUTS II cohorts, but at baseline and follow-up in papers

from ALSPAC, the Bogalusa Heart Study, DONALD, NGHS girls, Project EAT and the RAINE study. Cohorts chose different growth references to classify overweight/obesity and in some cases, papers from the same cohort study used different growth references.

Prevalence of overweight/obesity, as reported by papers from each cohort in order of age, is summarised in Figure 4-9. For each paper, the growth reference or cut-off used to classify overweight/obesity is listed in the data label.

Whichever growth reference was used, cohorts with younger children generally had a lower prevalence of overweight/obesity (approximately 1 in 5) than the cohorts with older adolescents (approximately 1 in 4). Studies that surveyed young adults at follow-up reported the highest prevalence of overweight/obesity (approaching 1 in 2).

The lowest overweight/obesity prevalence (~ 10% or less) was in the DONALD cohort, based on German national reference data with an overweight cut-off at the 90th percentile. Using the same German reference data but with the IOTF overweight cut-off at the 85th percentile, reported overweight/obesity prevalence for 9 year olds from the DONALD cohort was much higher at ~ 17%, only slightly less than the ~ 20% overweight/obesity prevalence reported for 10 year old children in the UK ALSPAC cohort. Approximately 20% of 9 to 15 year olds in the USA GUTS II cohort were classified as overweight/obese, using the same IOTF reference.

The lowest overweight/obesity prevalence (~ 10% or less) was in the DONALD cohort, based on German national reference data with an overweight cut-off at the 90th percentile. Using the same German reference data but with the IOTF overweight cut-off at the 85th percentile, reported overweight/obesity prevalence for 9 year olds from the DONALD cohort was much higher at ~ 17%, only slightly less than the ~ 20% overweight/obesity prevalence reported for 10 year old children in the UK ALSPAC cohort. Approximately 20% of 9 to 15 year olds in the USA GUTS II cohort were classified as overweight/obese, using the same IOTF reference.

Most USA cohorts applied the CDC 2000 reference, which classified 22% of 10 year olds from the USA Bogalusa Heart study as overweight/obese, and ~ 20% of 12 year old boys and girls from the USA GUTS cohort as overweight/obese. Overweight/obesity may be under-estimated in GUTS II and GUTS due to their reliance on self-reported height and weight. Overweight/obesity prevalence based on measured height and weight was much higher (28%) among 9/10 year old black and white girls from the USA NGHS.

Figure 4-8 Cohorts' mean BMI in order of age

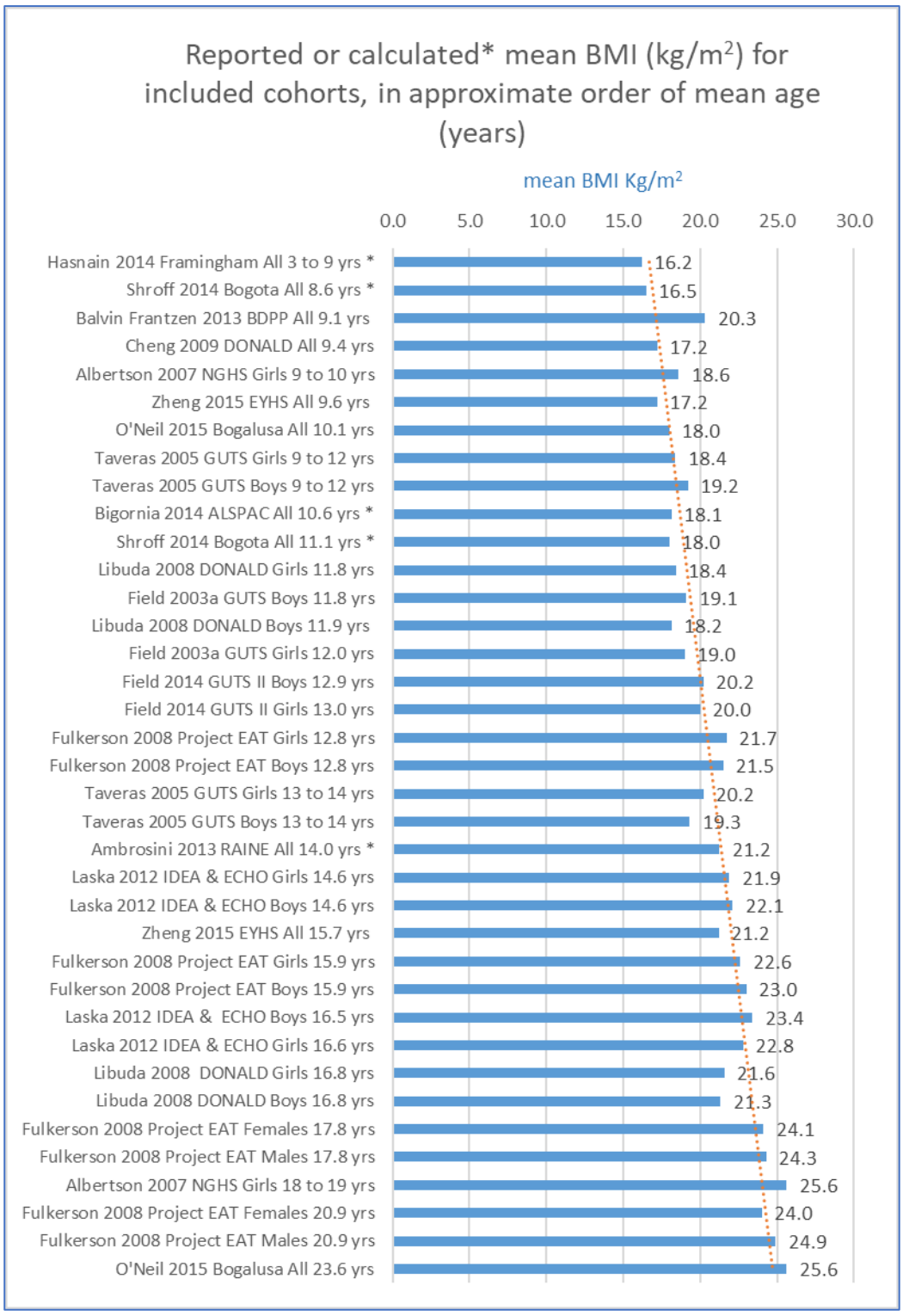
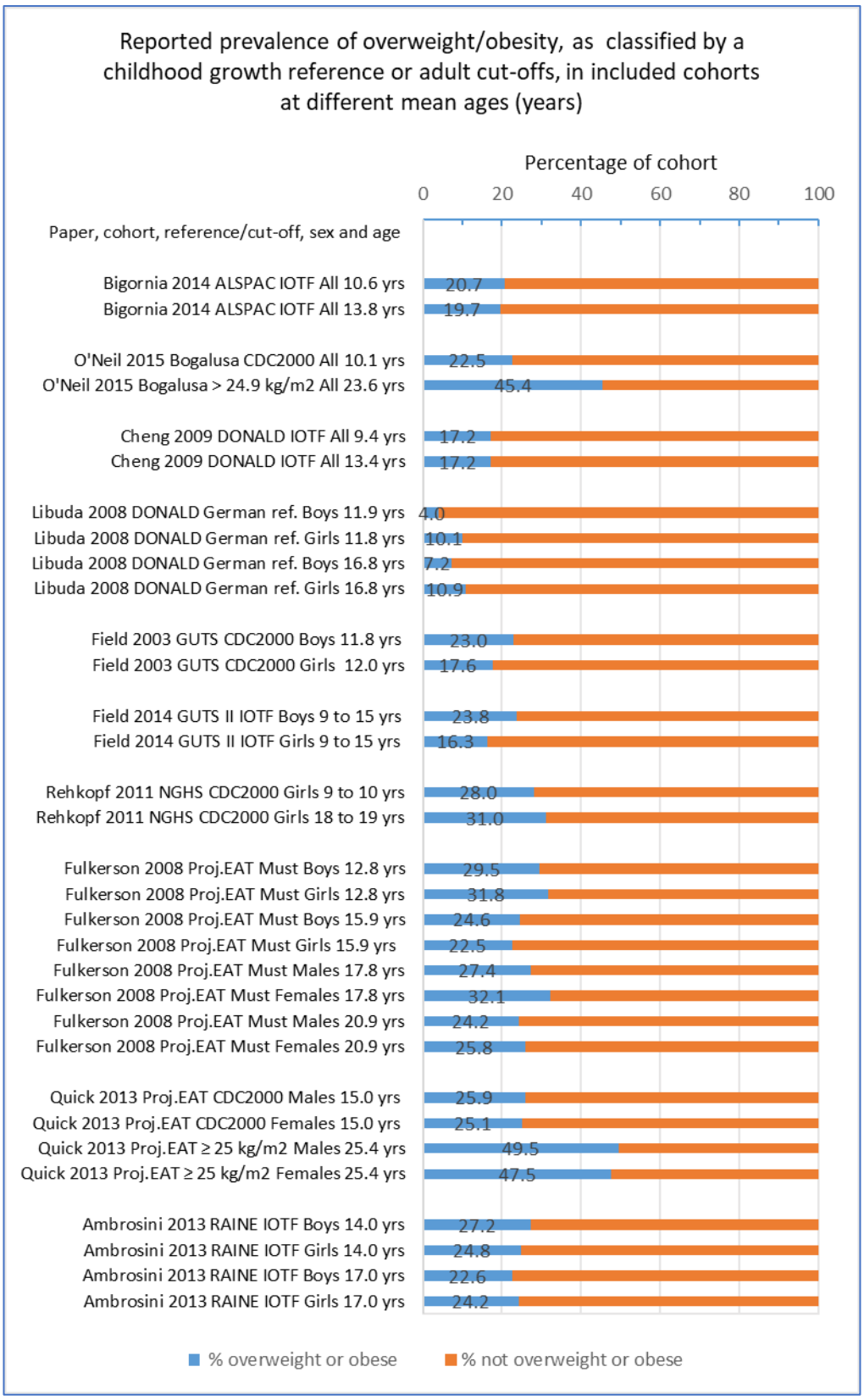


Figure 4-9 Reported prevalence of overweight/obesity (%) by cohort and age.



In older cohorts, such as the Australian RAINE study, overweight/obesity prevalence in teenagers aged 14 years was approximately 26% based on the IOTF reference. Similarly, in the USA Project EAT cohort, overweight/obesity prevalence in 15 year olds was 25% based on the CDC 2000 reference. When Must's growth reference was applied in Project EAT at almost 16 years old, a similar proportion of adolescents were classified as overweight/obese.
Two USA cohorts, the Bogalusa Heart study and Project EAT, used the adult cut off for overweight (BMI ≥ 25kg/m^2), finding that over 45% of young adults were overweight.

Although the prevalence of overweight/obesity tended to be higher in older-aged cohorts, overweight/obesity prevalence in each cohort did not always increase between childhood and adolescence. The prevalence of overweight/obesity fell during teenage for boys in the RAINE study yet stayed constant for teenage girls and did not change very much in the DONALD study or the ALSPAC cohort. In Project EAT, overweight/obesity prevalence fluctuated depending on baseline age and sex. In girls from the NGHS cohort, overweight/obesity prevalence increased between the ages of 9/10 years old 18/19 years old, from 28% to 31%.
Overweight/obesity prevalence between adolescence and young adulthood doubled in the Bogalusa Heart Study and in Project EAT. In part this dramatic rise may be due to the mismatch between the CDC 2000 growth reference and adult cut-offs at the 20 year age boundary.

4.8 Discussion

This examination of dietary assessment methods and adiposity measures employed by included cohorts, confirms the findings of the Newcastle-Ottawa quality assessment. None of the included cohort studies provide the highest quality evidence (due to the inherent bias in subjective dietary assessment) but the better quality studies (ALSPAC, IDEA and ECHO and the NGHS) reported methods well, including steps to reduce bias and measurement error. Lower quality studies (Project EAT, GUTS and GUTS II) were less well executed and sometimes less well reported. Never-the-less all papers offer useful insights about dietary exposures and adiposity outcomes in children.

Most included papers cited study protocols or carefully described their methods of measuring diet and adiposity and the steps taken to minimise measurement error/bias. In some instances, key items were omitted. For example, several papers did not give dates of study measures, few papers used summarised participant numbers or loss to follow-up clearly, and although quantitative DATs were used, portion sizes were not always stated. Some of these unreported items were later recommended by the Strengthening the Reporting of

Observational Studies in Epidemiology statement (STROBE) (Vandenbroucke et al., 2014) or by the extension statement for nutritional epidemiology (STROBE-nut) (Lachat et al., 2016).

Included cohorts were mainly established in the 1980s, 1990s and early 2000s and used the paper-based methods of dietary assessment available then. Only the IDEA & ECHO study established in 2006/7 used a 24-HDR with direct data entry linked to a nutrient database. Other cohorts relied upon manual coding of FD, 24-HDR and FFQs by trained research staff.

Baseline mean energy intakes in each cohort were broadly in line with UK estimated average requirements, indicating that average measures of diet were plausible for younger children, but with some signs of under-reporting by older teenagers. Even with parental help, the food diaries of over half of the 10 year old children in ALSPAC under or over-reported TEI and in a 24 HR used by the EYHS almost 10% of 9 year olds under reported TEI. In the RAINE study, an FFQ returned implausible reports of TEI from over a third of teenagers. This demonstrates how difficult it is to measure children's diets and the need for better methods.

In recent times large food composition databases have been built and linked to on-line DATs, allowing direct data entry and rapid nutritional assessment of reported dietary intakes. This greatly reduces the researcher burden (Carter et al., 2015; Carter et al., 2016). One on-line 24-HDR (myfood24), when compared with a traditional multiple pass 24-HDR, had comparable validity against biomarkers (Wark et al., 2018) and showed good agreement when tested in British adolescents (Albar et al., 2016). Other technology has also come to the fore, making use of computers, the internet, smartphones, cameras and barcode scanning for more objective dietary assessment (Burley et al., 2015 , Cade, 2017). It is hoped that these technologies will lower the respondent burden (as well as that of the researcher) and improve measurement accuracy, but more validation studies (ideally against reference measures such as doubly labelled water or biomarkers, rather than food diaries or 24-HDR) are needed (Lachat et al., 2016). All DATs contain sources of error and some biases will remain.

In the cohorts that reported mean energy intakes by sex, boys tended to consume more energy than girls at the same age. There is a strong argument for disaggregating data and conducting analyses separately by sex, if the sample size is large enough. However, many cohorts reported and analysed both sexes together, even after puberty when differences in growth by sex are more apparent. Many children and adolescents had diets that were high in fat, exceeding recommendations. This is consistent with a review of children's diets in Europe (Lambert et al., 2004) which found that in many countries, including the U.K., children obtained over 40% of their energy from fat.

A range of adiposity measures were reported. Skin fold thicknesses and DXA were helpful for comparing percent body fat outcomes within a cohort, but due to methodological heterogeneity it was not possible to make direct comparisons between cohorts. Many cohorts measured waist circumference, but this measure was only useful for comparing adiposity outcomes within a cohort, as no sex and age specific growth reference for WC has been established. Surprisingly, only one paper (Albertson et al., 2009) used waist-to-height ratio, although many papers included in this systematic review were published after waist-to-height ratio was shown to be useful in UK children (McCarthy and Ashwell, 2006). In future, childhood cohorts may decide to use waist-to-height ratio in preference to SFT or DXA, as it is an easily obtained measure of central adiposity that can be compared across different studies.

Most cohorts took measures of height and weight from which BMI could be calculated. Mean BMIs reported at different ages in different cohorts were plausible, although they were usually above the median BMI for age shown in the UK 1990 growth reference curves. Three cohorts (GUTS, GUTS II and Project EAT) relied on self-reports of height and weight. It has long been recognised that adults exaggerate their height and under report weight and are more likely to under-report if they are taller/heavier and to over-report if they are shorter/thinner (Schlichting et al., 1981). Similar tendencies have been observed in adolescents from Germany (Brettschneider et al., 2011) and from Wales (Elgar et al., 2005), causing an under estimation of BMI, particularly among teenagers with overweight and obesity. In younger school children (11 to 13 years) from the USA, boys and girls also tended to under-report their weight, with a bias towards under-reporting among the taller, heavier children (Shannon et al., 1991). Self-reported heights deviated considerably from measured values. 10% of children gave implausible answers or no response at all. Shannon et al speculated that the lack of a quantitative perception of height and weight might be due a growth spurt that happened too quickly for the children to keep up-to-date. In Project EAT a validation exercise conducted at the 10 year follow-up with a sub-sample of participants (63 male and 62 female) found that correlations between BMI based on self-reported height and weight versus measured height and weight improved in adulthood compared with adolescence (Berge et al., 2015). However, there was evidence of systematic measurement error, related to age, ethnicity and SES; the authors concluded that objective measures of height and weight were preferable, particularly at younger ages. In contrast the GUTS and GUTS II studies argued that self-reports were cost-effective measures that classified adolescent overweight/obesity correctly most of the time, based on validation studies in other cohorts. Internal validation of self-reported versus measured height and weight in GUTS and GUTS II would have been preferable, given the bias and random error seen elsewhere.

Adiposity outcome measures (at follow-up or change) were not always well chosen. In the Framingham Children's study, girls and boys were aged 3 to 5 years at enrolment and BMI at follow-up was the outcome measure (Hasnain et al., 2014). Following advice (Cole et al., 2005) BMIz at follow-up may have been a better option, especially as children were not all the same age at measurement. GUTS used change in BMIz as the outcome, although their own study group later recommended change in BMI as easier to understand (Berkey and Colditz, 2006). Many studies did use change in BMI, which is logical when children are the same age and sex, as with NGHS girls. Papers from the DONALD study considered dietary exposure at baseline and change in BMI-SDS, also giving an annualised "concurrent estimate" of change in dietary exposure and change in BMI-SDS, which recognises that there are changes in the exposure between baseline and follow-up which may influence the outcome. The NGHS, Project EAT and the RAINE study used categorical adiposity outcomes but applied different growth references, making it hard to compare levels of risk.

Using different instruments to measure diet and adiposity at follow-up in adulthood compared with baseline in childhood caused difficulties in Project EAT. Measuring diet with the YAQ at baseline and Willett's adult FFQ at follow-up did not give a valid measure of absolute change in dietary intakes (Larson et al., 2012), but this was overlooked by Quick et al. Bridging the transition between childhood and adult definitions of overweight was also a challenge, resolved by using Must's growth reference (Cutler et al., 2012; Fulkerson et al., 2008). The IOTF growth reference would have been a better choice, with greater utility for comparing overweight/obesity prevalence with other cohorts.

Chapter 5 Narrative review: quantified food and drink intakes and subsequent adiposity.

5.1 Summary

Quantified intakes of specific foods or drinks and later adiposity outcomes were investigated by over 60% of included papers, across all cohorts except the BDPP. Investigated foods and drinks were:

- whole grains
- dairy foods including milk as a beverage
- fruit and vegetables including juice as a beverage
- fish
- energy dense convenience foods
- sweet and savoury snack foods
- sugar sweetened and diet beverages

This chapter (Chapter 5) narratively reviews each food/drink in turn, firstly explaining how the study defined the dietary exposure and comparing reported quantified intakes by age, sex and cohort/nationality and trends over time, before focussing on the reported longitudinal associations (if any) between that specific dietary exposure and future adiposity. In the qualitative synthesis of reported associations, summarised in tables, model adjustment for energy intake was considered, as well as sample sizes and study quality (based on the Newcastle Ottawa score).

Studies concurred that higher intakes of whole grains were beneficial and higher intakes of sugar-sweetened beverages were adverse in terms of adiposity outcomes, but few studies agreed on significance or direction of influence for other dietary exposures. Most exposures were investigated in only a few cohorts, which restricted opportunities for quantitative synthesis.

5.2 Introduction

The 35 included papers reported a diverse range of dietary exposures. Their findings are organised according to whether papers focussed on quantified intakes of similar individual foods and drinks (Chapter 5) or whether they took a broader perspective, investigating diet quality, dietary patterns, eating habits or multiple dietary and non-dietary predictors of overweight and obesity (Chapter 6). Findings overlap as individual foods and drinks featured in dietary patterns or diet quality scores or were investigated in conjunction with eating behaviours. Intakes of specific foods and drinks and eating habits were also examined alongside many other potential overweight/obesity risk predictors.

Almost two thirds of included papers investigated associations between quantified intakes of specific drinks, foods or food groups and adiposity outcomes, with reported intakes based on food diaries, 24 hour diet recalls or semi-quantitative food frequency questionnaires.
In the larger USA cohorts (**NGHS**, **Project EAT** and **GUTS**) many different foods and drinks were examined and authors presented separate papers for each. Similar foods and food groups were usually investigated in papers from two or three different cohorts, but sugar sweetened beverages (SSB) were examined in ten different cohorts, sometimes in comparison with other beverages such as milk, juice or diet soda. This presented some opportunity for a meta-analysis of the results for SSBs, which is presented in Chapter 7.

In order to answer the research question, "To what extent does diet during childhood or adolescence influence future indicators of overweight or obesity?", the focus of the narrative is the reported **longitudinal associations** between **dietary exposures and adiposity outcomes**. Longitudinal findings are described and summarised in tables which show the sample size in analyses. Some studies also reported cross-sectional associations between diet and adiposity, which are mentioned in the narrative.

Approaches to longitudinal analysis varied. Exposures were considered either at baseline and/or as change over time, with adiposity outcomes considered at follow-up and/or as change over time. Most, but not all, studies adjusted for energy intake as well as other confounders. Only one paper (Quick et al., 2013), from Project EAT, excluded participants who already had overweight or obesity at baseline from their longitudinal analyses. Other studies included all participants, adjusting for baseline weight status in analyses, but reverse causality cannot be ruled out: Is the dietary exposure associated with risk of future overweight or does pre-existing overweight influence the individual to modify their dietary exposure?

Table 5-1 Cohort population characteristics, diet and adiposity measures

Cohort	Country	Est.	Cohort size	Sex	Age (yrs.)	D.A.T.	Height & weight	Growth ref	BMI	BMIz	o/w or obese	WC	SFT	%BF
ALSPAC	UK	1990/92	14,701	Girls & Boys	7	**3 day FD**	Measured	IOTF, CDC	√					DXA
BDPP	USA	2001	706	Girls & Boys	9	3 x 24-HDR	Measured	CDC	√					
Bogalusa	USA	1973	4,000	Girls & Boys	10	1 x 24-HDR	Measured	CDC	√			√	√	
Bogotá	Colombia	2006	3,202	Girls & Boys	5 to 12	***FFQ***	Measured	WHO	√			√	√	
DONALD	Germany	1985	1,400	Girls & Boys	3 to 18	**Weighed 3 day FD**	Measured	German national		√			√	
EYHS	Denmark	1997	590	Girls & Boys	9	1 x 24-HDR	Measured	LMS Cole (IOTF)	√	√		√	√	
FAMS	USA	2000	349	***Girls only***	9 to 14	**3 day FD**	Measured	CDC		√		√	√	DXA
Framingham	USA	1987	103	Girls & Boys	3 to 9	**3 day FD**	Measured	Not shown	√			√	√	DXA
GUTS	USA	1996	16,882	Girls & Boys	9 to 14	***FFQ***	***Self-report***	CDC	√	√				
GUTS II	USA	2004	10,919	Girls & Boys	9 to 15	***FFQ***	***Self-report***	IOTF	√					
IDEA & ECHO	USA	2006/7	723	Girls & Boys	9 to 15	3 x 24-HDR	Measured	Not shown	√					Bio-imp
NGHS	USA	1987	2,379	***Girls only***	9 to 10	**3 day FD**	Measured	CDC	√	√	√	√	√	Bio-imp
Project EAT	USA	1998	4,746	Girls & Boys	15	***FFQ***	***Self-report***	CDC, Must			√			
RAINE Study	Australia	1989/91	2,868	Girls & Boys	14	***FFQ***	Measured	IOTF	√		√	√		

The 14 included cohorts are outlined in Chapter 3 and fully described in Appendix E. For reference a summary of each cohort's characteristics is presented in Table 5-1 showing country of origin and when each cohort was established, cohort size, sex of children in the cohort with the mean age or age range at baseline, the dietary assessment tool(s) used, plus the selected growth reference and reported measures of adiposity. Note that the NHLBI Growth and Heath Study (NGHS) and the Female Adult Maturation Study (FAMS) are single sex cohorts of girls only.

5.3 Whole grains

Whole grain intakes were investigated as part of carbohydrate quality in the **DONALD** cohort (Cheng et al., 2009), as an element of diet quality in the **NGHS** (Berz et al., 2011) and as one of many predictors of overweight/ obesity in **Project EAT** (Quick et al., 2013). Longitudinal findings for intakes of whole grains and adiposity outcomes are summarised in Table 5-2

5.3.1 Whole grain intakes

In the **DONALD** study a multitude of whole grains were considered including amaranth, barley, buckwheat, bulgur, popcorn and corn bran, oats, millet, kamut wheat, whole rice, rye, spelt, triticale, whole wheat, wheat germ and wheat bran. Whole grain intakes were investigated at puberty take-off and for the next 4 years. Puberty take off (mean age 9.4 years S.D. 1.2 years) occurred between 1988 and 2003 for 215 children in the rolling cohort who had complete nutritional and anthropometric data. This cohort had the highest reported intakes of whole grain. Mean intake increased with age, from 62.4g/ day (SD 28.9) at puberty to 71.6g/ day (SD 34.4) at 4 year follow-up. Most children consumed at least some whole grain.

For their modified Dietary Approaches to Stop Hypertension (DASH) diet quality score the **NGHS** used the definition set out in the 2005 Dietary Guidelines for Americans, which describes whole grains as foods "made from the entire grain seed (kernel), which consists of the bran, germ, and endosperm." Whole grain intake among NGHS girls, averaged over approximately 20 days between the ages of 9 and 17 years old, was low. Based on a serving size of 28g, the 50th percentile of intake was 0.5 servings/ day. Even girls at the 95th percentile of intake failed to meet the recommended 3 servings per day, consuming just 1.2 servings/ day or 34g/day, barely half that consumed by German adolescents from the DONALD study.

Whole grains were considered as a potential predictor of obesity in the **Project EAT** cohort but were not defined. Baseline intakes of whole grain measured were given only for participants *not overweight at baseline*. Average whole grain intakes were approximately 1 serving per day in normal weight males and females (mean age 15.0 years). Intake increased by age 25 years, possibly doubling, although intakes at baseline and follow-up were measured with different FFQs, so the absolute change is uncertain.

5.3.2 Whole grain intakes and adiposity outcomes

The two USA studies found that higher intakes of whole grains had some protective effect against future adiposity for adolescent girls.

In the **NGHS**, girls in the highest category of whole grain intake (≥ 1 serving per day) had lower annual gains in BMI and hence the smallest mean BMI (24.1) at final follow-up.

In **Project EAT**, participants were only included in analyses if they were *not overweight at baseline*. For adolescent females, each extra serving per day of whole grains at baseline reduced the risk of overweight ten years later by 29%. Baseline whole grain consumption did not predict overweight in males not overweight at baseline. The measure of change in whole grain intake was uncertain, and not significantly associated with future overweight in males or females.

The German **DONALD** study had the shortest follow-up. Whole grain intakes were twice that in the USA cohorts and overweight prevalence was lower (17%). No cross-sectional relationships between whole grain intakes and adiposity were apparent and changes in whole grain intakes (g/day) were not linked with concurrent changes in body fat % or BMI-SD scores in any model, including those which adjusted for residual energy intake.

Table 5-2 Whole grain intakes and adiposity outcomes

Study & Paper	NOS stars	Population	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline & f' up	Adiposity outcomes	Findings
DONALD Cheng 2009	7/9	**215** Boys 99 Girls 116	9.4	4	Change in whole grain intake (g/day)	Overweight at 9.4 yrs.: 17.2% Overweight at 13.4 yrs.: 17.2%	Change in % body fat: β = 0.09 s.e. = 0.13 P = 0.5 Change in BMI-SDS: β = 0.005 s.e. = 0.008 P = 0.5	Not significant
NGHS Berz 2011	8/9	**2,327** girls	9.5	9	Whole grain, av. intake between ages 9 to 17 yrs.: < 0.25 serving/day vs ≥ 1 serving/day	Mean BMI at 9.5 yrs.: ~ 19.2 kg/m^2 Mean BMI at 18.6 yrs.: ~ 25.2 kg/m^2	BMI at follow-up = 25.5 (SD 0.26) vs 24.1 (SD 0.38), P = 0.002	Higher average whole grain intake: **↓BMI in girls** Significant
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	Baseline whole grain intake (servings/day)	Overweight at 15 yrs.: ~25%, All o/w at 25.4 yrs.: Males 56.1%, Females 47.5% If not o/w at start, o/w at 25.4 yrs.: Males 45.4%, Females 34.2%	Risk of o/w : Males not o/w at start O.R. = 0.96 95% CI 0.78 to 1.19) Females not o/w at start O.R. = 0.71, 95% CI 0.54 to 0.93	Higher whole grain intake at baseline **↓risk of overweight in females** *not overweight at baseline* Significant
SUMMARY Teenage females with higher average whole grain intakes had lower BMI outcomes in young adulthood. For females not already overweight, higher baseline intakes of whole grain were associated with a reduced risk of overweight ten years later.								

5.4 Dairy foods, milk and flavoured milk

One paper from **ALSPAC** considered the influence of dairy foods including milk on future adiposity (Bigornia et al., 2014). Milk as a beverage (Noel et al., 2011) and flavoured milk (Noel et al., 2013) were also considered in ALSPAC.

Among **NGHS** girls low fat dairy food was one element of the modified DASH diet quality score (Berz et al., 2011). Milk was one of many beverages examined in the cohort by a second paper (Striegel-Moore et al., 2006).

Additionally, milk as a beverage was investigated in the **Framingham Children's Study** (Hasnain et al., 2014).

Longitudinal findings for dairy food intakes, milk and flavoured milk intakes are summarised in Table 5-3 and Table 5-4.

Dairy foods also contributed to dietary patterns in the **NGHS** (Ritchie et al., 2007).

5.4.1 Dairy intakes

In the **ALSPAC** cohort Bigornia et al used the USDA's definition of dairy products, (white milk, flavoured milk, cheese, yogurt, dairy ice-cream and dairy desserts or puddings) where a serving is the calcium equivalent of 1 cup of milk. Foods made from milk but retaining little calcium (cream or butter), are not part of this definition. Full-fat dairy products were made with whole milk. Reduced-fat dairy products were made with semi-skimmed milk (1.7% fat), plus skimmed milk, reduced-fat cheese or reduced fat yogurt. Milk and dairy products in mixed dishes were not included in the total. Total dairy intakes including milk were reported as g/ day. Milk was the largest part of dairy intake for both sexes. At age 10 years mean intake of total dairy products including milk was higher for boys (336g/ day) than for girls (265g/ day), and this difference by sex persisted. Across quartiles of intake, as full-fat dairy increased, reduced-fat dairy decreased and vice versa, but approximately one third of children did not consume reduced-fat dairy products on days surveyed.

In the **NGHS** mean intakes of low fat dairy products (skimmed milk, yogurt, cottage cheese, with 2% fat or less) were low. 55% of girls consumed less than 1 serving/day of low-fat dairy, equivalent to 1 cup of milk (Berz et al., 2011). Dietary patterns in the **NGHS,** representing cumulative intake, included component foods of dairy origin (plain milk, flavoured milk, yogurt, cheese/cheese spread and cheese sandwiches). White girls had higher cumulative intakes of most dairy foods than black girls. Averaged intakes of plain milk were lowest in black girls who followed a "Snack-type foods" dietary pattern (92g/day) and highest in white girls who followed a so-called "Healthy" dietary pattern (209g/day). White girls tended to eat more cheese than black girls.

5.4.2 Beverage milk intakes

Younger children drank more milk than older children in every cohort. Boys drank more milk than girls, and milk as a beverage was more popular among white girls than black girls. Not all children consumed flavoured milk, but intakes peaked in mid to late childhood, and diminished thereafter. By teenage milk tended to be replaced by other drinks, predominantly sugar sweetened beverages.

In the **ALSPAC** cohort milk consumption comprised full-fat, reduced fat and non-fat (skimmed) cow's milk, both plain and flavoured. Flavoured milk (with added sugars) was also considered separately. Beverage milk intakes were reported as servings/ day, with 1 serving equal to 244g plain milk. or 250g flavoured milk. At baseline 10 year old boys (Mean 1.04 servings/ day) consumed more milk than girls (Mean 0.79 servings/ day). By teenage three years later, boys total milk intake was similar, but girls drank slightly less (0.70 servings/ day) and reduced fat milk accounted for a greater share than before.
Plain and total milk intakes were highly correlated, but flavoured milk drinkers had lower intakes of plain milk. At 10 years old 380 children (17%) reported drinking flavoured milk and 50 had more than 1 serving/ day. By 13 years old only 13% of children (302) had flavoured milk in the 3 days surveyed(Noel et al., 2013)

The **Framingham Children's study** looked at total fluid milk intake at four age periods (3 to 5 years, 6 to 9 years, 10 to 12 years and 13 to 17 years), including soymilk and rice beverages with plain and flavoured milks. At 3 to 5 years most children drank milk, usually plain (median 6.9oz/ day). At 7 to 9 years plain milk intakes were similar, but flavoured milk intake rose to its highest level (1.5oz/ day). By age 10 to 12 years, some children did not report drinking milk of any type. Median intakes of plain milk dropped with age falling to 3.9oz/ day by ages 13 to 17 years, even as total beverage intakes increased. Plain milk's share of all beverages fell from 34% at ages 3 to 5 years to 14% at ages 13 to 17 years. SSBs share more than doubled from 24% to 56.0%.

In the **NGHS** girls' cohort, milk as a beverage was defined as all kinds of cow's milk including flavoured milk. From the start (age 9.5 years) mean milk intakes were higher among white girls (352g/ day) than black girls (244g/ day). Among black girls, intake of milk fell rapidly to 145g/ day by age 18.6 years. White girls sustained their milk intake until the age of 11.5 years, but then intake fell steadily to 242g/ day by age 18.6 years. Here too, milk as a beverage was displaced, as regular (non-diet) soda intake increased every year, although mean intakes of juice, tea and coffee did not rise until the girls' early teens.

5.4.3 Dairy food intakes and adiposity outcomes

In pre-teen children higher intakes of full-fat dairy foods including milk, predicted lower BMI and reduced total body fat mass three years later, but reduced fat dairy foods including milk were not associated with adiposity outcomes. Over a longer period, higher intakes of low-fat dairy products including milk in girls were linked to more favourable BMI outcomes by late adolescence/early adulthood. Dairy foods without milk were not investigated.

In **ALSPAC** intakes of all dairy products (including milk) and intakes of full-fat and reduced fat dairy products at age 10 years were categorised into quantiles. Quintile of total body fat mass (TBFM) measured by DXA, and risk of overweight at age 13 years were reported.
Children in the highest quartile of total dairy intake appeared to be at lower risk of excess TBFM and for overweight at follow up, but in models which adjusted for dietary reporting bias, these relationships were attenuated and not significant.

Children in the highest quartile of full-fat dairy intake, versus those in the lowest quartile, had significantly smaller gains in BMI over three years (+2.5kg/m^2 vs +2.7kg/m^2). Full-fat dairy intake seemed protective against overweight, but this was not significant in the adjusted model. The highest versus the lowest consumers of full-fat dairy foods had a lower risk of excess TBFM in all models.
Baseline prevalence of overweight was slightly lower among non-consumers and very low consumers of reduced-fat dairy compared with the highest quantile of intake (19.4% vs 21.8%), which could indicate that some children with overweight favoured reduced-fat dairy products. No associations between baseline intake of reduced-fat dairy products and future adiposity were detected in ALSPAC.

In contrast, in the **NGHS** cohort low-fat dairy intake (including milk), although generally low, was one of the strongest individual predictors of BMI within the DASH diet quality score. Girls in the highest category of low fat dairy intake compared with girls in the lowest category had significantly lower BMI gains over time (Berz et al., 2011).

The “Healthy” dietary pattern among white girls in the NGHS cohort (with the highest mean intakes of yogurt and plain milk and relatively low intakes of flavoured milk) mitigated increases in waist circumference (Ritchie et al., 2007).

Table 5-3 Dairy intakes and adiposity outcomes

Study & Paper	NOS stars	Population	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline & f' up	Adiposity outcomes	Findings
ALSPAC Bigornia 2014	8/9	**2,455** Boys 1,154 Girls 1,301	10.6	3	Highest quartile of **total dairy** consumption (Mean 563g/ day SD 155) at 10 yrs. vs. lowest quartile (Mean 88g/ day SD 54)	Overweight at 10.6 yrs. ~21%	Risk of excess TBFM O.R. 0.73, 95% CI 0.46 to 1.16 P-trend = 0.28 Risk of o/w. O.R. 0.69, 95% CI 0.41 to 1.15 P-trend = 0.24	Not significant
		2,455	10.6	3	Highest quartile of **full-fat dairy** consumption (Mean 348g/ day SD 176) g/ day) at 10 years vs. lowest quartile (Mean 9g/day SD 10)	Overweight at 10.6 yrs. ~21%	Risk of excess TBFM O.R. 064, 95% CI 0.41 to 1.00 P-trend = 0.04 Risk of o/w O.R. 0.65, 95% CI 0.40 to 1.06 P-trend = 0.19. Gains in BMI. 2.5kg/$m^{2,}$ 95% CI 2.2 to 2.7 vs 2.7kg/m^2, 95% CI 2.5 to 3.0 P<0.01.	Higher full-fat dairy intake **↓ risk of excess TBFM, ↓ gains in BMI** Significant
		2,455	10.6	3	Highest category of **reduced-fat dairy** consumption (Mean 439g/ day SD 154) at 10 years vs. lowest category (Mean 9g/day SD 20)	Overweight at 10.6 yrs. ~21%	Risk of excess TBFM O.R. 0.77, 95% CI 0.47 to 1.25 P-trend = 0.23 Risk of o/w O.R. 0.85, 95% CI 0.50 to 1.44 P-trend = 0.42	Not significant

Study & Paper	**NOS stars**	**Population**	**BL age (yrs.)**	**F' up (yrs.)**	**Exposure**	**Adiposity at baseline & f' up**	**Adiposity outcomes**	**Findings**
NGHS Berz 2011	8/9	**2,327** girls Black 1,188 White 1,139	9.5	9	Low fat dairy ≥ 2.25 servings/ day vs < 1 serving/day	Mean BMI at 9.5 yrs.: ~ 19.2 kg/m^2 Mean BMI at 18.6 yrs.: ~ 25.2 kg/m^2	Mean BMI at follow-up = 23.2 (SD 0.55) vs 25.7 (SD 0.17), P < 0.01.	Higher intake of low fat dairy: **↓BMI in girls**, Significant
SUMMARY Higher intakes of full-fat dairy foods (including milk) are associated with lower BMI outcomes and reduced total body fat mass in pre-teen children. Higher intakes of low fat dairy foods were associated with lower BMI outcomes in girls/young women, but no significant association was found between reduced-fat dairy foods (including milk) and adiposity outcomes in pre-teen children.								

5.4.4 Beverage milk intakes and adiposity outcomes

Only the smallest cohort found any evidence that childhood beverage milk consumption was associated with reduced body fatness later. In the **Framingham Children's study**, higher milk intakes in young children (3 to 9 years old) were associated with lower % body fat and a smaller sum of four skinfolds in later adolescence (15 to 17 years) at the $P \leq 0.05$ significance level. Children in the highest tertile of milk intake tended to have lower BMI and waist circumference at follow-up although these trends were not significant. With such a small sample size (n = 98) and multiple testing, apparent benefits may be erroneous.

In the **ALSPAC** cohort, after adjustments including total energy intake, neither total milk intake, full-fat milk nor reduced-fat milk intake at age 10 years predicted % body fat at age 13 years. No associations between 3 year change in milk intake (10 to 13 years) and 2 year change in % body fat (11 to 13 years) were found

As only 17% of children in the ALSPAC cohort drank flavoured milk at age 10 years, intake was modelled as a dichotomous variable. Flavoured milk consumers had higher intakes of total energy, saturated fat, carbohydrate and calcium than non-consumers, but there was little difference in their baseline prevalence of overweight or obesity at age 10 years, or in % body fat at 11 and 13 years. Among ALSPAC children with overweight or obesity at baseline, there were smaller (less favourable) changes in body fat % in flavoured milk consumers compared with non-consumers. When the stratified analysis was restricted to plausible reporters (n=846) the difference in body fat % outcomes between flavoured milk consumers versus non-consumers was even greater and still significant.

In the **NGHS** drinking milk was associated with increased calcium and sucrose intakes with less fructose, but longitudinal changes in beverage milk intake (each additional 100g/day) did not predict girls' BMI change.

.

Table 5-4 Beverage milk intakes and adiposity outcomes

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline & f' up	Adiposity outcomes	Findings
ALSPAC Noel 2011	7/9	2,270 Boys 1,030 Girls 1,240 **2,245** in analyses, 907 plausible reporters	10.6	3	Milk intake (total, full-fat or reduced fat) Total milk intake at baseline (mean 0.90 servings/day SD 0.73 servings/day)	Body fat% at 10.6 yrs. n/a Body fat% at 11.7 yrs. boys 22.7%, girls 27.8% Body fat% at 13.8 yrs. boys 18.9%, girls 29.1%	Body fat% at 13 yrs. and 2 year change, 11 to 13 years. No association with baseline milk intake or changes in intake after adjustments including TEI & plausible reporters	Body fat% and change Not significant
Framingham Children's Study Hasnain 2014	7/9	98	3 to 9	12.	Beverage intake – Milk Tertile 1 (mean 5.0oz.day SD 2.2oz./day) vs T3 (mean 13.9oz./day, SD 3.2oz./day	Mean BMI at age 3 to 9 yrs. 16.2 kg/m^2	BMI p = 0.09 WC p = 0.13 Mean % body fat by DXA 30.0% vs 22.6% P = 0.01: Mean ∑4SF (mm) 72 vs 55 P = 0.05	BMI, WC Not significant ↓ % body fat Significant ↓ ∑4SF Significant
NGHS Striegel-Moore 2006	8/9	2,371 girls Black 1,210 white 1,161			Beverage intake - Milk		For each +100g/ day change in milk intake, after adjustments including TEI BMI change = - 0.002 SE 0.006, P >0.05.	BMI change Not significant
ALSPAC Noel 2013	8/9	2,270 Boys 1,030 Girls 1,240 normal wt. 1,715 ow/obese 449	10.6	3	Flavoured milk consumers vs non-consumers	Mean body fat % at 10.6yrs. 25.5% SD 9.2	Change in % body fat from 11 to 13 years: Normal weight, P = 0.36 Ow/obese, P = 0.002	% body fat Not significant

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline & f' up	Adiposity outcomes	Findings
	8/9	**708** normal weight plausible reporters	10.6	3	Flavoured milk consumers vs non-consumers		Change in % body fat from 11 to 13 years: -1.35 95%CI -2.77 to 0.07 vs -1.33 95%CI -2.54 to -0.12, P = 0.96	% body fat Not significant
	8/9	**138** ow/obese plausible reporters Boys 68 Girls 70	10.6	3	Flavoured milk consumers vs non-consumers		Change in % body fat from 11 to 13 years: -0.16 95%CI -3.8 to 3.5 vs -3.4 95%CI -6.5 to -0.42, P = 0.02	↑ % body fat Only significant ***for ow/obese*** at baseline
SUMMARY In a small cohort, higher intakes of milk in early childhood were associated with lower body fat by mid-teens, but in larger cohorts no significant associations were found. In small numbers of plausible reporters, overweight/obese children who drank flavoured milk had less favourable changes in body fat.								

5.5 Juice, fruit and vegetables

Beverage juice intakes and adiposity outcomes were investigated in four cohorts, **GUTS** (Field et al., 2003b), the **DONALD** study (Libuda et al., 2008), the **Framingham Children's study** (Hasnain et al., 2014)and the **NGHS** (Striegel-Moore et al., 2006). Longitudinal findings are summarised in Table 5-5.

Fruit intakes and vegetable intakes and their influences on adiposity outcomes were investigated in three cohorts, in **GUTS** (Field et al., 2003b), in the **NGHS** as elements within the DASH diet quality score (Berz et al., 2011), and as potential predictors of overweight in **Project EAT** (Quick et al., 2013) Longitudinal findings are summarised in Table 5-6 and Table 5-7. Additionally, fruit and vegetables featured in dietary patterns identified within the **NGHS** cohort (Ritchie et al., 2007) and in Project EAT (Cutler et al., 2012).

5.5.1 Juice intakes

Fruit juice intakes were similar for boys and girls in mixed sex cohorts. White girls drank more fruit juice than black girls, who favoured sugar sweetened fruit drinks. Several studies observed that fruit juice intake increased with age, although in USA cohorts sugar sweetened beverages took a far greater share of total beverage intakes. Adolescents in USA cohorts consumed less fruit juice than their peers in the German **DONALD** Study.

In the **DONALD** study "fruit juice" was 100% fruit juice. Diluted and sugar-sweetened fruit drinks, were included with regular soft drink. 244 children returned at least 4 of 6 weighed dietary records between the ages of 9 and 18 years old. Baseline intakes of fruit juice (mean age 11.9 years) were similar for both sexes (Boys: mean 178g/ day, Girls: mean 180g/ day) but some children did not consume any fruit juice on days surveyed. At the last assessment (mean age 16.8 years) fruit juice consumption was reported by more individuals and both sexes increased their intake, with greater increases for boys. (Boys: mean 277g/ day, Girls: mean 247g/ day). These mean intakes of juice were the highest reported by any cohort, contributing about 4% of total energy intake. (Libuda et al., 2008)

Juice was not defined in the **GUTS** paper but appears to be fruit juice alone. In the GUTS cohort baseline intakes of juice were also similar for girls (mean age 12.0 years) and boys (mean age 11.8 years) at 0.8 servings per day and 0.9 servings per day respectively. Based on a serving size of approximately 250g these mean values equate to 200g/ day for girls and 225g/day for boys. (Field et al., 2003b)

In the **Framingham Children's Study** "Fruit and vegetable juice" included unsweetened fruit juice and small intakes of sweetened fruit and vegetable juices. (Part-juice drinks were grouped with "sugar-sweetened beverages".) In this small cohort (n = 103), median fruit/

vegetable juice intake in early childhood (ages 3 to 5 years) was approximately 165g/day. Intake dropped throughout childhood to a median of 92g/ day at ages 10 to 12 years, rising slightly to 101g/day by ages 13 to 17 years. As Framingham children reached adolescence, fruit/vegetable juice accounted for a decreasing share of their total beverage consumption (29% at ages 3 to 5 years, 12% at ages 15 to 17 years) and contributed less to their TEI (5% at ages 3 to 5 years, 2.3% at ages 15 to 17 years). (Hasnain et al., 2014)

In the **NGHS** girls' cohort beverages containing juice were coded either as "fruit juices" - bottled, canned, fresh or frozen fruit or vegetable juice, or as "fruit drinks" – sweetened non-carbonated fruit flavoured drinks, punches or "ades", containing less than 100% juice. At age 9.5 years white girls drank slightly more fruit juice (mean 111g/ day) than black girls (mean 108g/ day), a tendency throughout the study. Intakes of fruit juice changed little until the early teens, then rose throughout adolescence to reach the highest level at age 18.6 years for white girls (mean 129g/ day) and black girls (mean 120g/ day).
Conversely black girls drank more sugar sweetened fruit drinks than white girls. Black girls' daily intake of fruit drinks increased over time (mean 135g/ day at age 9.5 years, mean 204g/ day at age 18.6 years) whereas white girls mean consumption of fruit drinks held steady at ~ 85g/day. Dietary patterns in the NGHS cohort included fruit juice and vegetable juice and also give an insight to the quantities consumed (Ritchie et al., 2007). White girls who followed the "Convenience" pattern had the lowest juice intake averaged over the 10 years of the study (mean 104g/ day). The highest juice intakes were in the "Sweets and cheese" pattern identified in black girls (mean 162g/ day) and the "Sweets and snack type foods" pattern identified in white girls (mean 160g/ day).

5.5.2 Fruit and vegetable intakes

In the three cohorts that investigated fruit and vegetables, most participants did not meet the recommended 5 servings of fruit and vegetables a day.
The semi-quantitative FFQ employed in **GUTS** had 11 questions about fruit and juice and 19 questions about vegetables. Models considered Fruit and Vegetables (with potatoes, but not French fries) together and separately. Baseline mean intakes of fruit (not including juice) for girls and for boys were 1.0 serving/day. Baseline mean intakes of vegetables for girls and boys were 1.6 servings/day and 1.5 servings/day respectively. About 23% of children had 5 servings of fruit and vegetables a day. (Field et al., 2003b)

In the **NGHS** fruit intakes and vegetable intakes were two of seven food groups that made up a modified DASH diet quality score used in the cohort. Girls' median fruit intake across the 10 years of the study was 1.1 serving/ day, median vegetable intake was 2.0 servings/ day. Even

girls at the 95th percentile failed to meet the DASH recommendations of 4 to 5 servings of fruit and 3 to 4 servings of vegetables daily. Only 16.5% of girls consumed ≥2 servings/ day of fruit, less than 13% of girls consumed ≥ 3 servings/ day of vegetables.

This was reflected in **NGHS** girls' dietary patterns, which had food groupings including fruit (fresh, canned and dried), green salad (tossed salad, other vegetable salad), legumes (beans, chili) and vegetables (fresh, canned and frozen but not potatoes).

The "Customary" pattern followed by most black girls had the lowest mean intakes of fruit (59g/ day) and comparatively low intakes of legumes and other vegetables. The "Convenience" pattern followed by 45% of white girls also had low mean intakes of fruit, below one portion a day, and the lowest mean intakes of legumes (7g/ day) and other vegetables (25g/ day). The highest mean intakes of fruit (115g/ day) and other vegetables (55g/ day) were in the "Healthy" pattern followed by 12% of white girls.

In **Project EAT** fruit servings/day and vegetable servings/day at baseline and change were considered as predictors of overweight incidence. Among adolescents who were *not overweight at baseline* mean intakes of fruit were the same for girls and boys (2.2 servings/day). Mean intakes of vegetables were also similar for girls and boys (1.8 servings/day and 1.6 servings/day respectively). Intakes at baseline and 10 years later were measured with a different FFQ, so absolute changes are uncertain, but at follow-up reported intake of fruit appeared unchanged among females and reduced among males. Both females and males had apparent increases in vegetable intakes.

Dietary patterns identified at baseline in **Project EAT** included a "vegetable" pattern and a "fruit" pattern (Cutler et al., 2012). Quantified fruit and vegetable intakes for these dietary patterns are not presented (Cutler et al., 2012)but an earlier paper (Cutler et al., 2009) showed that children who were high scorers for the "Vegetable" and "Fruit" patterns were more likely to meet Healthy People 2010 targets for vegetable and fruit intake . In the highest scoring quintile for the "Vegetable" pattern, close to half of children consumed ≥ 3 servings vegetables/ day, but close to none in the lowest scoring quintile. In the highest scoring quintile for the "Fruit" pattern, over two thirds of children consumed ≥ 2 servings fruit/ day, with less than a tenth in the lowest scoring quintile.

5.5.3 Juice intakes and adiposity outcomes

There was mixed evidence about juice intakes and future adiposity.

In the studies which investigated baseline intakes of fruit juice, only the **Framingham Children's study** reported any beneficial effect. Children in the highest tertile of juice intake versus those in the lowest tertile at ages 3 to 9 years had an 8cm smaller mean waist circumference and a 22mm smaller sum of 4 skinfolds (favourable) by ages 15 to 17 years. Childhood juice intakes did not predict BMI or % body fat.

In the **DONALD** cohort baseline intakes of juice did not predict body fat % or change in BMI-SDS. In the **GUTS** cohort no relationship between baseline fruit intake and annual change in BMI z-score over the 3 year study was seen in the simplest models. In models which adjusted for TEI, each additional serving of juice at baseline did show a tiny + 0.003 annual increase (unfavourable) in BMIz for girls, but results were not significant for GUTS boys.

Two studies investigated change in juice intake. In the **NGHS** adolescent girls' BMI outcomes were not predicted by each additional 100g/day of juice or fruit drinks. In contrast, in the **DONALD** cohort, each MJ increase in fruit juice consumption over the 5 years predicted + 0.096 gain in BMI-SDS for girls and this was significant. The same association was not observed in boys. However, change in fruit juice consumption in the DONALD study, which had the highest reported intakes, did not predict change in % body fat in either sex.

Table 5-5 Juice intakes and adiposity outcomes

Study & Paper	NOS stars	Popln	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
DONALD Libuda 2008	7/9	244 Boys 125 Girls 119	11.9	5	100% fruit juice Baseline intake MJ/day	Mean BMI at baseline, boys 18.1kg/m^2, girls 18.4kgm^2, Mean BMI at 16.8 yrs. boys 21. kg/m^2, girls 21.6kg/m^2	BMI-SDS %body fat	Not significant
	7/9		11.9	5	100% fruit juice Change in intake MJ/day		Change in BMI-SDS boys: β = -0.002, P = 0.964 girls: β = +0.096, P = 0.01 Change in %body fat, not significant	↑ Fruit juice, ↑ BMI-SDS change in girls Significant
Framingham Children's Study Hasnain 2014	7/9	98	3 to 9 years	12 years	Fruit and vegetable juice Baseline intake Tertile 1 (mean 56g or 1.9oz.day SD 1.0oz./day) vs T3 (mean 301g or 10.2oz./day, SD 2.8oz./day)	Mean BMI at baseline (age 3 to 9 yrs.) 16.2 kg/m2	At 15 to 17 years BMI P = 0.06 % body fat by DXA P = 0.12: Mean waist circumference (cm): 84 vs 76 P = 0.03 Mean ∑4SF (mm) 74 vs 52 P = 0.04	BMI, % body fat, Not significant ↑ Fruit & vegetable juice, ↓ Waist circumference, ↓ ∑4SF Significant
NGHS Striegel-Moore 2006	8/9	2,371 girls Black 1,210 White, 1,161	9.5	10	Beverage intake – fruit or vegetable juice +100g/ day change	n/a	BMI change = 0.005 S.E. 0.007, P>0.05	Not significant
NGHS Striegel-Moore 2006	8/9	2,371 girls Black 1,210 White, 1,161	9.5	10	Beverage intake – fruit drinks +100g/ day change	n/a	BMI change = 0.009 S.E. 0.007 P>0.05.	Not significant

Study & Paper	NOS stars	Popln	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
GUTS Field 2003a	5/9	14,918 Boys 6,715 Girls 8,203	Boys 11.8 Girls 12.0	3	Fruit and vegetable juice	Ow/obese at baseline Boys 23% Girls 18%	Annual BMIz change Boys: β per serving = 0.002, 95% CI 0.000 to 0.005 Girls: β per serving = 0.003, 95% CI 0.001 to 0.005	↑ Annual BMIz change in girls Significant
SUMMARY In a small cohort, higher intakes of juice in early childhood were associated with a smaller waist circumference and sum of skin folds by mid-teens. In a larger, older cohort higher baseline intakes of juice were linked to a higher BMIz score at follow-up for girls. An increase in juice intake was linked to a higher BMIz score at follow-up for girls only.								

5.5.4 Fruit intakes and adiposity outcomes

There was uncertain evidence about fruit intakes and future adiposity for girls. No association was found for boys.

Among girls from **GUTS** baseline intakes of fruit (without juice) did not predict annual change in BMIz score in the simple model, but when adjusted for TEI each additional serving of fruit predicted an annual increase in BMIz score, which was significant but very small, + 0.003.

In contrast, girls from the **NGHS** in the highest fruit intake category had significantly smaller BMI gains, with the lowest mean BMI at final follow-up. The "Healthy" dietary pattern identified in this cohort, with the highest mean intakes of fruit, was related to smaller increases in waist circumference. In **Project EAT**, for adolescent girls who were not overweight at baseline, fruit servings/ day at baseline (or 10 year change, although this measure is not valid) did not predict overweight at 10 year follow-up (Quick et al., 2013).

Among boys from **GUTS** baseline intakes of fruit (without juice) did not predict annual change in BMIz score and in **Project EAT** for adolescent boys who were not overweight at baseline, fruit servings/ day at baseline (or 10 year change) did not reliably predict overweight at follow-up.

5.5.5 Vegetable intakes and adiposity outcomes

Two of the three cohorts that investigated vegetable intakes reported that there may be a protective effect against future overweight, at least for boys.

In **GUTS**, higher baseline intakes of vegetables predicted a lower BMIz score at follow-up for boys, although after adjustments for TEI this association was attenuated and no longer significant. No significant associations between baseline vegetable intake and BMIz outcomes were seen in girls.

Similarly among **NGHS** girls, vegetable intakes did not predict BMI outcomes.

Similarly, among **NGHS** girls, vegetable intakes did not predict BMI outcomes. Girls' BMI at final follow-up were similar across all three categories of vegetable consumption (based on at least 2 sets of 3 day food records) and the small differences were not statistically significant.

In **Project EAT**, for adolescents who were not overweight at baseline, vegetable servings/ day at baseline did not predict overweight at follow-up. Although absolute changes in intake are uncertain (due to using different FFQs at baseline and follow-up) the apparent increase in vegetable servings/ day during the 10 years, when adjusted for energy intake, reduced the risk of overweight in males, but not females.

Table 5-6 Fruit intakes and adiposity outcomes

Study & Paper	NOS stars	Population	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
GUTS Field 2003a	5/9	14,918 Boys 6,715 Girls 8,203	Boys 11.8 Girls 12.0	3	Fruit servings/day, without juice	Ow/obese at baseline Boys 23% Girls 18%	Baseline fruit intakes and annual BMIz change, adjusted for TEI, Girls: β per serving = 0.003, 95% CI 0.001 to 0.004 Boys: β per serving = 0.002, 95% CI- 0.000 to 0.003	↑ Fruit, adjusted for Energy, ↑ Very small annual BMIz change Significant for girls
NGHS Berz 2011	8/9	2,327 girls	9.5	9	DASH food group score – fruit Fruit intake ≥2 servings per day vs Fruit intake <1 serving per day,	Mean BMI at baseline ~ 19.2 kg/m2, Mean BMI at follow-up ~ 25.2 kg/m2	≥2 servings per day, mean BMI at follow-up = 23.6 (SD 0.32,) vs <1 serving per day, mean BMI at follow-up = 26.0, (SD 0.19), P < 0.001	↑ Fruit ↓BMI in girls Significant

Study & Paper	NOS stars	Population	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	Fruit servings/day	Overweight at baseline (15 yrs.) ~25%, at follow-up (25.4 yrs.) males 56.1%, females 47.5%	Baseline fruit, risk of ow at follow-up adjusted for TEI Girls: OR 0.98, 95% CI 0.87 to 1.10 Boys: OR 1.03, 95% CI 0.92 to 1.17 10 year Change in fruit risk of ow at follow-up adjusted for TEI Girls: OR 0.98, 95% CI 0.89 to 1.09 Boys: OR 1.04, 95% CI 0.91 to 1.18	Not significant after adjustment for TEI
SUMMARY The association of fruit intake with future adiposity is uncertain.								

Table 5-7 Vegetable intakes and adiposity outcomes

Study & Paper	NOS stars	Population	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
GUTS Field 2003a	5/9	14,918 Boys 6,715 Girls 8,203	Boys 11.8 Girls 12.0	3	Vegetable servings/day	Ow/obese at baseline Boys 23% Girls 18%	Baseline vegetable intakes and annual BMIz change, adjusted for TEI. Girls: β per serving = 0.005, 95% CI -0.001 to 0.005 Boys: β per serving = -0.001, 95% CI -0.004 to 0.001	↓BMIz in boys Not significant after adjustment for TEI
NGHS Berz 2011	8/9	2,327 girls	9.5	9	DASH food group score – vegetables Vegetable intake ≥3 servings per day vs Vegetable intake <2 servings per day,	Mean BMI at baseline ~ 19.2 kg/m2, Mean BMI at follow-up ~ 25.2 kg/m2	≥3 servings per day, mean BMI at follow-up = 25.1, (SD 0.36) vs <2 servings per day, mean BMI at follow-up = 25.2 (SD 0.19), P > 0.99	Not significant

Study & Paper	**NOS stars**	**Population**	**BL age (yrs.)**	**F' up (yrs.)**	**Exposure**	**Adiposity at baseline**	**Adiposity outcomes**	**Findings**
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	Vegetable servings/day	Overweight at baseline (15 yrs.) ~25%, at follow-up (25.4 yrs.) males 56.1%, females 47.5%	Baseline veg, risk of ow at follow-up, adjusted for TEI, Girls: OR 1.03, 95% CI 0.90 to 1.18 Boys: OR 1.12, 95% CI 0.97 to 1.29 10 year change in veg , risk of ow at follow-up, adjusted for TEI Girls: OR 1.03, 95% CI 0.94 to 1.12 Boys: OR 0.88, 95% CI 0.78 to 0.99, P<0.05	↑Vegetables ↓risk of overweight in males (not overweight at baseline) Significant
SUMMARY Higher vegetable intakes may have some protective effect against future adiposity for boys.								

5.6 Fish

Fish consumption and its influence on future adiposity was examined in a single sex girls' cohort in Hawaii, the Female Adolescent Maturation Study (**FAMS**). (St-Jules et al., 2014). See Table 5-8.

5.6.1 Fish intakes

In **FAMS** baseline fish frequency and intakes were measured in 200 female adolescents. During the three days surveyed, 100 girls ate fish, 100 did not. Two thirds of Asian girls ate fish, compared to one third of White girls. Asian girls consumed fish more often and in greater quantity (median 24g/ week, IQ range 0 to ~112g/ week) than White girls (median 0g/ week, IQ range 0 to ~ 11g/ week). Serving sizes rarely exceeded 3 oz. (85g). The US Department of Agriculture recommends two servings (2 x 4 oz. = 8 oz. or ~ 220g.) of fish/ week for adolescents. Only 19% of Asian girls, 12% of Mixed ethnicity and 4% of White girls in FAMS met this advice.

5.6.2 Fish intakes and adiposity outcomes

The evidence of any beneficial effect of fish intake on adiposity outcomes was slight.
In **FAMS** 103 girls had anthropometric measures including BMIz score, waist circumference and fat mass ratios based on SFT, at baseline and at follow-up. During the 2 year study BMIz scores stayed relatively constant but an increase in trunk-to-peripheral fat ratio was observed, consistent with normal development in adolescent girls. Body fat % measured by DXA at follow-up varied widely (Mean 29% SD 8%).
Higher baseline fish intake was linked with a smaller waist circumference after 2 years. In a model which adjusted for ethnicity as well as age, physical activity, energy intake, pubertal stage and baseline anthropometric values, baseline fish intake still had an inverse association with change in waist circumference, although the difference was very small (- 0.013cm P = 0.026).

Table 5-8 Fish intakes and adiposity outcomes

Study & Paper	NOS stars	Population	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
Female Adolescent Maturation Study St. Jules 2014	7/9	200 girls Asian 68 White 51 Mixed ethnicity 81 **103** in analyses	11.5	2	Fish intake (oz/week)	Mean BMIz at baseline Asian 0.17 White 0.21 Mixed ethnicity 0.35	Baseline fish intake adjusted for Energy intake, 2 year change in waist circumference (cm): Girls: Co-eff = -0.013, P = 0.026	↓ change in waist circumference in girls only Significant
SUMMARY In a small cohort of girls, higher baseline fish intake was associated with marginally smaller gains in waist circumference at follow-up, even after adjustments for ethnicity. No evidence for boys.								

5.7 Convenience foods

The influence of convenience food intake on future adiposity was investigated in the **DONALD** study (Alexy et al., 2011) which subdivided convenience foods by their energy density. Findings are described below and summarised in Table 5-9.

5.7.1 Convenience food intake

In the **DONALD** study convenience foods (CFs) were defined as "pre-prepared savoury products, frozen, canned or instant, hot or cold, all-in-one-meals or courses, purchased in a store and eaten in the home". Fast foods and foods eaten outside the home were not included in this definition. Food intake was assessed using weighed 3-day food records. 585 children aged 3 to 18 years between 2004 and 2008 provided 1,890 food records. Over 5,300 convenience foods were found, with 89% of food records reporting at least one. Energy density (ED) for each product was calculated as the energy content (kJ) per gram. Convenience foods were then grouped as high EDCFs (>6 kJ/g) or low EDCFs, depending on whether they were above or below the median ED of reported CFs. On average CFs contributed 6% of TEI and most were low energy density. Typically, high EDCFs were pizza or meat dishes and cold sauces. Total CF and high EDCF intakes increased with age, accounting for an increasing share of TEI. High EDCF intakes were associated with a decrease in carbohydrate and protein and an increase in fat, predominantly poly unsaturated fatty acids as most dishes used vegetable oil as an ingredient rather than hard, saturated fats. High EDCF intakes were also linked with reduced intakes of dairy, grain and sweet foods, but not fruit, vegetables or beverages.

5.7.2 Convenience food intakes and adiposity outcomes

In the **DONALD** Study change in CF intake during the 5 year study period was measured in 363 children who had at least two food records. Cross-sectional and longitudinal analyses showed that total CF intakes were not associated with BMI-SDS or body fat % calculated from SFT. However, in a longitudinal linear regression model which adjusted for residual energy intake, baseline intakes of high EDCF did predict adverse changes in body fat % in boys.
Change in high EDCF intake among boys did not predict either BMI-SDS or body fat%. No longitudinal associations between high EDCF intake and adiposity outcomes were evident in girls.

Table 5-9 Convenience foods and adiposity outcomes

Study & Paper	NOS stars	Population	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
DONALD Alexy 2011	7/9 7/9	585 Boys 296 Girls 289 **363** in analyses Boys 190 Girls 173	3 to 18	5	Convenience foods	Not given	Baseline or change in convenience foods did not predict BMI-SDS or % body fat.	Not significant
		As above	3 to 18	5	Energy dense convenience foods (>6kJ/g)	Not given	Baseline high EDCF and change in % body fat Girls: β = 0.012, s.e. 0.031, P = 0.6953 Boys: β = 0.104, s.e. 0.104, P = 0.0098	↑% body fat in boys Significant
SUMMARY High intakes of convenience foods impaired dietary quality but were not associated with adiposity outcomes. Energy dense convenience foods were associated with small increases in % body fat in boys.								

5.8 Sweet and savoury snack foods

The influence of snack food intakes on adiposity outcomes was examined in two USA cohorts, **GUTS** (Field et al., 2004) which considered sweet and savoury snacks, and the **Bogalusa Heart Study** (O'Neil et al., 2015) which focussed on candy (confectionery). Their longitudinal findings are summarised in Table 5-10.

Snack foods also featured in dietary patterns identified in three studies, the **Bogotá Schoolchildren** cohort (Shroff et al., 2014), **Project EAT** (Cutler et al., 2012) and the girls only **NGHS** (Ritchie et al., 2007).

5.8.1 Snack food intakes

There is no widely accepted definition of snack foods, but in the Growing Up Today study they were described as foods which, "tend to be energy dense and of little nutritional value" which are "readily available" and "commonly consumed" (Field et al., 2004). Such foods may be eaten instead of or in addition to main meals, between meals or as part of a meal.

In **GUTS** snack food intake (not including SSBs) was assessed in girls and boys aged 9 to 14 years old at baseline, in 1996, 1997 and 1998. The YAQ asked 27 questions about snack foods including potato chips, corn chips, nachos, popcorn, pretzels, peanuts and nuts, fruit rollups, Graham crackers, wheat thins, pop tarts, cake, snack cakes, Danish or pastries, donuts, cookies, brownies, pie, chocolate, chocolate candy, candy without chocolate, jello, pudding, frozen yogurt, ice-cream, milkshake, popsicle and seeds. In 1997 and 1998 the survey also asked whether children ever ate "low fat or no fat" snacks.
At baseline girls ate 3 servings of snacks/ day on average. Boys ate slightly more, on average 3.2 servings of snacks/ day. For both sexes snacks made up about 18% of TEI (Girls: 365 calories/ day, Boys: 420 calories/ day). Throughout the study boys consumed more snacks than girls, but intake of snacks declined slightly during the study period. In the 1997 survey, 5% of children always ate reduced fat snack foods and 55% of children ate reduced fat snacks sometimes, with girls tending to eat reduced fat snacks more often than boys.

In the **Bogalusa Heart Study** candy included chocolate bars or packets, chocolate candy bars and sugar candy. Candy is a source of added sugars in the diet and chocolates are a source of saturated fats, so candy is energy dense. Single 24-HDR in 1973/74, 1976/77 and 1978/79 provided baseline data. 92% of 10 year old children surveyed ate candy and their mean (SD) candy consumption was approximately 46g (45) g per day. Tertiles of candy consumption were 0 to 19.5 g/ day, 20 to 54.3 g/ day, and 54.8 to 281.5g/ day.
355 participants (61% female) were followed up as young adults, either in 1989/1991 (mean

age 23 years) using 24-HDR, or in 1995/1996 (mean age 29 years) using an FFQ which assumed a 40 g serving size for candy. Two-thirds of adults (aged 19 to 38 years) ate candy and their mean (SD) candy consumption was approximately 20 (30) g per day.

The "Snacking" pattern in **Bogotá School Children** characteristically contained "high energy, low nutrient-density" foods such as candy, ice-cream, packed fried snacks, soda and sugary drinks. In the "Sweet and salty snack food" pattern seen in boys and girls from **Project EAT** heavily loaded factors were chocolate bars, cake, brownies, potato chips and nachos. Neither dietary pattern was broken down into quantified intakes of component foods.

In the **NGHS** highest mean intakes of snack-type foods in the eight identified dietary patterns were sweetened drinks 486g/day, candy 17.5g/ day (much lower than mean intakes in children in the BHS, measured a decade before), crackers 10.5g/ day, chips 12.5g/ day, pretzels 7g/ day, nuts and popcorn 9g/ day. Black girls had significantly higher mean intakes of sweetened drinks and candy than white girls; white girls had significantly higher mean intakes of crackers and pretzels than black girls.

5.8.2 Snack food intakes and adiposity outcomes

Evidence of an association between snack food intakes and future adiposity was slight. Eating reduced fat snacks, at least sometimes, was beneficial for boys.

In the **Bogalusa Heart Study** at baseline approximately 22.5% of 10 year olds had overweight or obesity, but by adulthood this had more than doubled. The BHS looked at childhood candy consumption by tertile of intake and BMI outcomes in young adulthood. No significant associations were found in models which adjusted for baseline BMI and TEI. O'Neill et al concluded that childhood candy consumption was not predictive of future adiposity.

In **GUTS**, girls and boys had a similar mean BMIz at baseline. Snack food intake and annual change in BMIz score was modelled using servings per day of snack foods, calories per day from snack food or as the percentage of daily calories from snack food, with consistent results. In girls baseline snack food intake inversely predicted BMIz score in the simplest model, but the association was attenuated and no longer significant after controlling for TEI. No association was observed in boys. Annual changes in servings per day of snack food intake were not associated with annual changes in BMIz score in either sex.

Table 5-10 Sweet and savoury snacks and adiposity outcomes

Study & Paper	NOS stars	Population	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
Bogalusa Heart Study O'Neil 2015	7/9	355 Males 138 Females 217	10	Mean 13 yrs.	Candy intake Tertile 1, from 0 to 19.5g/day Tertile 3 from 54.8 to 281.5 g/day	Ow and obesity at baseline 22.5% Ow and obesity at follow-up (19 to 38 yrs., mean 23.4 yrs.) 45.4%	BMI in young adulthood, adjusted for TEI Tertile 3 vs tertile 1: β = 0.74 s.e. 0.60 P = 0.214	Not significant
GUTS Field 2004	5/9	14,977 Boys 6,774 Girls 8,203	Boys 11.9 Girls 12.0	3	Snack food servings/day	BMIz at baseline Boys Mean 0.2 SD 1.1 Girls Mean 0.1 SD 1.0	Annual change in BMIz Boys: β = -0.003 P >0.05 Girls β = -0.007 P <0.05 Adjusted for TEI Boys: β = -0.004 95% CI -0.014 to 0.007 P >0.05 Girls β = -0.006 95% CI -0.013 to 0.001 P >0.05	Not significant for boys ↓BMIz change in girls Not significant after adjustment for Energy
SUMMARY Neither candy nor snack foods were linked with adiposity outcomes, after adjustment for Total Energy intake (TEI).								

Boys from GUTS who ate reduced fat snack foods sometimes, compared with never, had significantly lower BMIz outcomes (β = -0.041, P =0.03), but no protective effect was seen for girls.

Field et al. concluded that snack food was not an important independent predictor of weight gain in children and adolescents, with the caveat that as many snack foods have little nutritional value, it would be prudent to recommend "moderate" intakes.

In the **NGHS** white girls following the "Sweets and snack-type foods" dietary pattern (with the highest mean intakes of sweetened drinks, candy, crackers, pretzels, nuts and popcorn, and also the highest mean intakes of carbohydrate and sucrose) had the largest mean values for waist circumference at final follow-up (75.1cm SE 0.40cm). White girls following a "Healthy" pattern with comparatively low intakes of snack foods, had the smallest final waist circumference (73.2cm SE 0.85cm) which was significant (P = 0.037).

Ritchie et al surmised that dietary patterns characterised by high intakes of nutrient poor, energy dense, snack-type foods are related to higher adiposity outcomes.

5.9 Sugar sweetened and diet beverages

The influence of sugar sweetened beverage intakes on future adiposity was reported by 10 included papers, originating from five USA cohorts and four cohorts from other countries. Sugar sweetened beverages (SSBs) were investigated alone or in comparison with other beverages, including artificially sweetened diet drinks. Adiposity outcomes included changes in BMI, BMIz score, waist circumference, body fat percentage or skinfold thicknesses and the risk of future overweight/obesity.

In the USA the **Framingham Children's Study** (Hasnain et al., 2014) looked at quantiles of intake of milk, juice, SSBs and diet beverages.

GUTS II investigated servings per day of regular soda, diet soda and sports drinks (Field et al., 2014). SSBs and diet soda were also considered in **IDEA and ECHO** (Laska et al., 2012). Intakes of regular and diet soda, milk. fruit juice, fruit drinks, tea and coffee were explored in the **NGHS** girls only cohort (Striegel-Moore et al., 2006) while in **Project EAT** servings per day of SSBs was one of many predictors of future overweight (Quick et al., 2013).

Elsewhere, the Colombian **Bogotá School Children** study (Shroff et al., 2014) considered soda frequency as part of a snacking dietary pattern.

The German **DONALD** study, investigated energy intake from regular soft drinks (RSD) and other "energetic beverages" including fruit juice (Libuda et al., 2008).

In the Danish arm of the European Youth Heart Study (**EYHS**) continuous daily SSB intakes (Zheng et al., 2015) and categories of SSB intake (Zheng et al., 2014) were examined. The

Australian **RAINE** study compared baseline intakes of SSBs, diet drinks, milk and juices but only analysed risk of overweight/obesity based on tertiles of SSB intake (Ambrosini et al., 2013).

Longitudinal findings for SSBs including sports drinks, and for diet beverages and adiposity outcomes are summarised in Table 5-11 and Table 5-12 (Milk and juice are in Table 5-4 and Table 5-5 respectively.)

Two other papers from the **NGHS** looked at SSBs in the broader context of dietary patterns (Ritchie et al., 2007) and as an eating behaviour which may predict obesity (Rehkopf et al., 2011). One paper from **GUTS** reported SSB intakes in association with eating fried food away from home (Taveras et al., 2005).

5.9.1 Definitions of sugar sweetened and diet beverages

Each study categorised beverages in a slightly different way but most papers provided descriptions of drinks included in each category.

Sugar sweetened beverages (SSBs) were not defined in **Project EAT** or **GUTS.**(Taveras et al., 2005) Elsewhere SSBs were described as "drinks sweetened with a caloric sweetener, such as sugar" or as "energetic beverages"

Most studies (**Bogotá School Children, DONALD, EYHS, Framingham Children's Study, IDEA & ECHO, RAINE**) included regular soft drinks or RSD (i.e. non-diet, non-alcoholic), either carbonated or un-carbonated and sugar sweetened cordials, squash and fruit drinks containing less than 100% juice within their definition of SSBs. The Framingham Children's study and IDEA & ECHO also placed sugar-sweetened tea and coffee with SSBs, but the EYHS and NGHS put tea and coffee (sweetened or not) in a separate category.

Carbonated soft drinks were referred to as "soda" by USA cohorts. In the **NGHS** regular soda and fruit drinks (non-carbonated fruit flavoured drinks, excluding sports drinks) were investigated as two separate categories of SSB. In the **GUTS II** study soda was sub-divided into three categories: regular (non-diet) soda, diet soda and sports drinks, which were described as drinks containing approximately 50 calories, 110 mg of sodium and 41g sugar per 8 oz. serving, which are sold in large (20 to 32 oz.) bottles and promoted as part of an active lifestyle.

Diet drinks were described as "unsweetened or artificially sweetened beverages" or "low energy soft drinks".

Three cohorts (**DONALD, EYHS** and **RAINE**) defined diet drinks only to distinguish them from non-diet SSBs, the exposure of interest.

Four USA cohorts investigated diet drinks and adiposity outcomes. The **Framingham Children's study** and **IDEA & ECHO** included artificially sweetened drinks (carbonated and non-carbonated soft drinks and fruit drinks as well as unsweetened or artificially sweetened tea

and coffee) in their definition of diet drinks. The **NGHS** girls cohort and **GUTS II** considered only carbonated diet drinks (diet soda).

5.9.2 Sugar sweetened beverage intakes

In the USA cohorts, children's consumption of sugar sweetened beverages was ubiquitous from mid-childhood, rising through adolescence to take an ever greater share of increasing total beverage intake. SSB intakes above 500g/day or 2+ servings per day were reported by high consumers and older teenagers, contributing as much as 9% of TEI, but there were indications that SSB intake decreased by early adulthood. Boys reported higher mean intakes of SSBs than girls and black girls reported higher mean intakes than white girls.

In the small **Framingham Children's Study** baseline median consumption of SSBs of 3 to 5 year olds was 4.5 oz./ day (~133g/ day), or 4% of TEI. Some young children did not consume SSBs on days surveyed. SSB accounted for half of all beverages at 10 to 12 years old as intakes of milk and juice declined. By 13 to 17 years old median intake of SSBs had increased to 18.0 oz./ day (~533g/ day) or 9% of TEI.

In the **NGHS** girls SSB intakes increased with age. At age 9.5 years mean intakes of regular soda were similar for white (136g/ day) and black girls (134g/ day), but black girls consumed more sweetened fruit drinks. Tea and coffee intake stayed low until age 12.5 years.
Nine years later (mean age 18.6 years), mean intakes of regular soda were higher for white girls (377g/ day) than black girls (339g/ day). Conversely, mean intakes of sweetened fruit drink were far lower for white girls (87g/ day) than black girls (204g/ day). Mean intakes of tea/coffee were 106g/day for white girls and 47g/day for black girls. Dietary patterns in this cohort confirmed that black girls had significantly higher mean intakes of all sweetened drinks (regular soda and other calorically sweetened drinks) than white girls.

GUTS was the largest cohort. Higher SSB intakes in children aged 9 to 14 years were associated with eating fried food away from home (FFA), a proxy for fast food (Taveras et al., 2005). Children who never ate FFA averaged ~ 2 servings SSB a day. Children who ate FFA 4 to 7 times a week had closer to 3 servings of SSB a day, on average.

In **GUTS II**, a different cohort of children aged 9 to 16 years at baseline, regular soda and sports drink intakes were higher among boys than girls, but intakes of regular soda decreased over time. 14% of boys reported more than 1 serving/day (8oz or ~240g) of regular soda in 2004, falling to 10% by 2008. 7% of girls had more than 1 serving/day of regular soda in 2004, falling to 4% by 2008. Sports drinks were more popular among boys than girls, with 4% of boys reporting more than 1 serving/day of sports drink in 2004, rising to 7% in 2008. Less than 2% of girls reported more than 1 serving/day of sports drink at any survey.

SSB intakes in the slightly older **IDEA & ECHO** cohorts were measured at baseline (14.6 years) in 2006/7 and 2007/8 and two years later. Males drank more SSBs than females at baseline (males mean 1.02 servings per day vs females mean 0.70 servings per day, serving size not stated). Changes in SSB intake over the two years were small and not significant.

In **Project EAT** junior and senior high school students (mean age 15 years in 1988/99) were followed up after 10 years. Baseline intakes, *for the 75% of participants who were not overweight at baseline*, showed that males drank more SSBs than females (Males mean 1.3 servings per day SD 0.9 vs females mean 1.0 servings per day SD 0.9, serving size not stated). As intakes at baseline and follow-up were measured with a different FFQ, absolute change in SSB intakes in Project EAT is uncertain. However, in the not overweight sample, SSB intakes appeared to fall for both sexes by young adulthood. Males reported slightly above 1 serving/day, and females had slightly below 1 serving/day, on average.

In childhood and adolescent cohorts outside the USA, consumption of SSBs was less widespread. Between 11% and 50% of each cohort did not report SSBs on the days surveyed. In younger cohorts the highest category of SSB intake was 1+ servings per day, but more children reached this level of intake by teenage. In the older cohorts boys had higher intakes of SSBs than girls, and increased their intake during their later teens, whereas girls seemingly stabilised or even decreased their intake (sometimes switching to diet drinks). Boys aged 17 years in **DONALD** had the highest mean intake of SSBs at 455g/day. Reported mean intakes were generally lower than in USA cohorts at the same age, with a wider variance.

In **Bogotá School Children** study in Colombia children aged 5 to 12 years old were surveyed. Soda was investigated separately as a food item within the "Snacking" pattern. At baseline 11% of children were non-consumers of soda, 19% consumed soda less than once/month. 10% consumed soda more than once per day.

In the Danish **EYHS**, at baseline (mean age 9.6 years) 47% of children were non consumers of SSB, while 13% consumed more than one 12 oz. (~330g) serving per day. Intakes of SSBs at baseline showed a wide variance (mean 154.0g/day SD 204.9g/day) and were lower than mean intakes reported at similar ages in the Framingham Children's study (~190g/day) and in the NGHS (white girls ~214g/day). By first follow-up (mean age 15.7 years) half the children were non-consumers of SSB but the percentage who consumed more than one SSB serving per day (49 children) doubled to 26%.

In the Australian **RAINE** study 11% of the teenagers surveyed at baseline (n = 1,667, age 14 years) and follow-up (n= 1,294, age 17 years) did not consume SSBs. Boys consumed more SSBs than girls at age 14 years (mean 324g per day vs mean 288g per day at age). At 17 years SSB intakes increased in boys (mean 390g per day) but decreased in girls (mean 246g per day).

Among consumers, average SSB intake was 335g/day or ~1.3 servings/day, providing ~ 5% of TEI.

In the German **DONALD** study, data from 1,316 dietary records from 244 children in the age range 9 to 18 years who had returned at least four dietary records were used to assess beverage intakes. 80% of boys' records and 72% of girls' records showed consumption of regular soft drinks. Intakes of RSD were higher for boys (mean 277g/day SD 296g/day at mean age 11.9 years) than girls (mean 243g/day SD 273g/day at mean age 11.8 years) but showed a wide variance. At five year follow-up, girls' intake of regular soft drinks was stable (mean 240g/day SD 290g/day P<0.001 at mean age 16.8 years). Boy's intake of regular soft drinks increased (mean 455g/day SD 498g/day P <0.05 at mean age 16.8 years). This level of intake is similar to 16.5 year old white girls from the **NGHS** (~427g/day). Averaged over 5 years, RSD represented one quarter of all beverage intake in **DONALD** (compared with half of all beverages in the **Framingham Children's study**) and, like the Australian **RAINE** study, contributed about 5% of TEI.

5.9.3 Diet beverage intakes

The youngest children tended not to drink diet drinks. Mean intakes were less than half a serving /day in all cohorts, tending to rise with age. Girls had diet drinks more often than boys.

In the **Framingham Children's Study** very few children drank diet drinks at ages 3 to 5 years old. By ages 13 to 17 years old mean intake was still zero, although at the 95th percentile diet drinks made up one third of total beverage intake.

White girls in the **NGHS** had higher mean intakes of diet drinks than black girls. White girls steadily increased their diet soda intakes from age 9.5 years (mean 22.g/ day) to 18.6 years (mean 82g /day), the highest mean diet drink intake reported. Black girls' mean intakes of diet soda stayed consistently low at no more than 11.5g/day.

In **GUTS II** approximately 6% of adolescent girls at baseline were daily consumers of low calorie soda compared with only 4% of boys. At older ages girls' intakes of regular soda reduced and they were more likely to be consumers of diet soda.

In the **IDEA & ECHO** cohort too, males consumed less diet soda than females at age 14.6 years and at 16.6 years. For all participants, average intakes of diet soda at baseline were 0.17 servings per day (~40g/ day assuming an 8oz. serving size). Changes in diet soda consumption at follow-up were small and not significant (Laska et al., 2012).

In Australia, diet beverages intakes in the **RAINE** study were similarly low (mean ~ 30g/day) for girls and boys at age 14 years and at 17 years.

In the German **DONALD** study diet soft drinks were consumed infrequently and in small quantities (Boys mean 25g/day at age 11.9 years, girls mean 15g/day at age 11.8 years). Although intake of diet drinks increased at follow-up (Boys mean 32g/day at age 16.8 years, girls mean 41g/day at age 16.8 years) the changes were small and only significant for girls. (Libuda et al., 2008)

5.9.4 Sugar sweetened beverage intakes and adiposity outcomes

In all but two of the nine cohorts which investigated sugar-sweetened beverages, such drinks were linked with adverse adiposity outcomes. In some studies associations were no longer significant after adjusting for energy intake, which suggests that the contribution of SSBs to energy intake is a factor.

The two studies which found no association between SSB intakes and adiposity outcomes were from USA cohorts with the youngest and the oldest aged participants at baseline. In the **Framingham Children's study**, some of the youngest children (ages 3 to 9 years) did not consume SSBs at all. When tertiles of SSB intake were compared, no significant associations between baseline SSBs intakes and any measure of body fat in adolescence (15 to 17 years) were seen (Hasnain et al., 2014). It is likely that the study was statistically underpowered to find such associations; the authors conceded that the small sample size (98 children) and homogeneity of the cohort were limitations.

In **Project EAT**, over one third of participants who were not overweight at baseline were overweight as adults by the 10 year follow-up. The exact time at which participants became overweight is unknown (overweight status was assessed at baseline and 10 year follow-up only) but during this period the reported frequency of SSB intake (servings/day) decreased. (This introduces the possibility of reverse causality - did some participants decrease their SSB intakes as a way of controlling their weight?) As intakes measured by the youth FFQ at baseline and the adult FFQ at follow-up were only moderately correlated (Larson et al., 2012), the difference between the two is not be a reliable measure of the apparent decrease. Among participants who were not overweight at baseline, neither baseline SSB intake in adolescence nor 10 year change in SSB intake, adjusted for energy intake, predicted the risk of overweight in young adulthood. **Project EAT** used self-reported measures of height and weight, which, as the authors acknowledged, may have introduced reporting bias. (Quick et al., 2013).

.

Table 5-11 Sugar sweetened beverages and adiposity outcomes

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
Bogotá School children Shroff 2014	7/9	**961** Boys 468 Girls 493	8.6	2.5	Soda intake at baseline ≥ 1 time/day compared with Never	Ow and obesity at baseline = 22%	Annual change in BMI = +0.20kg/m^2 per year 95% CI 0.04 to 0.36, P trend = 0.01 SKF ratio = +0.014 per year 95% CI – 0.002 to 0.03, P trend = 0.24 WC = 06cm per year 95% CI -0.1 to 1.4, P trend = 0.04	↑ in BMI Significant trend ↑ in waist circumference Significant trend
DONALD study Libuda 2008	7/9	**244** Boys 125 Girls 119	9 to 18 Boys mean 11.9 Girls age 11.8	5	Regular soft drink intake (MJ/ day)	Ow and obesity at baseline Boys 7% Girls 11%	Baseline intake, change in BMI-SDS and change in % body fat	Not significant
DONALD study Libuda 2008 continued	7/9	**244** Boys 125 Girls 119	9 to 18 Boys mean 11.9 Girls age 11.8	5	Regular soft drink intake (MJ/ day)	Ow and obesity at baseline Boys 7% Girls 11%	5 year change in intake, adjusted for RESIDUAL Energy Change in BMI-SDS boys: β = 0.009, P = 0.71 girls: β = 0.055, P = 0.08 Change in % body fat: boys: β = 0.046, P = 0.87 girls: β = - 0.45, P = 0.19	Not significant
EYHS Zheng 2015	7/9	**358** Boys 157 Girls 201	9.6	6	SSBs – 100g/day regular soft drinks,	BMIz at baseline Mean 0.4 SD 1.1	Baseline intake and 6 year change in: BMIz β = 0.05 SE 0.02	↑ BMIz change, ↑∑4SF change Not significant after

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
					lemonade and fruit drinks, 100g/day	BMIz at follow-up Mean 0.5 SD 1.0	P=0.02 ∑4SF β = 0.86mm SE 0.37 P=0.02 WC β = 0.22cm SE 0.15 P=0.14	adjustment for Energy
EYHS Zheng 2014	7/9	358 **283** in analyses Boys 125 Girls 158	9.6	12	SSBs - regular soft drinks, fruit drinks and cordials, > 1 12oz.serving/day compared with none	Mean BMI at baseline 17.3 kg/m^2	Baseline intake and 12 year change in: BMI β = 1.42 SE 0.68 P=0.29 WC, β = 0.80cm SE 2.02 P=0.69 ∑4SF β = 0.98mm SE 3.63 P=0.79	Not significant
EYHS Zheng 2014	7/9	358 **187** in analyses Boys 89 Girls 98	9.6	6	Increase in SSB intake compared with no change	Mean BMI at baseline 17.3 kg/m^2	Increase in intake and 6 year change in: BMI β = 0.91 kg/m^2 SE 0.57 P=0.09 WC β = 2.72cm SE 1.53 P=0.04 ∑4SF β = 3.54 mm SE 3.97 P = 0.38.	↑ BMI change, ↑ WC change Not significant after adjustment for Energy
	7/9	358 **187** in analyses Boys 89 Girls 98	15	6	SSBs - regular soft drinks, fruit drinks and cordials, > 1 12oz.serving/day compared with none	Mean BMI at baseline 21.2 kg/m^2	Baseline intake and 6 year change in: BMI β = 0.92 kg/m^2 SE 0.54 P = 0.046 WC β = 2.69 cm SE 1.45 P = 0.04 ∑4SF β = 3.20 mm SE 3.90 P = 0.42.	↑ BMI change ↑ WC change Not significant after adjustment for Energy

Study & Paper	**NOS stars**	**Popln.**	**BL age (yrs.)**	**F' up (yrs.)**	**Exposure**	**Adiposity at baseline**	**Adiposity outcomes**	**Findings**
Framingham Children's Study Hasnain 2014	7/9	98	3 to 9	12	SSB tertiles T1 mean 2.8 oz./day vs T2 mean 5.8 oz./day vs T3 mean 10.7 oz./day	Mean BMI 16.2 kg/m^2	BMI P = 0.42 % body fat by DXA P= 0.93 WC P = 0.35 ∑4SF P = 0.98	Not significant
GUTS II Field 2014	4/9	**7,559** Boys 3,438 Girls 4,121	Boys 12.9 Girls 13.0	2 to 3	Regular soda servings/day	Mean BMI Boys 20.2 kg/m^2 Girls 20.0 kg/m^2 Ow and obesity at baseline Boys 24% Girls 16%	Baseline intake and BMI change Boys: β = 0.05 95% CI -0.06 to 0.16 Girls: β = - 0.00 95% CI -0.10 to 0.10	Not significant
		As above	As above	As above	Regular soda change in servings/day	Mean BMI Boys 20.2 kg/m^2 Girls 20.0 kg/m^2 Ow and obesity at baseline Boys 24% Girls 16%	Change in intake and BMI change Boys: β = 0.08 95% CI -0.06 to 0.22 Girls: β = 0.10 95% CI -0.03 to 0.22	Not significant
		As above	As above	As above	Sports drink servings/day	As above	Baseline intake and BMI change Boys: β = 0.15 95% CI -0.04 to 0.34 Girls: β = 0.23 95% CI 0.05 to 0.41	Not significant

Study & Paper	**NOS stars**	**Popln.**	**BL age (yrs.)**	**F' up (yrs.)**	**Exposure**	**Adiposity at baseline**	**Adiposity outcomes**	**Findings**
GUTS II Field 2014	**4/9**	**7,559** Boys 3,438 Girls 4,121	Boys 12.9 Girls 13.0	2 to 3	Sports drink change in servings/day	Mean BMI Boys 20.2 kg/m^2 Girls 20.0 kg/m^2 Ow and obesity at baseline Boys 24% Girls 16%	Change in intake and BMI change Boys: β = 0.29 95% CI 0.07 to 0.50 Girls: β = -0.09 95% CI -0.30 to 0.10	↑ in sports drink ↑ BMI change in boys Significant No energy adjustment
IDEA & ECHO Laska 2012	8/9	693 Male 327 Female 339 **535** in analyses Male 276 Female 286	14.6	2	SSB intake servings/day	Mean BMI at b'line Males 22.1kg/m^2 SD 5.1 Females 21.9kg/m^2 SD 4.9 Mean BMI at f'up Males 23.4kg/m^2 SD 5.1 Females 22.8kg/m^2 SD 4.6	Change in intake (servings/day) adjusted for TEI, <u>Change in BMI</u> Males β = 0.27 s.e.0.10, P <0.05 Females β = - 0.05 s.e.0.17, P = 0.75 <u>Change in % bf</u> Males β = 0.73 s.e.0.21, P <0.05 Females β = 0.04 s.e.0.35, P = 0.91	↑BMI ↑% body fat in Males only, after adjustment for Energy Significant
NGHS Striegel-Moore 2006	8/9	**2,371** girls Black 1,210 White, 1,161	9.5	10	Regular soda intake +100g/ day change	n/a	BMI change = estimate 0.011 SE 0.005 P <0.05.	↑ regular soda ↑BMI change in girls Significant

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	SSB servings/day	Overweight at 15 yrs.: ~25%, All o/w at 25.4 yrs.: Males 56.1%, Females 47.5% If not o/w at start, o/w at 25.4 yrs.: Males 45.4%, Females 34.2%	Baseline SSB and risk of ow, adjusted for TEI Boys: OR 1.09, 95% CI 0.91 to 1.31 Girls: OR 1.14, 95% CI 0.95 to 1.38	Not significant
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* Males 756 Females 887	15	10	SSB change in servings/day	As above	Change in SSB and risk of ow, adjusted for TEI Boys: OR 1.06, 95% CI 0.94 to 1.21 Girls: OR 1.09, 95% CI 0.93 to 1.27	Not significant

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
RAINE Ambrosini 2013	6/9	1,667 Boys 867 Girls 800 **1,009** in analyses Boys 494 Girls 545	14	3	SSB intake change Move to top tertile of SSB intake (>1.3 servings per day, 331 to 2,876 g/ day) compared with reference, stay in lowest tertile (0 to 0.5 servings per day, 0 to 130g/ day)	Overweight/ obese at baseline Boys ~27% Girls ~25% Overweight/ obese at follow-up Boys ~23% Girls ~24%.	<u>Overweight/ obesity</u> Boys OR = 0.8, 95% CI 0.3 to 2.1, P = 0.76. Girls OR = 3.8, 95% CI 1.5 to 9.3, P = 0.004. <u>BMI change</u> Boys.+0.8%, 95% CI -1.3% to 2.9%, P = 0.46 Girls +3.6%, 95% CI 1.5% to 5.8%, P = 0.001 <u>WC change:</u> Boys +1.4% change 95%CI 0.2% to 2.3% P= 0.019 Girls +0.9% change 95%CI -0.2% to 2.0% P = 0.09	↑ SSBs ↑ overweight/ obesity risk in girls only. Significant after adjustment for dietary pattern scores ↑ BMI in girls only Significant after adjustment for dietary pattern scores ↑ waist circumference in boys only Significant after adjustment for dietary pattern scores
SUMMARY Sugar sweetened beverages were associated with adverse adiposity outcomes in boys and in girls in most studies.								

Papers from three other USA cohorts reported significant associations between increased SSB intakes and BMI outcomes, for girls and/or boys.

Among 2,371 girls from the **NGHS**, in models which adjusted for other beverages as well as age, race and energy intake, each 100g/ day increase in regular soda intake predicted a 0.011 kg/m^2 increase in BMI (Striegel-Moore et al., 2006).

In the **IDEA & ECHO** study, no longitudinal associations were seen among females, but among males a one serving/day increase in SSB intake between baseline and 2 year follow-up, adjusted for energy intake, predicted an additional 0.3 kg/m^2 in BMI as well as an extra 0.7% body fat, significant at $P = 0.05$. At $\alpha = 0.003$, used to correct for the number of tests, only the association between increased SSB intake and males' body fat% remained significant. (Laska et al., 2012).

In the large **GUTS II** cohort, in models which controlled for age, BMI at the start, TV viewing and physical activity (but not energy intake), baseline intake and changes in regular soda intake did not predict BMI change in either girls or boys. However, intakes of sports drinks, a type of SSB marketed in large portion sizes, did have associations with future BMI. Baseline intakes of sports drinks predicted greater BMI at follow-up for girls (although < 2% girls had > 1 serving/day). Change in intake of sports drinks predicted BMI outcomes for boys. In a final model which took account of baseline *and* change in sports drink intakes, each additional serving/day predicted an increase in BMI of approximately $0.3kg/m^2$ for both girls and boys (Field et al., 2014)

In cohorts outside the USA, there was also evidence that higher or increased SSB intakes were linked with larger BMI outcomes. In several studies SSB intakes were associated with greater gains in waist circumference, even after adjustment for energy intake, suggesting that SSBs may contribute to central adiposity.

In the **Bogotá School Children** study in Colombia, after adjustments including TEI, children in the highest category of soda intake (≥ 1 time/day) compared to those in the lowest category of soda intake (Never) at ages 5 to 12 years had significantly higher annual gains in BMI ($+0.20kg/m^2$) and waist circumference (+0.06cm), a measure of central adiposity. However, baseline soda intake was not associated with change in SFT ratios (Shroff et al., 2014).

In the chosen sample from the German **DONALD** cohort overweight/ obesity prevalence was comparatively low at baseline (7%) and follow-up (10%). Baseline consumption of RSD was not linked with BMI-SDS or % body fat in boys or girls, in either cross-section or longitudinal analyses. Increases in all energetic beverage intake (*regular soft drinks and fruit juice*) during the 5 year study, after controlling for residual energy, predicted increased BMI-SDS for girls, but not for boys. When energetic beverages were considered separately, small BMI-SDS gains

for girls were attributed to increased fruit juice consumption. (See 5.5.3)

For each MJ increase in regular soft drink intake during the 5 year study, although girls BMI-SDS increased by +0.055 this was not significant (P= 0.08). No association was found between change in regular soft drink intake and change in BMI-SDS for boys. Change in regular soft drink consumption did not predict concurrent change in % body fat in either sex (Libuda et al., 2008).

In the Danish arm of the **EYHS**, SSB intake at age 9 years predicted higher BMIz and higher sum of four skinfolds at six year follow-up, controlling for baseline age and adiposity measures, sex, physical activity, socio-economic and pubertal status. Results were similar after further adjustment for TEI but no longer significant (Zheng 2015). SSB intake at age 9 years did not predict any measure of adiposity by the twelve year follow-up, probably as many dietary and lifestyle habits change over such a long period (Zheng 2014).

Teenagers in the highest category of SSB intake (>1 serving/day) at age 15 years, compared with non-consumers, had larger gains in BMI (+0.92)and waist circumference (+ 2.69cm) by age 21 years. Children who increased their SSB intakes between the ages of 9 and 15 years, compared with those whose SSB intakes stayed the same, also experienced greater gains in BMI (+ 0.91) and waist circumference (+2.72cm) by 21 years, although no association was seen between increased SSB intakes and sum of four skin folds . After adjustment for TEI associations were no longer significant. (Zheng 2014).

The **RAINE** study used a model which adjusted for age, pubertal stage, physical fitness, dietary misreporting, maternal education, family income and baseline BMI plus healthy and Western dietary pattern scores. Girls who moved into the highest tertile of SSB intake (>1.3 servings per day) between 14 years and 17 years of age, had a significantly higher BMI (+3.6%) at age 17 years and a significantly greater risk of overweight/obesity (OR 3.8) than girls who stayed in the lowest tertile of SSB intake (0 to 0.5 servings/day). Neither of these associations were found in boys.

Teenagers who moved into the highest tertile of SSB intake had a higher waist circumference at follow-up (Girls +4.2%, boys +2.3%), but after full adjustment including dietary pattern scores, associations were attenuated and only remained significant for boys (+1.4%) (Ambrosini et al., 2013)

5.9.5 Diet beverage intakes and adiposity outcomes

The cohort with the largest sample size (**GUTS II**) found that baseline intake of diet soda, without energy adjustment, was predictive of a greater increase in BMI, but for girls only. No associations between diet beverages and future adiposity were seen in other cohorts.

In **GUTS II** in longitudinal analyses which adjusted for variables including baseline BMI, but not energy intake, each serving/day of diet soda at baseline predicted an increase of 0.19kg/m^2 in BMI at follow-up for adolescent girls. Results for boys were not significant. (Field et al., 2014)

In the **Framingham Children's study**, 44 of the 98 children in the sample did not consume diet beverages at baseline. It is doubtful whether this small sample had enough power to detect change in adiposity. In longitudinal analyses (adjusted for baseline age and anthropometry, energy from fat, television and video time, other beverage intakes, mother's education and BMI) no significant trends were observed between tertiles of diet beverage intake at baseline (3 to 9 years) and any measure of adiposity at follow-up (15 to 17 years). (Hasnain et al., 2014)

In the **IDEA & ECHO** cohort, there were cross-sectional associations between baseline diet soda and BMI and % body fat in females, which were partly attributed to reverse causality (girls who are overweight may choose to drink diet soda to control their weight). In longitudinal analyses the two year change in diet soda intake (servings/day), which was small and not significant, did not predict BMI or % body fat outcomes in males or females. (Laska et al., 2012)

Similarly, in the larger **NGHS** cohort, where white girls steadily increased their diet soda intake, but black girls drank comparatively little, there was no evidence of a longitudinal association between change in diet soda intake (100g/day) and BMI outcomes. Models adjusted for age, race, energy intake and consumption of other types of beverage. (Striegel-Moore et al., 2006).

Table 5-12 Diet beverages and adiposity outcomes

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
GUTS II Field 2014	4/9	**7,559** Boys 3,438 Girls 4,121	Boys 12.9 Girls 13.0	2 to 3	Low calorie/diet soda servings/day	Mean BMI Boys 20.2 kg/m^2 Girls 20.0 kg/m^2 Ow and obesity at baseline Boys 24% Girls 16%	BMI change Boys: β = 0.16 95% CI -0.02 to 0.34 Girls: β = 0.19 95% CI 0.08 to 0.29	Baseline diet soda ↑ BMI change in girls. Significant No energy adjustment
		As above	As above	As above	Low calorie/diet soda change in servings/day	As above	BMI change Boys: β = 0.20 95% CI -0.02 to 0.42 Girls: β = -0.06 95% CI -0.20 to 0.09	Not significant
Framingham Children's Study Hasnain 2014	7/9	**44**	3 to 9	15 to 17	Unsweetened/ diet beverages T1 mean 0 oz./day vs T2 mean 0.4 oz./day vs T3 mean 2.3 oz./day	Mean BMI 16.2 kg/m^2	BMI P = 0.44 % body fat by DXA P = 0.58 WC P = 0.40 ∑4SF P = 0.27	Not significant

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
IDEA & ECHO Laska 2012	8/9	693 Male 327 Female 339 **535** in analyses Male 276 Female 286	14.6	2	Diet soda intake servings/day	Mean BMI at b'line Males 22.1kg/m^2 SD 5.1 Females 21.9kg/m^2 SD 4.9 Mean BMI at f'up Males 23.4kg/m^2 SD 5.1 Females 22.8kg/m^2 SD 4.6	Change in intake (servings/day) adjusted for TEI, <u>Change in BMI</u> Males β = -0.09 se 0.24 P = 0.72 Females β = 0.10 se 0.24 P = 0.67 <u>Change in % bf</u> males β = 0.09 se 0.79 P = 0.91 females β = 0.55 se 0.36 P = 0.13	Not significant
NGHS Striegel-Moore 2006	8/9	**2,371** girls Black 1,210 White, 1,161	9.5	10	Diet soda intake +100g/ day change	n/a	BMI change = estimate -0.010 SE 0.013	Not significant
SUMMARY In the largest cohort sample, baseline intakes of diet soda were linked to a greater increase in BMI for girls only. Diet soda was not significantly associated with adiposity outcomes elsewhere.								

5.10 Discussion

This systematic review retrieved over twenty papers that reported associations between children's' quantified food and drink intakes and later adiposity outcomes. The range of included food and drink intakes was broad, but specific foods were often investigated in only one or two cohorts, as discussed below.

5.10.1 Whole Grains

Average intakes of whole grains in the **Project EAT** and **NGHS** cohorts were below the USA recommendation of 3 to 5 servings a day, at 1 serving/day or less. The UK National Diet and Nutrition Survey (NDNS) 2008-11 shows that median intakes for children/teenagers in the UK are also low,13g/day based on dry weight. UK children/teenagers with higher intakes of whole grain had higher intakes of total energy and dietary fibre and obtained a greater proportion of their energy intake from carbohydrates rather than fat (Mann et al., 2015).

Project EAT and **NGHS** found that higher intakes of whole grains mitigated against future adiposity for adolescent girls. This accords with a systematic review of whole grain intakes in adults, which saw significantly lower weight gains over time for adults who had 48 to 80g/day (3 to 5 servings) compared with those who never or rarely consumed whole grain (Ye et al., 2012). Whole grains contribute to dietary fibre, shown to help satiety in adults, which may be part of the mechanism that promotes better weight regulation (Pereira and Ludwig, 2001).

Mean intakes of whole grain were 63g/day in the **DONALD** cohort at puberty, so it was surprising that no association with body fat or BMIz was found. Possibly the study was underpowered to detect any effect as the sample was small (n = 215). The authors cautioned that participants were typically of high socio-economic status (> 55% of fathers of the children had ≥12 years education) so may not represent an "at-risk" population. Baseline whole grain and dietary fibre intakes already met recommendations for the age group; maybe there was little room for an increase that would benefit body composition.

5.10.2 Dairy foods and milk

Included studies investigated dairy foods and milk together (dairy products) or milk alone. Higher intakes of full-fat dairy products were associated with more favourable adiposity outcomes in **ALSPAC** children, while low fat dairy products including milk were linked to more favourable BMI outcomes for teenaged girls in the **NGHS**. Higher milk intakes in younger children in the **Framingham Children's study** were associated with lower body fat outcomes, but with such a small sample size (n=98) and multiple testing of other adiposity outcomes

(BMI, sum of 4 skinfolds, waist circumference) that were not significant at p = 0.05, this seems uncertain. In other cohorts no association between milk intakes and adiposity outcomes were seen after adjustments for energy intake. In **ALSPAC** flavoured milk only increased the risk of adverse body fat outcomes for children who were already overweight.

These findings broadly agree with evidence reported elsewhere. A systematic review in adults and children found inconsistent results, albeit suggestive of a protective effect of dairy intakes on risk of overweight/obesity (Louie et al., 2011). A systematic review and meta-analysis of 22 childhood studies saw a net neutral effect of dairy intake on adiposity in children to the age of 11 years, with a modest protective effect in teenagers (Dror, 2014). A recent critical review found that milk and dairy were consistently not associated, or were inversely associated, with adiposity outcomes in children. Adjustments for energy tended to convert inverse relationships to neutral. The authors concluded that there is little reason to limit children's intake of milk and dairy products (Dougkas et al., 2019).

Dairy products contribute vital micronutrients and protein to the diets of growing children, but also add fat, while sweetened yogurts often contain high quantities of added sugar (Moore et al., 2018). Dairy products are more energy dense than liquid milk, but as pointed out by Dror, few studies have disaggregated milk from dairy to consider dairy products alone. It would be interesting to find out whether there is a U-shaped longitudinal relationship between dairy products (without milk) and adiposity and if there is an optimal intake.

5.10.3 Juice, fruit and vegetables

Four studies investigated juice intake. Juices add vitamins and minerals to the diet and are perceived as healthy, but there are concerns that excessive consumption may also add too much sugar to the diet, increasing obesity risk. The findings from the **DONALD** study seem to support this concern as additional servings of juice were associated with increased (adverse) gains in BMI-SDS in girls. Children in this German cohort had the highest juice intakes by follow-up (~ 250g/day), double that seen in NGHS girls (~125g/day) but close to intakes reported in GUTS (~210g/day), where no associations between juice intakes and adiposity outcomes were in evidence. Higher juice intakes in younger children in the **Framingham Children's study** were linked with smaller (beneficial) waist circumference but the sample size was small (n=98) and this seems less certain.

Three included studies from the USA looked at fruit intakes and vegetable intakes and future adiposity, with different results. **Project EAT** found no significant associations between baseline fruit intakes or vegetable intakes and future overweight. Measures of absolute change in intake in **Project EAT** were invalid, so the reported association between increased

vegetable intake and reduced risk of overweight in males is unreliable. The larger **GUTS** cohort found that higher fruit intake resulted in a very small increase in girls BMI z score three years later, and higher vegetable intakes predicted more favourable (lower) BMIz scores for boys. In both studies, random error and bias may have impaired measurement accuracy, as baseline diet was measured with a YAQ, and BMI was calculated from self-reported height and weight. The higher quality **NGHS** used annual 3 day FDs and calculated BMI from measured height and weight, and had a longer follow-up, so may offer a more reliable estimate. Girls in the **NGHS** cohort who consumed ≥ 2 servings of fruit/day compared with < 1, had significantly lower BMI outcomes by final follow-up, but vegetable intakes did not predict BMI.

Increasing fruit and vegetables intake is often put forward as way to avoid excess adiposity/obesity, based on the satiating effect of increased fibre, the moderation of dietary glycaemic load and replacement of more energy dense foods. A systematic review of adult cohort studies found that adults in the highest categories of intake had significantly reduced risks of adiposity, fruit (OR =0.83), vegetables (OR = 0.83), fruit and vegetables (OR = 0.91) (Schwingshackl et al., 2015).

Although longitudinal studies of adults confirm that higher fruit and/or vegetable intakes are associated with slower weight gain, significant inverse relationships were only seen in two of four children's cohorts included in another systematic review, possibly because of short follow-up or inaccuracies in measuring diet (Ledoux et al., 2011).

5.10.4 Fish

Fish intakes were investigated in only one cohort (**FAMS**). The evidence of any association with adiposity outcomes was unconvincing. Higher fish intake was linked with very small differences (beneficial) in waist circumference, but not with any other measures of adiposity. Asian girls consumed much more fish than white girls. Some of the difference in waist circumference might be explained by ethnicity, even after adjustment.

After more adjustments including parental BMI, education level and factors related to energy balance and fat partitioning, the association was no longer significant. The authors ventured that null findings were due to overfitting, which is a consideration, especially with such a small sample size (n = 100).

5.10.5 Convenience foods

Convenience foods were investigated in only one cohort (**DONALD**). Higher baseline intakes of CF with a high energy density, such as pizzas, were associated with reduced dietary quality and

with small, adverse changes in body fat % in boys. However, the results are based on a modest sample (n = 585), and children in the **DONALD** study are of a higher socio-economic status than the German population in general. As the authors acknowledge, these children probably do not represent the highest intakes of energy dense convenience foods, or extremes of weight status. They advised that families should check nutrition labels, choosing lower energy density convenience foods if possible, and should compensate for high energy density CF by including low energy density foods such as fruit and vegetables in the diet.

The pre-prepared convenience foods considered in the **DONALD** study also fit the description of "ultra-processed" foods (manufactured food products that are ready-to-heat or ready-to-eat). A systematic review of ultra-processed foods (snacks, fast foods, soft drinks and sweetened beverages, sweets, chocolates and ready-to-eat cereals as well as convenience foods) found that most studies saw a positive (adverse) association with body fat during childhood and adolescence (Costa et al., 2018). Costa et al observed that many studies adjusted for TEI, noting that as ultra-processed foods are often energy dense, this may have over-adjusted for the exposure. Instead they advocated adjusting for residual energy, i.e. Energy from all other sources apart from the exposure of interest. The **DONALD** study was the only study that did this.

5.10.6 Snacks

GUTS looked at energy dense snack foods in general, while the **Bogalusa Heart Study** looked specifically at candy or confectionery. Neither study found longitudinal associations with adverse adiposity outcomes.

The large **GUTS** cohort used a FFQ with a comprehensive list of snack foods, asking about usual intake in the past year. On average children reported having 3 snacks/day, representing 18% of TEI. Higher baseline snack intake, without adjustment for TEI, inversely predicted BMIz score for girls (so beneficial).

To date, it seems that no comprehensive systematic review of children's snacking and adiposity outcomes has been published, but similar aged USA children were surveyed in NHANES 1999-2004. Dietary data were obtained from 24-HDR, defining snacks by eating occasions(Keast et al., 2010). Snacks accounted for 21% of daily energy intake, which lends credibility to the snack quantities reported in **GUTS**. The NHANES survey found that children who snacked, compared with 19% who did not snack, were less likely to have overweight or obesity, although reverse causality cannot be ruled out as the NHANES survey was cross-sectional.

Candy certainly fits the description of an energy dense, nutrient poor, snack type food. Mean intakes of candy (46g/day, about the same as a UK Snickers bar) in the **Bogalusa Heart Study** in the 1970s were far higher than the average 35g /day reported by NHANES 1999-2004. The higher level of candy intake may be due to measurement error, or it may be that candy consumption the 1970s really was much higher, with relatively few participants exposed to moderate intakes of candy. Recognising these limitations, O'Neill et al called for evidence from larger, more nationally representative cohorts and more up-to-date datasets. A later systematic review observed no longitudinal associations between confectionery consumption in children/adolescents and outcomes of overweight or obesity (Gasser et al., 2016).

Undoubtedly confectionery adds sugar to the diet, but if children do not eat much candy, what do they eat instead? In the **NGHS** the highest mean candy intakes (equivalent to ~ 2 Snickers bars per week) were found in the dietary patterns with the highest sucrose intakes, which had adverse associations with final waist circumference outcomes among white girls. Black girls with the lowest mean candy intake (15.8g/day) compensated with the highest mean intakes of sugar-sweetened and diet drinks, coffee/tea, yogurt, crackers and pretzels. White girls with the lowest mean candy intake (12.4g/day) had the highest mean intakes of yogurt, cereals, breakfast grains, soups, fruit, green salad, vegetables and (not fried) potatoes. This so called "Healthy" pattern was protective against obesity, which lends some support to O'Neill's suggestion that "modest" amounts of candy can be added to an otherwise nutrient rich diet without long-term adverse health effects, although the authors did not mention dental health. By follow-up in adulthood, fewer participants in the **Bogalusa Heart Study** consumed candy, and the average quantity consumed was far lower. A recent systematic review of added sugars and sugary foods, including 24 longitudinal studies, observed a similar trend, finding a significant decrease in confectionery intake between adolescence and early adulthood (Winpenny et al., 2017).

5.10.7 Sugar sweetened beverages

The narrative synthesis of sugar sweetened beverages strongly indicates that high intakes of caloric drinks in children and adolescents are linked with adverse adiposity outcomes. This is in agreement with a comprehensive systematic review and meta-analysis of evidence from cohort studies and RCTs which concluded that SSB consumption "promotes weight gain" in adults and in children (Malik et al., 2013). SSBs may contribute to weight gain by adding surplus energy to the diet in the form of sugar, but with low satiety effects that result in little compensatory reduction in energy intake at later meals (Malik et al., 2006). An example of this was reported by the **NGHS** cohort; for each extra 100g of regular soda, girls' average TEI

increased by 82 calories, but regular soda contributed only half of those calories (Striegel-Moore et al., 2006).

In another systematic review of SSB intakes in children under 6 years old, five studies found associations with increased BMI, waist circumference or increased risk of overweight in later childhood, but two studies found no association (Pérez-Morales et al., 2013). In response to such contradictory findings, a review of systematic reviews of SSBs and obesity among children was conducted (Keller and Bucher Della Torre, 2015). Nine reviews concluded that SSBs and obesity outcomes were associated, four reviews did not. The best quality reviews were among those with "discrepant" results, but no review found an inverse relationship between SSBs and obesity. Keller and Bucher Della Torre subsequently appraised the methodological quality in cohort and experimental studies that looked at SSB consumption and obesity risk in children and adolescents. They found that high quality studies saw an association (5 studies) or had mixed results (4 studies). The other 23 studies had at least one methodological flaw, finding positive associations (7 studies), mixed results (9 studies) or no association (7 studies) (Bucher Della Torre et al., 2016). Methodological issues included measuring dietary intake with "sub-optimal" non-quantitative FFQs and under-representation of weekend days (children's SSB intake may be higher then). Attrition impacted internal and external validity, particularly in cohort studies, as participants who gained more weight were more likely to be lost to follow-up, which weakened the association between the exposure (SSBs) and the outcome of interest. Other difficulties included the heterogeneity in defining SSBs and not indicating portion size in analyses making it hard to make direct comparisons. We encountered similar shortcomings in SSB studies in this systematic review.

5.10.8 Diet beverages

Diet beverages were only associated with adiposity outcomes in the **GUTS** study. BMI at baseline was correlated with baseline intakes of diet soda for girls (r = 0.25) and boys (r = 0.21), so even though the models adjusted for baseline BMI there is a possibility of reverse causality. No associations were seen in other cohorts, perhaps because baseline diet beverage consumption was low, especially at younger ages, and increases in intake were small.

5.10.9 Conclusion

Several studies reported significant associations between food and drink intakes and adiposity outcomes, agreeing on the direction of influence. Higher intakes of whole grains and dairy foods were beneficial. Higher intakes of sugar-sweetened beverages were adverse. Evidence for other foods and drinks and adiposity outcomes was conflicting (juice, fruit, vegetables, diet beverages) with a lack of consensus between studies about the direction of influence. Some foods were not directly comparable (snack foods) or were investigated in only one cohort (fish, convenience foods).

Energy density was a recurring theme. Foods and drinks which may add surplus sugar or fat to the whole diet if consumed in excess (SSB, flavoured milk, juice, convenience foods, snack foods) were sometimes, but not always, shown to be associated with adverse adiposity outcomes.

Chapter 6 Narrative Review: Dietary patterns, eating habits and multiple predictors, and subsequent adiposity

6.1 Summary

Papers from seven included cohorts took a wider view of diet, exploring dietary patterns or eating habits/usual eating behaviours, and their relationships with future adiposity. Two papers considered non-dietary as well as dietary variables as potential predictors of overweight and obesity.

This chapter (Chapter 6) firstly describes the methods used to determine each type of exposure:

- Dietary patterns (cluster analysis, principal components analysis, factor analysis, and reduced rank regression).
- Eating habits (specific questions or derived from dietary assessment data).
- Multiple predictors (logistic regression or regression tree analysis).

The longitudinal associations (if any) between specific exposures and future adiposity are then narratively synthesised, with results summarised in tables. Some studies adjusted for baseline weight status in their analyses or excluded those who were already overweight at baseline.

Dietary patterns showed that overall diet quality may be a factor in adiposity outcomes, and that risk is not best explained by single foods or drinks. Energy dense dietary patterns (typically high in fat and sugar and low in dietary fibre) and irregular eating or eating habits that promote energy density, predicted adverse adiposity outcomes, but explained only part of the risk. Other important predictors of overweight and obesity included socio-economic, psychological and environmental factors, and physical activity.

6.2 Introduction

Instead of investigating associations between quantified intakes of specific foods or drinks and adiposity outcomes, studies in seven of the 14 included cohorts took a wider view, exploring dietary patterns or eating habits/usual eating behaviours, and their relationships with future adiposity. Two papers considered many dietary and non-dietary exposures as potential predictors of overweight and obesity.
Investigated exposures were:

- Dietary patterns and diet quality scores
- Eating habits
- Multiple dietary and non-dietary predictors

6.3 Dietary patterns

Dietary patterns (DP) are used to consider a matrix of foods which provide nutrients in combination, allowing for synergistic, additive or antagonistic effects on the human body. They may better explain dietary health risks than single foods or nutrients in isolation. Dietary patterns can be derived á priori (beforehand) or á posteriori (afterwards) (Alles et al., 2012).

Á priori DPs are theoretically derived. How well individuals meet a pre-defined "healthy" diet is usually summarised by a score, with higher scores indicating higher diet quality or diversity, or greater adherence to dietary recommendations. The focus of diet quality scores tends to be a selection of foods and nutrients rather than the whole diet (Ambrosini, 2014).

Á posteriori DPs are empirically derived from observed food intakes in a population. Most methods are exploratory and make no assumptions about the benefit or otherwise of the DPs that are identified (Alles et al., 2012). Á posteriori methods include:

- Cluster analysis (CA)
- Principal components analysis (PCA)
- Factor analysis (FA)
- Reduced rank regression (RRR)

CA is a classification method which sorts individuals into the smallest possible number of non-overlapping groups or clusters with similar food intakes (Devlin et al., 2012).
PCA and FA generate several dietary patterns within the studied population, each with factor loadings or weightings that characterise the food/nutrient profile that is typical for that DP. The extent to which an individual's dietary intake matches the identified DP, compared with other individuals, is measured with a z-score which can be used as a predictor for modelling

health risks (Ambrosini, 2014).

RRR is semi-exploratory and uses prior knowledge of diet-disease relationships, seeking patterns of food intake which take account of variations in predefined intermediate variables (such as nutrient intakes) that may have a role in the pathophysiology of a disease (Ambrosini, 2014). This makes RRR dietary patterns better at predicting disease/ health outcomes than either PCA or FA methods (Hoffmann et al., 2004). Like PCA and FA, RRR uses factor loadings which characterise the food/nutrient profile for identified DPs in the population. Individuals are given z-scores, reflecting how well they match each dietary pattern.

Dietary patterns were described in five papers, from four included cohort studies. The five studies used different methods to derive dietary patterns.

In the **NGHS** girls cohort one paper (Berz et al., 2011) employed an á priori approach, investigating effects of adherence to a modified DASH (Dietary Approach to Stop Hypertension) diet quality score on BMI during adolescence and at 10 year follow-up.

Four papers from four cohorts reported á posteriori dietary patterns and investigated their influence on adiposity outcomes. DPs were derived by different methods in each cohort.

ALSPAC used RRR (Ambrosini et al., 2012), **NGHS** girls used CA (Ritchie et al., 2007).

PCA was used in the **Bogotá School Children** cohort (Shroff et al., 2014) and in **Project EAT** (Cutler et al., 2012). No included paper used factor analysis.

Longitudinal findings are summarised in Table 6-1.

Table 6-1 Diet quality scores and dietary patterns and adiposity outcomes

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
NGHS Berz 2011	8/9	2,327 girls Black 1,188 White 1,139	9.5	10	Modified DASH **diet quality** score	BMI Approx. 19kg/m^2	Quintile 5 (highest scores) had lowest BMI at follow-up Q1 BMI 26.3 SD 0.28 Q2 BMI 24.9 SD 0.28 Q3 BMI 25.2 SD 0.28 Q4 BMI 25.3 SD 0.29 Q5 BMI 24.4 SD 0.30 Q1 to Q4 vs Q5, P <0.05	↑ DASH score ↓BMI in girls, adjusted for TEI Significant
ALSPAC Ambrosini 2012	7/9	6,772	7, 10 and 13	At ages 11, 13 and 15	Increase in Energy Dense dietary pattern z-score **Reduced Rank Regression**	Overweight/obese Boys 7.5 yrs. 12.1% Girls 7.5 yrs. 17.5%	1 SD unit increase in Energy Dense z score and later Fat Mass Index z score At all ages, +0.04 SD units, 95% CI 0.01 to 0.07 At 11 years, +0.10 SD units, 95% CI 0.07 to 0.12 At 13 years, +0.09 SD units, 95% CI 0.06 to 0.11 At 15 years, +0.09 SD units, 95% CI 0.06 to 0.12	↑ Energy Dense DP z-score ↑ Fat Mass Index Significant

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
NGHS Ritchie 2007	7/9	2,371 girls Black 1,211 White 1,160	9.5	10	8 Dietary patterns (4 for Black girls, 4 for White girls) **Cluster Analysis**	Not shown	BMI, %body fat and waist circ. Black girls: Not significant White girls: WC at f' up "Healthy" DP = + 9.7 cm SEM 0.85 vs. "Sweets & Snacks" DP = + 11.7 SEM 0.40, P = 0.037	Less Energy Dense pattern ↓ WC in white girls Significant
Bogotá School Children Shroff 2014	7/9	961	5 to 12 Mean 8.6	2.5	"Snacking" dietary pattern **Principal Component Analysis**	Mean BMI Approx. 16.5kg/m^2 Overweight 18% Obese 4%	Quartile 4 (highest adherence to "snacking" DP) had highest gains in: BMI +0.09kg/m^2 per year 95% CI -0.01, 0.19, SKF +0.012mm per year 95% CI 0.001, 0.022 Q1 vs Q2 vs Q3 vs Q4 BMI P trend = 0.05 SKF P trend = 0.05 WC P trend = 0.59	↑ "Snacking" pattern ↑BMI ↑SKF Significant
Project EAT Cutler 2012	4/9	Younger boys 528 Older boys 1,257 Younger girls 534 Older girls 1,253	Boys 12.9 or 15.9 Girls 12.8 or 15.8	5	"Starchy food" dietary pattern **Principal Component Analysis**	Overweight/obese Younger boys 31% Younger girls 26% Older boys 25% Older girls 23%	Overweight/obesity Younger boys with high scores O.R. 0.70 95%CI 0.50 to 0.98 P ≤0.05 Not significant in older boys or in girls.	↑ "Starchy food" pattern ↓risk of ow/ obesity in younger boys. Significant

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
Project EAT Cutler 2012	4/9	As above	As above	5	"Fruit" dietary pattern **Principal Component Analysis**	Overweight/obese Younger boys 31% Younger girls 26% Older boys 25% Older girls 23%	Overweight/obesity Younger boys with high scores O.R. 1.47 95%CI 1.13 to 1.92 P ≤0.05 Not significant in older boys or in girls.	↑ "Fruit" pattern ↑risk of ow/ obesity in younger boys. Not sig. after adj. for baseline wt. status
Project EAT Cutler 2012	4/9	4,746 middle and high school children	Boys 12.9 or 15.9 Girls 12.8 or 15.8	5	"Vegetable" dietary pattern **Principal Component Analysis**	As above	Overweight/obesity Older girls with high scores O.R. 0.85 95%CI 0.75 to 0.97 P ≤0.05 Not significant in boys or in younger girls.	↑ "Vegetable" pattern ↓risk of ow/ obesity in older girls. Not sig. after adj. for baseline wt. status
	4/9	As above	As above	5	"Sweet and salty snack food" dietary pattern **Principal Component Analysis**	As above	Overweight/obesity Older boys with high scores O.R. 0.85 95%CI 0.74 to 0.98 P ≤0.05. Not significant in younger boys or in girls	↑ "Sweet and salty snack food" pattern "↓risk of ow/obesity, in older boys. Not sig. after adjustment for baseline wt. status
SUMMARY: Dietary scores and dietary patterns with lower energy density (including higher intakes of vegetables, fruit, whole grains and low-fat dairy products) had more favourable adiposity outcomes.								

6.3.1 Dietary pattern methodologies

Modified DASH diet quality score in **NGHS** girls (Berz et al., 2011)

The original DASH score (Levitan et al., 2009) with ten food groups reflects adherence to the Dietary Approach to Stop Hypertension pattern of eating. Among NGHS girls three food groups (Added sugars, discretionary fats/oils and alcohol) added little to analyses, so were removed. This left a modified DASH score with seven food groups: fruit, vegetables, low-fat dairy products, total grains, whole grains, lean meats, and nuts, seeds and legumes.

At least two sets of 3 day diet records between 9 and 17 years were used to estimate average energy and nutrient intakes and total servings of major food groups and sub-groups. Adherence to each food group was scored, then totalled to give a modified DASH score for each girl. Assuming linearity between age and BMI, mixed models estimated mean BMI at each age by quintile of modified DASH score and by categories of intake for each food group.

Energy dense dietary pattern from RRR in **ALSPAC** (Ambrosini et al., 2012)

In an earlier study (Johnson et al., 2008) an energy dense, high fat, low fibre DP in the ALSPAC cohort at age 5 and 7 years was linked with a greater fat mass at 9 years. Data from 3 day food diaries at 7, 10 and 13 years old were used to derive energy dense DPs again, hypothesising that they would predict greater body fatness at 11, 13 and 15 years. The DP which explained the most variation (45%) in energy density, fibre density and % energy from fat was used in analyses. Z-scores were assigned to show how well each child's dietary intake matched this DP. Linear regression models were used to examine associations between dietary pattern z-scores at 7, 10 and 13 years and fat mass index (FMI) z score at 11, 13 and 15 years. Change in DP z-score between 7 to 13 years and change in FMI z-scores between 11 to 15 years were also modelled.

Eight dietary patterns from CA in **NGHS** girls (Ritchie et al., 2007)

Eight x 3 day food records collected annually between ages 9.5 and 18.5 years were used to derive 40 food groupings. Dietary patterns representing cumulative dietary intakes over the study duration were identified from food groupings, specifying four discrete clusters each for Black girls and White girls. Average nutrient intakes, BMI, % body fat and waist circumference at baseline, at follow-up, and change for each of the eight identified DPs were compared.

"Snacking" dietary pattern from PCA in **Bogotá School Children** (Shroff et al., 2014)

In an earlier study (McDonald et al., 2009), four DPs were derived from a baseline 38 item FFQ data in this cohort. McDonald et al found that one pattern, "Snacking", was significantly associated with overweight prevalence at baseline. Factor loadings for "Snacking" were

chocolate bar/cookies 0.64, ice-cream/ popsicles 0.61, guava/coconut candy bars 0.58, candy/lollipop 0.58, packed fried snacks 0.52, soda 0.50, and packed fruit punches 0.38. Factor scores for the "Snacking" dietary pattern were categorised into quartiles, i.e. four levels of adherence to this DP. The relationship between "Snacking" DP quartiles and change in adiposity measures (BMI, skinfold thickness ratio and waist circumference) between baseline and 2.5 year follow-up was examined using linear mixed effects models. Adiposity changes in relation to specific foods within the pattern were also investigated.

Four dietary patterns from PCA in **Project EAT** (Cutler et al., 2012)
Dietary patterns were derived from a 152 item YAQ employed at Time 1 (1998/99 school year) and Time 2 (2003/04). Identified DPs and factor loadings were described in an earlier paper (Cutler et al., 2009). Factor scores for each pattern were categorised as quintiles.
BMI was calculated from self-reported height and weight. If BMI values were ≥ 85th percentile the participants were categorised as overweight/obese (Must et al., 1991).
For each sub-group (younger and older boys and younger and older girls) the longitudinal risk of overweight/obese compared with normal weight, for a one quintile increase in factor score for each dietary pattern, was calculated using logistic regression. Participants reporting implausibly high or low energy intakes were excluded from analyses. Models did not adjust for energy intake, which was thought to be on the causal pathway.

6.3.2 Dietary patterns and adiposity outcomes

In the **NGHS** most girls did not meet DASH recommendations for fruits, vegetables, whole grains or low-fat dairy products included in the DASH diet quality score. Higher DASH scores were linked to higher TEI and higher intakes of most food groups. However, girls in the highest quintile of DASH scores also had higher mean physical activity and lower mean sedentary behaviours, indicating that their lifestyle helped them to achieve a better energy balance. Black girls and girls with lower SES were more likely to have low DASH scores (Berz et al., 2011).
Girls in Quintile 5 (highest DASH scores) had significantly smaller BMI gains and the lowest mean BMI (24.4) at age 19 years compared with other quintiles. Girls in Quintile 1 (lowest DASH scores) had greater BMI gains over time and by age 19 years their mean BMI (26.3) exceeded the threshold for adult overweight. Girls with higher intakes of fruit, low-fat dairy products and whole grains experienced less weight gain than girls with lower intakes of these food groups, as presented in Chapter 5.

In **ALSPAC** 73% of 7,285 children at age 7 years returned food diaries with plausible energy intakes. 7,471 food diaries were completed by children with parental help at age 10 years (63% plausible) and 6,106 at age 13 years more independently (39% plausible). As expected during puberty, girls had a greater average total fat mass than boys, and on average boys were taller than girls. ALSPAC girls had a greater prevalence of overweight/obesity than boys, at all ages.

The derived energy dense, high fat, low fibre dietary pattern had characteristically low intakes of fruit and vegetables. The foods at the bottom (fruit, vegetables, legumes, high fibre breakfast cereal) and top (confectionery, chocolate) of the factor loadings explained most of the variations in DP z-score.

The energy dense DP significantly predicted future Fat Mass Index, at all ages. Each 1 standard deviation increase in energy dense DP z-score at age 7 years, predicted a greater FMI z score at 11 years, 13 years and 15 years. Energy dense DP z-score at age 10 years and at 13 years similarly predicted FMI z score, albeit more weakly at older ages, which may be a consequence of dietary measurement error. In longitudinal models of change, an increase in energy dense DP z-score predicted a corresponding increase in FMI z-score. (Ambrosini et al., 2012)

In the **NGHS**, intakes of 40 different food groupings were averaged over the 10 years of the study (Ritchie et al., 2007). Although mean intakes of plain breads, plain grains, cereal, nuts/ popcorn, fish/ poultry (not fried) and pizza were similar for all girls, for some food groupings there were significant differences in mean intakes between black and white girls. Four discrete dietary patterns were identified for each ethnicity.

Black girls in the **NGHS** had higher mean intakes of 16 food groupings: sweetened drinks, juice, other breakfast grains, baked desserts, other desserts, candy, chips (crisps), eggs, fried fish/poultry, red meat, processed meats/sandwiches, burger sandwiches, ramen, legumes, other vegetables and fried potatoes.

Identified dietary patterns were:

- “Customary”, 53% of black girls
- “Snack-type foods”, 23% of black girls.
- “Meal-type foods”, 22% of black girls.
- “Sweets and cheese”, <2% of black girls.

In each DP for black girls mean fat intakes were ~ 36% of energy and mean saturated fat intakes were ~ 13% of energy, exceeding the Dietary Reference Intake guidelines (Total fat, 20 to 35% of energy intake, Saturated fat <10% of energy intake).

Adiposity measures for black girls at final follow-up (age 18-19 years) were not significantly different by dietary pattern after adjusting for baseline BMI, puberty, pregnancy, parental education, physical activity and TV/video watching.

White girls in the **NGHS** had higher mean intakes of 18 food groupings: coffee/ tea, diet drinks, plain milk, flavoured milk, yogurt, cheese, ice cream, sweet rolls, crackers, pretzels, other meat sandwiches, peanut butter sandwiches, cheese/spread sandwiches, mixed dishes, soups, fruit, green salad, and potatoes (not fried).
Identified dietary patterns were:

- "Convenience", 45% of white girls
- "Sweets and snack-type foods", 33% of white girls.
- "Fast food", 10% of white girls.
- "Healthy", 12% of white girls.

For white girls the highest intakes of energy and salt and lowest intakes of dietary fibre were in the "Fast food" DP, with mean fat and saturated fat intakes equivalent to DPs in Black girls. The "Healthy" DP had the lowest fat content (31.5% of Energy) and the highest dietary fibre of all identified DPs, but not the lowest mean TEI. The lowest mean TEI was found in the "Convenience" DP, which had higher % of total fat and sugar than the "Healthy" DP and the lowest mean intakes of Vitamin C. The "Convenience" DP conferred no benefit in terms of BMI, body fat % or waist circumference.
The "Healthy" DP was the only pattern that was beneficial in terms of adiposity outcomes. Compared with the "Fast food" DP, white girls following the "Healthy" DP tended to have a lower % body fat at follow-up, (27.7% vs 29.7%, P = 0.06) although the difference was not significant at the $P \leq 0.05$ level. Compared with the "Sweets and snack-type foods" DP the "Healthy" DP was significantly associated with smaller gains in waist circumference at follow-up (+ 9.7 cm vs + 11.7 cm, P = 0.04).

In the **Bogotá School Children** cohort, only investigations of the "Snacking" DP were reported (Shroff et al., 2014). The corresponding author confirmed that none of the other 3 dietary patterns, "Cheaper protein", "Traditional/starch" or "Animal protein", were associated with the outcomes reported in Shroff's study.
In total 975 children in the Bogotá School Children cohort had dietary pattern data, of whom 961 had anthropometric measures at baseline and at follow-up. Cross-sectionally, children's baseline BMIz scores were lowest in the first quartile of adherence to the snacking pattern (lowest adherence/ least snacking).

At 2.5 year follow-up, higher adherence to the snacking pattern was associated with greater changes in BMI and skinfold thickness in unadjusted models, but not with change in waist circumference. In models which adjusted for TEI, compared with those in the lowest quartile of adherence to the snacking pattern, the BMI of those in the highest quartile increased by an extra 0.09kg/m^2 per year. Children in the highest quartile of snacking also had greater increases in their subscapular: triceps skinfold thickness ratio, which is indicative of greater truncal adiposity.

Children with higher intakes of fried food snacks tended to have greater gains in waist circumference, while children with higher intakes of ice-cream, popsicles and candy tended to have greater gains in BMI and skinfold thickness ratio, but these associations were not significant at $P \leq 0.05$.

Soda was the only factor within the "Snacking" DP which significantly influenced adiposity change. As presented in Chapter 5, in adjusted models, children who drank soda daily, compared with those who never drank soda, had a 0.6cm/year extra gain in waist circumference (P trend = 0.04) and an additional 0.20 kg/m^2 gain in BMI (P trend = 0.01).

In **Project EAT** (Cutler et al., 2012) the four dietary patterns identified at baseline and the foods or factors that were loaded heavily in each pattern were listed as:

- "Vegetable": Zucchini (courgette), squash, eggplant (aubergine), kale and greens, spinach, carrots, peas and lima beans.
- "Fruit": Oranges, grapefruit, apples, apple sauce, pears, grapes, bananas, strawberries, cantaloupe and melons, peaches, plums and apricots.
- "Starchy food": English muffins/bagels, grilled cheese, pancakes, crackers, pretzels, macaroni and cheese, spaghetti with sauce, mashed potatoes, lasagne.
- "Sweet and salty snack food": Chocolate bars, other candy bars, candy with chocolate, brownies, cake, potato chips (crisps) and nachos.

At follow-up 5 years later, dietary patterns were re-investigated in half of the original cohort. Only younger girls kept all four previously identified dietary patterns. The "Starchy food" pattern did not persist in younger boys. Among older boys and girls (now young adults, mean age 20.4 years) the "Fruit" and "Vegetable" patterns merged into one DP. A new "Fast Food" dietary pattern was identified in boys and older girls, heavily loaded with hamburgers, French fries, fried food and non-diet soda.

Cross-sectional analyses at baseline found that high scores for the "Vegetable" DP were linked with lower overweight/ obesity in older teenage girls. High scores for the "Snack food" DP

were also linked with lower overweight/ obesity in older boys and younger boys. High scores for the "Fruit" DP were linked with higher overweight/ obesity in younger boys
Cross-sectional analyses at follow-up also found that high scores for the "Vegetable" DP in younger girls and the new "Vegetable and fruit" DP in older girls were linked with a lower prevalence of overweight/ obesity, but high scores for the "Snack food" DP were no longer linked with lower overweight/ obesity in older boys after adjustment for physical activity.

In longitudinal analyses of associations between DPs at baseline and overweight/ obesity outcomes five years later, a similar picture emerged.
Younger boys had a 47% increased risk of future overweight/ obesity for each one quintile increase in the "Fruit" DP score.
Older girls had a 15% reduced risk of future overweight/ obesity for each one quintile increase in the "Vegetable" DP score.
Older boys had a 15% reduced risk of future overweight/ obesity for each one quintile increase in the "Snack food" DP score.
After adjustment for baseline weight status, only one association remained significant. Each one quintile increase in the "Starchy" pattern score gave younger boys a 30% reduction in risk of future overweight/ obesity. This association was not observed in any other sub-group.

6.3.3 Discussion of dietary patterns

Twenty different dietary patterns were referred to by included papers. One diet quality score and fourteen DPs were described, capturing some of the variety that exists between diets of different individuals. Some DPs also demonstrated that health risks cannot always be explained in terms of single foods.

The modified DASH diet quality score used á priori in the **NGHS** girls cohort (Berz et al., 2011) and the Energy dense dietary pattern derived á posteriori in **ALSPAC** (Ambrosini et al., 2012) were both founded on an understanding of diet and health outcomes. The two are almost opposite measures, as many foods with negative factor loadings in the Energy dense DP (fruit, vegetables, high fibre breakfast cereal, legumes, refined grains, fruit juices, and low fat milk) were positively scored elements of the DASH score.
Teenaged girls from **NGHS** with higher DASH scores experienced smaller BMI gains by final follow-up. In **ALSPAC** children of both sexes who had higher DP z-scores (more energy dense diets) experienced greater gains in fat mass. Children who increased their dietary energy density had corresponding increases in Fat mass. This suggests that improving diet quality and reducing energy density could reduce obesity risk in children. It also demonstrates that dietary patterns can shift over time.

A cluster analysis of dietary patterns in **NGHS** was again based on girls' averaged food intakes, from 3 day diet records across all 10 years of the study. Different dietary patterns were observed between and within ethnic groups (Ritchie et al., 2007). Most identified DPs were high in fat so were energy dense. Only the so-called "Healthy" DP identified in white girls (lower in fat so "healthier", although this DP did not meet all dietary guidelines) showed any benefit in terms of adiposity.

The white girls who followed the "Healthy" DP (with the highest mean intakes of vegetables, salad, fruit, plain grains, cereal, breakfast grains, plain milk and yoghurt) were likely among the **NGHS** girls who adhered most closely to the DASH diet quality score.

One disadvantage of using cumulative measures of food intake in **NGHS** is that changes in dietary patterns are not revealed. In support of aggregating the data into cumulative DPs, the authors stated that fluctuations in DPs as girls progressed to young adulthood meant that comparable dietary patterns at the different stages of adolescence could not be identified.

Two studies used principal components analysis, exploring dietary patterns based on food intakes from FFQs. Factor loadings in the "Snacking" DP identified in the **Bogotá School Children** cohort (Shroff et al., 2014) were sweets, snacks and beverages high in sugar and/or fat, which are energy dense. Children who followed the "Snacking" DP closely at baseline experienced larger increases in BMI and skinfold thickness, indicative of greater adiposity gains.

In **Project EAT**, the associations of DPs and adiposity outcomes were inconsistent (Cutler et al., 2012). High adherence to the "Vegetable" pattern by girls and to the "Snack food" pattern by boys seemed helpful but, after adjustment for baseline weight status, did not predict future overweight. Younger boys who followed a "Fruit" pattern had higher baseline BMI (perhaps that was why they ate more fruit) but were not at increased risk of future overweight after adjustment for baseline weight status. Project EAT relied upon self-reported height and weight, so it may be that overweight status was underestimated. Implausible reporters were excluded, but dietary measurement error cannot be ruled out.

When **Project EAT** re-examined dietary patterns, at the 5 year follow-up, a new "Fast Food" DP was observed, while other DPs merged, demonstrating again that dietary patterns evolve over time, as children progress to adulthood.

Although the five studies used different methods and their findings about how dietary patterns influence adiposity outcomes were not identical, there were broad similarities:

- Dietary patterns characterised as "Energy dense" (low in fibre but high in fat and sugar), were associated with an increased risk of future overweight/ obesity.
  "Energy dense" DPs had comparatively high intakes of foods such as chocolate, confectionery, cakes, biscuits, ice-cream, crisps, fried foods and sugar sweetened drinks.
- Dietary patterns characterised as "Healthy" (higher in fibre and lower in fat and sugar) were associated with a decreased risk of future overweight/ obesity.
  "Healthy" DPs had comparatively high intakes of vegetables, fruit, whole grains and low fat dairy foods.

Evidence from diet quality scores and dietary patterns explored among girls from the NGHS suggests that diet quality, not just quantity, is a driver of overweight. It may be helpful to consider, if a child's intake of a particular food or drink is high, which alternative foods or drinks are *not* being consumed?

It was also evident that the dietary patterns of children and adolescents are dynamic, not static. This finding is supported by a study of adult women, of whom only 55% sustained the same dietary pattern over a 5 year period (Greenwood et al., 2003). Such fluidity presents an opportunity to help young people to modify their dietary patterns in ways which will improve, rather than worsen, diet quality, thereby reducing their risk of adverse adiposity outcomes.

6.4 Eating habits

Eating habits are usual actions or behaviours, measured by frequency rather than by quantified intakes of foods or drinks. Some studies ascertained eating behaviour frequencies based on information from dietary assessment, while other studies asked direct questions, separately from the dietary assessment, about whether or how often, participants engaged in an eating behaviour.

Eating habits were investigated in five USA cohorts.
Family meals, breakfast/cereal frequency, other meals and eating frequency and their influence on adiposity were extensively investigated in **NGHS** girls (Affenito et al., 2005; Albertson et al., 2007; Albertson et al., 2009; Barton et al., 2005; Franko et al., 2008; Rehkopf et al., 2011; Ritchie, 2012) and in **Project EAT** (Berge et al., 2015; Fulkerson et al., 2008; Quick et al., 2013), where dieting behaviour was also considered as a predictor of future overweight. Breakfast eating was also examined in the **BDPP** (Balvin Frantzen et al., 2013) and in the combined **IDEA and ECHO** cohort, alongside the frequency of purchasing fast food (Laska et al., 2012). Frequency of eating fried food away from home (Taveras et al., 2005) and dieting habits (Field et al., 2003a) were investigated in **GUTS**.

6.4.1 Family meals

Family meals and its opposite, eating alone, were potential predictors of BMI percentile change considered in the **NGHS** (Rehkopf et al., 2011) Family meal frequency and overweight/obesity outcomes were examined twice in **Project EAT**, with a 5 year follow-up (Fulkerson et al., 2008) and a 10 year follow-up (Berge et al., 2015). Results are presented in Table 6-2.

6.4.1.1 Family meal methodologies

In the **NGHS** family meals were measured by parental response during an interview at the start of the study. Parents were asked about the frequency of “family eating dinner together” and how often their child “eats dinner alone”. Response options were: Never or less than once a week, 1 -3 times/ week, 4-7 times/ week and 8 or more times/ week.

Table 6-2 Family meals and adiposity outcomes

Study & Paper	NOS stars	Population	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
NGHS Rehkopf 2011	6/9	**2,150** girls in analyses Black 1,044 White 1,106	9.5	10	Predictor: Family eats dinner together Predictor: Eats alone	Baseline: Overweight 15% Obese 13% Follow-up: Overweight 16% Obese 18%	BMI %ile change: Family eats dinner together P = 0.18 Eats alone P = 0.33 Onset of overweight or obesity: Not sig.	Not significant predictors
Project EAT Fulkerson 2008	5/9	**806** children in middle school Males 367 Females 439 > 30% white	12.8	5	Family meal frequency at baseline, Never vs Ref (3 –7/week)	Baseline ow: Males 30% Females 25% Follow-up ow: Males 27% Females 32%	O/w at follow-up Males: OR = 1.8 95%CI = (0.5, 6.3) Females: OR = 2.6 95%CI = (0.9, 7.5)	Not significant
		1,710 children in high school Males 763 Females 947 > 50% white	15.9	5	Family meal frequency at baseline, Never vs Ref (3 - 7/week)	Baseline ow: Males 25% Females 23% Follow-up ow: Males 24% Females 26%	O/w at follow-up Males: OR = 0.9 95%CI = (0.4, 1.7) Females: OR = 1.0 95%CI = (0.6, 1.8)	Not significant

Study & Paper	NOS stars	Population	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
Project EAT Berge 2015	5/9	**2,117** Males 953 Females 1,164	12.8 and 15.9 together	10	Family meal frequency at baseline, vs Ref (Never)	Baseline not given but see Fulkerson 2008 above. Follow-up: Overweight: 51% Obese: 22%	<u>O/w at follow-up</u>: 5+ /week OR 0.63 95%CI = (0.46,0.87) 3 – 4 /week OR 0.60 95%CI = (0.42,0.85) 1 -2 /week OR 0.55 95%CI = (0.38,0.79) <u>Obese at follow-up</u>: 5+ /week OR 0.68 95%CI = (0.47,0.99) 3 - 4/week OR 0.50 95%CI = (0.33,0.76) 1 -2/week OR 0.67 95%CI = (0.44,1.02)	↓risk of o/w and obesity Significant No adjustment for Energy intake
SUMMARY: Compared with never eating family meals at baseline, eating three or more family meals a week was protective against future obesity.								

Using a tree-based regression method the accuracy of both variables in predicting BMI percentile change (based on measured height and weight) or onset of overweight or obesity (using CDC references) was compared with 40 other potential predictors in the NGHS girls' cohort. (Rehkopf et al., 2011)

In **Project EAT** 4,746 ethnically diverse adolescents were recruited through middle schools and high schools in 1998/99. They were asked, "During the past seven days, how many times did all, or most, of your family living in your house eat a meal together?". Responses were collapsed to three: Never, Infrequent (1 - 2 times/week) and Frequent (3 or more times/week). 53% of the original cohort completed follow-up surveys after 5 years and 45% of the original cohort completed follow-up surveys after 10 years. Attrition did not occur at random, so data was weighted in analyses, adjusting for non-response. For consistency self-reported height and weight was used to calculate BMI each time.

At the five year follow-up (Fulkerson et al., 2008), adolescent overweight was based on age and sex specific BMI cut-offs. To ensure continuity between adolescent and adult measures, adult overweight was defined as BMI >85th percentile using Must's classification.

Each sub-group (middle school females, high school females, middle school males, and high school males) was analysed separately. The reference category was Frequent family meals. Models controlled for age, race and socio-economic status, physical activity, sedentary behaviour and energy intake.

At the ten year follow-up (Berge et al., 2015) baseline adolescent overweight was based on a BMI >85th percentile for age and sex using updated CDC reference data. Adult overweight was defined as BMI ≥25kg/m^2, with obesity defined as BMI > 30kg/m^2. All respondents (n=2,117) were included in one analysis in a model which controlled for sex, age, race, SES and baseline overweight or obese status. The reference category was changed to No family meals.

6.4.1.2 Family meal frequency and adiposity outcomes

In the **NGHS** the mean frequency of "family eats dinner together" at age 9 to 10 years was 1.2 times/ week (SD 0.4). The mean frequency of "eats dinner alone" was 1.4 times/ week (SD 0.6). Neither "family eats dinner together" nor "eats alone" were significant predictors of change in BMI or of onset of overweight or obesity at the $P \leq 0.05$ level. (Rehkopf et al., 2011).

In **Project EAT** almost two thirds of the older teenagers and three quarters of the younger teenagers reported frequent family meals at baseline. Around 15% of students reported never eating family meals. This tendency was higher among the older teenagers with 19% of high school females and 12% of high school males never eating family meals. (Fulkerson et al.,

2008).

By 5 year follow-up prevalence of overweight was highest among the former middle-school students. In cross-sectional analyses middle school females who reported never eating family meals were more likely to be overweight at baseline (OR 3.1 95% CI 1.3 to 7.3) than middle school females who consumed frequent family meals. Cross-sectional associations were not found in other subgroups. In longitudinal analyses, stratified by age and sex, baseline family meal frequency was not significantly associated with overweight status at five year follow-up.

By the ten year follow-up of 2,117 participants in **Project EAT**, 51% were overweight and 22% were experiencing obesity, now defined using adult guidelines (Berge et al., 2015). The longitudinal analysis was not stratified, so the sample size was larger, offering more power to detect an effect.

Compared with never eating family meals at baseline, *all* levels of eating family meals at baseline were protective against overweight ten years later in young adulthood. The highest frequencies of eating family meals at baseline reduced the risk of overweight the most, by ~ 37%.

Compared with never eating family meals at baseline, eating 3 or more family meals/ week was also protective against future obesity.

6.4.2 Breakfast and cereal eating

The influence of eating breakfast and/or cereal on future adiposity was reported by eight papers from four different cohorts (**BDPP**, **NGHS**, **IDEA and ECHO**, and **Project EAT**). Five papers considered the habitual frequency of eating breakfast. Three papers looked more specifically at the frequency of eating cereal.

Breakfast eating in children from the Bienestar Diabetes Prevention Programme (**BDPP**) was explored before focussing on frequency of eating ready-to-eat cereal (RTEC), which is often consumed with milk (Balvin Frantzen et al., 2013). The first author of the BDPP study was employed by DairyMAX, a regional dairy council in America.

Breakfast eating among girls from the NHLBI Growth and Health Study (**NGHS**) was investigated by 2 papers (Affenito et al., 2005) (Albertson et al., 2007) and frequency of eating cereal was investigated by 2 more (Albertson et al., 2009; Barton et al., 2005). Research was supported by General Mills Ltd., a USA based food company and manufacturer of branded cereals. A fifth **NGHS** paper considered breakfast eating as a potential predictor of BMI percentile change (Rehkopf et al., 2011).

Change in breakfast frequency was investigated in the **IDEA & ECHO** cohort (Laska et al., 2012) while in **Project EAT** baseline breakfast frequency and change were considered as potential predictors of overweight (Quick et al., 2013)

Longitudinal findings for breakfast eating and cereal eating and adiposity outcomes are summarised in Table 6-3. Cereal also contributed to dietary patterns identified in NGHS girls (Ritchie et al., 2007)

6.4.2.1 Breakfast and cereal eating methodologies and definitions

All the cohorts included in the systematic review used quantitative or semi-quantitative dietary assessment methods. In the BDPP and NGHS this information was used to assess *frequency* of having breakfast or eating cereal, expressed as the number or percentage of days surveyed when breakfast or cereal was consumed.

In the Mexican-American **BDPP** cohort breakfast was the first food and/or drink in the morning named as "breakfast" in 3 multiple pass 24 hour dietary recalls at the start of fourth grade (mean age 9.1 years) and at the end of the fifth and sixth grades (Balvin Frantzen et al., 2013). Breakfast foods included juice, eggs, bread, tacos, sausage, biscuits, tortillas and pancakes as well as cereal. The focus was frequency of eating ready to eat cereal (RTEC), defined as processed cereal that needs no preparation before eating, often consumed with milk.

In the **NGHS** breakfast was defined as any eating between 5 – 10am weekdays or 5 – 11am at weekends (Affenito et al., 2005). Frequency of breakfast eating was measured by 8 annual 3 day food records between 1987 and 1997. Cereal was described loosely as "cereal breakfasts", mostly fortified with essential nutrients (Barton et al., 2005) but in a later paper cereal was defined as either RTEC (often recorded as a brand name) or cooked cereal such as oatmeal (Albertson et al., 2009). Albertson et al excluded data from Study years 1 and 2.

In the **IDEA and ECHO** studies breakfast was a meal with at least 50 calories that participants called "breakfast" in up to 3 telephone-administered 24 hour dietary recalls, at baseline and 2 year follow-up (Laska et al., 2012).

In **Project EAT** breakfast was not defined. Breakfast frequency in the last week was self-reported at baseline and at 10 year follow-up. Other aspects of dietary intake were assessed using semi-quantitative food frequency questionnaires (Quick et al., 2013).

Table 6-3 Breakfast and cereal and adiposity outcomes

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
BDDP Balvin Frantzen 2013	7/9	**625** children Boys 309 Girls 316	9.1	3	Days of **Ready to eat Cereal** consumption, from 9 days measured	BMI %iles Boys: 68.2 SD 30.0 Girls: 69.0 SD 29.3	BMI %ile change β1 = -1.977 s.e. = 0.209 p= 0.001	Greater frequency of RTEC: **↓ BMI percentile** Significant
IDEA & ECHO Laska 2012	8/9	693 Male 327 Female 339 **535** in longitudinal analyses	14.6	2	Increase in **breakfast** (days/ week)	Baseline mean BMI Male 22.1 kg/m^2 Female 21.9 kg/m^2 Follow up mean BMI Male 23.4 kg/m^2 Female 22.8 kg/m^2	Change in BMI adjusted for TEI: Males: β = -0.19 s.e.= 0.48 P = 0.69 Females β = -0.26 s.e. = 0.46 P = 0.57	Not significant
		As above	As above	As above	As above	As above	Change in % body fat adjusted for TEI: Males: β = -1.47 s.e. = 1.27 P = 0.25 Females: β = -0.18 s.e. = 0.86 P = 0.83	Not significant
NGHS Affenito 2005	8/9 *	2,379 girls Black 1,213 White 1,166	9.5	10	**Breakfast** frequency (number of days 0,1,2,3)	Baseline mean BMI: Black girls 18.6 kg/m^2 White girls 17.6 kg/m^2	BMI χ^2 [1] = 14.05 P < 0.005 BMI, adjusted for TEI χ^2 [3] = 3.10 P = 0.38	↓BMI in girls Not significant after adjustment for Energy

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
NGHS Albertson 2007	8/9 *	2,371 girls Black 1,210 White 1,161 **<1,900** in model?	9.5	10	**Breakfast** "history" (% days eating breakfast of days surveyed in Study years 1 to 5, 7, 8 and 10)	Baseline mean BMI: 2,352 girls 18.6 kg/m^2	BMIz in Year 10: Est.= -0.0013 s.e. = 0.001 P >0.1 BMIz in Year 10, adjusted for baseline BMIz: est.= -0.0026 s.e. = 0.007 P <0.0001	BMIz in girls Not significant ↓BMIz for girls with high BMIz at baseline Significant
NGHS Albertson 2009	8/9	**2,313** girls Black 1,187 White 1,126	11.5	7	Cumulative % of days with **cereal** consumption	Mean BMI at 11.5 yrs. n/a	% body fat at 18.6 yrs. Est. = - 0.04 s.e.= 0.01 p = 0.01	Consuming cereal on more days: **↓% body fat in girls** Significant
NGHS Barton 2005	7/9	2,379 girls Black 1,213 White 1,166 Between **772 to 2,034** at each survey	9.5	9	Days eating **cereal** (0, 1, 2 or 3 days)	Mean BMI at 9.5 yrs. ~18kg/m^2 Mean BMI at 18.6 yrs. ~25+kg/m^2	BMIz change: β = - 0.015 p <0.001	Each additional day eating cereal: **↓BMIz in girls** Significant
					1, 2 or 3 days eating **cereal** vs. not eating cereal	As above	Risk of o/w O.R. = 0.93, 0.90 or 0.87 respectively p<0.05	Eating cereal vs not eating cereal: **↓risk of overweight in girls** Significant

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
NGHS Rehkopf 2011	6/9	**2,150** girls in analyses Black 1,044 White 1,106	9.5	10	Predictor: Eats **Breakfast**	Baseline: Overweight 15% Obese 13% Follow-up: Overweight 16% Obese 18%	BMI %ile change: Eats breakfast P <0.01 Onset of overweight or obesity: Not sig.	Significant predictor of BMI %ile change (direction not shown)
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	Predictor: **Breakfast** (times/week) at baseline	Overweight at 15 yrs.: ~25%, If not o/w at start, o/w at 25 yrs.: Males 45%, Females 34%	Females, overweight at follow up: O.R. = 0.91, 95% CI 0.86 to 0.97 Males, overweight at follow up: O.R. = 0.95, 95% CI 0.90 to 1.01	↓risk of overweight in females *not overweight at baseline* Significant
SUMMARY: Eating breakfast was a significant predictor of BMI change in girls and may be linked with lower BMI or BMIz outcomes for girls/females. Eating cereal or ready to eat cereal on more days was associated with lower adiposity outcomes in two USA cohorts. Eating cereal versus not eating cereal reduced girls' risk of overweight. NOTE: Papers from the NGHS (Albertson 2007, Albertson 2009 and Barton 2005) declared funding/support from General Mills Inc.								

6.4.2.2 Breakfast and cereal frequency

Most children consumed breakfast some, if not all the time. Frequency of breakfast eating and eating cereal tended to decline during adolescence. Cereal was a popular choice of breakfast food. There was evidence that eating cereal (often with milk) improved diet quality, increasing mean intakes of dietary fibre, vitamins and minerals including calcium, and lowering mean intakes of fat and cholesterol.

In the **BDPP** study in fourth grade 64% of the children recalled eating any kind of breakfast on all 3 days surveyed, but three years later this dropped to 42%. Most RTECs contained added sugar but were not associated with mean TEI in the cohort. RTEC was usually eaten with added milk. (Balvin Frantzen et al., 2013).

In the **NGHS**, girls ate breakfast on 70% of days surveyed, with 65% of girls eating breakfast at least one day in three. Only 3 girls never had breakfast. Older girls were more likely to skip breakfast sometimes (Albertson et al., 2007).
At 9 years old 77% of white girls and 57% of black girls ate breakfast on all 3 days, but by age 19 years this had fallen to 32% and 22% respectively (Affenito et al., 2005). Girls who ate breakfast more often had higher energy intakes, and were more likely to engage in walking, running, cycling or sport (Albertson et al., 2007).
Cereal was eaten at any time, with 90% of girls reporting cereal consumption at least once and 18% eating cereal on most days surveyed between years 3 and 10. Over 40% of girls at age 9 years had cereal but by 19 years this had fallen to < 20% (Barton et al., 2005). Cereal intake peaked at an average 58g/day at 15 years old, falling to 52g/day by 19 years old (Albertson et al., 2009). Mean intakes of cereal were similar among black and white girls (Ritchie et al., 2007).

In **IDEA and ECHO** cohort the mean frequency of breakfast consumption at 14 years old, expressed as the percentage of recall days (from 2 or 3) that participants reported eating "breakfast", was 91% of days for boys and 88% of days for girls. By follow-up two years later the mean frequency of breakfast eating on recall days fell to 83% for boys and 84% for girls (Laska et al., 2012).

An exception to this falling trend was reported in the **Project EAT** cohort which followed participants into young adulthood. The reported mean frequency of breakfast eating among males (not overweight at baseline) decreased between the ages of 15 years and 25 years from 4.3 times/week to 3.5 times/week, whereas in females (not overweight at baseline) the

reported mean frequency of breakfast eating rose by a small but significant amount, from 3.7 times/week to 4.3 times/week. (Quick et al., 2013).

6.4.2.3 Breakfast and cereal frequency and adiposity outcomes

More frequent breakfast eating was beneficial for baseline measures of adiposity and may be linked with lower BMIz outcomes, at least for girls, but breakfast eating habits did not predict overweight or obesity.

In the **NGHS** girls who reported eating breakfast consistently had a lower BMI than girls who skipped breakfast on some or all days (Affenito et al., 2005) (Barton et al., 2005).
In a model which controlled for race, age and their interactions, using Type III Wald χ^2 significance tests, the number of days eating breakfast was a significant predictor of BMI, but after adjustments for energy intake, physical activity and parental education, the independent effect of eating breakfast was no longer significant (Affenito et al., 2005). The number of days eating breakfast was not predictive of BMIz scores or the risk of overweight (Barton et al., 2005).
Breakfast history (the % of days surveyed when breakfast was eaten) was only a predictor of BMIz among girls with obesity at baseline (Albertson et al., 2007).
In the same **NGHS** cohort, out of 8 eating behaviours considered, eating breakfast was the most important predictor of change in BMI percentile, but did not predict onset of overweight or obesity (Rehkopf et al., 2011).

In the **IDEA and ECHO** cohort, there was cross-sectional evidence that boys and girls who consumed breakfast more frequently had lower BMI and lower body fat % than those who ate breakfast less frequently. However longitudinal regression analyses, either with or without adjustment for energy intake, found no significant association (at either $\alpha = 0.003$, used to correct for the number of tests, or the more usual $P \leq 0.05$) between breakfast consumption and later BMI or body fat % outcomes, (Laska et al., 2012)

In **Project EAT**, breakfast eating was investigated as a predictor of overweight for adolescents in the cohort who were not overweight at baseline (mean age 15.0 years). By young adulthood 10 years later 34% of female and 45% of male respondents were overweight based on self-reported height and weight. For adolescent females not overweight at baseline, eating breakfast more times per week at baseline reduced the risk of overweight in young adulthood by ~ 9%. However baseline breakfast consumption did not predict overweight outcomes in males and the direction of influence of 10 year change in breakfast frequency on future overweight for either sex was uncertain (Quick et al., 2013)

Eating cereal or ready to eat cereal more often was associated with lower adiposity outcomes at follow-up. In the **BDPP**, after controlling for age, sex, ethnicity and TEI, for every extra day of RTEC consumption (out of the 9 days surveyed, 3 x 24-HDR on 3 occasions) a child's BMI decreased by 2 percentiles over the 3 years (Balvin Frantzen et al., 2013).

A similar effect was reported in the **NGHS**. Each extra day of eating cereal predicted a BMIz change of - 0.015. Compared with not eating cereal, girls who ate any type of cereal at any time (not just at breakfast) on 1, 2 or 3 days had around a 10% reduced risk of overweight. (Barton et al., 2005)

In the second **NGHS** paper, girls who ate cereal on more days had relatively lower body fat %, measured by bio-impedance, at the study's end. Each percentage increase in the number of days consuming cereal between study years 3 and 10, after adjustment for body fat %, TEI and Physical activity, was associated with a small yet significant decrease in body fat of -0.04% by study year 10. (Albertson et al., 2009)

White girls in the **NGHS** following the "Healthy" dietary pattern, which had the highest mean intakes of cereal and other breakfast grains, also had the smallest final waist circumference at age 19 years (Ritchie et al., 2007)

6.4.3 Other meals and eating frequency

In **Project EAT**, baseline lunch and dinner frequency and change in frequency were considered as independent predictors of overweight, as well as breakfast frequency. (Quick et al., 2013)

The cross-sectional relationship between meal frequency and BMI was explored in the **NGHS** girls cohort (Franko et al., 2008), followed by a later paper that considered the total number of eating episodes (meals and snacks) and longitudinal adiposity outcomes (Ritchie 2012). "Eats snack food" was also considered as a predictor of BMI percentile change (Rehkopf et al., 2011).

Meal and eating frequencies and adiposity outcomes are summarised in Table 6-4.

Table 6-4 Meal and eating frequency and adposity outcomes

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	Lunch meal frequency (times/week)	Overweight at 15 yrs.: ~25%, If not o/w at start, o/w at 25 yrs.: Males 45%, Females 34%	Increased freq. and ow in males O.R. = 1.13, 95% CI 1.03 to 1.23	↑risk of overweight, in males *not overweight at baseline* Significant
		As above	As above	As above	Predictor: Dinner meal frequency (times/week) at baseline	As above	Baseline freq. and ow in females O.R. = 0.88, 95% CI 0.81 to 0.95	↓ risk of overweight in females *not overweight at baseline* Significant
		As above	As above	As above	Dinner meal frequency (times/week)	As above	Increased freq. and ow in males O.R. = 1.14, 95% CI 1.00 to 1.29	↑risk of overweight, in males *not overweight at baseline* Significant
NGHS Franko 2008	8/9	2,375 girls Black 1,209 White 1,166	9.5	10	Meal frequency, (Each additional day of 3+ meals/ day) across Study years 3, 4, 5, 7, 8 and10.	Not shown	*Cross-sectional analyses* BMIz, all girls Est = -0.05, 95%CI -0.3,-0.6 across Study years 3, 4, 5, 7, 8 and10.	*Cross-sectional analyses:* Lower BMIz in girls and therefore ↓risk of o/w in Black girls Significant

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
NGHS Ritchie 2012	7/9	2,372 girls Black 1,210 White 1,162 **2,146** in analyses	9.5	10	Fewer eating episodes (meals and snacks) in Study years 1 and 2	Mean BMI All 18.6 kg/m^2 Black 19.2 kg/m^2 White 17.9 kg/m^2 Mean WC All 65.2cm Black 66.7cm White 63.4cm	Larger increases in BMI P = 0.006 Larger increases in WC P = 0.020 Model adjusted for TEI	↑BMI in girls ↑WC in girls Significant
NGHS Rehkopf 2011	6/9	**2,150** girls in analyses Black 1,044 White 1,106	9.5	10	Predictor: Eats snack food	Baseline: Overweight 15% Obese 13% Follow-up: Overweight 16% Obese 18%	BMI % change: Frequency of snacks P = 0.03 Onset of overweight: Not sig Onset of obesity: Significant.	Significant predictor of BMI %ile change (direction not shown) and onset of obesity in girls
SUMMARY: Higher baseline meal frequency mitigated future overweight in females, while fewer eating episodes added to adiposity risk in girls. For males (who had a higher baseline meal frequency than females) increased meal frequency predicted overweight.								

6.4.3.1 Other meal methodologies

In **Project EAT** lunch and dinner frequency in the last week were self-reported at baseline and at 10 year follow-up. (Quick et al., 2013)

In the **NGHS** meal frequency and number of eating episodes were based on the annual 3 day food records which contained information about food and drink, type of meal (breakfast, snack, lunch etc.) and time of intake. Dietitians rated how many days (out of 3) that a girl had three or more meals a day for each of the annual food records collected in Years 1 to 5 and Years 7, 8 and 10. The relationship of meal frequency and BMIz scores at each time were documented (repeat cross-sectional analysis). (Franko et al., 2008)

In the next **NGHS** study, girls' meals and snacks each day were counted to calculate the number of eating episodes (any amount of food or drink eaten at a single time). Average eating frequency during the first two years (ages 9 to 11 years) was used as the baseline measure, categorised as 1 to 3, 3.1 to 4, 4.1 to 6 and > 6 eating episodes/day. Frequency of eating was compared with change in BMI and waist circumference by age 18/19 years old. (Ritchie, 2012)

6.4.3.2 Other meal and eating frequencies and adiposity outcomes

In **Project EAT**, on average females ate lunch 5.3 times/week and dinner 5.9 times/week at the age of 15 years and reported that they increased this frequency to lunch 5.6 times/week and dinner 6.3 times/week by age 25 years. For females not overweight at baseline, a higher baseline dinner meal frequency was associated with a reduced risk of future overweight.
On average males in **Project EAT** ate lunch 6 times/week and dinner 6.4 times/week at the age of 15 years and reported that they decreased this lunch frequency to 5.4 times/week by age 25 years. Males' dinner frequency did not change significantly. For males not overweight at baseline, those who did increase lunch and dinner frequency (from a high base) increased their risk of overweight. (Quick et al., 2013)

In the **NGHS** a quarter of girls aged 9 to 10 years did not have 3 meals or more on any of the 3 days surveyed. By the final survey at 18 to 19 years this increased to half of all girls surveyed, showing that teenagers eating habits change over time. 15% of girls had 3+ meals at the start, but only 6% had 3+ meals by the end. Repeat cross-sectional analyses indicated that girls who did have 3+ meals on all 3 days had lower BMIz scores than girls who did not have 3+ meals every day. Black girls who had 3+ meals on all 3 days were less likely to be overweight. (Franko et al., 2008)

White girls in the **NGHS** ate meals and snacks more often each day than Black girls and at 9 to 11 years old they had significantly more eating episodes (White girls mean 5.2 episodes/day vs Black girls mean 4.7 episodes/day P <0.0001.

As girls grew older, they tended to reduce their snack frequency, but changes in meal frequency were small. Across the 10 year study, **NGHS** girls averaged 4.2 eating episodes a day (SD 1.0) made up of ~2.5 meals and ~1.7 snacks each day. More meals and snacks were eaten at weekends compared with week days (Ritchie, 2012).

In the **NGHS**, lower baseline eating frequencies led to greater gains in adiposity.

White girls with lower snack frequencies and fewer total eating episodes had larger 10 year gains in BMI and waist circumference. Black girls with lower meal frequency and lower snack frequency also had larger 10 year gains in BMI and waist circumference. Fewer total eating episodes at the start were associated with greater gains in waist circumference for Black girls.

In analyses of all girls, after adjustments for baseline adiposity, race, parental education, physical activity, television/video viewing, and TEI, lower initial eating frequencies were significantly associated with 10 year increases in BMI and waist circumference.

After adjustment for dieting, results were attenuated but still significant; the authors acknowledged that self-reports of dieting (ever mentioned during 10 years) may not account for all reverse causality (Ritchie, 2012). Some girls in the NGHS did have dieting concerns, even at age 11 years (Rehkopf et al., 2011).

In the same **NGHS** cohort, "Eats snack food" predicted BMI percentile change and onset of obesity but did not predict onset of overweight. (Rehkopf et al., 2011).

6.4.4 Fast food

The influence of fast food on future adiposity was reported by 4 papers, each from a different cohort.

In **GUTS** frequency of eating fried food away from home was the exposure (Taveras et al., 2005), whereas in the **IDEA and ECHO** cohorts the dietary behaviour under consideration was frequency of fast food purchases. (Laska et al., 2012)

Fast food as an eating behaviour was considered as a predictor of overweight or obesity in the **Project Eat** cohort (Quick et al., 2013) and of BMI percentile change in the **NGHS** girls cohort (Rehkopf et al., 2011).

Longitudinal findings are summarised in Table 6-5

Table 6-5 Fast food frequency and adiposity outcomes

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
GUTS Taveras 2005	6/9	14,355 Boys 6,610 Girls 7,745	9 to 14	3	Fried Food Away from home, Increase in 1 year from Never or < 1 time/week (Ref) vs 4 to 7 times/week)	Mean BMI Boys 9 to 12 yrs. 19.2 kg/m^2 Boys 13 to 14 yrs. 19.3 kg/m^2 Girls 9 to 12 yrs. 18.4 kg/m^2 Girls 13 to 14 yrs. 20.2 kg/m^2	Increase in BMI after 1 year β= +0.21kg/m^2, 95% CI 0.03 t0 0.39 kg/m^2	↑BMI in children Significant
IDEA & ECHO Laska 2012	8/9	693 Male 327 Female 339 **535** in analyses Male 276 Female 286	14.6	2	Increase in Fast food purchases/week	Baseline mean BMI Male 22.1 kg/m^2 Female 21.9 kg/m^2 Follow up mean BMI Male 23.4 kg/m^2 Female 22.8 kg/m^2	Change in % BF adjusted for TEI: Males: β = 0.17 s.e.= 0.22 P = 0.43 Females β = 0.33 s.e. = 0.14 P = 0.03	↑ % Body fat in females Significant at P≤0.05 Not significant at P≤0.003
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	Increase in Fast food frequency (times/week)	Overweight at 15 yrs.: ~25%, If not o/w at start, o/w at 25 yrs.: Males 45%, Females 34%	O/w Males: O.R. = 1.02, 95% CI 0.95 to 1.10 Females: O.R. = 1.15, 95% CI 1.04 to 1.27	↑risk of overweight, in females *not overweight at baseline* Significant

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
NGHS Rehkopf 2011	6/9	**2,150** girls in analyses Black 1,044 White 1,106	9.5	10	Predictor: Eats Fast Food	Baseline: Overweight 15% Obese 13% Follow-up: Overweight 16% Obese 18%	BMI %ile change: Eats fast food P = 0.32 Onset of overweight or obesity: Not sig.	Not a significant predictor
SUMMARY: Baseline intakes of fast food at younger ages did not predict BMI change in girls. Increasing fast food frequency during adolescence predicted adiposity gain and was associated with increased risk of overweight in females.								

6.4.4.1 Fast food methodologies

In **GUTS**, the number of times a week of eating fried food away from home (E.g. French fries, chicken nuggets) was found to be a reasonable proxy for fast food consumption from establishments such as McDonald's, Burger King or Taco Bell. (Taveras et al., 2005). At each annual questionnaire children were asked how often they ate fried food away from home (FFA), with 4 response options: Never or < 1 time/week, 1 to 3 times/week, 4 to 6 times/week, daily.

In the **IDEA and ECHO** cohorts, adolescents were asked how many times in the past month they had bought food from restaurant where food is ordered at a counter or drive through window. Named examples of fast-food facilities were given. (Laska et al., 2012) This method assumes that food purchased is eaten.

In **Project EAT**, respondents were asked how often they ate something from a fast food restaurant such as McDonalds or Burger King, as adolescents and 10 years later as young adults. (Quick et al., 2013)

In the **NGHS**, girls aged 9 to 10 years old were asked about fast food frequency during an in-person interview. Response options were: Never or < once a week, 1 - 3 times/ week, 4 - 7 times/ week and 8 or more times/ week. The accuracy of the fast food variable in predicting future BMI percentile change or onset of overweight or obesity was compared with 40 other potential predictors (Rehkopf et al., 2011).

6.4.4.2 Fast food frequency and adiposity outcomes

In **GUTS** at baseline 6% of boys and 3.5% of girls had FFA 4 times/week or more. Few children had FFA every day, but older children (13 to 14 years) consumed FFA more frequently than younger children (9 to 12 years). Based on semi-quantitative FFQ data, children who had FFA "4 to 7 times/week" at baseline had significantly higher energy intakes than children who had FFA "Never or < 1 time/week" (2,446 kcal vs 2,024 kcal, $P < 0.0001$). The diets of high frequency FFA consumers had more SSBs, red and processed meats and whole dairy foods and less low-fat dairy foods, fruit and vegetables than low frequency FFA consumers. Cross-sectional analyses found that boys baseline BMI (based on self-reported height and weight) tended to be higher if they were more frequent consumers of FFA.

By the 3 year follow-up the proportion of teenagers in GUTS who had FFA 4 times/ week or more had doubled, to 12.7% of boys and 7.5% of girls. In longitudinal analyses, children who increased their FFA frequency from "Never or < 1 time/week" to "4 to 7 times/week" one year

later had significantly greater annual BMI gains (+0.21kg/m^2) than children who had FFA "Never or < 1 time/week" at both times.(Taveras et al., 2005)

In the slightly older **IDEA and ECHO** high school cohorts, 14.6 year old boys and girls purchased fast food less than once a week on average, but again frequency increased with age. By 16 years old fast food purchases rose by an average +0.24 times/week, reaching ~ 1.2 times/week for boys and for girls. Unlike the GUTS study, no cross-sectional associations between fast food and BMI or % body fat were observed. At a significance level of $P \leq 0.05$, increased fast food purchase frequency predicted greater gains in % body fat in girls, after adjustment for TEI, but this was not significant if a more stringent $\alpha = 0.003$ was applied to correct for the number of tests (Laska et al., 2012).

Adolescents in **Project EAT** were also teenagers at baseline (~ 15 years old). The average fast food frequency was 1.6 times/week for females and 1.8 times/week for males. By the 10 year follow-up young women reported a lower average fast food frequency of 1.3 times/week, while young men increased their average fast food frequency to 2.1 times/week. Young women (not overweight at baseline) who increased their fast food frequency were more likely to be overweight at 25 years old. The same tendency was not observed in men, but the authors noted that young men who increased their lunch and dinner frequency were at higher risk of future overweight. They postulated that the increase in fast food between baseline and follow-up seen in young men may be part of "unhealthy food choices" that lead to weight gain (Quick et al., 2013).

In the **NGHS** the mean frequency of "eats fast food" for young girls aged 9 to 10 years was 1.6 times/ week (SD 0.8). "Eats fast food" was not a significant predictor of BMI percentile change, onset of overweight or onset of obesity. (Rehkopf et al., 2011)

6.4.5 Dieting

One paper focussed specifically on dieting as an exposure in the **GUTS** cohort (Field et al., 2003a). Dieting and weight control behaviours were also considered as independent predictors of overweight in **Project EAT** (Quick et al., 2013).

Longitudinal findings for less frequent eating and dieting are shown in Table 6-6

Table 6-6 Dieting and adiposity outcomes

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
GUTS Field 2003	5/9	14,972 Boys 6,769 Girls 8,203	9 to 14 Mean age Boys 11.9 Girls 12.0	3	Dieting, Never (Ref), Infrequent vs Frequent	Mean BMI Boys 19.1 kg/m^2 Girls 19.0 kg/m^2	Increase in BMIz Boys: infreq. dieters: β= 0.07, 95% CI 0.07, 0.08 freq. dieters: β= 0.07, 95% CI 0.06, 0.08 Girls: infreq. dieters: β= 0.04, 95% CI 0.04, 0.05 freq. dieters: β= 0.06, 95% CI 0.05, 0.06	↑BMIz in girls and boys Significant
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	Dieting at baseline (% yes)	Overweight at 15 yrs.: ~25%, If not o/w at start, o/w at 25 yrs.: Males 45%, Females 34%	O/w Males: O.R. = 2.01, 95% CI 1.31 to 3.08 Females: O.R. = 1.29, 95% CI 0.95 to 1.74	↑risk of overweight in males, *not overweight at baseline* Significant
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	Increase in Dieting (% yes)	Overweight at 15 yrs.: ~25%, If not o/w at start, o/w at 25 yrs.: Males 45%, Females 34%	O/w Males: O.R. = 3.25, 95% CI 2.24 to 4.72 Females: O.R. = 3.01, 95% CI 2.17 to 4.17	↑risk of overweight, *not overweight at baseline* Significant
SUMMARY: Dieting behaviours may promote weight gain in both sexes.								

6.4.5.1 Dieting methodologies

The **GUTS** paper referred to "dieting to lose weight" and "dieting to control weight" (Field 2003). Dieting was measured annually with the question, "During the past year, how often did you diet to lose weight or keep from gaining weight?" Children were categorised as non-dieters (Never), infrequent dieters (less than once a week) or frequent dieters (2 to 6 times/ week, every day). Children were also asked about frequency of binge eating, defined as "eating a very large amount of food in a short amount of time and feeling out of control during the eating episode". Three one year periods were investigated. Linear models were used to compare change in BMI z score (based on self-reported height and weight) of infrequent and frequent dieters in **GUTS** with those who never dieted. Girls and boys were considered separately. Simple models adjusted for baseline age and age at the end of the year, height, height change and Tanner stage of puberty. Other models additionally adjusted for activity, inactivity and energy intake. The final model also adjusted for binge-eating. (Field et al., 2003a)

In **Project EAT** Quick et al used regression models to test which variables predicted overweight at 10 year follow-up. In adolescence and again in young adulthood respondents were asked, "How often have you gone on a diet during the last year? By diet we mean changing the way you eat so you can lose weight." Respondents were classified as either non-dieters or dieters. Questions were also asked about binge eating and unhealthy weight control behaviours, such as meal skipping or the use of laxatives in the past year. (Quick et al., 2013)

6.4.5.2 Dieting frequency and adiposity outcomes

In **GUTS** dieters gained more weight over the three years than non-dieters.
4.5% of girls and 2.2% of boys were frequent dieters at baseline, 25% of girls and 13.6 % of boys were infrequent dieters. Non-dieters tended to be slightly younger (mean age 11.7 years), whereas infrequent and frequent dieters were slightly older (mean ages 12.4 years and 12.8 years respectively). Over the next two years girls dieting habits increased to 30% infrequent dieters and 8% frequent dieters but remained steady among boys. Dieters were more likely to be binge-eaters, but rates of binge-eating stayed low (~ 2% of girls and < 1% of boys). At baseline the mean BMI of non-dieters was 18.1 kg/m^2 but infrequent and frequent dieters were significantly heavier (mean BMI 20.5 kg/m^2 and 21.8 kg/m^2 respectively). Differences in BMI remained after adjustment for age and Tanner stage and may explain why children chose to diet.

Dieting was not a successful strategy for weight control. After adjustment for TEI, the additional change in girls BMIz score was +0.04 kg/m^2 for infrequent dieters and +0.06 kg/m^2 for frequent dieters, compared with non-dieters. In boys after adjustment for TEI, the

additional change in BMIz score was +0.07 kg/m^2 for infrequent and frequent dieters compared with non-dieters. Boys who practised binge eating also gained more weight than non-dieters. (Field et al., 2003a)

In **Project EAT**, among adolescents *who were not overweight* at baseline 49% of girls and 16% of boys reported any dieting in the last year, with just over half of girls and one quarter of boys reporting unhealthy weight control behaviours (meal skipping). Around 9% of girls and 2.4% of boys reported binge eating or extreme weight control behaviours, which is a higher rate than in the slightly younger adolescents from **GUTS**. Dieting, binge eating, and extreme weight control behaviours all increased significantly by young adulthood 10 years later, by which time 34% of females and 45%of males (not overweight at baseline) were classified as overweight. Any dieting in the last year increased to ~ 55% of females and ~ 25% of males.
In **Project EAT** dieting at baseline in males (OR 2.01 95%CI 1.31 to 3.08) and binge eating at baseline in males (OR 4.02 95%CI 1.11 to 14.6) increased their risk of overweight at follow-up. Baseline dieting and binge eating in females did not predict future overweight. However, change (increase) in either dieting or binge-eating in males and in females were associated with higher risk of overweight at follow-up. Unhealthy weight control behaviours at baseline in girls (OR 1.76 95%CI 1.29 to 2.41) and increases in unhealthy or extreme weight control behaviours in male or females also increased the risk of overweight (Quick et al., 2013).

6.4.6 Discussion of eating habits

All 14 included papers that explored eating habits in children and adolescents were from USA cohorts. The described eating habits and associations with adiposity outcomes may not apply to other settings. There was more evidence for girls than boys, as seven of the papers originated from the **NGHS** girls' cohort. The content of some **NGHS** papers overlapped, particularly in the case of breakfast and cereal eating, where funding was received from a cereal manufacturer.

Several papers reported cross-sectional associations of eating habits with adiposity. Eating habits that seemed helpful were eating breakfast (Laska et al., 2012), and higher meal frequency in girls (Franko et al., 2008). Detrimental eating habits were never eating family meals (Berge et al., 2015) and higher frequencies of eating Fried Food away from home (Taveras et al., 2005).

Longitudinally, there was also evidence that family meals, eating breakfast and higher eating frequency (meals and snacks) were eating habits that had beneficial associations with adiposity outcomes, although not all studies agreed. Fast food and dieting were linked with adverse adiposity outcomes.

Family meals were found to have a protective effect for adolescents of "all race/ethnicities, sex, age and SES status" in **Project EAT** (Berge et al., 2015), reducing overweight risk by up to a third after the 10 year follow-up. (At the 5 year follow-up no protective effect of family meals was seen, possibly because the sample sizes in stratified analyses were too small to find an effect.) The authors put forward three mechanisms to explain the protective effect of family meals on adult overweight/ obesity:

i. Family meals are "healthier" as fruit and vegetables are served.
ii. Family meals give a supportive environment which helps children regulate their eating.
iii. Adults are role models for helpful eating behaviours and recognition of satiety cues.

However, a previous systematic review of 15 studies of family meals found little evidence of an inverse relationship with childhood overweight (Valdés et al., 2013). In the **NGHS** "Eats family meals" was not an important predictor of BMI percentile change over an equivalent 10 year period among girls (Rehkopf et al., 2011). This may be because girls were younger at baseline (9 to 10 years old compared with 12 to 15 year in **Project EAT**) and more of them had family meals quite often. Another possibility is that, when asked about family meals, parents of **NGHS** girls selected higher frequency response options, introducing measurement error due to social desirability bias. (Categorical response rates were not given.)
Changes in family meal frequency were not investigated, but it seems likely that frequency declines as teenagers become more independent, either choosing or needing to opt-out of family meals at times.

Eating breakfast also seemed to be a protective habit, at least for females. It was the most important eating habit for predicting **NGHS** girls' change in BMI percentile, based on responses from the girls themselves (Rehkopf et al., 2011) and eating breakfast more often at baseline was associated with a reduced risk of future overweight in young adult women (Quick et al., 2013) and lower BMI outcomes in girls (Affenito et al., 2005). Breakfast eating was also associated with higher energy intakes and higher physical activity levels girls (Albertson et al., 2007), which suggests that girls who regularly ate breakfast were balancing energy intake and expenditure, which helped them to maintain a healthy weight. There was no equivalent evidence for boys.
Breakfast frequency tended to decrease through adolescence, with more breakfast "skipping" (Albertson et al., 2007; Laska et al., 2012). There were indications that adolescent females increased breakfast frequency by young adulthood, (Quick et al., 2013), but this could be explained by dietary measurement error due to social desirability bias.
Eating cereal more often (not necessarily for breakfast) at baseline was linked to lower body

fat outcomes (Albertson et al., 2009) and a lower risk of overweight in girls (Barton et al., 2005). Similarly Mexican-American children had lower BMI outcomes if they ate ready to eat cereal for breakfast more often (Balvin Frantzen et al., 2013). This could be because RTEC with milk was a less energy dense breakfast than traditional breakfast choices, not a benefit of RTEC per se.

Recent systematic reviews of children's breakfast habits and body weight included mainly cross-sectional studies, (Rampersaud et al., 2005; Szajewska and Ruszczynski, 2010) and (de la Hunty et al., 2013). Reviews confirmed that breakfast "skipping" is common in USA and European children and that, although breakfast eaters have higher TEI than breakfast "skippers", they are less likely to be overweight. Missing breakfast is more prevalent among girls, lower socio-economic groups and older children/adolescents. Reasons given for skipping breakfast included lack of time, not feeling hungry and weight concerns. Regular breakfast eating was advocated for children/adolescents, including high-fibre grains and cereals that improve glucose/insulin regulation and increase satiety, thereby promoting healthy weight gain. Review authors also called for more longitudinal studies.

Eating frequency, measured as eating meals more often at baseline, offered some protection against future overweight for females from **Project EAT** (Quick et al., 2013). This agreed with evidence from the **NGHS** girls cohort, where fewer eating episodes at baseline resulted in greater gains in adiposity (Ritchie, 2012), although reverse causality could account for some of this effect. "Eats snacks" at baseline was an important predictor of BMI percentile change and onset of obesity for **NGHS** girls (Rehkopf et al., 2011), but the study design (regression tree) does not make it clear if "Eats snacks" more often or less often is associated with obesity. As teenaged girls advanced towards adulthood, eating three or more meals a day and total eating episodes tended to reduce (Franko et al., 2008 , Ritchie, 2012). The reduction was attributed to skipping breakfast, increasing levels of freedom and self-determination, disrupted meal routines, and a desire to control weight by restricting meals and snacks. Females from **Project EAT** reported that they increased their lunch and dinner frequency by young adulthood (Quick et al., 2013), but no significant associations with overweight were seen. Males who increased lunch and dinner frequency were at greater risk of overweight, which was partly attributed to additional fast food meals. Eating more frequently may offer benefits in terms of adiposity outcomes, but only if greater frequency does not lead to excess energy intake.

Fast food habits were adverse for adiposity outcomes. Until their early teens few children had FFA or fast food very often (Taveras et al., 2005 , Laska et al., 2012). Perhaps because of this,

"Eats fast food" when **NGHS** girls were only 9 or 10 years old was not a significant predictor of BMI percentile change (Rehkopf et al., 2011). However fast food frequency tended to rise during the teenage years, to once a week or more. Increased fast food frequency during adolescence was significantly associated with greater BMI gains (Taveras et al., 2005), and tentatively linked to higher body fat percentages in females (Laska et al., 2012). Most women in **Project EAT** reported decreased fast food frequency compared to their mid-teens (social desirability bias?), but young adult females who increased their fast food frequency were at higher risk of overweight. (Quick et al., 2013)

Taveras et al observed that high frequency FFA consumers had higher energy intakes than low consumers. They noted that portion sizes in fast food establishments are large and suggested that the palatability of fat in fried food may lead to overconsumption, driving excess weight gain. (Taveras et al., 2005)

Dieting in the **GUTS** and **Project EAT** cohorts was shown to be detrimental. Dieting was a concern even in 9 to 10 year old girls in the **NGHS** (Rehkopf et al., 2011) but children under the age of 12 years tended to be infrequent dieters (Field et al., 2003a). Dieting prevalence increased with age, and girls were more likely to diet than boys. Despite lower energy intakes and higher levels of physical activity than non-dieters, dieters gained more weight, which suggests misreporting of one or both measures (Field et al., 2003a). Field et al proposed that dieting may improve metabolic efficiency (so less energy is needed) and that restricted diets may lead to episodes of over eating, leading to a positive energy balance and weight gain. Worryingly, about a third of the adolescents in **Project EAT** who were a healthy weight at baseline reported dieting and unhealthy weight control behaviours. Teenage boys who dieted at baseline were more likely to experience overweight as young adults, and individuals who increased dieting also seemed to increase their risk of overweight 10 years later (Quick et al., 2013) but, as it is not known exactly when each individual became overweight, reverse causality cannot be ruled out. Does dieting cause overweight or does overweight cause dieting?

6.5 Multiple predictors of overweight and obesity

Two papers investigated multiple dietary and non-dietary factors as potential predictors of overweight and obesity. Quick et al 2013 set out to find predictors of overweight incidence among male and female participants in **Project EAT** as they progressed from adolescence to early adulthood (Quick et al., 2013). Rehkopf et al 2011 assessed the relative importance of 41 baseline predictors of change in BMI percentile and the onset of overweight or obesity among adolescent girls from the **NGHS** (Rehkopf et al., 2011).

Individual foods and drinks and eating behaviours investigated by these two papers have already been presented in Chapter 5 but are briefly summarised here with other non-dietary predictors of overweight/obesity. Significant longitudinal findings are shown Table 6-7.

6.5.1 Methodologies and potential predictors

In **Project EAT** 4,746 ethnically diverse junior and senior high school students were surveyed in 1998/99. At 10 year follow-up 45% responded. Quick et al hypothesised that 3 personal, 6 socio-environmental and 15 behavioural factors at baseline (or changes in those factors) would predict overweight incidence at 10 year follow-up:

- Personal factors: Depressive symptoms, weight concerns and body satisfaction. Body satisfaction was measured with a modified version of the Body Shape Satisfaction scale (Pingitore et al., 1997)
- Socio- environmental factors: Home availability of high calorie snacks or healthful foods, parental concerns about weight, being teased about weight, having peers who dieted and perceived parental overweight.
- Behavioural factors: Energy and food intakes (servings/day of fruit, vegetables, whole grains and sugar sweetened beverages) measured by semi-quantitative YAQ and Willet's FFQ. Meal frequency (fast food, breakfast, lunch, dinner), Weight control behaviours (Meal skipping, laxatives, dieting), Binge eating and Physical activity (Moderate and vigorous PA, sedentary behaviours) from questionnaires.

Table 6-7 Multiple predictors of overweight and obesity

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
NGHS Rehkopf 2011	6/9	**2,150** girls in analyses Black 1,044 White 1,106	9.5	10	41 predictors: Dietary intake, eating behaviours, physical activity, psychological, social and familial factors	Baseline Overweight 15% Obese 13% Follow-up Overweight 16% Obese 18%	BMI percentile change (direction of change not evident)	20 significant (P<0.05) predictors, in order of importance: Body dissatisfaction, drive for thinness, unhappiness with physical appearance, household income, education level of primary care-giver, perfectionism, bulimia scale, anxiety, emotional eating index, interoceptive awareness, ineffectiveness, number of siblings, race, **eats breakfast**, **time to eat**., parent depression, parent BMI, **frequency of snacks**, **eats with soda on table**, self-worth.
NGHS Rehkopf 2011	6/9	As above	As above	As above	As above	As above	Onset of overweight (change from < 85th to ≥85th and <95th percentile of BMI)	Significant (P<0.05) predictors, in order of importance: Household income, ineffectiveness, race, perfectionism, number of siblings, male in household, parent physical activity, perceived stress, interoceptive awareness.
NGHS Rehkopf 2011	6/9	**2,150** girls in analyses Black 1,044 White 1,106	9.5	10	41 predictors: Dietary intake, eating behaviours, physical activity, psychological, social and familial factors	Baseline Overweight 15% Obese 13% Follow-up Overweight 16% Obese 18%	Onset of obesity (change from < 85th to ≥ 95th percentile of BMI)	Significant (P<0.05) predictors, in order of importance: Interoceptive awareness, household income, race, emotional eating index, distrust, **frequency of snacks.**

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	24 predictors: Personal, behavioural, socio-environmental and demographic factors	Overweight at 15 yrs.: ~25%, All o/w at 25 yrs. Males 56.1%, Females 47.5% Not o/w at baseline but o/w at 25 yrs. Males 45.4%, Females 34.2%	↑ incidence of overweight	Baseline predictors (P<0.05): Females: Body dissatisfaction, weight concerns, fasting, weight-related teasing, parental concerns about weight, perception that biological parents were overweight. Males: Weight concerns, dieting, binge eating, parental concerns about weight.
Project EAT Quick 2013	5/9	As above	As above	As above	As above	As above	↑ incidence of overweight	10 year change (P<0.05): Females: Body dissatisfaction, depressive symptoms, weight concerns, **increased fast food freq.**, fasting, purging, dieting, binge eating, weight-related teasing. Males: Body dissatisfaction, weight concerns, **increased lunch freq., increased dinner freq.,** fasting, purging, dieting, binge eating, weight-related teasing, peer dieting.
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	24 predictors: Personal, behavioural, socio-environmental and demographic factors	Overweight at 15 yrs.: ~25%, All o/w at 25 yrs. Males 56.1%, Females 47.5% Not o/w at baseline but o/w at 25 yrs. Males 45.4%, Females 34.2%	↓ incidence of overweight	Baseline predictors (P<0.05): Females: Body satisfaction, **higher whole grain intake, higher breakfast freq., higher dinner freq.** Males: None were significant.

Study & Paper	NOS stars	Popln.	BL age (yrs.)	F' up (yrs.)	Exposure	Adiposity at baseline	Adiposity outcomes	Findings
Project EAT Quick 2013	5/9	**1,643** *not o/w at baseline* (out of 2,134) Males 756 Females 887	15	10	24 predictors: Personal, behavioural, socio-environmental and demographic factors	Overweight at 15 yrs.: ~25%, All o/w at 25 yrs. Males 56.1%, Females 47.5% Not o/w at baseline but o/w at 25 yrs. Males 45.4%, Females 34.2%	↓ incidence of overweight	10 year change (P<0.05): Females: Body satisfaction, moderate and vigorous physical activity. Males: Body satisfaction, **increased vegetable intake?** moderate and vigorous physical activity.
SUMMARY Studies show that overweight and obesity is a multi-factorial issue. Dietary intake and eating behaviours (in **bold** in the table) explain only part of the risk and are not the most important predictors. Note that predictors are not necessarily causal factors.								

BMI in **Project EAT** was calculated from self-reported height and weight. Adolescent overweight was based on a BMI ≥ 85th percentile, using the CDC 2000 reference. Adult overweight was based on BMI ≥ 25kg/m^2. At the 10 year follow-up 47% of young women and 56% of young men were overweight. Adolescents who were overweight at baseline were excluded from analyses. Among those who were not overweight at baseline, 34% of females and 45% of males were overweight ten years later. Multivariable logistic regression was used to estimate odds ratios for overweight at follow-up for each of the 24 predictors at baseline and for the 10 year change in each predictor. Models of food intakes (fruit, vegetables, whole grains and SSBs) were adjusted for TEI. (Quick et al., 2013)

In the **NGHS** Rehkopf et al initially considered 142 risk factors for adolescent weight gain, measured in Black and in White girls. This was reduced to 41 potential predictors including dietary intake and eating behaviours, physical activity, psychological and social risk factors and parental health, mostly measured at baseline when the girls were between 9 and 10 years old.

- Dietary intake: Total calories, calories from fat and from protein, measured by a 3 day FD.
- Eating behaviours: Eats breakfast, eats snack food, eats fast food, eats while watching television, eats with soda on the table*, family eats dinner together*, eats dinner alone*, time to eat*, from baseline interview or parental report*.
- Physical activity: Estimated as metabolic equivalents or METS/week.
- Psychological: Traits related to eating behaviour (Body dissatisfaction, bulimia, drive for thinness/concern with dieting, interoceptive awareness/ability to distinguish between hunger and satiety) were measured by a 64 item Eating Disorders inventory (EDI) when the girls were 11 to 12 years old (Garner et al., 1983). Other psychological factors were evaluated using validated scales at baseline interview.
- Social: Number of siblings*, race/ethnicity, male in household*, category of household income*, category of education level of mother/primary care-giver*, self-report or parental report*.
- Parental health: Parental depression evaluated using a validated scale, parent's BMI* from self-reported height and weight of mother/primary care-giver, parents health*, parents' physical activity*, importance of exercise* from parental report*.

NGHS girls' BMI was calculated from height and weight measured by trained examiners. Approximately 10% of the recruited cohort of 2,379 girls were missing BMI at baseline or follow-up, so were excluded from analyses. At the start (age 9 to 10 years) 15% of girls had overweight and 13% had obesity, based on the CDC 2000 reference. Ten years later 16% of young women in the analyses had overweight and 18% were classified as obese.

Using a tree-based regression method (Random forest) the relative ability of each risk factor to predict future BMI percentile change, onset of overweight or onset of obesity was assessed. All 41 predictor variables were ranked in order of their mean decrease in (predictive) accuracy (MDA). A higher MDA shows that variable is a more important predictor. (Rehkopf et al., 2011) Although the baseline predictors are ranked in order of importance, with this method the *direction of influence on BMI percentile change is not clear*.

6.5.2 Predictors of overweight and obesity

In **Project EAT**, several baseline dietary factors (Energy intake, fruit, vegetable or sugar sweetened beverage intakes, fast food frequency, lunch frequency and dinner frequency) were not associated with future overweight. In females (not overweight at baseline), higher intakes of whole grains, higher breakfast frequency and higher dinner frequency at baseline were protective against future overweight.

Changes in intake of whole grains, fruit and SSBs were not associated with future overweight in **Project EAT**. An increase in vegetable intake was reported to reduce the risk of overweight in males, but this is not certain as dietary intakes in Project EAT were measured by the YAQ at baseline and Willett's FFQ at 10 year follow-up. The difference between the FFQs is not a valid measure of change as the two FFQs were only moderately correlated for absolute intakes (Larson et al., 2012).

Males who increased how often they ate lunch and dinner increased their risk of overweight, perhaps because additional meals were energy dense fast foods. Increased fast food frequency in females significantly increased their risk of future overweight.

Non-dietary factors linked with the risk of future overweight included personal factors (body dissatisfaction or concerns about weight), socio-environmental factors (parental concerns about weight/related behaviours and weight related teasing) and weight control behaviours (fasting, purging, dieting or binge eating), either at baseline and/or increased over time.

The only non-dietary factors which decreased the risk of future overweight were increased physical activity between baseline and follow-up, body satisfaction at baseline in females and an increase in body satisfaction for females and males. Overall, young adults in Project EAT reported a decrease in body satisfaction between baseline and follow-up.(Quick et al., 2013)

In the **NGHS** girls' cohort Rehkopf et al found that 20 of the 41 variables considered were significantly associated with BMI percentile change, based on a p value ≤ 0.05. Fewer variables

predicted onset of overweight or onset of obesity. Sensitivity analyses, restricted to those girls who were not overweight at baseline, or using baseline measures at 11 years (instead of 9 years) yielded similar results. (Rehkopf et al., 2011)

Energy intake, calories from fat and calories from protein did not predict BMI percentile change or onset of overweight or obesity in the **NGHS** and neither did physical activity. Among the eight eating behaviours considered, only four predicted BMI percentile change. Frequency of eating breakfast was the most important eating behaviour predictor, accounting for a small but still significant mean decrease in accuracy (MDA 0.44), followed by time to eat (MDA 0.41), frequency of snacks (MDA 0.24) and eats with soda on the table (MDA 0.22). No eating behaviours predicted onset of overweight and only frequency of snacks predicted onset of obesity (MDA not given).

Psychological factors were among the most important and significant predictors of BMI percentile change in the **NGHS**, with the largest values for mean decrease in accuracy. Significant predictors were body dissatisfaction (MDA 5.97) drive for thinness (MDA 4.32), bulimia (MDA 1.04) and interoceptive awareness (MDA 0.82). Only interoceptive awareness (possibly a lack of it?) predicted onset of overweight.

Household income was the most important significant social predictor of BMI percentile change (MDA 2.01) and predicted onset of overweight and onset of obesity. Other significant social predictors were parent education level (MDA 1.66), number of siblings (MDA 0.64) and race (MDA 0.60). Number of siblings predicted onset of overweight. Race predicted onset of overweight and onset of obesity.

Parental depression (MDA 0.37) and parent's BMI (MDA 0.25) were significant predictors of child's BMI percentile change but did not predict onset of overweight or obesity. (Rehkopf et al., 2011)

6.5.3 Discussion of multiple predictors

The **NGHS** cohort and the **Project EAT** cohort are both USA based cohorts with over 2,000 participants and a 10 year follow-up, but there are some key differences between the cohorts and the methodologies used which make comparison difficult.

The **NGHS** included girls only, aged 9 or 10 years at baseline in 1987 or 1988, most of whom were followed annually for all 10 years of the study. At baseline 28% had overweight or obesity, rising to 34% at follow-up. Although the intention was to identify the best predictors of future overweight or obesity, all girls with BMI measures at both times were kept in the analyses, even if they were classified as overweight or obese at baseline.

Project EAT included adolescents of both sexes who were recruited a decade later in 1998/99 at a mean age of 15 years and followed up only once, ten years later. Attrition rates were high (over 50%) which introduced bias. Participants who responded to baseline and follow-up surveys were more likely to be female, white and of higher socio-economic status.

Baseline prevalence of overweight was 25%, rising to 51% 10 years later, markedly higher than in NGHS girls. **Project EAT** used the same CDC 2000 reference for overweight as the NGHS, but BMI was based on self-reported height and weight. Quick et al acknowledged that this may have caused bias but pointed to high correlations between measured and self-reported BMI in the cohort at baseline (male r = 0.88 and female r = 0.85) and in a sub-sample at the 10 year follow-up (male r = 0.95 and female r = 0.98). Never-the-less these correlations show that self-reported height and weight generally underestimated BMI, so the true prevalence of overweight in **Project EAT** was likely higher. Unlike the NGHS, analyses in **Project EAT** were restricted to participants reportedly not overweight at baseline.

Although Rehkopf and Quick considered similar behavioural, social and psychological factors as potential predictors of future overweight and obesity, they did not use the same methods to measure exposures or use the same analytical techniques to explore their association with future adiposity. Only baseline exposures were considered as predictors in **NGHS** analyses, while baseline and change in exposures and their association with risk of overweight were considered by **Project EAT**. (Though described as such, change in exposure is not a predictor in the true sense, as change can only be measured retrospectively.) One advantage of Quick's statistical method, multivariable logistic regression using odds ratios, is that the direction of influence of each considered predictor is clear, whereas the tree-based regression method used by Rehkopf, which helpfully ranked potential predictors, does not explicitly show the direction of influence on BMI percentile change. Additionally, for categorical variables such as

race (which did predict onset of overweight and obesity in **NGHS** girls) the category which had the highest risk is not self-evident and the authors did not provide any clarification.

Neither study found that total energy intake predicted overweight.
Measurement of baseline dietary intake in **NGHS** by 3 day food diaries, allowed Rehkopf to consider the percentage of total energy from fat and protein, yet macronutrient intakes were not helpful predictors. **Project EAT** employed a semi-quantitative FFQ, so Quick investigated foods rather than macronutrients. Baseline whole grain intake in girls (and maybe change in vegetable intake in males?) predicted overweight in the cohort. This suggests that, in preference to TEI (which is difficult to measure), specific foods may be more useful as predictors of overweight.

Both studies considered physical activity. In the **NGHS**, baseline physical activity was not a significant predictor of BMI percentile change, overweight or obesity onset ten years later. Similarly, in **Project EAT** baseline physical activity and sedentary behaviours were not associated with future overweight, but a change (increase) in moderate to vigorous physical activity did predict a lower risk of future overweight in both girls and boys. It seems plausible that change in physical activity has a greater influence on adiposity outcomes 10 years later than physical activity at baseline, but few longitudinal studies of adolescents have reported this relationship (Rauner et al., 2013).

Energy expenditure and energy intake are recognised factors in energy balance, so it is perhaps surprising that physical activity and more of the dietary variables did not predict adiposity outcomes. This could be due to measurement errors introducing bias to the study findings, or because studies were underpowered to find certain predictors. By restricting the analysis to those not overweight at baseline, **Project EAT**'s sample size reduced by over 25% to 1,643 and in **NGHS** relatively few girls of healthy weight at baseline developed overweight or obesity during the 10 year study. There were few "new events" of onset of overweight or obesity.

Each study asked about eating behaviours or habits. Some eating behaviours correlate with dietary intake, socio-environmental or psychological factors investigated by the other study; the two studies broadly agreed about the influence of some eating habits on adiposity outcomes.
As already described in the section on Eating Habits (see page 192 onwards) in the **NGHS** "Eats breakfast" was a significant predictor of BMI percentile change for girls, while adolescent girls in **Project EAT** who had a higher frequency of breakfast eating at baseline reduced their risk of overweight 10 years later.

In **Project EAT**, baseline frequencies of fast food were not associated with future overweight and "Eats fast food" at baseline was not a significant predictor of BMI percentile change, overweight or obesity onset in **NGHS** girls. (This may be because at younger ages most children ate fast food infrequently, as previously discussed.)
Unhealthy weight control behaviours increased the risk of overweight at follow-up in **Project EAT** and "Bulimia" (binge eating or purging) predicted BMI percentile change in **NGHS** girls.

There was conflicting evidence about the predictive value of other eating habits.
Girls in **Project EAT** with a higher frequency of eating dinner at baseline reduced their risk of overweight 10 years later, yet in the **NGHS** neither "Eats dinner alone" nor "Family eats dinner together" were significant predictors. "Eats snack food" predicted BMI percentile change and obesity onset in **NGHS**, yet home availability of high-caloric snack foods did not predict overweight in **Project EAT** (perhaps they were not consumed, even if available). "Eats with soda on the table", as reported by the parent, predicted BMI percentile change in **NGHS** girls, but baseline SSB intakes were not associated with future overweight in **Project EAT**.

In **Project EAT** socio-environmental factors were explored as predictive variables, but all regression models adjusted for socio-economic status, tacitly acknowledging its influence on future adiposity. Personal (psychological) factors including body satisfaction/ dissatisfaction and weight concerns were shown to be significantly associated with future overweight, although in the case of change (increase) in these factors, the authors again acknowledged the possibility of reverse causation. They suggested that obesity prevention/treatment interventions for adolescents should include strategies to support healthy eating behaviours and promote a positive body image, while limiting negative comments about weight.

In the **NGHS** household income and psychological factors including body dissatisfaction were the most important predictors of BMI percentile change, overweight or obesity onset. The authors cautioned that although they had identified important predictors, these predictors are not necessarily causal factors. However, relationships were examined longitudinally, using baseline variables (selected because the literature suggested an association with adolescent weight gain) and BMI outcomes at follow-up.

6.6 Conclusion

The studies of dietary patterns included in this systematic review, confirmed some findings from studies of specific foods and drinks, detailed in Chapter 5.
Many foods and drinks which were individually associated with an increased risk of overweight (SSB, flavoured milk, energy dense convenience foods, energy dense snack foods) also featured heavily in dietary patterns that were associated with adverse adiposity outcomes, including the "Energy Dense" DP identified in **ALSPAC** and the "Snacking" DP in **Bogotá School Children**. Similarly, foods which were individually associated with a reduced risk of overweight (whole grains, dairy foods/milk, vegetables) improved DASH diet quality scores and featured in dietary patterns which were linked to more beneficial adiposity outcomes, such as the "Healthy" DP in **NGHS** girls and possibly the "Vegetable" DP in **Project EAT**.
The adverse effect of the "Fruit" DP in **Project EAT** was doubtful after adjustment for baseline weight status, which mirrored the uncertainty about juice and fruit as single predictors.

This systematic review found that some eating habits or behaviours (family meals, breakfast eating, eating frequency) are protective against future adiposity. These helpful habits are interlinked. For example, eating more frequently during childhood and adolescence may indicate more organised and well-regulated eating habits, including family meals *and* snacking. It is interesting that higher frequency of snacking (eating between meals) seems to be a helpful habit, as snack foods, sugar sweetened beverages and high adherence to a "Snacking" DP characterised by high sugar/high fat snack foods and sugary drinks, increased the likelihood of increased changes in BMI. This hints that snacking as an eating habit, does not necessarily equate to consuming energy dense foods and drinks.
Other eating habits (breakfast skipping, eating with soda, high fast food frequency and dieting) predicted adverse adiposity outcomes, and often were linked to other behaviours detrimental to health, such as infrequent exercise or unhealthy weight control behaviours.

Such non-dietary factors were considered in the multiple predictor studies by Rehkopf et al and Quick et al, which demonstrated that childhood/adolescent overweight and obesity is a complex, multi-factorial issue. Dietary intakes and eating behaviours explain only some of the risk in some children, some of the time. They are not even the most important predictors of overweight/ obesity. When choosing baseline variables to include in a simple dietary assessment tool to predict future obesity risk in young people, we need to consider socio-economic, psychological, environmental and physical activity predictors in addition to specific foods and eating habits.

Chapter 7 Meta-analysis of sugar sweetened beverage intakes and adiposity outcomes.

7.1 Summary

The associations between quantified intakes of most specific foods or drinks and later adiposity outcomes were reported by only a few papers from a few cohorts. The exception was sugar sweetened beverages. Quantified SSB intakes and adiposity outcomes were reported by 10 papers from 9 different cohorts. SSB studies were similar enough to justify quantitative synthesis.

This chapter (Chapter 7) sets out the definition of sugar sweetened beverage and the selection of SSB studies for data extraction. As there was methodological heterogeneity between studies, conservative random-effects meta-analyses were used. Results of meta-analyses are presented as forest plots and discussed.

Few SSB studies used the same adiposity outcome measures, which limited options for pooling results. Exploratory meta-analyses which pooled all reported adiposity outcome measures suggest that higher intakes of SSBs might be linked to greater adiposity outcomes. Two meta-analyses of SSB intakes and change in BMI, reported in three cohorts using β coefficients, also suggest that a greater increase in SSB intake is associated with a greater increase in BMI, but this was not significant. A dose-response meta-analysis was not feasible with the available data.

7.2 Introduction

The systematic review of childhood and adolescent cohorts measuring whole diet and subsequent adiposity included 35 papers from 14 common studies or cohorts. Most dietary exposures were reported by only a few cohorts. Perhaps reflecting public health concerns, the dietary exposure investigated most frequently was sugar-sweetened beverages, which were considered in various ways by 13 papers from 10 cohorts. This gave some opportunity for quantitative synthesis, using meta-analysis to generate a pooled effect-size, which could potentially strengthen the evidence.

Meta-analysis is a statistical way of combining results from multiple studies that have addressed the same research question. The larger, combined sample size offers more statistical power, giving a more robust estimate than is possible from each individual study. The increase in statistical power decreases the probability of making a Type II error (wrongly failing to reject the null hypothesis). Alternatively, an increase in statistical power can be described as an increased ability to detect an effect, *if* that effect is real. To be worthwhile, a minimum of three studies from different populations or cohorts that have used comparable effect-size measures are needed.

7.3 Methods

7.3.1 Selection of studies

For the purposes of this meta-analysis "sugar sweetened beverage" (SSB) includes:

- Regular carbonated or non-carbonated soft drinks, sodas, sports drinks, energy drinks, cordials, squash or fruit juice drinks sweetened with a caloric sweetener such as sugar.

All papers which reported sugar-sweetened beverages, drinks or sodas (as intake or in a wider context such as part of a dietary pattern) and adiposity outcomes were considered. Results in full-texts were examined to find effect-size measures suitable for meta-analysis.

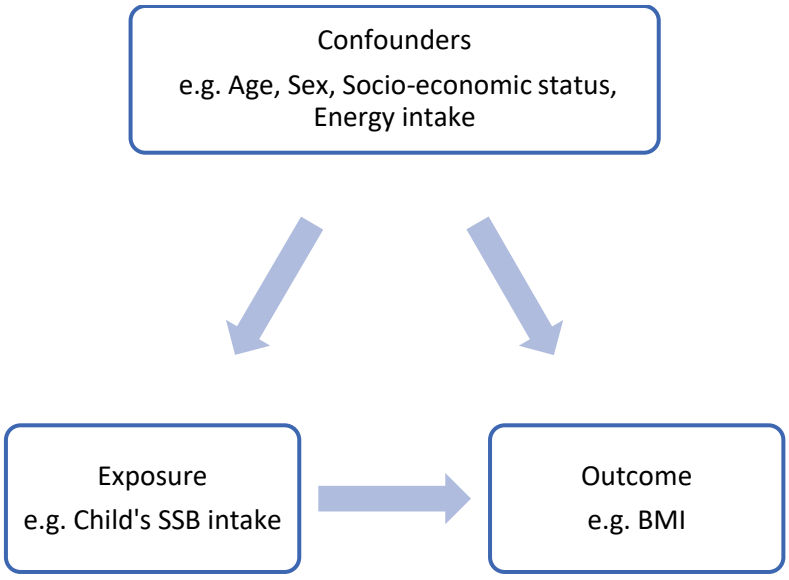
7.3.2 Data extraction and synthesis

As previously described (See Chapter 2), study characteristics, cohort participant details, dietary assessment and reported dietary exposures, adiposity assessment and reported adiposity outcomes were extracted from each paper. Additionally, from those papers which reported effect-size measures of the associations between SSB intake and adiposity outcomes, the following information was extracted to Excel:

- SSB intake at baseline and follow-up, change in SSB intake, SSB serving size.
- Adiposity outcomes, continuous: Mean difference between the most exposed and least exposed group, or β coefficient, with s.e., SD, low/high 95% confidence interval (CI) or P value.
- Adiposity outcomes, categorical (Overweight/obese vs normal weight): OR, with s.e., SD, or low/high CI.
- How models were adjusted.
- Sample size in analyses

When examining the relationship between an exposure, such as SSB intake (the independent variable) and an outcome, such as overweight (the dependent variable), it is helpful to consider other variables which may influence both the dependent variable and the independent variable, acting as confounders of the relationship of interest. (See Figure 7-1.) A confounder may be a cause of the exposure, or the cause of the outcome in unexposed people. However, as a confounder is not itself affected by the exposure, it is not on the causal pathway from the exposure to the outcome.

Figure 7-1 Confounding variables



Confounding occurs when there are differences in the outcome in the exposed and the unexposed populations that are due to factors other than the exposure. Variables that are thought to act as confounders can be factored into the experimental design of randomised controlled trials, but in observational studies this is not possible (Harrell, 2001). Due to confounding factors, observational studies may give an estimate of an association that does not reflect the true underlying relationship, as individuals exposed to the factor being investigated may vary in other aspects that are also important in relation to the risk of

developing the outcome. Hence it is usual practice to adjust for known confounding factors in the analysis (Egger et al., 2001).

As energy intake is part of the mechanism whereby dietary intakes influence adiposity outcomes, it is sometimes included as a confounder in regression models to investigate the association between a dietary exposure and an adiposity outcome, although inaccuracies in the measurement of confounding factors may add residual confounding (Egger et al., 2001). Adjusting for energy intake often attenuates (reduces) the effect size or makes an association no longer significant.

Wherever available, data was extracted from two models:

- the most adjusted model that did not adjust for energy intake
- the final energy adjusted model

Included papers were grouped by the type of effect-size measure they had employed in their analyses (OR, mean difference or β coefficients), to determine whether there was data from at least three cohorts to use in a meta-analysis.

In order to run meta-analyses in STATA using β coefficients, the standard error of the β coefficient is required. If the β coefficient s.e. value was not given, as was the case for three papers (Field et al., 2014; Laska et al., 2012; Libuda et al., 2008), the s.e. variable was generated in Excel as follows:

- calculating the β coefficient standard error from the extracted low and high 95% confidence intervals using the formula:
  β coefficient se = (((high CI – estimate)/1.96) + ((estimate – low CI)/1.96))/2
- deriving the Z score from the extracted P value (assuming standard normal distribution) using the Excel formula:
  Z = ABS(NORMSINV(P value/2))
  and then using Z in the formula:
  β coefficient se = β coefficient/Z

Some mixed cohorts gave results separately for girls and boys. These results were combined using the metan command (consistently using random effects) to give one pooled estimate and s.e. for boys and girls together, which could then be used in a meta-analysis with other mixed cohorts.

To facilitate sorting of the data into subgroups for random-effects meta-analysis, extracted information was coded as shown in Table 7-1 on the next page.

Table 7-1 Codes applied to extracted variables

Variable type	Description	Code
Sex	Boys	1
	Girls	2
	Mixed	3
	Boys and Girls combined (random effects meta-analysis)	4
Dietary Analysis Tool	Food frequency questionnaire	1
	24 hour recall	2
	Food record/diary	3
Adiposity measure	BMI	1
	BMIz or BMI-SD	2
	Waist circumference	3
	Skinfold thickness	4
	Body fat %	5
Analysis type	SSB intake at baseline, adiposity outcome at follow-up	1
	SSB intake change, adiposity outcome at follow-up	2
	SSB intake at baseline, adiposity outcome change	3
	SSB intake change, adiposity outcome change	4
	Mean of SSB intake at baseline and follow-up, adiposity outcome change	5
Model adjusts for energy	Yes, adjusted for energy	1
	No, did not adjust for energy	0

7.3.3 Random effects meta-analysis

Meta-analysis was carried out using STATA 13 and STATA 15 software (Stata Corp).

Pooled estimates with forest plots were generated using the metan command in STATA. The forest plots provide a graphical summary of the meta-analysis.

In a fixed effect meta-analysis the summary effect calculated by STATA is an estimate of the true effect-size, assuming that all the studies in the analysis are measuring the same effect-size, common to all studies.

Although the SSB studies were similar enough to make a synthesis of the data potentially worthwhile, there were differences between their participants and methodologies, so allowances must be made for the possibility that the true effect-size could vary somewhat

from study to study. Not only is there within-study error, there is also variation in the true effects between studies.

In the more conservative random effects meta-analysis which gives wider confidence intervals, STATA assigns study weights to minimise both types of variance and calculates a summary effect that is an estimate of the mean of all the relevant true effects (Borenstein et al., 2011). Random effects, rather than fixed effect, meta-analysis was used throughout.
Too few papers gave mean difference or O.R. as the effect-size measure for meaningful meta-analysis. All papers considered for meta-analysis reported associations between SSB exposure and change in continuous adiposity outcomes using β coefficients. Papers considered SSB exposure in three different ways:

- SSB intake at baseline
- SSB intake change over time
- Mean of SSB intake at baseline and follow-up.

If two approaches were presented, data from the model which considered SSB intake change over time was preferred.

As an exploratory step to understand the data, random effects meta-analyses of *all* adiposity outcome measures were run, first using data from models which did not adjust for Energy, and then using data from models which did adjust for Energy. Separate meta analyses were conducted for the mixed/combined cohorts, followed by girls only and boys only. The six resulting exploratory meta-analyses and forest plots were stratified by adiposity outcome measures for comparison only. Note that as some cohorts featured more than once in the same meta-analysis (as they had more than one type of adiposity outcome measure) the summary estimate is not a valid estimate of the true effects.

Using the codes applied to extracted information, data was sorted into subgroups, by sex, adiposity measure, and whether data was available from a not energy adjusted and/or an energy adjusted model, to determine whether there was sufficient evidence (from three or more studies) for a more formal meta-analysis. If enough data was available a meta-analysis was run for each individual adiposity outcome measure, using information extracted from the most adjusted model in each paper.

Heterogeneity was evaluated using the I^2 statistic, which indicates the proportion of total variation attributable to between-study heterogeneity. An I^2 below 25% was considered as low heterogeneity, with an I^2 between 25% and 50% indicating moderate heterogeneity. An I^2 above 50 % to 75% indicated substantial heterogeneity; an I^2 above 75% was judged as high. Based upon previous research (Malik et al., 2013) it was anticipated that heterogeneity would

be above 50%, a level at which caution should be exercised when reaching conclusions based upon the data (Higgins et al., 2003). It is advised that the I^2 statistic is interpreted cautiously when there are fewer than 10 studies in the meta-analysis.

The likely presence of bias in meta-analyses (such as publication bias where only significant "positive" findings are published) can be explored using funnel plots and Egger's test of asymmetry (Egger et al., 1997) if there are enough studies (Threapleton et al., 2013). If there are too few studies in the meta-analysis the ability of Egger's test to detect bias is compromised and results will be uncertain.

7.3.4 Dose response or standardised difference meta-analyses

Reported SSB serving sizes and study lengths were surveyed to assess whether a dose-response meta-analysis was feasible, as done by Malik and colleagues in their comprehensive systematic review and meta-analysis of SSB exposure and change in BMI outcomes (Malik et al., 2013). Replicating their methods required many assumptions about the limited data available. A dose-response meta-analysis was not pursued.

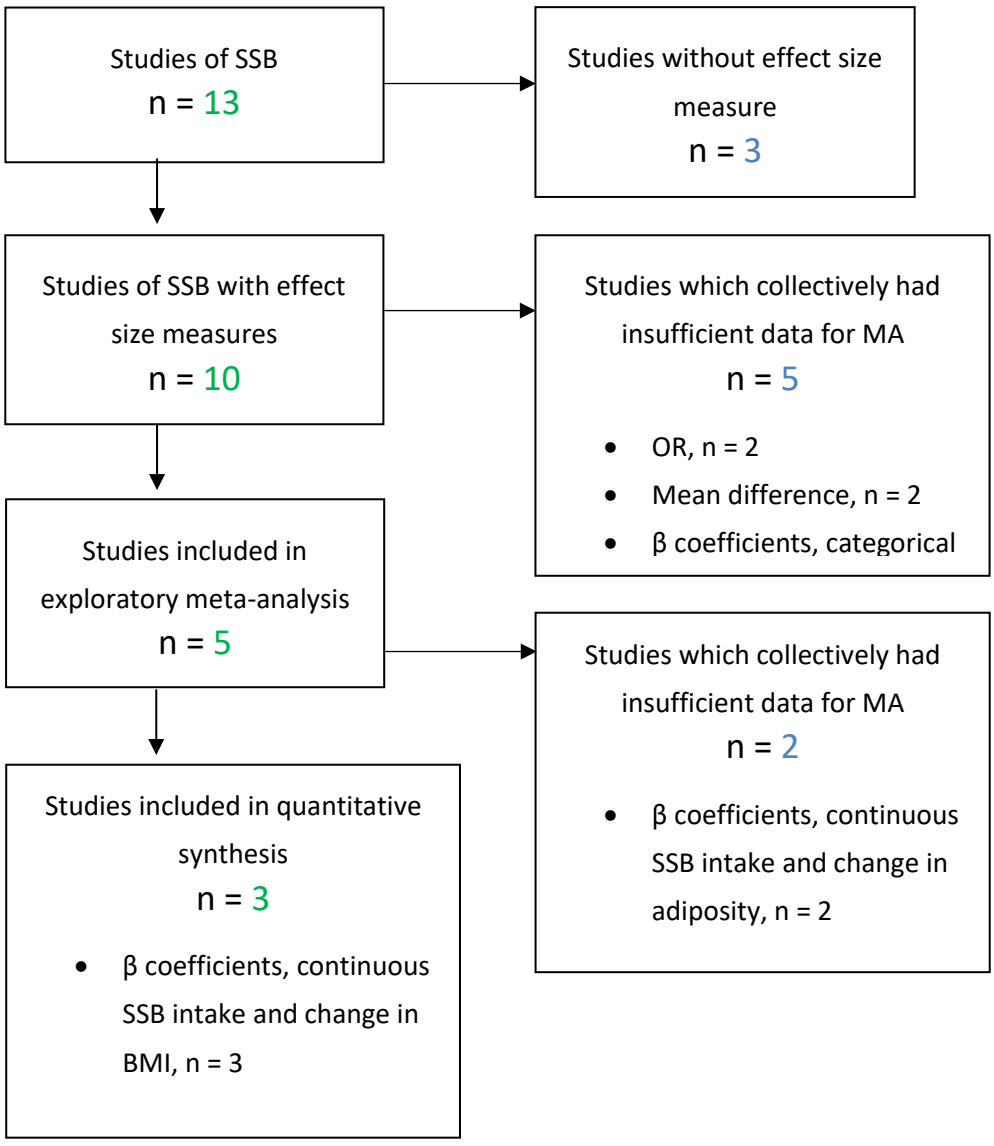
A further option is to use a standardised difference meta-analysis that allows the pooling of studies that use different scales to measure the same effect, but as only two included studies reported mean difference this was not worthwhile.

7.4 Results

7.4.1 Suitable studies for meta-analysis

Thirteen papers investigated SSBs and adiposity intakes in some way, but not all presented suitable effect sizes or had enough data collectively to be included in meta-analyses. All SSB papers are narratively reviewed in Chapters 5 or 6. The numbers of SSB papers considered for meta-analysis are shown in the PRISMA style flow-chart in Figure 7-2.

Figure 7-2 SSB papers considered for meta-analysis



Three papers without effect size measures considered SSBs in the context of dietary patterns (Ritchie et al., 2007), as obesity predictors (Rehkopf et al., 2011) or in association with fried foods away from home (Taveras et al., 2005).

Ten papers reported SSB intakes and adiposity outcomes using an effect-size measure. SSB exposure was reported as intake at baseline, as change in intake over time, or as the mean of intake at baseline and follow-up. Adiposity outcomes were reported as dichotomous variables (overweight/obese or not), or as continuous variables E.g. change in BMI.
Most studies adjusted for TEI. The **DONALD** study used models which adjusted for residual energy (i.e. Energy from all sources other than SSBs). The **Framingham Children's study** presented models that were adjusted for the percentage of energy from fat. The **RAINE** study did not present models that were energy adjusted, as associations were unchanged after adjustment for TEI. Only the **GUTS II** cohort did not adjust models for energy.

Five of the ten papers had insufficient collective data for meta-analysis as the effect size measure was only used in two papers/cohorts:

- Two papers (Quick et al., 2013; Ambrosini et al., 2013) reported **odds ratios** (OR) for SSB intake and dichotomous adiposity outcomes (overweight/obese or not).
  Project EAT investigated continuous measures of SSB intake at baseline and change and overweight outcomes. Not significant.
  The **RAINE** study used tertiles of SSB intake and overweight outcomes. Girls who moved into the top tertile of SSB intake had an increased risk of overweight.
- Two papers (Shroff et al., 2014) (Hasnain et al., 2014) reported the **mean difference** in continuous measures of adiposity at follow-up between categories of SSB intake at baseline. **Bogotá Schoolchildren** employed 5 categories of frequency of SSB intake at baseline. Higher SSB frequency positively associated with change in BMI and change in waist circumference.
  The **Framingham Children's study** used tertiles of baseline SSB intake. No association with BMI, body fat percentage, waist circumference or sum of 4 skinfolds at follow-up.
- Two papers (Zheng et al., 2014) (Ambrosini et al., 2013) presented results as **β coefficients** for the associations between **categorical measures of SSB intake** and continuous adiposity outcomes.
  The **EYHS** used 3 categories of SSB intake based on servings/day. Consumers had significantly larger increases in BMI and waist circumference than non-consumers.
  The **RAINE** study used tertiles of SSB intake based on quantified servings/day. Girls who moved into the top tertile of SSB intake had a greater percentage change in BMI.

Five papers from five different cohorts gave **β coefficients** for the associations between **continuous measures of SSB intake** and continuous adiposity outcomes (change over time). (Libuda et al., 2008) DONALD study, (Zheng et al., 2015) Danish EYHS, (Field et al., 2014) GUTS II, (Laska et al., 2012) IDEA & ECHO and (Striegel-Moore et al., 2006) NGHS. Adiposity outcomes reported were change in BMI, BMIz, waist circumference, skinfold thickness or body fat %. Most types of adiposity outcome were reported by only one or two papers, but all five papers were included in exploratory meta-analyses.

Three papers from different cohorts (GUTS II, IDEA & ECHO, and NGHS girls) reported **β coefficients** for the associations between **continuous measures of SSB intake** and change in BMI. Extracted data was included in a meta-analysis, to generate a pooled estimate.

7.4.2 Exploratory random effects meta-analyses

The β coefficient (with corresponding s.e., SD, low/high CI or P value) was extracted from four papers (Field et al., 2014), (Laska et al., 2012),(Libuda et al., 2008) and (Zheng et al., 2015) from four cohorts (**GUTS II**, **IDEA & ECHO**, **DONALD** & **EYHS**) from presented models which ***did not adjust for energy*** intake.
Similarly the β coefficient was extracted from four papers (Striegel-Moore et al., 2006) (Laska et al., 2012),(Libuda et al., 2008) and (Zheng et al., 2015) from four cohorts (**NGHS girls**, **IDEA & ECHO**, **DONALD** and **EYHS**) from models which ***did adjust for energy*** intake.

Six exploratory random effects meta-analyses were tried in STATA, using β coefficients for SSB exposure and adiposity outcome measures (Change in BMI, BMIz, waist circumference, skinfold thickness or body fat %).

- Mixed/combined sex not adjusted for energy intake.
- Mixed/combined sex, adjusted for energy intake
- Girls only, not adjusted for energy intake
- Girls only, adjusted for energy intake
- Boys only, not adjusted for energy intake
- Boys only, adjusted for energy intake

Corresponding forest plots, stratified by adiposity measure, are shown on the following pages. Extracted data from each study is represented by a solid square with horizontal arms that show the 95% CI. The relative sizes of the different squares reflect the weightings assigned by STATA to minimise variance. The summary of the pooled data is represented by a diamond; the widest points of the diamond indicate the 95% CI of the summary estimate.

For the mixed/combined sex studies, whether energy adjusted or not, change in BMIz was the most important outcome measure which was allocated the greatest weight (See Figure 7-3 and Figure 7-4). The diamonds which represent the summary effects lie to the right of zero on the scale, which suggests that higher intakes of SSBs might be linked to greater adiposity outcomes. However, in some instances the 95% confidence touch or cross zero, showing that the direction of the effect is uncertain and not significant.

The overall summary effect is indicated by the dotted line and the bottom diamond. As some cohorts are represented more than once (as they had more than one adiposity outcome measure) the overall pooled results are not valid, serving as an informal exploration only. Heterogeneity between mixed/combined sex studies was moderate, as indicated by I^2 values of 27.4% if not adjusted for Energy, and 46.4% if adjusted for Energy. With only four studies in each meta-analysis, these I^2 values should be interpreted with caution.

For girls only, when using data from not energy adjusted models (See Figure 7-5) change in BMIz was the outcome measure given the greatest weight. When using data from energy adjusted models (See Figure 7-6) change in BMI received the greatest weight, strongly influenced by the inclusion of the large girls only NGHS cohort which chose this measure of adiposity.

In the forest plot from not energy adjusted models (Figure 7-5) the overall dotted line of effect lies slightly to the right of zero, which suggests that higher intakes of SSBs might be linked to greater adiposity outcomes in girls, but when β coefficients from energy adjusted models are used (Figure 7-6) the dotted line shifts to zero, indicative of no effect. As before, some cohorts are represented more than once, so these meta-analyses can only be regarded as exploratory. The I^2 statistic was zero in both meta-analyses (very low heterogeneity) but this may be explained by the dominance of one study from only three studies in each meta-analysis. Again, I^2 values should be interpreted with caution.

For boys only, when using data from not energy adjusted models (See Figure 7-7) the overall line of effect lies to the right of zero, which again suggests that higher intakes of SSBs might be linked to greater adiposity outcomes in boys. However, 95% confidence intervals cross zero so this is uncertain and as some cohorts are represented more than once in the plot, the results are exploratory. With an I^2 value of 64%, heterogeneity appears to be substantial, but there are only three studies in the meta-analysis.

Only two of the four studies presented adiposity outcome data for boys from energy adjusted models (See Figure 7-8) which is too few, even for an exploratory meta-analysis.

Figure 7-3 Forest plot: Mixed sex cohorts using data from models not adjusted for Energy

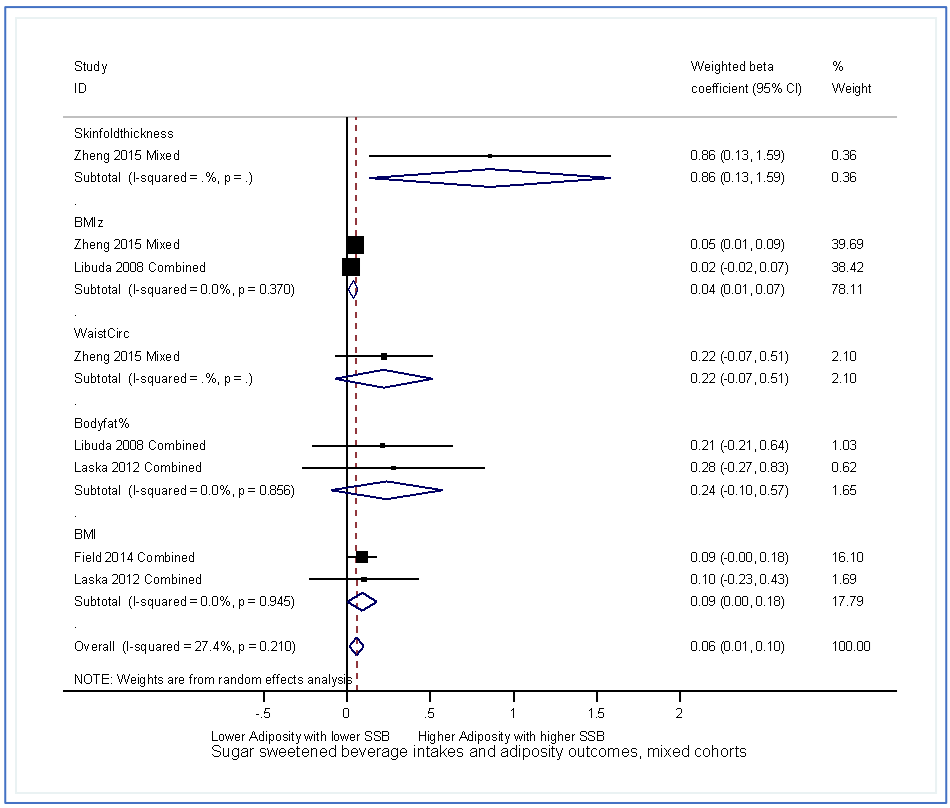


Figure 7-4 Forest plot: Mixed sex cohorts using data from models adjusted for Energy

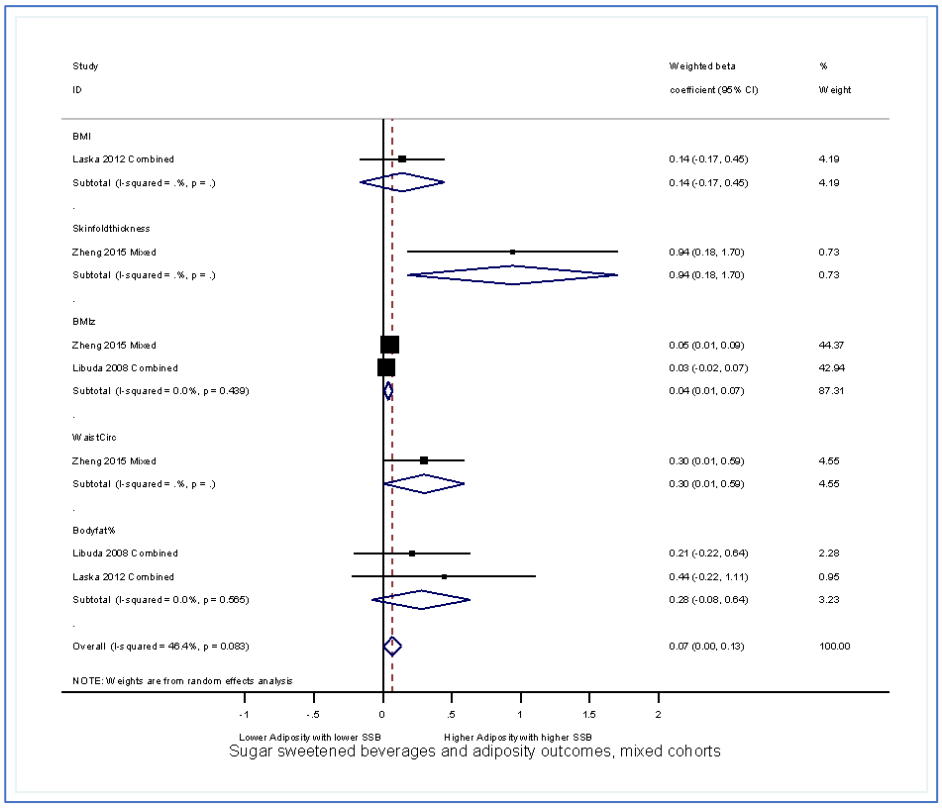


Figure 7-5 Forest plot: Girls only using data from models not adjusted for Energy

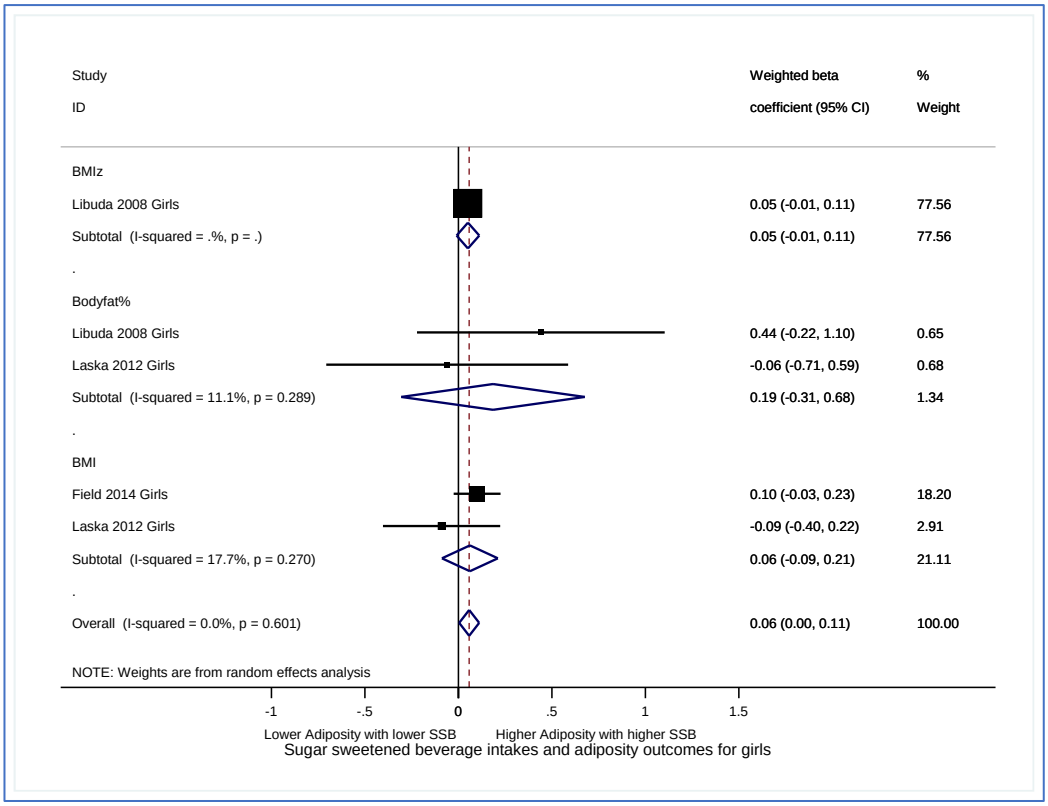


Figure 7-6 Forest plot: Girls only using data from models adjusted for Energy

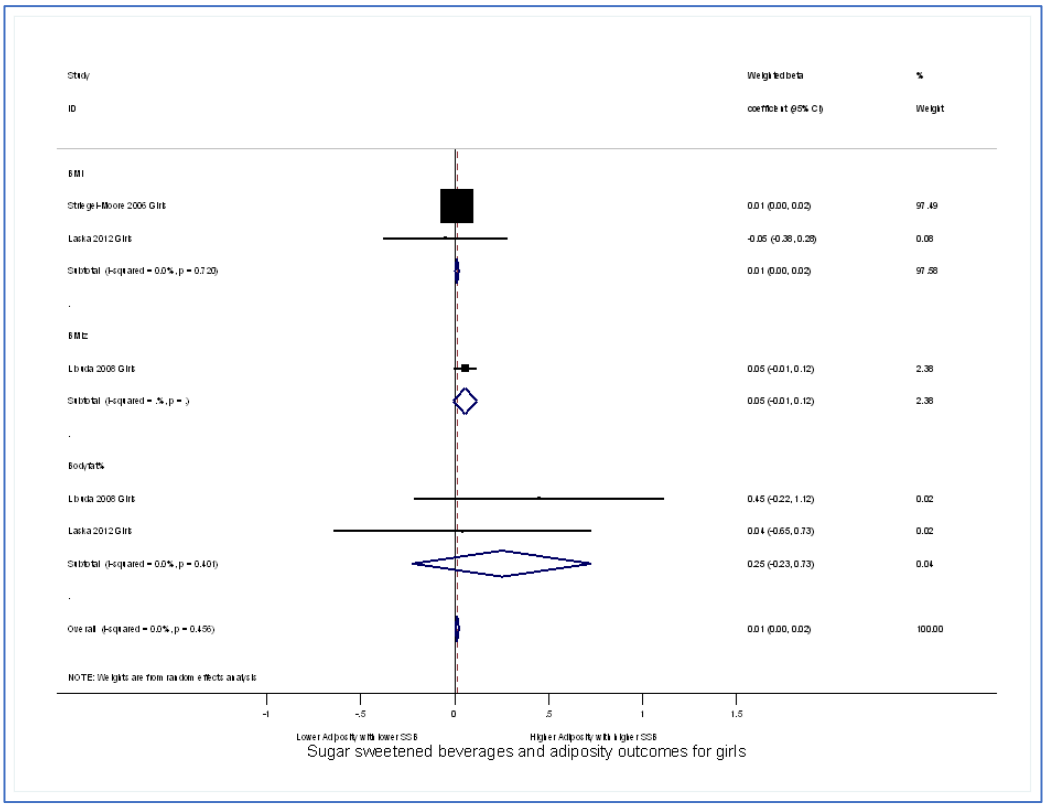


Figure 7-7 Forest plot: Boys only using data from models not adjusted for Energy

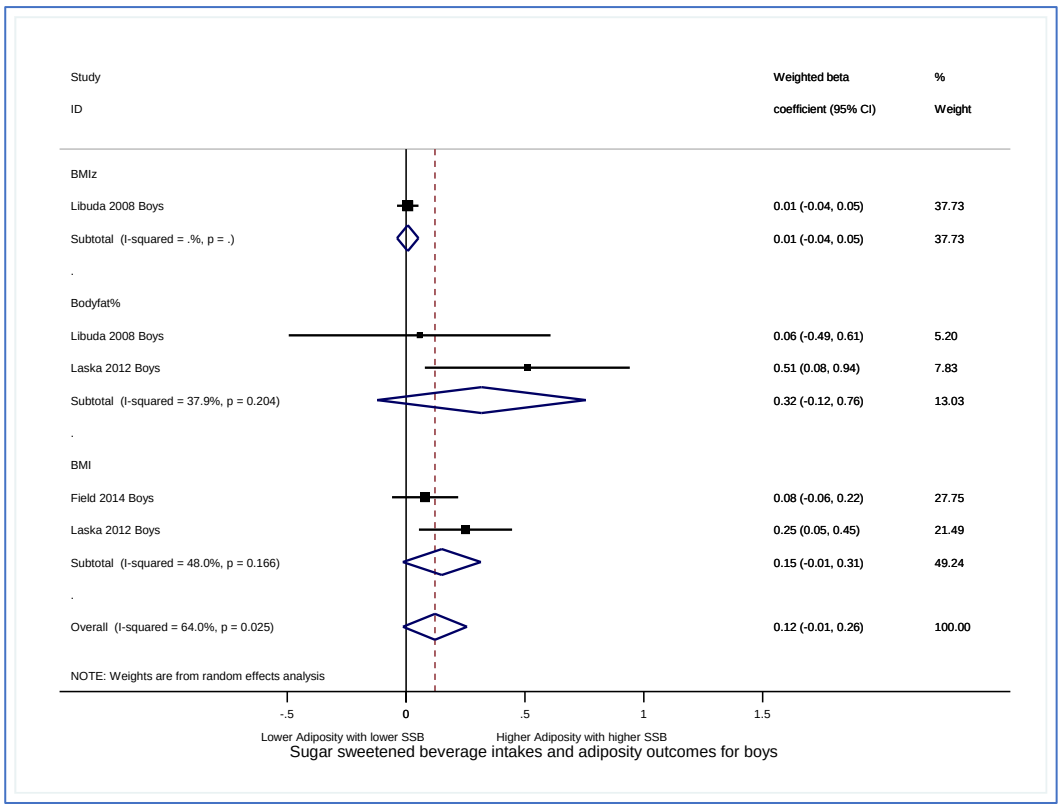
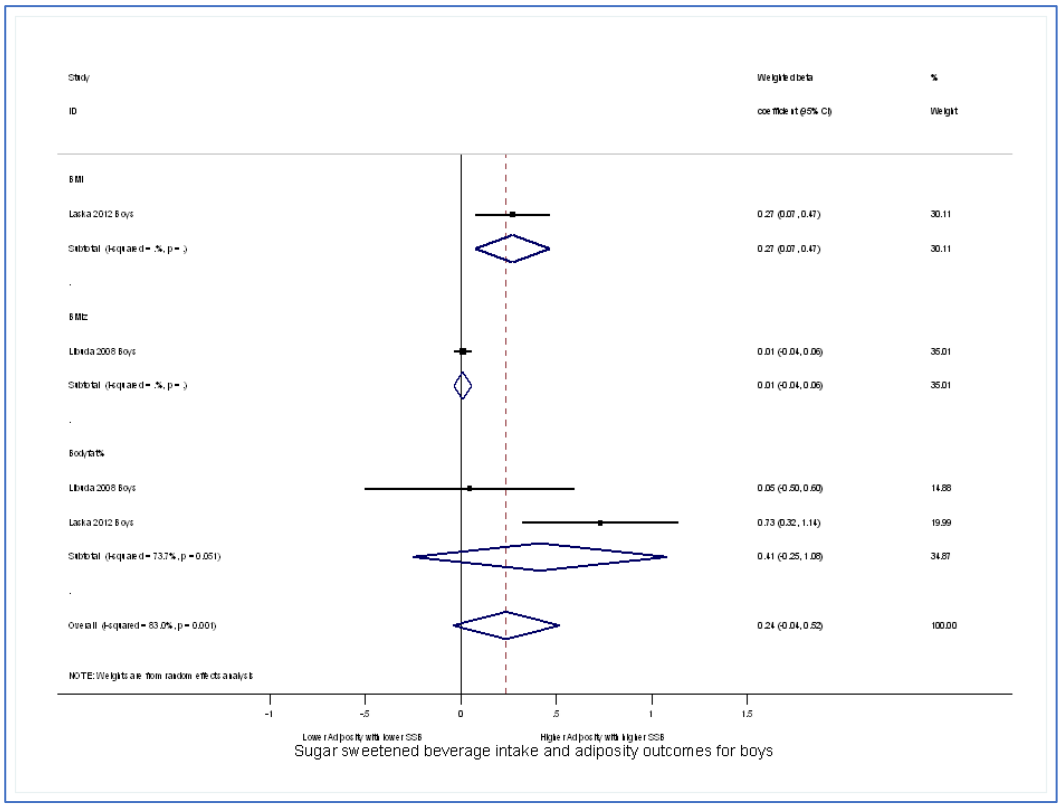


Figure 7-8 Forest plot: Boys only using data from models adjusted for Energy



7.4.3 Opportunities for formal random effects meta-analyses

The methodological heterogeneity of the five cohort studies restricted opportunities for formal meta-analysis of outcomes, as most comparable measures were employed by only two included papers. A breakdown of the evidence from each cohort is summarised in Table 7-2. The five papers which considered associations between SSB exposure and change in continuous adiposity outcomes using β coefficients each took a different approach:

- **NGHS** was girls only, **EYHS** presented results for a mixed sex cohort. **DONALD, GUTS II** and **IDEA & ECHO** presented results for girls and boys separately.
- All five cohorts obtained height and weight information at baseline and follow-up. **DONALD** and **EYHS** reported change in BMI-SDS or BMIz. **NGHS, GUTS II** and **IDEA & ECHO** reported change in BMI.
- **EYHS** also reported change in waist circumference and change in sum of four skinfolds.
- **DONALD** and **IDEA & ECHO** also reported change in body fat percentage (calculated from skinfold thickness or measured by bio impedance).
- **GUTS II** only presented models that were adjusted for energy intake. **NGHS** only presented models that were not adjusted for energy. **DONALD, EYHS** and **IDEA & ECHO** presented results from energy adjusted and not energy adjusted models.

Table 7-2 Evidence for formal meta-analysis of SSB intake and change in adiposity

Participants in cohort	Adiposity measures	Not adjusted for Energy models	No of papers	Adjusted for Energy models	No of papers	Maximum no. of papers giving data
Mixed/ combined	ΔBMI	GUTS II, IDEA & ECHO	2	IDEA & ECHO	1	2 = insufficient data
	ΔBMIz	EYHS, DONALD	2	EYHS, DONALD	2	2 = insufficient data
	ΔWaist circ.	EYHS	1	EYHS	1	1 = insufficient data
	ΔSkinfolds	EYHS	1	EYHS	1	1 = insufficient data
	ΔBody fat %	DONALD, IDEA & ECHO	2	DONALD, IDEA & ECHO	2	2 = insufficient data
Girls only	ΔBMI	GUTS II, IDEA & ECHO	2	NGHS, IDEA & ECHO	2	3
	ΔBMIz	DONALD	1	DONALD,	1	1 = insufficient data
	ΔWaist circ.	-	0	-	0	No data
	ΔSkinfolds	-	0	-	0	No data
	ΔBody fat %	DONALD, IDEA & ECHO	2	DONALD, IDEA & ECHO	2	2 = insufficient data
Boys only	ΔBMI	GUTS II, IDEA & ECHO	2	IDEA & ECHO	1	2 = insufficient data
	ΔBMIz	DONALD	1	DONALD	1	1 = insufficient data
	ΔWaist circ.	-	0	-	0	No data
	ΔSkinfolds	-	0	-	0	No data
	ΔBody fat %	DONALD, IDEA & ECHO	2	DONALD, IDEA & ECHO	2	2 = insufficient data

DONALD: Libuda et al 2008, EYHS: Zheng et al 2015, GUTS II: Field et al 2014(Field et al., 2014)(Field et al., 2014), IDEA & ECHO: Laska et al 2012, NGHS: Striegel-Moore et al 2006

7.4.4 Meta-analysis and forest plots, not rescaled

Three papers from three different cohorts considered SSB exposure (albeit with different definitions of SSBs, different DATs and different serving sizes) and reported change in BMI using β coefficients (Field et al., 2014), (Laska et al., 2012), (Striegel-Moore et al., 2006) There were known differences between the three cohorts:

- **NGHS** girls were aged 9 to 10 years old at baseline. Striegel Moore et al reported annualised change in BMI (based on measured height and weight) for each 100g/day serving of regular soda, with change in intake measured by annual 3 day food records over a ten year period.
- **GUTS II** children were between 9 and 16 years old (girls' mean age 13.0 years, boy's mean age 12.9 years) at baseline. Field et al reported two to three year change in BMI (based on self-reported height and weight) for each 8 oz./day serving of regular soda, with no adjustment for energy intake, with change in intake in the same period measured by food frequency questionnaires. (A serving of 8 American fluid ounces is approximately 240 ml or 250g.)
- Adolescents in the **IDEA & ECHO** cohort were from school grades 6 to 11 (mean age 14.6 years for both sexes). Laska et al reported two year change in BMI (based on measured height and weight) for each serving /day of sugar sweetened beverage including tea and coffee, with the mean intake at baseline and follow-up measured by 24 hour recalls. Serving size was not specified.

There was insufficient data to generate a pooled estimate for boys (as Striegel Moore et al investigated the NGHS girls only cohort) but pooled estimates were generated for change in BMI in girls and for change in BMI in whole cohorts (mixed/combined sex or girls only), using β coefficients from the most adjusted models. The limited data was not rescaled to equivalent SSB serving sizes or study lengths.

Two forest plots of SSB intake and change in BMI were generated in STATA, using data from the most adjusted models, including energy adjustment where available.

- for girls only, using β coefficients. See Figure 7-10.
- for the whole cohort, β coefficients. See Figure 7-11.

An example of a STATA do file, with commands used for the meta-analysis and forest plot of SSB intake and change in BMI in girls only, is shown below, see Figure 7-9.

Figure 7-9 Example STATA do file for meta-analysis of SSB and change in BMI, girls only

```
*/metan betacoefficient allbetacose if sex4combinedrandomma3mixed2girls ==2, randomi /*
*/ label(namevar=DataRowdescription) lcols() effect(Weighted beta coefficient) /*
*/ boxsca(25) xlabel(-0.5,0,0.5)  /*
*/ favours(Lower change in BMI with lower SSB  # /*
*/ Higher change in BMI with higher SSB ) classic texts(300) force /*
*/ graphregion(fcolor(white)) xtitle("Sugar sweetened beverages and change in BMI for girls only, with no scaling for serving size or study length", size(small)) /*
*/ plotregion(fcolor(white)) /*
*/ diamopt (lwidth(vthick) lcolor(red)) /*
```

The meta-analysis of SSB intake and change in BMI in girls had 6,778 participants.
The meta-analysis of SSB intake and change in BMI in the whole cohorts had 10,492 participants.
In both meta-analyses the **NGHS** girls only cohort, a large, high quality study with the longest follow-up period, received the greatest weight. In the girls only meta-analysis the high quality **NGHS** cohort study dominated. Pooling results from **GUTS II** and **the IDEA & ECHO** cohorts made very little difference to the **NGHS** effect size.
Both forest plots suggest that a greater increase in SSB intake is associated with a greater increase in BMI, but this is not significant as the confidence intervals cross zero.
Heterogeneity, shown by I^2, was very low for girls only cohorts (2.6%), in part because of the dominance of the **NGHS** cohort. Heterogeneity between the whole cohorts was moderate (43.3%), which agrees with what we know, but with only 3 studies the I^2 values should be treated with caution.
With so few studies a sensitivity analysis (drop one study approach) was unnecessary.
There were too few studies in the meta-analyses to assess the extent of publication bias by checking funnel plots for asymmetry (Egger et al., 1997)

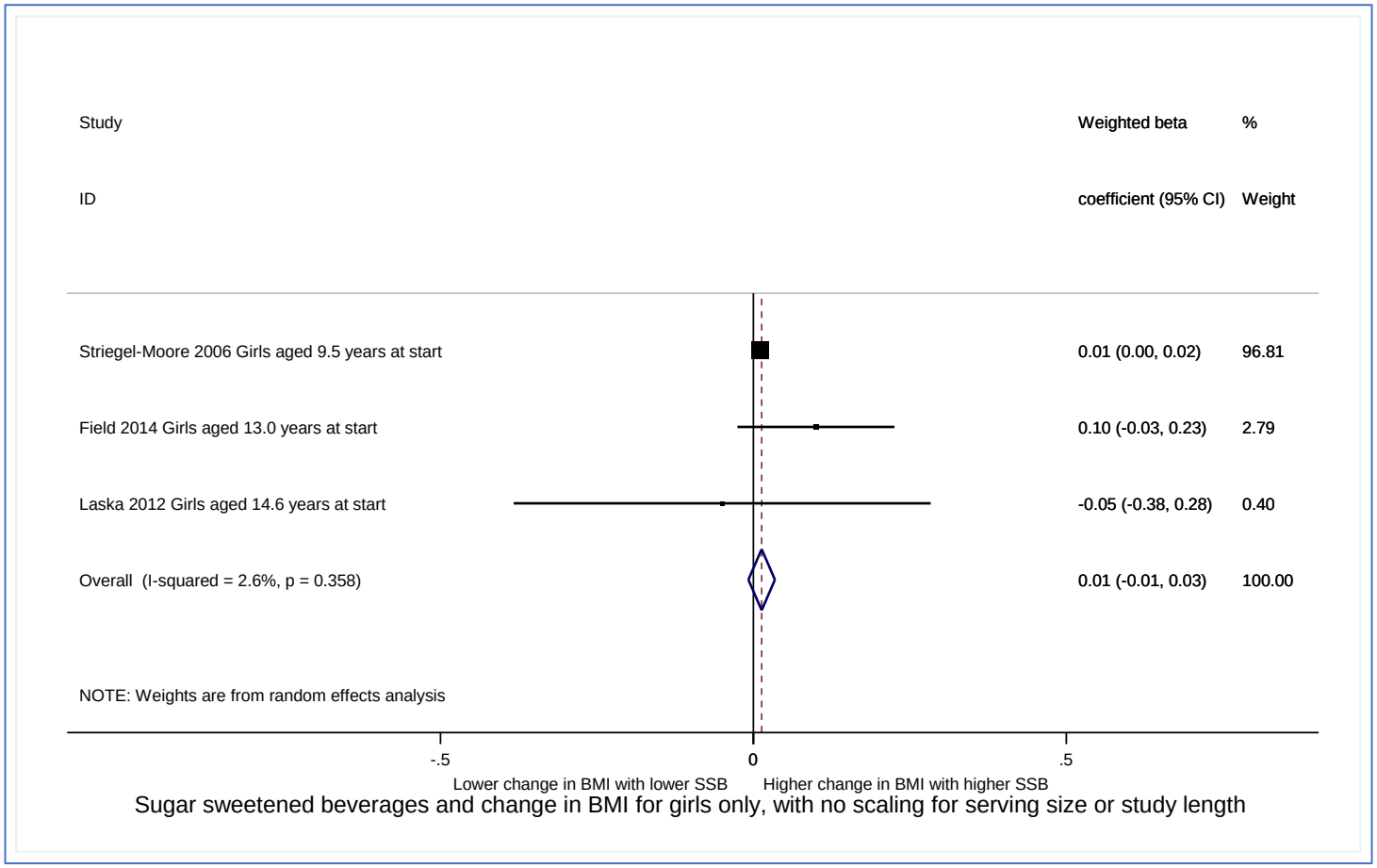


Figure 7-10 Forest plot: SSB intake and change in BMI for GIRLS only, most adjusted models, no re-scaling, n = 6,778

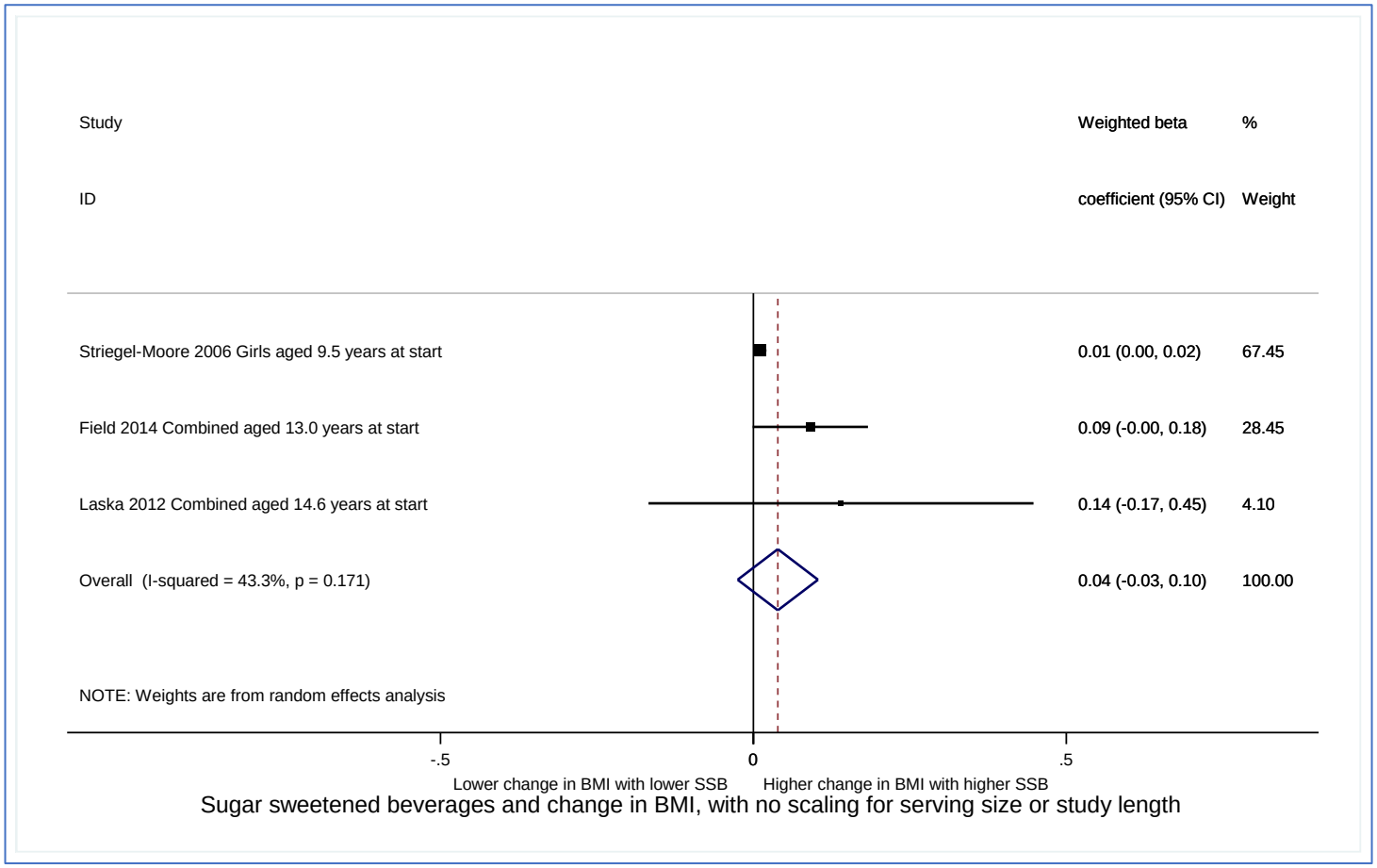


Figure 7-11 Forest plot: SSB intake and change in BMI in 3 whole cohorts, most adjusted models, no re-scaling, n = 10,492

7.5 Discussion

Meta-analysis is a useful way to pool findings from comparable studies and strengthen the evidence. The papers included in the Systematic Review examined a broad range of dietary exposures and adiposity outcomes in children and adolescents, using similar measures but widely varying analytical approaches. As most dietary exposures were examined in only a few cohorts, there was little opportunity for meta-analysis. (If meta-analyses had been the primary objective of this research and thesis, a more focussed literature search strategy or strategies would have been necessary.)

The exception was sugar sweetened beverage intake and adiposity outcomes, investigated in nine of the included cohort studies. The ten papers which analysed SSB intakes and adiposity outcomes were heterogenous, particularly in their choice of analysis. Sometimes it was hard to know if investigators had looked at baseline SSB intake or change in SSB intake vs change in adiposity, as written methods were ambiguous. Results tables were more informative; coding extracted data and checking the reported effect-size measure helped to group like-with-like papers for meta-analysis. However, few studies used the same measures and the number of studies in each comparable group was disappointingly small.

A recent systematic review of SSBs and adiposity outcomes (Malik et al., 2013) included three of the papers (Laska et al., 2012; Libuda et al., 2008; Striegel-Moore et al., 2006) in this systematic review, as well as papers that we had excluded (children too young, follow up < 2 years). Their meta-analyses of 15 cohort studies in children showed a positive association between SSB consumption and change in BMI, with no evidence of publication bias. Malik at al converted SSB serving sizes to a standardised 12oz serving, transformed children's BMIz measures to BMI (using LMS equations and CDC growth charts) and calculated the annualised change in BMI per unit increase in SSB intake. Conversions and "re-scaling" calculations were set out in supplementary material.

Attempts to follow the methods of Malik et al., in the hope of pooling studies in a similar dose-response meta-analysis, were unsuccessful. Some papers gave BMIz outcomes only for a mixed cohort (so not possible to convert to BMI) while others did not state a SSB serving size. Without major assumptions, it was not feasible to do a dose-response meta-analysis with the available data. Change in BMI was the only adiposity outcome reported by the minimum of three studies, and two meta-analyses were run. Although suggestive of higher SSB intakes leading to higher increases in BMI, the pooled summary effects were not significant for either girls only or the whole cohorts. With so few studies and insignificant results, post-hoc techniques such as sensitivity analysis or checks for publication bias were not worthwhile.

Children's BMI is not directly comparable across different age groups, as explained in Chapter 4. As children grow their BMI inevitably changes, but the total extent of BMI change (even if it is annualised) is dependent on the child's age at baseline. The children from the three cohorts included in the meta-analyses were different ages at baseline:

- **NGHS** girls were aged 9 to 10 years old
- **GUTS II** children were between 9 and 16 years old
- Adolescents in the **IDEA & ECHO** cohort were aged between 11/12 and 16/17 years.

It is not best practice to combine BMI into a mean value across different ages within a study. Similarly, change in BMI (unadjusted for age and sex) is not a suitable measure for a pooled estimate when children's baseline ages vary between studies. This is a weakness of the meta-analysis that was attempted. (It follows that outcome measures such as change in waist circumference, skinfold thicknesses or body fat percentage, where no adjustment for age and sex can be made, are also unsuitable for pooled estimates from cohorts with differing baseline ages.) For a meta-analysis of cohorts of children of differing ages, combining either children's BMI z scores, change in BMIz, or dichotomised overweight/obese versus normal weight outcomes (using BMI percentiles and cut-offs stipulated by a growth reference) is more appropriate.

Researchers must decide upon the best adiposity measure to use for children who are still growing, based upon their research objectives. In the literature, BMIz is recommended for assessing a child's adiposity on one occasion (Cole et al., 2005), while BMI is better for measuring change and is more easily understood (Berkey and Colditz, 2006). The latter advice is helpful if comparing adiposity outcomes within a study of children of similar ages.

In the interests of meta-analysis and strengthening the evidence, researchers should consider using and reporting either BMIz scores, change in BMIz or overweight/obesity outcomes for children, alongside any other chosen measure of adiposity outcome. Most cohort studies know the age and sex of each child when height and weight is measured (or self-reported), so calculating a BMI z score or classifying children as overweight/obese versus normal weight using a named growth reference is feasible. For international comparisons the IOTF growth reference offers the greatest utility.

Chapter 8 Tool development: Pre-specification of candidate predictor variables and preparation of the ALSPAC data

8.1 Summary

After identifying prognostic/risk factors in late childhood which increase or decrease the likelihood that a child will develop obesity as a teenager, the aim is to translate them into a short questionnaire and predictive risk algorithm or score, named the Children's Obesity Risk Assessment (CORA).

This chapter (Chapter 8) considers predictive risk algorithms in different settings and the recommended ways to develop, validate and report them, before describing how evidence from the Systematic Review was used to pre-specify childhood predictors of early adolescent obesity. Potential predictors include diet/food exposures and non-diet exposures.

Assumptions about the causal and temporal relationships between childhood diet and obesity outcomes and competing exposures were checked with a directed acyclic graph (DAG), prior to devising an evidence based questionnaire to assess the child's categorical exposure to each potential predictor of future obesity.

20 of 24 potential predictors were matched to candidate variables in a dataset from the Avon Longitudinal Study of Parents and Children, a birth cohort established by the University of Bristol in the early 1990s. Close attention was given to the timing of the measurement of variables to ensure that each candidate variable genuinely preceded the obesity outcome at age 13+ years. Missing observations were imputed where feasible, so that relationships with obesity outcomes could be examined in as much of the cohort as possible. Dichotomous candidate variables were coded in accordance with their expected direction of influence on obesity outcomes.

There were 5,486 eligible respondents (singleton children who completed 3 day diet diaries and did not have obesity at baseline, aged 10+ years). They were randomly split 3:1 into a derivation sample (n = 4,114) and a validation sample (n = 1,372) in preparation for model development and internal validation, presented in Chapter 9.

8.2 Predictive risk tools – an introduction

8.2.1 Risk algorithms in clinical and population settings

Risk algorithms can be used to predict an outcome or make a prognosis. In clinical medicine prognostic models or predictive risk tools are routinely used to estimate an individual's risk and guide decision-making about how to prevent or treat their disease. Clinical risk algorithms are helpful for clinicians when there are several factors that contribute to the risk, the range of risk is wide and baseline risk influences the decision to take action. Such tools commonly rely on clinical measures. For example, QRISK2 is a multivariable risk algorithm that estimates an individual's 10 year risk of cardiovascular disease using baseline blood cholesterol ratios and systolic blood pressure as prognostic/risk factors as well as age, smoking status, ethnicity, BMI, family history of coronary heart disease, deprivation score, treated hypertension and diagnosed rheumatoid arthritis, atrial fibrillation, Type 2 diabetes or chronic renal disease (Collins and Altman, 2012).

Public health professionals have recognised that multivariable risk algorithms could also be used to generate estimates of disease risk in populations as well as in individuals. Such estimates could inform decision making in population settings, having applications for resource allocation and for assessing the impact and equity of community-wide prevention strategies (Manuel et al., 2012) Population-based risk algorithms are developed and assessed in the same way as clinical risk algorithms, but use routinely collected or easily obtained non-clinical measures of baseline exposure. A recent example is CVDPoRT, the Cardiovascular Disease Population Risk Tool, modelled with data collected from 104,000+ adults in Ontario who were respondents to Canadian Community Health Surveys between 2001 and 2007, followed up from 2001 to 2012. CVDPoRT is designed to estimate 5 year incidence of a cardiovascular disease event in a community setting, without the help of clinical measures. Predictor variables initially considered for CVDPoRT were age and sex, socio-demographics, general health and chronic conditions and health behaviours including smoking, alcohol and leisure physical activity plus average daily consumption of fruits and vegetables, potatoes and juice. (Manuel. et al., 2018)

8.2.2 Developing risk algorithms

Numerous methodological papers have been written about how to develop and test clinical prognostic models. The British Medical Journal published a series of four papers on prognosis and prognostic research in 2009 (Moons et al., 2009a; Royston et al., 2009; Altman et al., 2009; Moons et al., 2009b) with a further series in 2013 from the Prognosis Research Strategy

(PROGRESS) group (Hemingway et al., 2013; Hingorani et al., 2013; Riley et al., 2013; Steyerberg et al., 2013). The PROGRESS group later issued a paper on improving the transparency of prognosis research (Peat et al., 2014). The group made a case for writing (and registering) a protocol or research plan before data acquisition or analysis, even when the goal is to find new prognostic factors through exploratory or data-driven analyses.

A systematic review about reporting and methods in clinical prediction research (Bouwmeester et al., 2012) found that many studies had an unclear study design or did not follow methodological recommendations. Similarly, a systematic review about methods and reporting of the external validation of prediction models (Collins et al., 2014), found that most included articles were unclear or poorly reported. The authors concluded that it was unsurprising that "the majority of developed prediction models are not used in practice". To address these short-comings and improve reporting, so that prediction models can be adequately assessed, methodologists have issued the TRIPOD statement (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.) (Collins et al., 2015)

Much of the guidance about methodology and reporting of clinical prognostic models can also be applied to population based prediction models.

There are 3 main stages to fully developing any risk algorithm:

- Model development
- Validation
- Assessment of impact.

8.2.3 Model development

Risk algorithms are usually developed using multivariable regression, ideally with prospective cohort data (Moons et al., 2012). Logistic regression is a suitable technique for short-term dichotomous or binary (yes or no) outcomes. The starting point for the model is an individual with a certain state of health (e.g. Not obese) and a combination of predictor values, while the endpoint is an estimate of their risk of a specific outcome (e.g. Obese) within a specific time period (e.g. 3 years). Risk can be expressed as absolute risk, relative risk or as a risk score

The goal is to create a discriminating and well-calibrated algorithm that is easy-to-use and avoids over fitting, which happens when a model has too many parameters for the information in the data. It can be tempting to over fit a model when choosing from many candidate predictors in a dataset of restricted size, but statistical overfitting increases Type 1 error (more false positive findings), thus overstating the model's predictive ability. As a consequence, a

complex model with many variables and interactions may perform less well in different settings. Care must also be taken with missing data, checking model assumptions and modelling continuous predictor variables. Categorisation of continuous variables is practical in use, but inevitably loses information and reduces power. (Steyerberg et al., 2013)

8.2.4 Validation

The performance of a predictive model based on a development dataset is often "optimistic". Before putting a model into practice, it must be tested with data not used in the development process, to ensure that the model still gives valid predictions. Algorithms tend to be more reliable if they are built in a large, high-quality dataset, follow a study protocol with a pre-specified analysis plan, such as the protocol used to develop CVDPoRT (Taljaard et al., 2014) and are validated in independent datasets. (Steyerberg et al., 2013)

Internal validation.

As a minimum, the model should be evaluated using data from the same population as the development dataset. Internal validation gauges the model's reproducibility by assessing the quality of the predictions and the stability of the selection of predictors. The simplest approach to internal validation is to use a split sample, allocating a random portion of the dataset for model development, known as the "training" or "derivation" dataset, and reserving another portion, the "test" or "validation" dataset, for internal validation. However, a split sample can be an inefficient use of the data. In small datasets where all the information is needed for model building, resampling and repeated training methods are recommended (Kuhn and Johnson, 2016). Examples of resampling methods include *k*-fold cross-validation and boot strapping.

In *k* fold cross-validation, the dataset is spilt into *k* subsets of roughly equal size, and the model is fitted with all but one subset which is held out as the test dataset. This process is repeated *k* times with each subset held out in turn.

For bootstrapping, random samples are taken from the dataset with replacement, until the sample size is as large as the original dataset. Some samples are selected several times, while others are not selected at all. Bootstrap samples are randomly drawn multiple times (200+) with selected samples used to build and subsequently recalibrate an original model, and corresponding unselected or held out samples used to test the predictions (Steur et al., 2011). Bootstrapping can create an overly optimistic model, with very low error rates (Kuhn and Johnson, 2016).

External validation.

Ideally, the model should be evaluated using data from a different setting (different time

and/or different location). External validation is a more exacting test which gauges the model's transportability, (Steyerberg and Vergouwe, 2014) but is beyond the reach of this study. External validation should preferably be conducted by independent investigators, rather than by the researchers who developed the model, to avoid any potential conflict of interest or investigator bias.

If the distribution of measurements for predictors and outcomes in the different setting lie outside the ranges used in model development, the model's performance may be compromised. It is therefore helpful to report the ranges and categories for predictors in the model, and to whom the model is applicable, so that changes in the model's performance in the different setting/population can be understood and if needed the predictive model can be adjusted. (Collins et al., 2014).

A model's predictive ability in each setting (training dataset and internal or external validation dataset) is judged by the model's calibration and discrimination.

Calibration is the level of agreement between the predicted and observed outcomes, which can be assessed by a goodness-of-fit (g-o-f) test and by plotting predictions on the x-axis and observed outcomes on the y-axis. Perfect predictions lie on a line with an intercept of 0 and a slope of 1. When assessing calibration graphically the intercept shows the systematic tendency of predictions to be too high or too low, while the plotted line shows if predictions are exaggerated, with low predictions too low and high predictions too high.

Discrimination is the ability of a model or tool to differentiate between individuals with and without the outcome of interest. It can be quantified with the concordance (c) statistic, which for binary outcomes is equal to the area under the receiver operating characteristic (AUROC) curve. This curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) for consecutive cut-offs for the predicted risk. The area under the curve represents the probability that, for two randomly chosen individuals, one with the condition and one without, the predictive tool will assign a higher likelihood of having the condition to the individual who does have that condition. An AUROC equal to 0.5 is no better than chance. An AUROC of 0.7 and above is considered "acceptable" (Hosmer, 2013) and may be "clinically useful".

Overall measures of discrimination and calibration such as a **Brier score** can also be employed (Bouwmeester et al., 2012). The Brier score is an aggregate measure of disagreement between the observed outcome and the prediction – the average squared error difference. Brier scores lie between 0 for a perfect forecast and 1 for a completely wrong forecast. The score for a non-informative model, where prediction is no better than chance, would be $(0.5)^2$ or 0.25. Smaller scores are desirable.

There is always a trade-off between calibration and discrimination. As one improves, the other gets worse. For clinical decision making, distinguishing between high and low risk individuals is vital, so clinical risk algorithms are tailored to be as discriminating as possible. In population settings, a better calibrated tool may be preferred for more accurate prediction of the number of future cases and the allocation of limited resources. However, discrimination is still important, as results can influence whether interventions are targeted narrowly (concentrated risk –a small section of the population carries most of the risk) or applied community-wide (diffuse risk – most individuals in the population have a similar baseline risk). A lack of discrimination may add to inequities, with predicted risk being over estimated in advantaged social groups and under estimated in disadvantaged ones (Manuel et al., 2012).

As part of their statistical tutorial paper, "Towards better clinical prediction models", Steyerberg and Vergouwe, recommended that clinical usefulness be assessed as well (Steyerberg and Vergouwe, 2014). They introduced an ABCD framework for validation; Alpha or intercept and Beta or gradient assessment of calibration, the C-statistic measure of discrimination and Decision curve analysis of clinical usefulness.)

Clinical usefulness is assessed using decision curve analysis over a range of decision thresholds, to judge whether better informed decisions can be made with the model rather than without it.

There are four possible results when model predictions are compared with observed outcomes – true and false positives or true and false negatives, which can be presented in a classification table. (See Table 8-1).

Sensitivity (true positive success rate) and **specificity** (true negative success rate), the **Positive predictive value** (PPV) and the **Negative predictive value** (NPV) calculated from the classification tables, should be reported for the model run in the training dataset and in the test or validation dataset.

Table 8-1 Example classification table for a logistic regression model

	Actual or observed outcome		
Predicted outcome	Obese	Not obese	Total
Obese	A True positive	B False positive	A + B
Not obese	C False negative	D True negative	C + D
Total	A + C	B + D	

Sensitivity or success rate is the ability of the tool to correctly predict a *true positive* outcome
Sensitivity = A / (A + C)

Specificity is the ability of the tool to correctly predict a *true negative* outcome.
Specificity = D / (B + D)

The positive predictive value is the probability that someone predicted to have the outcome really does have the outcome.
PPV = A / (A + B)

The negative predictive value is the probability that someone predicted *not* to have the outcome really does *not* have the outcome.
NPV = D / (C + D)

The aim is to find the decision threshold or probability cut-off which optimises sensitivity and specificity, while not misclassifying too many cases. (Even if the model is well fitted, there is a tendency for more of the predicted outcomes to be classified to the larger group of the binary options, which in this case would be "Not obese".) The optimal balance is found by plotting sensitivity and specificity against probability cut-offs, to see where they cross.

A model's predictive performance may diminish over time due to underlying changes in the population or the model may need to be recalibrated for use in a different setting or additional predictive variables may be found. Existing models can be updated, recalibrated, or combined, rather than creating a new model each time (Steyerberg et al., 2013). Systematic Reviews show that there has already been duplication of effort in developing predictive models for the same kind of outcomes, albeit in different populations and settings where a different model may be required. (Collins et al., 2014, Ziauddeen et al., 2018).

8.2.5 Assessment of impact

There is a cost to implementing a new predictive model or tool. Wide scale implementation will only be worthwhile if predictions lead to decisions that improve outcomes or make treatment/interventions more cost-effective. There is also the possibility that treatment/interventions may mistakenly be withheld if a model wrongly identifies an individual/population as at low risk. A comparative study, looking at decision making and outcomes with and without the predictive tool, is needed to provide clear evidence of a positive or a negative impact. Perhaps due to cost, impact studies are few and evidence is surprisingly scarce, even for tools in clinical use (Steyerberg et al., 2013).

8.3 Pre-specification methods and analysis

8.3.1 Data request and ethics

The CORA tool was derived and internally validated using secondary data from the ALSPAC (Avon Longitudinal Study of Parents and Children) birth cohort, which was provided by the University of Bristol. In the Systematic Review ALSPAC achieved the maximum 8 stars out of 9 in the customised Newcastle-Ottawa quality assessment scale for cohort studies, so was identified as a high quality cohort. Principal investigator Dr Charlotte Evans and 3 research postgraduates from the University of Leeds, including CR, submitted an online research proposal in November 2016. The proposal B2798 "Adolescent diet and cardio-metabolic health" was approved by the ALSPAC Executive Committee.

With guidance from a data buddy, researchers explored ALSPAC documentation, the data dictionary and variable catalogues available from http://www.bristol.ac.uk/alspac/researchers/data-access/data-dictionary/.
Researchers identified potential variables for their individual projects and submitted a joint data request, using exact variable names and labels, in May 2017.

Subject to terms set out in the ALSPAC access policy version 7.0 (payment, signed confidentiality agreements, secure data transfer and restricted access permissions) the final version of the requested, anonymised data in STATA was received in October 2017. Researchers wrote a data management plan to help on-going compliance with the ALSPAC access policy.

As the proposed study used secondary data, separate ethical approval was not required. The ALSPAC Law and Ethics Committee and local research ethics committees approved the original studies that generated the data. Parents provided informed consent at the time of their recruitment and for subsequent measures.

8.3.2 ALSPAC study design and participants

The ALSPAC cohort, known as "Children of the 90s", is an ongoing, observational birth cohort, established 1990 – 1992. Pregnant women were recruited from Bristol and areas around the River Avon, with later additional recruitment of those who fitted the original eligibility criteria. There were 20,248 potentially "eligible" pregnancies from which 14,541 mothers were recruited, with 706 recruited later. There were 14,775 live born children from these pregnancies.

At recruitment mothers in the ALSPAC cohort were broadly representative of the UK population, although ethnic minority groups were underrepresented. Compared to the 1991 census mothers in ALSPAC more likely than mothers elsewhere in Avon and Great Britain to be married and more likely to have a car and live in owner-occupied rather than rented homes, reflecting the demographic profile of the catchment area (Fraser et al., 2013) Mothers who did not respond to later surveys were generally younger, less likely to be degree educated and more likely to have a lower SEC background and to have had two or more children already. Ongoing attrition has caused more affluent groups to be over-represented. Slightly more boys than girls have been lost and children lost have tended to have lower SES and slightly higher rates of parental obesity than the children who stayed in the study (Hughes et al., 2011).

During adolescence 12,776 individuals were still enrolled, of whom 75% responded to at least one survey. Children were surveyed throughout infancy, childhood, adolescence and into young adulthood by measures at clinics and by questionnaires completed by the mother or answered by the child directly. Children's diet was quantified using 3 day food diaries at age 10+ years and 13+ years, enlisting parental help if needed. Short food frequency questionnaires and other questionnaires were also employed between these ages.

Height (cm) and weight (kg) were measured by trained personnel during the Focus 10+ clinic (age approximately 10 years 6 months), the Teen Focus 2 clinic (age approximately 13 years 6 months) and the Teen Focus 3 clinic (age approximately 15 years 6 months). Measured height and weight were used to calculate Body Mass Index (kg/m^2) at each time point. Most children who attended clinics had these measures, from which obesity status could be derived.

8.3.3 Outcome

The outcome of interest is adolescent obesity at age 13+ years, based upon a BMIz score ≥ 1.64 (BMI at or above the 95th centile) using UK 1990 age and sex specific growth references (Cole et al., 1995). In another paper Cole et al concluded that, in growing children, "BMI z-score is optimal for assessing adiposity on a single occasion", whereas BMI or BMI% are better for measuring change (Cole et al., 2005). The UK 1990 reference was shown to have a moderately high sensitivity (88%) and high specificity (94%) for obesity in the ALSPAC cohort, albeit at a younger age of 7 years. In the same cohort, also at age 7 years, the International Obesity Task Force definition for obesity had a lower sensitivity, which differed significantly between girls (72%) and boys (46%) (Reilly et al., 2000)

The UK 1990 reference is also used by the National Child Measurement Programme in England (NHSDigital, 2019) to define BMI classifications based on centiles as follows:

- BMI centile ≤ 2 Underweight
- BMI centile > 2 and < 85 Healthy weight
- BMI centile ≥ 85 and < 95 Overweight
- BMI centile ≥ 95 Obese
- BMI centile ≥ 99.6 Severely obese (a subset of obese)

BMI was calculated as weight/ (height)^2. BMI was converted to age and sex adjusted BMIz scores using the STATA command "zanthro", found via STATA help (Vidmar et al., 2013).

8.3.4 Eligibility criteria

Eligible respondents included singleton children who had completed a 3 day diet diary at age 10+ years in enough detail to calculate their total energy intake (TEI) and who did not have obesity at baseline (Focus 10+ clinic). All twins were excluded, as twin pregnancies are more likely to result in premature births and/or lower birthweights which may influence the child's subsequent growth trajectory. Singleton children who had obesity at baseline were also excluded.

8.3.5 Sample sizes

The ALSPAC dataset provided by the University of Bristol contains 15,445 observations and 687 variables, with 14,701 children who were alive at one year of age. Children in the cohort were invited to the Focus at 10+ clinic and 7,557 attended, of whom 7,462 completed 3 day diet diaries and had energy intake. Age and gender specific BMIz scores were derived for 7,461 of those children and 14% were categorised as obese at baseline. In total 7,522 children who attended the Focus 10+ clinic also came to the Teen Focus 2 and Teen Focus 3 clinics. Attendance at clinics is summarised in Figure 8-1.

Age and gender specific BMIz scores were derived for 6,116 children who attended Teen Focus 2 and 5,411 children who attended Teen Focus 3. Obesity prevalence at both follow-ups was approximately 13%, although they were not necessarily the same children each time.

Characteristics of children in the whole cohort and those who did not attend the Focus at 10+ clinic versus those who did are presented in Table 8-2.

Just over half of the children in the ALSPAC cohort were boys and 95% of children had a white ethnic background (i.e. both parents were white). Approximately 13% of mothers in the cohort reported that they were educated to degree level at the time of their pregnancy, a higher level than the general population, and almost half of mothers reported that they had never smoked. Adult smoking prevalence in the UK was approximately 30% in 1990. However, observations were missing for as much as 20% of the cohort.

Compared to the whole cohort, boys were slightly underrepresented in the Focus at 10+ clinic. Children's mothers were more likely to be educated to degree level and to have never smoked. Levels of missing data were noticeably lower among children who took part in the Focus 10+ clinic.

Figure 8-1 Venn diagram of 15,445 children in the ALSPAC cohort, with numbers attending F10+, TF2 and TF3 child clinics.

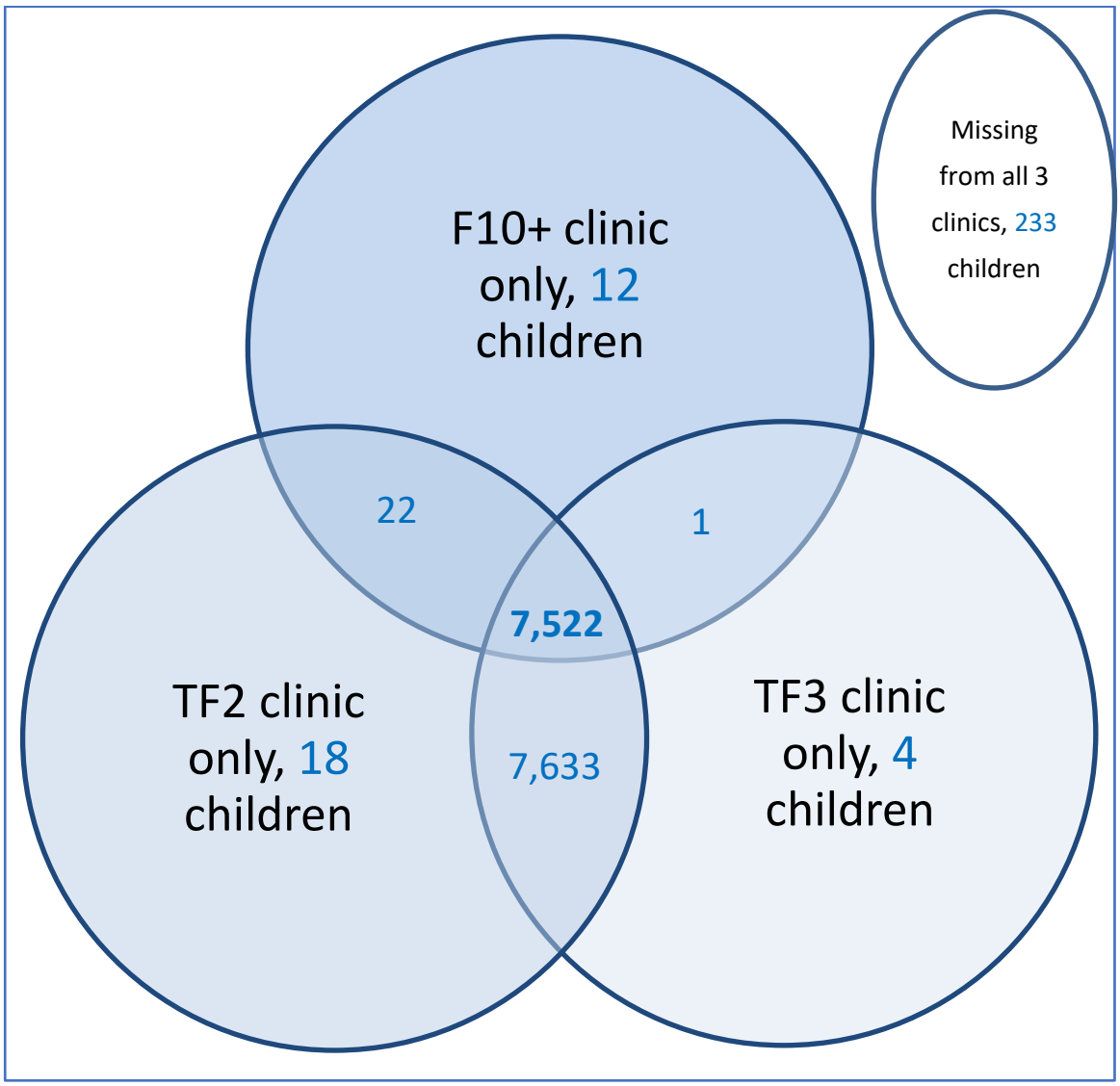


Table 8-2 Characteristics of the ALSPAC cohort

	Whole cohort		Not in Focus at 10+		In Focus at 10+	
	No. obs.		**No. obs.**		**No. obs.**	
Sample size, n	15,445		7,888		7,557	
Boys	7,635	51.4%	3,905	53.5%	3,730	49.4%
Girls	7,219	48.6%	3,398	46.5%	3,821	50.6%
Sex missing	*591*		*585*		*6*	
Age at Focus 10+ in days Mean (SD)	n/a		n/a		7,551	3,902 (96)
Age at Focus 10+ missing	*7,894*		*7,888*		*6*	
White Ethnic background	11,537	95.0%	5,013	93.6%	6,524	96.0%
Non-white Ethnic background	613	5.0%	343	6.4%	270	4.0%
Ethnic background missing	*3,295*		*2,532*		*763*	
Mothers education, degree	1,609	12.9%	515	9.2%	1,094	15.9%
Mothers education, other	10,884	87.1%	5,082	90.8%	5,802	84.1%
Mothers education missing	*2,952*		*2,291*		*661*	
Mother ever smoked, yes	6,739	50.9%	3,643	58.2%	3,096	44.3%
Mother ever smoked, no	6,510	49.1%	2,614	41.8%	3,896	55.7%
Mother ever smoked, missing	*2,196*		*1,631*		*565*	

8.3.6 Analysis plan

The analysis plan was developed and set out in a protocol after access to the ALSPAC dataset was granted. The protocol was not registered. In line with guidance from Steyerberg (Steyerberg, E.W., 2009) and from Harrell (Harrell, 2001), all predictors were pre-specified and matched to variables before running descriptive analyses of associations between predictors and outcomes. Data preparation, analyses, model fitting and testing were carried out using STATA IC 15 (StataCorp, 2017).

8.3.7 Potential predictors of future obesity

Evidence from the Systematic review of childhood and adolescent cohorts measuring whole diet and subsequent adiposity helped to identify potential predictors of future obesity and establish their direction of influence. All evidence from the systematic review, whether baseline exposure or change in exposure, with an adiposity outcome at follow-up or change in adiposity, was useful, regardless of how adiposity was measured or whether the analysis adjusted for energy intake or not.

Identified potential predictors included:

- Dietary exposures significantly associated with future adiposity, plus confounders of such associations.
- Dietary and non-dietary factors shown to be strong predictors of BMI change, overweight or obesity.

Some evidence in the Systematic Review came directly from research conducted in ALSPAC, but there was supporting evidence from other cohort studies. The main findings are recapped below and summarised in Table 8-3.

Higher baseline frequency/amounts or increased frequency/amounts of whole grains and dairy foods including milk (Bigornia et al., 2014; Berz et al., 2011) were associated with reduced adiposity risk. Higher milk intakes in the youngest children from the Framingham Children's study were also linked to reduced body fat outcomes (Hasnain et al., 2014), but were not significant among 10 year olds from ALSPAC (Noel et al., 2011) or in girls from the NGHS (Striegel-Moore et al., 2006)

There were some indications that higher vegetable intakes had a protective effect against future overweight, at least for boys in GUTS (Field et al., 2003b). "Vegetable" dietary patterns (Cutler et al., 2012) or a higher DASH adherence score (based on fruit, vegetables, low fat dairy, total grains, whole grains, lean meats and nuts, seeds and legumes) (Berz et al., 2011) were also helpful.

Table 8-3 Potential predictors of future obesity

	Potential predictor of obesity	Obesity more likely if:
Energy	Total Energy Intake	Higher TEI
Food intake	Whole grains	Lower intake
	Dairy foods (separate from milk)	Lower intake?
	Vegetables	Lower intake?
	Fruit, without fruit juice	Lower intake?
	Energy dense foods and snacks	Higher intake?
Drinks intake or frequency	Milk	Lower intake or frequency?
	Pure (100%) fruit juice	Higher intake or frequency?
	Sugar sweetened drinks	Higher intake or frequency
	Diet or unsweetened drinks	Higher intake or frequency?
Eating habits and dietary patterns	Breakfast frequency	Lower breakfast frequency
	Other meal frequency	Girls: lower meal frequency Boys: higher meal frequency
	Family meals	No family meals
	Eating between meals/snacking	Higher frequency?
	Fast food or take away meals	Higher frequency
	Dieting (or fussy eating)	Higher frequency
	Dietary patterns	Energy dense, snacking, fruit?
Family factors	Socio economic status	Lower SEC, family income or education level of parent
	Parental overweight (perceived or actual)	Mother/parent is overweight
	Smoking	Smoking in household
Child factors	Age	Older
	Sex	Female
	Ethnicity	Non-white
	Puberty status	Early puberty
	Body dissatisfaction	Dissatisfied with body image
	Physical activity level	Lower PAL
	Sleep duration	Shorter sleep duration
	Number of children in family	Only child
	Baseline weight status	Already overweight/obese

Higher baseline frequency/amounts or increased frequency/amounts of sugar sweetened beverages and possibly energy dense convenience food (Alexy et al., 2011) were associated with increased adiposity risk. Diet drinks were associated with unfavourable adiposity outcomes in GUTS II (Field et al., 2014), but were not significantly associated with adiposity outcomes in other cohorts. "Energy dense" dietary patterns were linked with unfavourable adiposity outcomes in ALSPAC (Ambrosini et al., 2012) as were dietary patterns with high fat content among NGHS girls (Ritchie et al., 2007).

There was conflicting evidence for intakes of fruit and fruit juice. Possibly some fruit/juice helps diet quality, but too much adds surplus energy in the form of fructose. A "Fruit" dietary pattern seemed to increase the risk of adverse weight outcomes for younger boys in Project EAT but this was no longer significant after adjustment for baseline weight status (Cutler et al., 2012).

Reduced fat snack foods were helpful in boys but "Eats snack food" was a significant predictor of obesity onset among NGHS girls (Rehkopf et al., 2011) and high adherence to a "Snacking" dietary pattern increased the risk of adverse weight outcomes among Bogotá School Children (Shroff et al., 2014)

Helpful eating habits that reduced adiposity risk were family meals (Berge et al., 2015)and eating breakfast (Affenito et al., 2005), yet "Family eats dinner together" and "Eats breakfast" were not significant predictors of obesity onset among young NGHS girls (Rehkopf et al., 2011)

Eating habits that increased adiposity risk or predicted overweight included eating fried food away from home (Taveras et al., 2005) and reduced eating frequency in girls (Franko et al., 2008). Dieting was also associated with greater BMI outcomes, with unhealthy weight control behaviours and dieting predicting overweight (Quick et al., 2013; Field et al., 2003a).

There was conflicting evidence for meal frequency – higher dinner frequency or 3+ meals a day was helpful for some girls, but increased lunch and dinner frequency was associated with adverse weight outcomes for boys (Quick et al., 2013). Possibly this was due to the quantity consumed rather than frequency.

Body satisfaction predicted lower adiposity outcomes, whereas body dissatisfaction and a perception that biological parents were overweight predicted higher adiposity (Quick et al., 2013). Other non-dietary factors that were significant predictors of BMI change (positive or negative) included household income, education level of parent or primary caregiver, number of siblings, race and parent's BMI (Rehkopf et al., 2011)

Many papers adjusted for energy intake, child's age, sex, ethnicity, pubertal status and physical activity levels in their regression models. Other common confounders of the relationship

between child's diet and adiposity outcomes were sedentary activity, family socio-economic status (socio-economic class, household income or parent's/mother's education level), parental overweight and smoking. Several papers adjusted for baseline weight status, dietary pattern scores or dietary misreporting.

8.3.8 Assumptions and directed acyclic graph

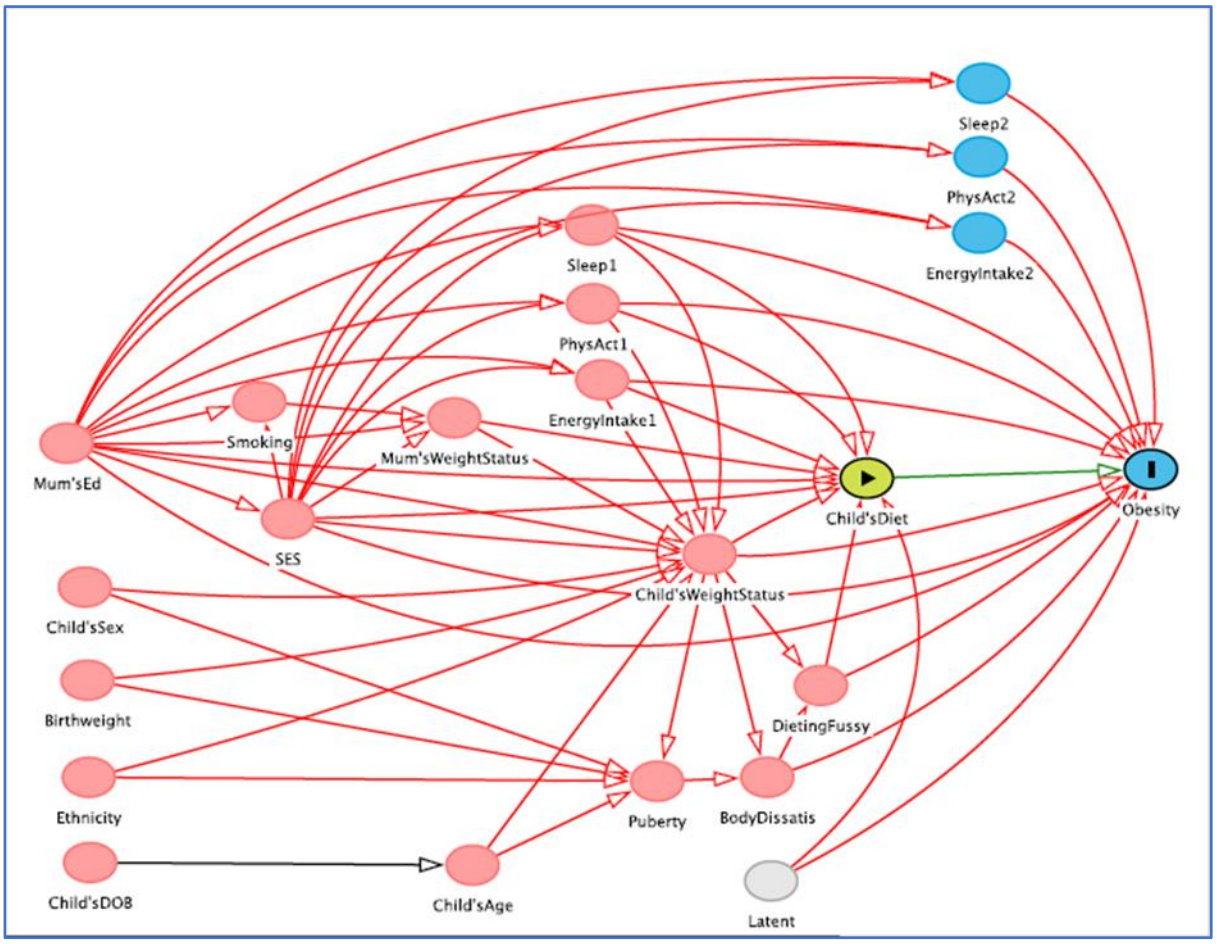
Regression models are useful tools for estimating the association between an exposure and an outcome, but the *causal* inferences that can be drawn from a model depend upon the experimental design, lack of measurement error, and the completeness of the set of variables believed to measure confounding (Harrell, 2001). Causal effects can be considered with a non-parametric causal diagram, which connects the different variables with unidirectional arrows (Arnold et al., 2020). Such conceptual diagrams are called directed acyclic graphs (DAGs).

The causal relationship between dietary exposures and obesity outcomes in children/ adolescents was explored using a DAG (See Figure 8-2), drawn using the online tool DAGitty (http://www.dagitty.net) (Textor et al., 2017). The link with the outcome of interest need not be directly causal for a potential predictor to have a strong predictive performance; often the addition of non-causal predictors may strengthen the accuracy and precision of a predictive model in a specific setting. (Moons et al., 2012). However, a "cause" which genuinely precedes the outcome will ordinarily serve as a predictor that is generalisable to other contexts. As the intention is to develop a predictive algorithm that can be transported beyond the development dataset, the focus of the DAG is on exposures shown to be associated with future obesity. The DAG in Figure 8-2 sets out the assumptions about causal inference and competing exposures (confounders) and establishes which of the (presumed to be causal) exposures and confounders precede the outcome of obesity.

The population of interest is English schoolchildren of both sexes in the ALSPAC cohort, aged ~ 10 years old at baseline. Most have not yet experienced puberty. The outcome of interest is obesity (or not) at 3 year follow-up. It is difficult to assess body fatness in growing children, but for the purposes of the predictive model the aim is to predict obesity at one time point, based on BMIz cut-offs for age and sex, calculated from height and weight. A study of young children, 29 to 68 months old, concluded that BMIz score was the optimal measure for "assessing adiposity on a single occasion" (Cole et al., 2005).

The main exposure is children's food and drink intakes and eating habits at baseline, simplified to "Child's diet". Other variables also influence the outcome. Relevant variables were arranged in temporal order, left to right, to establish whether they were confounders or mediators of the primary relationship, Child's Diet → Obesity.

Figure 8-2 Directed acyclic graph to explore causal relationships between children's diet and obesity outcomes.



Key

Causal pathway	
Confounder	
Main exposure	
Latent (unobserved or unavailable variables)	
Mediator	
Outcome	

"Child's diet" and competing exposures (confounders) which occurred or were established beforehand were considered to be potential predictors. Energy intake 2, Physical activity level 2, and Sleep 2 are mediators of the primary relationship which precede the outcome of obesity but occur after the main exposure. Mediators cannot be measured prospectively so are not a practical choice as predictors. As age and sex are used to calculate BMI z scores used establish obesity outcomes, they also cannot be predictors of obesity outcomes.

8.3.9 The questionnaire

An evidence based questionnaire about children's current (baseline) diet and other exposures was proposed as the basis of a risk tool or model to predict future obesity. Not all potential predictors are included in the questionnaire as children (or their parent/carer) might find it hard to answer quantitative questions about the amount of food consumed or dietary patterns or be unable/unwilling to answer questions about family income. However, children (or their parent/carer) can be asked about the child's eating habits and frequency of consuming specific foods and drinks, indicating whether a child's usual dietary pattern leans toward "healthy" or "energy dense".

Possible question items were listed, organised in 6 sections:

- **Child**: Age, sex, ethnicity, puberty.
- **Family**: Socio-economic status (parental occupation or education level, household income or IMD), number of children in household, parental overweight, smoking.
- **Eating habits**: Breakfast frequency, other meal frequency, family meals, eating between meals, fast food or take-away meals, eating frequency.
- **Drinks frequency** (before foods): How often? Water, milk, flavoured milk, non-dairy milks, 100% fruit juice, juice drinks, sugar sweetened drinks - squash, cordial or carbonated drinks, diet drinks, tea, coffee, other infusions, alcohol.
- **Food intakes**: How much/how often? Breakfast cereal, whole grains, dairy foods not including milk, milk, vegetables, fruit not including fruit juice, fish, snack foods, reduced fat foods, energy dense foods.
- **Health behaviours**: Dieting/ fussy eating, body dissatisfaction, sleep, physical activity.

This list is too long to be practical, but some items lack enough evidence, while certain eating habits and specific food intakes potentially correlate so could be captured by one question E.g. Breakfast eating and cereal intake. A shorter "prototype" questionnaire was devised to ask 10 year old children (or their parent/carer) about the child's exposure to potential predictors of future obesity. See Appendix F.

The language style and format of questions and response options was modelled on Child questionnaires employed in the ALSPAC cohort but has not yet been tested with the target audience. Examples of foods and drinks are based on examples in the World Cancer Research Fund's Adult diet quiz "Are you making yourself attractive to cancer?" (WCRF, 2017)

Cut-offs for food intakes and drink frequency categories in the questions were guided by categories used by studies in the Systematic Review, average portion sizes at different ages in the UK National Diet and Nutrition Survey (NDNS) and UK dietary recommendations. Quantitative or frequency thresholds were translated into frequency questions suitable for 10 year old children, with standardised response options in each section for ease of use. Categorical and binary response options in the questionnaire align with theoretical thresholds for lower or higher likelihood of future obesity, based primarily on evidence from the Systematic Review.

8.3.10 Matching potential predictors to candidate variables

The 24 potential predictors included in the questionnaire and outcomes were matched to candidate variables in the ALSPAC dataset, making sure that each matched variable preceded the obesity outcome at age 13+ years. At the same time inclusion and exclusion variables and outcome variables were identified. Variables were measured at clinics or by a series of Mothers questionnaires, Fathers questionnaires, Child-based questionnaires completed by mother/carer, Child questionnaires completed by child and Puberty questionnaires. For some predictors more than one candidate variable was identified, so measures were put in time order to help identify the most appropriate ones, with potential predictors/candidate variables ideally measured close to baseline (Focus at 10+ clinic). See Table 8-4.

There were no suitable candidate variables from a time near baseline and before 3 year follow-up (Teen Focus 2 clinic) for four potential predictors (breakfast frequency, family meals, diet drinks and dieting/fussy eating).

Transport to school (used as a simple measure/question of routine physical activity) and the eating habits 3 meals a day, eating between meals and fast food frequency were measured after baseline, but before follow-up. In the absence of other options these mediators were selected as proxy predictors.

Table 8-4 Variables in time order (Directly matched variables are highlighted)

Approx. age of child	Type of variable	Mother Q	Father Q	Child based Q (completed by mother)	Child Q (completed by child)	Puberty Q	Clinic	Matched to?
12 weeks gestation	Confounder or Predictor	Questionnaire D Mother's pre-pregnancy BMI						Not matched, later variable used
18 weeks gestation	Confounder or Predictor	Questionnaire B Ever smoked?	Questionnaire PB Father's edu					Not matched, later variables used
32 weeks gestation	Confounder or Predictor	Questionnaire C **Mother's edu** **Partner's edu**						Mothers education level
Birth ~ 8 weeks	**Exclusion criteria**	Questionnaire E **TWINS**						Twin or singleton
Birth ~ 8 weeks	Confounder or Predictor	Questionnaire E Birthweight Sex Ethnic group~~,~~ OR Ethnic background						Child's sex Ethnicity
9y 2m	Confounder or Predictor	Questionnaire P Mother's weight and height used to derive BMI						Mother's overweight

Approx. age of child	Type of variable	Mother Q	Father Q	Child based Q (completed by mother)	Child Q (completed by child)	Puberty Q	Clinic	Matched to?
9y 7m	Confounder or Predictor			Questionnaire KU Sleep times used to derive sleep duration				Sleep duration
10y 2m	Confounder or Predictor	Questionnaire Q No. children in household used to derive only child No. smokers in household used to derive smoking						Only child Adult smoking
10y 6m	**MAIN DIET EXPOSURE Inclusion criteria**						Clinic F10+ 3 day DD with **Total Energy Intake**	Whole grain, dairy, milk, vegetables, fruit & snacks
10y 6m	**Exclusion criteria**						Clinic F10+ Height, weight and age at F10	Baseline obesity status
10y 6m	Confounder or Predictor						Clinic F10+ IMD quintile	Not matched, used Mother's education
10y 8m	**MAIN DIET EXPOSURE**				Questionnaire CCH Drinks FFQ			Milk, juice & sugary drinks

Approx. age of child	**Type of variable**	**Mother Q**	**Father Q**	**Child based Q (completed by mother)**	**Child Q (completed by child)**	**Puberty Q**	**Clinic**	**Matched to?**
10y 8m	Confounder or Predictor				Questionnaire CCH Body image, Body preference used to derive Body satisfaction	Questionnaire PUB3 Puberty – armpit hair		Puberty Body satisfaction
MEASURES AFTER MAIN DIET EXPOSURES								
11y 2m	Mediator	Questionnaire R Household income Spend on food Smoking freq.						Not matched, used Mother's education
11y 6m	Mediator						Clinic F11 Daily MVPA	Not matched, used Active travel
11y 8m	Mediator (PROXY predictor)			Questionnaire KW Sleep time Transport to school used to derive Active travel		Questionnaire PUB4 Puberty		Active travel to/from school
13y 1m	**PROXY DIET EXPOSURE**			Questionnaire TA 3 meals a day? Snacks all day? Fast food?				3 meals a day, eating between meals, fast food
13y 1m	Mediator			Questionnaire TA Body image		Questionnaire PUB5 Puberty		Not matched, used Body image at 10y

Approx. age of child	**Type of variable**	**Mother Q**	**Father Q**	**Child based Q (completed by mother)**	**Child Q (completed by child)**	**Puberty Q**	**Clinic**	**Matched to?**
13y 6m	**OUTCOME**						ClinicTF2 Height, weight and age at TF2	Obesity outcome
13y 6m	DIET EXPOSURE Mediator						Clinic TF2 3 day FD Daily MVPA	Not matched, too close to outcome
MEASURES AFTER OUTCOME OF INTEREST								
13y 10m	Too late			Questionnaire TB Food avoidance	Questionnaire CCP Transport to school			Not matched, after outcome
14y 7m	Too late					Questionnaire PUB6 Puberty		Not matched, after outcome
15y 6m	Too late					Questionnaire PUB7 Puberty	ClinicTF3 Height, weight and age at TF3. Daily MVPA	Not matched, after outcome
16y 6m	Too late			Questionnaire TC Food avoidance	Questionnaire CCS Went on diet			Not matched, after outcome

Continuous candidate variables for total intakes of wholegrains, dairy foods (without milk), milk (as an alternative to the milk as a drink frequency variable), vegetables, fruit, and energy dense treats/snack-type foods were generated by combining quantitative intakes of appropriate foods from the 3 day diet diary. Dichotomous and categorical food intake variables were also generated for a "user friendly" version of the prediction model. Cut-offs for categorical and dichotomous food and drink intake quantities and frequencies in the ALSPAC cohort were matched to those in the questionnaire.

New candidate variables for only child, mother's overweight, adult smoking, sugar sweetened beverage frequency, body satisfaction, sleep duration meets recommendations and active travel were derived from pre-existing matched variables. Details of the preparation and imputation methods used for candidate variables are given in Appendix G.

8.3.11 Data cleaning and coding of candidate variables

The ALSPAC dataset was pre-cleaned by the University of Bristol. Nevertheless, continuous variables of interest were checked using descriptive statistics and histograms to find outliers and implausible values. Obvious errors were corrected. Categorical variables were tabulated, using histograms to check their frequency distributions. The combination of categories with zero or very small numbers of respondents was considered to ensure greater stability in regression analyses.

Most candidate predictors are categorical or dichotomous. In some instances, categories were collapsed to create a categorical or dichotomous variable in line with the questionnaire.

Dichotomous candidate variables were coded as dummy variables (0 and 1), with the category thought more likely to increase the risk of future obesity coded as 1. Coding of the candidate variables was done before examining associations with obesity outcomes in the ALSPAC cohort.

8.3.12 Missing data and imputation

Missing data is common in longitudinal studies and ALSPAC is no exception. Although three quarters of adolescents in ALSPAC responded to at least one survey, most did not respond to every survey that provided candidate predictors of obesity and not all teenagers had clinic measures from which to derive their obesity status at baseline and follow-up. Known reasons for missing data included loss to follow-up (child did not attend later clinics at TF2 and/or TF3) and non-response (questionnaires not completed/returned or individual questions not answered.) Possibly other observations were missing due to equipment failure or data entry errors.

Missing data can impact the analysis, especially if directly or indirectly related to other subject variables, including the outcome under investigation. The influence of missing values increases as the percentage of missing information increases. (Moons et al., 2012)

The number and percentage of missing observations for inclusion/exclusion variables, candidate variables and outcome variables (with the variables needed to derive them) for children in Focus at 10+ was checked. See Table 8-5.

In total 7,557 children attended the Focus at 10+ clinic and 7,462 completed a 3 day diet diary. BMIz and obesity status at baseline could not initially be derived for 96 children due to missing sex or missing observations for age, height and weight at Focus at 10+. These 96 children included 39 boys, 51 girls and 6 sex not known. Their mean age was 18 days older (3,920 days old, SD 106 vs 3,902 days, SD 96 days) than children with derived BMIz and they were more likely to have mothers who had ever smoked (46% vs 44%). Ethnic background and mother's education levels were similar in both groups.

The highest level of missing observations was for Armpit hair at age 10y8m (34%), but this question was asked again in later puberty questionnaires, so some missing observations could be reliably imputed if armpit hair was not present on a second, later occasion. All the other candidate variables had ≤ 30% missing observations.

Missing data mechanisms include:

- Missing completely at random (MCAR) where the reason for missing is unrelated to the values of any variables, missing or not. Other variables in the dataset and the unobserved value of the variable do not predict whether that value will be missing.
  E.g. Participant accidently skipped a question.
  Most missing data are not MCAR.
- Missing at random (MAR) where the reason for missing is unrelated to the missing values but may be linked with observed values of other variables. Other variables in the dataset can be used to predict missingness
  E.g. Girls may be more likely to decline to answer a question than boys.
- Missing not at random (MNAR) where the reason for missing is related to the value of that missing variable. The unobserved variable itself predicts missingness. E.g. Mothers with overweight were more likely to decline to answer questions about their weight.

Table 8-5 Missing observations before imputation

Variable type	Matched variables	Variable name	Missing at F10	Obs. in F10	Detail
Inclusion	Attended F10	In_f10	n/a	7,557	
Inclusion	Total energy intake at F10	fd10kj	95 1.3%	7,462	
Inclusion / exclusion	Pregnancy size	mz010	6 <1%	7,551	200 twins 7,351 single
	Child's sex	kz021	6	7,551	49.4% male
	Age at F10 (days)	fd003a	6	7,551	Mean 3,902 SD 96
	Height at F10 (cm)	fdms010	73	7,484	
	Weight at F10 (kg)	fdms026	46	7,511	
	Derived BMI at F10	zyF10bmi	91	7,466	
	Derived BMIz at F10	crF10bmizuk	96	7,461	
Inclusion / exclusion	Derived Child obesity status at F10	crF10obese	96 1.3%	7,461	1,078 obese, 6,383 not obese
Child Q	Ethnic background	c804	763 10%	6,794	96% white, 4% non-white
Child Q	Armpit hair at 10y 8m	pub370	2,568 34%	4,989	13.5% yes
Family Q	Mother's education. level	c645a	661 9%	6,896	16% had degree

Variable type	Matched variables	Variable name	Missing at F10	Obs. in F10	Detail
Family Q	No. children in household at 10y 2m	q3002	1,114 15%	6,443	12% only child in household
	Mother's self-reported weight (kg), child is 9y	p1290	2,099	5,458	
	Mother's self-reported height (cm,) child is 9y	p1291	1,901	5,656	
	Derived Mother's BMI	crMumsBMI	2,177	5,380	
Family Q	Derived Mother's o/w status	crMumoverw	2,177 28%	5,380	38% are overweight
Family Q	No. smokers in household at 10y 2m	q3031	1,279 17%	6,278	70.3% none
Proxy Eating habit Q	Number of real meals a day	ta8004	1,871 25%	5,686	70% have 3+ meals/ day
Proxy Eating habit Q	Teenager snacks all day or has meals, schooldays & weekends	ta8000 ta8002	1,770 23% 1,865 25%	5,787 5,692	68% don't snack school days, 48% don't snack weekend days
Proxy Eating habit Q	Eats in fast food restaurant.	ta8230	1,779 24%	5,778	6% eat in fast food restaurant 1 to 2 times a week or more

Variable type	Matched variables	Variable name	Missing at F10	Obs. in F10	Detail
Candidate, Drinks Q	Plain milk frequency	cch709	2,287 30%	5,270	
Candidate, Drinks Q	Fruit juice frequency	cch704	1,521 20%	6,036	
Candidate, Drinks Q	Cola freq. Fizzy drink freq. Sweetened fruit drink freq.	cch700 cch701 cch705	1,424 19% 1,464 19% 1,472 20%	6,133 6,093 6,085	
Candidates, Food & drink intake Qs	Derived intakes of Whole grain, Dairy, Milk, Vegetables, Fruit and Snacks/treats	fd10 various used to make combined variables, crfd10…	95 1.3%	7,462	
	Child's perceived body shape at 10y 8m	cch200	1,400	6,157	
	Child's desired body shape at 10y 8m	cch201	1,634	5,923	
Health behaviour Q	Derived body satisfaction at 10y 8m	crBODYSAT	1,644 22%	5,913	67% match/ are satisfied
	Usually wakes hours & mins	ku340a ku340b	1,095	6,462c	88% after 7am
	Usually asleep hours & mins	ku341a ku341b	1,100	6,457	7% after 10pm

Variable type	Matched variables	Variable name	Missing at F10	Obs. in F10	Detail
Health behaviour Q	Derived Sleep duration meets NHS advice	crkuSLEEPrec	1,095 15%	6,462	89% meet NHS advice for sleep
	Walks to school	kw7010	n/a	3,189	
	Walks home	kw7020	n/a	3,329	
	Bikes to school	kw7015	n/a	316	
	Bikes home	kw7025	n/a	307	
Proxy Health behaviour Q	Derived Active travel to/from school at 11y8m	crACTIVTR	0%	7,557	48% walk or bike to/from school
	Child's sex	kz021	6	7,551	
	Age at TF2 (weeks)	fg0011b	1,857	5,700	
	Height at TF2 (cm)	fg3100	1,865	5,692	
	Weight at TF2 (kg)	fg3130	1,873	5,684	
	Derived BMI at TF2	zyTF2bmi	1,873	5,684	
	Derived BMIz at TF2	crTF2bmizuk	1,877	5,680	
Outcome at Teen Focus 2	Derived Child obesity status at TF2	crTF2obese	1,877 25%	5,680	13% obese before exclusions

A complete case approach which excludes participants with missing values from the analysis risks selection bias, particularly if observations are MNAR and so non-ignorable. A complete case approach is also an inefficient use of available information and reduction in power, even if all the missing observations are MCAR (unlikely) such that a complete case analysis gives the same result as the full data set would if it had no missing observations. The number of children in the Focus at 10+ clinic who had measures for all the candidate variables (i.e. complete cases) was approximately 1,700 - a much reduced sample size. Complete case analysis was not attempted.

Instead single imputation methods were used to fill the gaps intelligently (best estimates) or by replacing missing values with the estimated mean from available cases (unconditional mean imputation). Imputation methods for all variables are given in Appendix G.

Missing **pregnancy size**, used to determine singleton/twin status, was deduced from the unique pregnancy identifier and birth order within that pregnancy. (These two variables had no missing observations for children in the Focus at 10+ clinic.)

Missing **sex** was deduced from responses to sex-specific puberty questionnaires.

Missing **age, height and weight** measures, needed to derive BMIz and obesity status at baseline and follow-up, were estimated from reported age, height and weight values for that individual child at earlier and later clinics.

8.3.13 Split sample

After imputation, individuals whose obesity status at ages 10.5 years, 13.5 years and 15.5 years could still not be derived were excluded from the analysis. This left 5,486 singleton children, aged 10+ years old, who had completed 3 day diet diaries and who did not have obesity at baseline.

Using a random number seed (a bank note serial number) for reproducibility, the remaining eligible respondents were randomly split 3:1 into a derivation sample (n = 4,114) and a validation sample (n = 1,372) in preparation for model development.

8.3.14 Pre-specified candidate variables

A summary of all 20 pre-specified variables, with their initial degrees of freedom (d.f.) allocation, number of observations and level of missing in the derivation sample is shown in Table 8-6.

Note that milk appears twice, measured by the drinks frequency questionnaire or alternatively by the 3 day diet diary. Only one of these milk variables can be used in a model. Similarly, only one variant (continuous, categorical or dichotomous) of each food intake variable can be used in a model.

Table 8-6 Pre-specified candidate variables for CORA

Variable	Scale	Categories (frequency) or range, mean and SD in derivation sample	d.f.	No. obs in deriv. sample (missing)
Child, family and socio-demographic variables Hypothesis: Obesity more likely				
Ethnic background	Dichotomous	Non-white (135) White (3,617)	1	3,752 (362 or 8.8%)
Armpit hair at 10y 8m	Dichotomous	Yes (352) No (3,762)	1	4,114 (0%)
Mother has degree	Dichotomous	Yes (658) No (3,153)	1	3,811 (303 or 7.4%)
Only child in household at 10y 2m	Dichotomous	Yes (398) No (3,241)	1	3,639 (475 or 11.6%)
Mother is overweight when child is 9y	Dichotomous	Yes (1,323) No (2,567)	1	3,890 (224 or 5.4%)
Smokers in household when child is 10y 2m	Dichotomous	Yes (1,080) No (2,514)	1	3,594 (520 or 12.6%)
Eating habits Hypothesis: Obesity more likely				
3 or more meals a day	Dichotomous	Yes (2,380) No (959)	1	3,339 (775 or 18.8%)
Snacks between meals	Dichotomous	Yes (2,294) No (1,820)	1	4,114 (0%)

Variable	Scale	Categories (frequency) or range, mean and SD in derivation sample	d.f.	No. obs in deriv. sample (missing)
Fast food once a week or more	Dichotomous	Yes (187) No (3,206)	1	3,393 (721 or 17.5%)
Drinks frequency Hypothesis: Obesity more likely				
Milk frequency	Dichotomous	< 1 serving a day (1,577) ≥ 1 serving a day (2,537)	1	4,114 (0%)
Juice frequency	Dichotomous	< 1 serving a day (1,379) ≥ 1 serving a day (2,735)	1	4,114 (0%)
Sugar sweetened beverages frequency	Dichotomous	< 1 serving a day (1,882) ≥ 1 serving a day (2,232)	1	4,114 (0%)
Food intakes Hypothesis: Obesity more likely				
Whole grain intake **1 serving = 40g**	Continuous	0 to 382g per day, mean 28g per day, SD 37g per day	1	4,114 (0%)
	Categorical	Zero (1,651) < 1 serving a day (1,329) ≥ 1 serving a day (1,134)	2	4,114 (0%)
	Dichotomous	< 1 serving a day ≥ 1 serving a day	1	4,114 (0%)
Dairy food intake, not including milk **1 serving based on yogurt = 125g**	Continuous	0 to 450g per day, mean 46g per day, SD 54g per day	1	4,114 (0%)
	Categorical	Zero (1,041) < 1 serving a day (2,667) ≥ 1 serving a day (406)	2	4,114 (0%)
	Dichotomous	< 1 serving a day ≥ 1 serving a day	1	4,114 (0%)

Variable	Scale	Categories (frequency) or range, mean and SD in derivation sample	d.f.	No. obs in deriv. sample (missing)
Milk intake (as an alternative to Milk frequency) **1 serving = 250g**	Continuous	0 to 1,370g per day, mean 211g per day, SD 181g per day	1	4,114 (0%)
	Categorical	Zero (529) < 1 serving a day (2,208) ≥ 1 serving a day (1,377)	2	4,114 (0%)
	Dichotomous	Zero servings a day > zero servings a day	1	4,114 (0%)
Vegetable intake **1 child sized serving = 50g**	Continuous	0 to 435g per day, mean 69g per day, SD 59g per day	1	4,114 (0%)
	Categorical	Zero (604) < 2 servings a day (2,460) ≥ 2 servings a day (1,050)	2	44,114 (0%)
	Dichotomous	< 2 servings a day ≥ 2 servings a day	1	4,114 (0%)
Fruit intake **1 serving = 80g**	Continuous	0 to 677g per day, mean 71g per day, SD 76g per day	1	4,114 (0%)
	Categorical	Zero (1,074) < 2 servings a day (2,540) ≥ 2 servings a day (500)	2	4,114 (0%)
	Dichotomous	< 2 servings a day ≥ 2 servings a day	1	44,114 (0%)
Energy dense treats and snack type food intake **1 serving = 30g**	Continuous	0 to 422g per day, mean 97g per day, SD 48g per day	1	4,114 (0%)
	Categorical	Zero (36) < 2 servings a day (897) ≥ 2 servings a day (3,181)	2	4,114 (0%)

Variable	Scale	Categories (frequency) or range, mean and SD in derivation sample	d.f.	No. obs in deriv. sample (missing)
	Dichotomous	< 2 servings a day ≥ 2 servings a day	1	4,114 (0%)
Other Hypothesis: Obesity more likely				
Satisfied with body	Dichotomous	Yes (2,405) No (914)	1	3,319 (795 or 19.3%)
Meets recommended sleep	Dichotomous	Yes (3,275) No (363)	1	3,638 476 or 11.6%)
Active travel to/from school	Dichotomous	Yes (2,062) No (2,052)	1	4,114 (0%)

8.4 Discussion

This chapter has described how 20 of 24 dietary and non-diet exposures in children, identified as potential predictors of future obesity, were successfully matched to candidate variables in preparation for developing a risk algorithm to predict obesity in early adolescence. Other prediction models for childhood overweight/obesity initially considered between 5 and 19 predictors (Ziauddeen et al., 2018) with one study judging 33 (Weng, S. F. et al., 2013), but none considered dietary factors as predictors. One study acknowledged that not being able to consider predictors including diet, physical activity and sleep (as those prognostic/risk factors were not available in both cohorts used for their model development) was a limitation (Pei et al., 2013).

The Avon Longitudinal Study of Parents and Children recruited 15,000+ pregnant women in the early 1990s and has studied mothers and their children ever since. The study, assessed as high quality using the Newcastle Ottawa Scale, has validated measures of children's food intake using 3 day food diaries, plus objective measures of children's height and weight from which BMI can be calculated. Over 7,500 children in the age range of interest (10 to 13 years old) attended clinics and 7,452 had quantified food intakes at 10 years old. The cohort was chosen because of these measures, which are core elements of the risk algorithm, and because the

cohort offered a large sample size as well as other variables that might serve as non-diet predictors.

The University of Bristol's access policy for ALSPAC data is well established, but the application still took twelve months from first proposal to data acquisition. With so many catalogued variables, it was a hard task to choose variables wisely and within budget, especially as the number of observations per variable was not always clear from the documentation. This was the first indication that missing observations might be an issue.

Other sources of data for tool development were considered. Cohorts in the Systematic review that used 3 day food diaries were far smaller (FAMS, Framingham Children's Study, numbers in the 100s), methodologically challenging (DONALD, rolling recruitment) or single sex (NGHS, girls only.) The largest cohorts (GUTS and Project EAT in the USA) were judged to be poorer quality than ALSPAC, relying on FFQs and self-reported height and weight, and Project EAT participants were already teenage at baseline.

The I. Family study, a European-wide project with 8 child cohorts that used 24 hour dietary recalls, seemed promising, but at the time of enquiry their data-access policy was not agreed.

While offering much of what was sought, the study design of the ALSPAC cohort had limitations. Due to the self-selecting recruitment strategy, the cohort was slightly over-representative of more affluent and better educated mothers, and ongoing attrition has intensified this bias. Loss to follow-up is not unusual in longitudinal studies. More of the lower socio-economic status families dropped out altogether or did not participate in follow-up clinics, and often those lost had potential predictors that we were interested in, such as children whose parents had higher rates of obesity (Hughes et al., 2011). From the outset there were only 5% of non-white participants, making it difficult to investigate ethnicity with confidence.

Three quarters of children who attended the Focus at 10+ clinic at 10 years old did not have a 100% return rate for subsequent questionnaires. The cumulative effect of missing observations was higher than anticipated with only ~ 1,700 out of 7,577 children having measures for all variables considered as predictors. These keen responders are unlikely to be representative of the whole cohort. Besides adding bias, a complete case analysis would restrict the number of predictors that could be fitted in a model, due to the reduced sample size and fewer obesity "events". Using a split-sample approach to validation restricts the model further, but time-constraints ruled out the use of *k*-fold cross-validation or boot strapping resampling methods. This was a limitation.

Imputation was clearly necessary. The simple imputation methods used (best estimates and

unconditional mean imputation) proved helpful for understanding the dataset and making assumptions explicit, but not all missing observations could be singly imputed, which was a further limitation. Since the singly imputed observations are estimates, their values contain random error. This extra source of error is overlooked by STATA in analyses, which will give standard errors and p-values which are too small.

Alternative multiple imputation methods generate multiple estimates, with a random component which adds variation that the software can factor into analyses, giving more accurate estimates of standard error. However, as some variables had missing >25% there was concern that multiple imputation methods may add uncertainty. Time constraints also dictated that multiple imputation methods were not employed.

Food intake predictors (Whole grains, Dairy, Milk, Vegetables, Fruit, Energy dense treats) were readily matched to quantified intakes of representative foods. New variables for Whole grain intake and Dairy intake could not capture all intake, as representative foods were not differentiated into whole grain or dairy-based variants. Elsewhere in ALSPAC full-fat and reduced-fat dairy foods have been considered separately (Bigornia et al., 2014), but only cow's milk was differentiated by fat content in our dataset. Comparing predicted energy requirements (from accelerometer data and body composition) with reported dietary intakes at age 10 years revealed that 44% of children in ALSPAC were plausible reporters, with 35% under-reporting and 21% over-reporting (Noel et al., 2011). Such dietary measurement error will add "noise" to a predictive model but using categorical food intake variables may off-set bias caused by extreme under or over reporting.

Drinks frequency predictors (Milk, Fruit juice, Sugary drinks) were matched to variables from a drinks frequency questionnaire at age ~ 10 years 8 months. The FFQ was completed by most children on their own, so may have less social desirability bias than the 3 day food diary completed with the help of a parent/carer and be more representative of habitual drinks intake. Comparison with quantified intakes of drinks showed that discrimination between the four frequency categories was poor, so they were collapsed into infrequent and frequent consumers. Children in higher drinks frequency categories also had higher quantified intakes of drinks on the 3 days surveyed, so drinks quantities from the food diary were used to impute missing responses.

Children's eating habits were reported by the mother when the child was ~13 years 1 month. This was shortly before the outcome to be predicted (obesity at 13 years 6 months) so Three meals or more a day, Eats in-between meals and Takeaway or Fast food frequency were used as proxy predictors, assuming that eating habits at 13+ years would have been similar at 10+ years. This assumption may be reasonable for Meal frequency and Snacking habits, but less so

for Takeaway or fast food, as it has been observed that fast food frequency is lower at younger ages (Taveras et al., 2005). In the absence of other candidates, the fast food variable was still considered, but levels of missing observations were high, and it did not seem reasonable to assume that all missing observations were from mothers of children who had Fast Food less than once a week.

Non-food predictors were the most difficult to match. Ingenuity, assumptions and imputations were necessary, as set out in Appendix G, which will all add error. For example, the question about Mother's highest education level, used as an indicator of socio-economic status, was asked during pregnancy. We assumed that mother's educational attainment stayed the same, but in some cases it will have changed.

Missing puberty/armpit hair observations were successfully imputed from the same measure at older ages, but we had not requested the same measure at younger age (which would have been helpful). We assumed that those still with missing observations after imputation were less likely to have armpit hair at such a young age, although a few may have started puberty.

Mother's overweight status was used as a stand-in for child's perception of parental overweight, assuming children would perceive it. The variable is generated from mother's self-reported height and weight, so calculated BMIs will contain reporting bias and may tend to be under-estimated, not capturing all instances of mother's overweight. We also applied a uniform BMI gain on pre-pregnancy BMI (also based on self-reported height and weight) for imputation – a big assumption over 9 years.

Mothers may have seen smoking as undesirable, which could account for so few replies about their own smoking frequency. However the presumed-to-be-honest answers given helped to corroborate the number of smokers in the household and to impute missing observations.

The body satisfaction variable simplistically assumes that any level of difference between perceived and desired body shape equates to body dissatisfaction, while the Active transport variable takes no account of distance travelled to/from school or of other forms of physical activity.

Four potential predictors could not be matched. There was no candidate variable for Diet drinks, or for Breakfast eating and Family meals (although they may partly be captured by Three meals or more a day). Food avoidance and dieting were measured in older teenagers, but not at younger ages, so there was no suitable candidate for Dieting/fussy eating.

Four matched candidate variables, Ethnic background, Fast food frequency, Number of children in household (for Only child status) and Meets sleep recommendation, had more missing observations than there were children in a category, with no further opportunities for

imputation. These variables were still considered as potential predictors, but ultimately were not tested in the model to avoid bias, as explained in the next chapter.

Baseline obesity status generated for 7,461 out of 7,557 children who attended the Focus at 10+ clinic showed that 15.7% were classified as obese. Obesity status generated for 5,680 children in F10 with height and weight measures from the Teen Focus 2 clinic showed that 13.0% were classified as obese at follow-up. These percentages are in line with obesity prevalence and the trend reported elsewhere in the ALSPAC cohort, 15.8% at age 11+ years and 12.4% at age 15+ years (Hughes et al., 2011).

Missing age, sex, height and weight were imputed so that obesity status at baseline and follow-up could be derived for children without this variable at F10 (n= 96) and TF2 (n = 1,877). Missing age and sex were verified by measures at other times. Missing height and weight were imputed with unconditional means, assuming that children had the same average gains over time. BMIz and obesity status derived from imputed height and weight values are "best estimates", unlike BMIz and obesity status derived from measured height and weight which are "known". Imputation gave 70 additional estimates of obesity status at baseline and 958 additional estimates of obesity status at follow-up. After imputation obesity prevalence fell to 14.5% out of 7,531 at age 10+ years and 13.3% out of 6,638 at age 13+ years, which suggests that imputed BMIz values are slightly under-estimated. Although the objective of keeping more children in the sample for model development was achieved, the imputations add uncertainty to the obesity outcome to be sought by the predictive model. In hindsight, a way of distinguishing between "known" and "best estimates" of obesity status could, and maybe should, have been employed.

After excluding twins, children without a 3 day food diary at F10 and all children with known or estimated obesity at baseline (aged 10+ years) there were fewer new cases of obesity (239 in a sample of 5,486, 4.4% incidence) than we anticipated. It would be easy to attribute this to disparities in loss to follow up in ALSPAC, especially as Hughes et al found that 35% children with obesity, 32% children with overweight and 30% children with a healthy weight at 11 years, were lost to follow-up by 15 years. However, the same investigation reported that the incidence of overweight and obesity reached a peak of 5% between ages 7 and 11 years but was less than 2% between age 11 to 15 years (Hughes et al., 2011), confirming that there are not very many new obesity "events" for a model to predict.

Chapter 9 Tool development: Model fitting, internal validation and risk score allocation

9.1 Summary

This chapter (Chapter 9) describes the methods used in the split-sample ALSPAC dataset to develop and internally validate a logistic regression model to predict obesity at age 13+ years, and to translate that model into a predictive risk score for use with the CORA questionnaire. Models in the derivation (n = 4,114) and validation (n = 1,372) samples are presented as tables with odds ratios.

Candidate variables for 16 of 24 potential predictors were tested in a sequence of models in the derivation sample. The final model had nine "useful" predictors (Puberty/armpit hair, Mother's overweight, Child's body satisfaction, Active travel, Categorical Vegetable intake, Dichotomous Dairy servings, Dichotomous Milk servings, Dichotomous Energy dense treats intake and SSB frequency) with 1 interaction term (Categorical Vegetable intake and SSB frequency).

Internal validation results for the final model are summarised as a table, showing model performance in the derivation sample, validation sample and combined cohort. In the derivation sample the final model offered acceptable discrimination (AUROC = 0.76) and calibration, correctly predicting 78% of outcomes at the optimum 5% decision threshold. The model performed less well in the smaller validation sample, correctly predicting 71% of outcomes. 10% of obesity predictions matched observed obesity outcomes.

The final model was run in the combined cohort with logit coefficients as the output. Coefficients were rounded to integers to give a risk score for each variable/predictor. 20% of children with a total risk score were in the highest risk quartile, scoring ≥ 10/25. Predictive metrics were calculated for the total risk score. Discrimination was reduced but still acceptable (AUROC = 0.72).

Eight potential predictors could not be tested due to high levels of missing observations or because there were no suitable candidate variable in the dataset. Their risk scores were estimated by comparison with "useful" predictors, using evidence from the Systematic Review and published obesity risk tools. Estimated risk scores were added to the revised CORA questionnaire.

9.2 Methods

The two main strategies for generating a predictive model are the full model approach and the predictor selection approach (Moons et al., 2012). A full model strategy selects candidate predictors based on prior knowledge and includes them all in the final model without further testing. We used the exploratory predictor selection strategy, including all the pre-specified candidate predictors in an initial multivariable model, before removing the ones that did not make a useful contribution.

9.2.1 Model specification

An analysis plan was set out before investigating exposure-outcome associations or any model fitting. The 20 pre-specified candidate variables were then checked and tested in a series of logistic regression models to predict the likelihood that a child who is not classified as obese at 10.5 years will have obesity as a teenager at 13.5 years old. Analyses and model fitting were carried out using STATA IC 15.0 (StataCorp, 2017).

In order to minimise the loss of predictive information, in the initial model development phase continuous food intakes were employed (Steyerberg and Vergouwe, 2014), although linearity of the continuous predictor-outcome association was not assumed. "User friendly" dichotomous and categorical food intake variables, which could support the ease of use of the final model, were also tried.

An important consideration is whether the data is adequate for a particular model. Are there are enough events per variable (EPV)? In logistic regression models 10+ EPV or events per degree of freedom are advised to avoid overfitting the model (Peduzzi et al., 1996). With EPV values below 10, models become harder to interpret as regression coefficients can be positively or negatively biased.

Each candidate variable was allocated an initial degree of freedom (d.f.). Continuous variables had one d.f. and categorical variables had one d.f. per category minus one. (See Chapter 8 Table 6). In the derivation sample set of 4,114 children 168 children (4.1%) had obesity at age 13.5 years, allowing up to 16 d.f., but only if all observations are kept in the model. As there are relatively few new obesity outcomes or "events", if all 20 candidate variables are included the model will be over-fitted. Additional estimated standard errors accumulate as more variables are added to a model, making an over-fitted model more dependent upon the observed data in the development dataset. Overfitting gives an optimistic view of the error rate of the prediction model created in the development dataset – the model will likely

perform less well in a different set of data. Instead the aim is to build a parsimonious model, with a minimized selection of covariates that still mimics the true outcome. In theory a reduced model is likely to be more stable, more transportable to a different setting and easier to use.

To help decide which candidate variables to include as predictors, simple correlations between all categorical candidate variables were checked in the derivation sample. Pairwise interactions were checked using two-way tables with Pearson's chi square test. Highly correlated (corr ≥0.6) variables should not be used together in the same model, as the second variable adds error with little extra predictive value.

All 20 pre-specified candidate variables were included in initial exploratory multivariable logistic regression models to estimate an odds ratio for obesity at approximately 13 years 6 months (Teen Focus 2 clinic) with p-values and 95% confidence intervals. As a learning exercise multivariable models were run in the derivation sample only, with either all continuous, all categorical or all dichotomous versions of the food intake variables for comparison. Model outputs were interpreted using guidance from UCLA Statistical Consulting Group, accessed online via STATA help (UCLA, 2019).

A univariable logistic regression model was then run in the derivation sample for each individual candidate variable. For categorical variables the significance level of the likelihood test was examined. Post-estimation commands were used to calculate the area under the receiver operating characteristic curve for each model, to assess the ability of each univariable model to differentiate between individuals with and without the outcome of interest. No candidate variable was eliminated at this stage.

A multivariable logistic regression model was then run in the derivation sample for each sub-group of candidate variables (Non-food variables, Eating habits, Drinks frequency and Food intake variables).

The sub-group and full multivariable models were run with stepwise removal and stepwise addition, at significance levels $p \leq 0.05$ and $p \leq 0.25$. Unlike stepwise addition, stepwise removal gives an assessment of the effect of all the candidate variables together, although the model may be over fitted at the start. A smaller significance level minimises the number of predictors but can miss potentially important ones, whereas using a larger significance level may keep less important predictors in the model (Moons et al., 2012).

Paradoxical associations may occur when observations are missing not at random, giving a misleading view of a variable's influence on outcomes within the model. Candidate variables were checked to ensure that no more than 20% of observations in the derivation sample were missing after imputation. (Imputation is described in Chapter 8 and in Appendix G.) If missing was less than 20%, yet higher than the frequency observed in a category, that candidate variable was also eliminated to avoid bias. For candidate variables with lower levels of missing, an alternative variable with missing as an extra category was generated. Such variables keep observations in the model by avoiding list-wise deletion of missing but use an additional degree of freedom.

Each remaining candidate variable was assessed, based upon:

i. inclusion in full models with stepwise removal/addition
ii. inclusion in sub-group models with stepwise removal/addition
iii. whether the variable achieved significance at $p = 0.05$ as a single predictor of future obesity

Candidate variables which met all three criteria ("best predictors") were included in a REDUCED model; candidate variables which failed to meet any criteria were eliminated. Remaining variables, whose predictive value was uncertain, were added to the REDUCED model in turn, aiming to maximise the model's Likelihood ratio, Pseudo R^2 and AUROC values. In single variable logistic regression models, servings/day of total milk (measured by the 3 day diet diary) was a significant predictor of future obesity, so was tried instead of milk as a drink (measured by the drinks frequency questionnaire). The candidate variables which most improved model performance were identified and added to the reduced model in combination until the best model performance was achieved, while not exceeding the d.f. permitted by the number of obesity outcomes kept in the model.

The direction of influence of some remaining candidate variables on obesity outcomes was not as expected. These variables were checked and revised if necessary.

Finally, interactions between pairs of remaining candidate variables were investigated if they seemed plausible. An interaction implies that the effect of one variable is not constant, given changes in the other variable. Interactions were checked between the pair alone (2 variables) and as an interaction term in the multivariable model, which used more d.f. If interactions were significant at $p \leq 0.05$ in the full model and improved model performance, two-way tables with obesity outcomes in each category were run to help understand the interaction. Limited d.f. meant that not all significant interaction terms could be added to the model

without overfitting. The interaction term that added most to the model's performance, without exceeding the available degrees of freedom, was kept.

9.2.2 Purposeful selection of covariates

As a cross-check, the "purposeful selection of covariates" method was also tried (Hosmer, 2013). Steps 1 to 5 of 7 iterative steps were followed in the derivation sample, using categorical variables only, to see if the same set of covariates were included in a final model without interaction terms. (Steps 6 and 7 pertain to checking interaction terms and assessing model adequacy and fit, so were not needed as part of this check.)

"Purposeful selection" keeps more observations than a complete case analysis (which drops all missing observations from the outset). However, after Step 1 (univariable analyses of all candidate variables) the method requires the same sample size at each step, to allow a fair comparison of the performance of each successive multivariable model. To do this, missing observations from the initial selection of covariates must be dropped. As a result, the "purposeful selection" method uses fewer observations than the approach described in section 9.2.1.

9.2.3 Model estimation

As the outcome (not obese or obese) is binary (0 or 1), a generalised linear model was estimated using logistic regression. This calculates the log-odds of event "success" (an outcome of obesity) using the equation:

$$\text{Log}\ (p/1\text{-}p) = \text{Constant} + (\text{estimate}_1 \times \text{predictor}_1) + (\text{estimate}_2 \times \text{predictor}_2) + (\text{estimate}_3 \times \text{predictor}_3) \ldots\ldots\ldots + (\text{estimate}_n \times \text{predictor}_n)$$

The right hand side of the equation is the linear predictor, where n equals the total number of predictors.

The default output in STATA is a coefficient in log-odds units, which can range between -infinity and + infinity, with no effect equal to a value of 0. For easier interpretation, the output can also be presented as an odds ratio (O.R), which is the exponential of the logit coefficient. O.R.s range between 0 and + infinity, with no effect equal to a value of 1. An O.R. below 1 is a "negative" effect, indicating a reduced likelihood of the outcome, while an O.R. above 1 is a "positive" effect, indicating an increased likelihood. Where the O.R. was in the opposite direction for a predictor/candidate variable, the reference category was changed to help interpretation of the final model.

The extent of overfitting in the final model was estimated by a heuristic shrinkage factor, based on the log likelihood ratio chi squared statistic χ^2 and the number of parameters or degrees of freedom used by the model. Shrinkage was calculated using the formula:

$(\chi^2 - \text{d.f.}) / \chi^2$

A value of 1 indicates no shrinkage, while values below 0.9 suggest overfitting. Some shrinkage is to be expected (Van Houwelingen and Le Cessie, 1990).

9.2.4 Assessment of model performance

The performance of the final model was assessed in the derivation and validation samples and in the combined cohort. The amount of shrinkage in the fitted model in each dataset was estimated.

Overall performance was measured using the Brier score, an aggregate measure of disagreement between the observed outcome and the prediction.

In line with the ABCD framework (Steyerberg and Vergouwe, 2014), discrimination (how well the model differentiates between children at high risk and low risk of future obesity), calibration (how well obesity outcomes predicted by the model and observed obesity outcomes agree) and clinical usefulness were measured.

Discrimination was measured using the area under the receiver operating characteristic (AUROC) curve which is equal to the concordance (c) statistic for binary outcomes. Calibration was assessed graphically and by goodness-of-fit tests. Calibration plots were generated using the command “pmcalplot” in STATA, found via STATA help. (Ensor et al., 2018)

Clinical usefulness was assessed using classification tables. Sensitivity (proportion of correctly predicted true positive outcomes), specificity (proportion of correctly predicted true negative outcomes), positive predictive values (PPV) and negative predictive values (NPV) were calculated at the decision threshold which optimised sensitivity and specificity, ascertained by plotting sensitivity and specificity against a range of probability cut-offs between 0 and 1 to find where they cross.

9.2.5 Model presentation

Models are shown in table format with odd ratios and the intercept. The final model without the interaction term is also given as an equation with regression coefficients and the intercept. A summary table of the predictive performance of the final model with the interaction term in the derivation, validation and combined cohort samples is presented.

9.2.6 Scoring

A risk score for each candidate predictor kept in the model was calculated based on logit coefficients from the final model, with and without the interaction term, in the combined sample. The smallest logit coefficient was standardised to be equal to 1, with the other regression coefficient weights standardised accordingly and rounded to whole numbers or integers that could easily be added together, supporting practical use of the model (Weng, S. F. et al., 2013) . Interactions were not assigned scores as some of them had logit coefficients with negative values. Instead the weights of the interacting variables were amended to reflect their increased importance in the model, while the other variables' scores were adjusted to keep them in the same order and proportionally similar to their coefficients in the model with interactions. The distribution of scores was divided into quantiles of risk, and the number of obesity outcomes in each quantile was examined.
Inevitably this simplification leads to some loss of information and a reduction in predictive performance. Logistic regression models, using either the risk score or the quantile of risk as the predictor of future obesity, were run in the combined sample, followed by post-estimation commands to estimate the AUROC (c-statistic) of the simplified models to compare with the original model performance.

Eight of the 20 potential predictors of obesity could not be tested in the model. This was due to:

- a lack of suitable candidate variables in the ALSPAC dataset (Breakfast frequency, family meals, diet drinks and dieting/fussy eating)
- missing observations (Child's ethnic background, only child, fast food once a week or more and meets sleep recommendation).

Using evidence from the Systematic Review, the strength of the association of these missing prognostic/risk factors with obesity was compared with that of non-missing prognostic/risk factors. Scores were estimated accordingly and allocated for future consideration.

9.3 Results

9.3.1 Participants

Almost half the children (n = 7,557) enrolled in the ALSPAC cohort attended the Focus at 10+ clinic (baseline) and 7,462 children returned 3 day diet diaries. In total 1,270 children were excluded (197 twins, 1,060 singleton children with obesity at baseline and 13 singleton children whose baseline obesity status was unknown and could not be derived), leaving 6,192 eligible respondents. Obesity status at the Teen Focus 2 clinic (follow-up) could not be established for 706 of these children (11.4%). Hence the combined derivation and validation samples for analyses consisted of 5,486 singleton children (48% boys) without obesity at baseline, with 239 new obesity events (41% boys) at follow-up three years later.

9.3.2 Pre-specified candidate variables and degrees of freedom

The 20 pre-specified candidate variables identified as potential predictors of future obesity are presented at the end of Chapter 8 in Table 6.
The variable with the highest missing initially (34%) was Armpit hair at 10y 8m, but all missing observations were successfully imputed. After imputations, the variables with the most missing data in the derivation sample were Eating habits, 3 meals or more a day (19%) and Fast food once a week or more (18%). Family and demographic variables had between 5% and 13% missing observations. Levels of missing in the validation sample were similar. There were no missing observations for Drinks frequency or Food intake variables in the combined sample.

After allocating degrees of freedom, the initial model with all 20 candidate predictors using either continuous or dichotomous food intake variables had 20 d.f. Using categorical food intake variables instead (with 3 categories each) increased this to 25 d.f.

9.3.3 Correlations

In the derivation sample no pair of candidate variables had a correlation above 0.6, but for some candidate variables there were fewer than 3,000 observations in the pairing due to missing observations for both variables.

Two-way tables with Pearson's chi2 were run for 28 x 27 combinations of categorical variables. Approximately 200 correlations out of 378 were significant at $p<0.05$. As expected, categorical and dichotomous variables for each food (Whole grain, Dairy, Milk, Vegetables, Fruit, and Treats) were perfectly correlated as they are based on the same data. Strongly and significantly correlated pairs, with a Pearson's chi2 value above 200, included Mother has

degree vs. Mother has A level or above (alternative dichotomous versions of Mother's education level) and Milk frequency vs. Milk intake. The perfectly and strongly correlated candidate variables were those generated from the same, or part of the same, information. Alternative versions of a candidate variable were not employed together in the same model, but their predictive capabilities were compared before deciding which version to use.

Other significant correlations, with a Pearson's chi2 < 200 but > 40 included:

- Mother's education level and Child's eating habits (3 meals a day, Snacks between meals), SSB frequency, food intakes (Whole grain, Vegetables, Fruit), Smokers in household and Mother's overweight.
- Smokers in household and Snacks between meals.
- 3 meals a day, Snacks between meals and Fast food
- Child's eating habits (3 meals a day, Snacks between meals) and SSB frequency, food intakes (Whole grain, Vegetables)
- Juice frequency and SSB frequency
- Juice frequency and Fruit intake
- SSB frequency and food intakes (Fruit, Whole grain)
- Whole grain intake and other food intakes (Fruit, Vegetables, Dairy, Milk)
- Dairy intake and other food intakes (Fruit, Vegetables)
- Vegetable intake and Fruit intake

9.3.4 Exploratory multivariable model

The results of fitting exploratory MULTIVARIABLE logistic regression models, using all 20 candidate variables with either continuous, categorical *or* dichotomous food intake variables, are detailed in Table 9-1, Table 9-2 and Table 9-3.
Model summaries and parameter estimates were interpreted as follows:

Model summary

- Log likelihood at final iteration.
  Logistic regression uses maximum likelihood. STATA runs a series of iterations of the model, starting with no predictors (an "empty" model), then adding all predictors, re-running the model until the log likelihood is maximised and the difference between iterations is negligible. STATA lists the log likelihood for each iteration, with the log likelihood for the final iteration in the model summary. The value by itself has no meaning but can be helpful when comparing nested models.

- Number of observations

  STATA defaults to list-wise deletion of cases with a missing value for any variable. Only 2,345 children were kept in these initial models, of whom 82 (3.5%) had obesity at follow-up.

- LR chi2 (n)

  The likelihood ratio (LR) chi-square test statistic is equal to -2 x the difference between the first and last log likelihood.

  The number in brackets shows the number of degrees of freedom or number of predictors in the model. The model with categorical food intake variables uses 25 d.f. and the models with either continuous or dichotomous food intake variables use 20 d.f., so all are over-fitted for the 82 obesity "events" kept in the model.

  This was confirmed by shrinkage calculations with the formula: $(\chi^2 - \text{d.f.}) / \chi^2$ Values < 0.9 indicate overfitting.

  - Model with continuous food intake variables (98.98 – 20) / 98.98 = 0.80
  - Model with categorical food intake variables (102.52 – 25) / 102.52 = 0.76
  - Model with dichotomous food intake variables (96.01 – 20) / 96.01 = 0.79

- Prob >chi2

  The probability of the model achieving the likelihood chi-square statistic, given that the null hypothesis is true (no effect of the independent variables or candidate predictors, on the dependent variable or obesity outcome) is less than 0.001 in each case, so the model is statistically significant at $p \leq 0.05$.

- Pseudo R2

  In ordinary least squares regression, R-squared indicates the proportion of total variability explained by the model, but logistic regression does not have an equivalent measure. Instead various pseudo R-squared measures have been developed to evaluate the *goodness-of-fit* of logistic models. McFadden's pseudo R-squared is the default value reported by STATA. Pseudo R-squared has little meaning on its own but can be used to compare models predicting the same outcome, run with the same data. A higher pseudo R–squared value indicates the model with better predictive ability. The initial model with categorical food intake variables has the largest pseudo R-squared and LR chi2 values compared with the models with continuous or dichotomous food intake variables, indicating that it is the better predictive model of the three.

- AUROC

  Post estimation commands were used to generate an AUROC value or c-statistic. All initial models have "acceptable" discrimination (≥ 0.7 and < 0.8) and so are better able than chance (0.5) to distinguish between two randomly selected individuals, one with the outcome of obesity and one without.

Parameter estimates

- Dependent variable = Obesity outcome at Teen Focus 2 (TF2) clinic

  Shown in the top left cell (crTF2obese)

- Independent variables

  Candidate variables included in the model are listed in the left hand column.

  The final item in the column, _cons, is the intercept.

- Odds Ratio

  Models were run with the logistic command to give odds ratios as the output. O.R. values above 1 indicate an increased risk of obesity at TF2, O.R. values below 1 indicate a decreased risk. O.R. values for each candidate variable vary slightly between the different models, depending upon whether continuous, categorical or dichotomous food intake variables were included in the model.

- Std.Error

  Standard errors associated with the O.R.s

- z and P|z|

  The z-statistic from a Wald chi-square test and 2-tailed p values are used to test the null hypothesis that the individual candidate variable (alongside the other candidate variables within this model) has no effect on the outcome of obesity, O.R. = 1. If a candidate variable is significant at $p \leq 0.05$ the null hypothesis can be rejected; the individual candidate variable does have some effect on the outcome and makes a useful contribution to this particular model.

- Lower and Upper 95% Confidence Intervals (CI) for the Odds Ratio

  CIs indicate how low and high the population value of the parameter might be. If the range includes 1.0, the O.R. is not statistically significant.

Table 9-1 MULTIVARIABLE model with all 20 candidate variables including continuous food intake variables, using 20 d.f.

Log likelihood at final iteration	-306.03
Number of obs. (obesity outcomes)	2,345 (82)
LR chi2 (20)	98.98
Prob > chi2	**<0.001**
Pseudo R2	0.1392
AUROC	0.796

crTF2obese	Odds Ratio	Std. Error	z	P>|z|	Lower 95% CI for OR	Upper 95% CI for OR
crEthnicity	1.74	0.87	1.10	0.271	0.65	4.66
crArmpit	3.38	0.95	4.32	**<0.001**	1.94	5.86
crMumDegree	1.25	0.45	0.61	0.541	0.61	2.53
crONLYCHILD	1.59	0.52	1.43	0.153	0.84	3.01
crSMOKERS	0.79	0.22	-0.86	0.387	0.46	1.35
crMumoverw	3.08	0.74	4.69	**<0.001**	1.93	4.94
cr3Meals	1.15	0.31	0.52	0.603	0.68	1.93
crSnacking	1.24	0.31	0.88	0.377	0.77	2.02
crFastFood	0.94	0.47	-0.13	0.898	0.35	2.52
crMILK	1.32	0.32	1.14	0.254	0.82	2.11
crJUICE	1.10	0.30	0.36	0.716	0.65	1.88
crSSB	1.14	0.30	0.52	0.601	0.69	1.90
crfd10WHLGR	1.00	0.00	0.31	0.757	0.99	1.01
crfd10DAIRY	1.00	0.00	2.67	**0.008**	1.00	1.01
crfd10VEG	1.00	0.00	-1.78	0.074	0.99	1.00
crfd10FRUIT	1.00	0.00	-0.22	0.825	1.00	1.00
crfd10TREATS	1.00	0.00	-0.74	0.460	0.99	1.00
crBODYSAT	3.39	0.80	5.18	**<0.001**	2.13	5.37
crkuSLEEPrec	0.22	0.16	-2.08	**0.038**	0.05	0.92
crACTIVTR	1.43	0.34	1.51	0.132	0.90	2.26
__ cons	0.01	0.00	-8.25	**<0.001**	0.00	0.02

Table 9-2 MULTIVARIABLE model with all 20 candidate variables including categorical food intake variables, using 25 d.f.

Log likelihood at final iteration	-304.26
Number of obs. (obesity outcomes)	2,345 (82)
LR chi2 (25)	102.52
Prob > chi2	**<0.001**
Pseudo R2	0.1442
AUROC	0.795

crTF2obese	Odds Ratio	Std. Error	z	P>|z|	Lower 95% CI for OR	Upper 95% CI for OR
crEthnicity	1.81	0.91	1.18	0.238	0.68	4.86
crArmpit	3.38	0.96	4.28	**<0.001**	1.94	5.91
crMumDegree	1.30	0.47	0.73	0.466	0.64	2.65
crONLYCHILD	1.48	0.49	1.19	0.234	0.78	2.82
crSMOKERS	0.81	0.22	-0.76	0.447	0.47	1.39
crMumoverw	2.99	0.72	4.53	**<0.001**	1.86	4.80
cr3Meals	1.11	0.30	0.38	0.706	0.65	1.87
crSnacking	1.24	0.31	0.86	0.388	0.76	2.03
crFastFood	1.03	0.52	0.05	0.957	0.38	2.79
crMILK	1.36	0.33	1.26	0.208	0.84	2.18
crJUICE	1.12	0.31	0.42	0.672	0.66	1.91
crSSB	1.14	0.30	0.51	0.613	0.68	1.90
crWHLGRcat < 1	1.27	0.36	0.84	0.399	0.73	2.20
crWHLGRcat ≥ 1	1.21	0.37	0.62	0.534	0.66	2.22
crDAIRcat < 1	1.35	0.42	0.96	0.337	0.73	2.49
crDAIRcat ≥ 1	2.50	1.04	2.20	**0.028**	1.11	5.63
crVEGcat < 2	0.54	0.17	-1.95	0.052	0.29	1.00
crVEGcat ≥ 2	0.43	0.16	-2.22	**0.027**	0.20	0.91
crFRUITcat < 2	1.01	0.29	0.02	0.983	0.57	1.77
crFRUITcat ≥ 2	1.10	0.47	0.22	0.829	0.48	2.52
crTREATScat < 2	0.29	0.25	-1.41	0.159	0.05	1.63
crTREATScat ≥ 2	0.20	0.18	-1.83	0.067	0.04	1.12
crBODYSAT	3.36	0.80	5.11	**<0.001**	2.11	5.35
crkuSLEEPrec	0.22	0.16	-2.09	**0.037**	0.05	0.91
crACTIVTR	1.47	0.35	1.63	0.103	0.92	2.35
cons	0.03	0.03	-3.64	**<0.001**	0.00	0.18

Table 9-3 MULTIVARIABLE model with all 20 candidate variables including dichotomous food intake variables, using 20 d.f.

Log likelihood at final iteration	-307.52
Number of obs. (obesity outcomes)	2,345 (82)
LR chi2 (20)	96.01
Prob > chi2	**<0.001**
Pseudo R2	0.1350
AUROC	0.795

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI for OR**	**Upper 95% CI for OR**
crEthnicity	1.61	0.81	0.94	0.345	0.60	4.30
crArmpit	3.46	0.98	4.40	**<0.001**	1.99	6.01
crMumDegree	1.32	0.48	0.77	0.439	0.65	2.68
crONLYCHILD	1.54	0.50	1.32	0.188	0.81	2.90
crSMOKERS	0.79	0.22	-0.84	0.398	0.46	1.36
crMumoverw	3.11	0.75	4.73	**<0.001**	1.94	4.97
cr3Meals	1.14	0.30	0.50	0.615	0.68	1.92
crSnacking	1.27	0.31	0.98	0.330	0.78	2.06
crFastFood	0.97	0.49	-0.05	0.958	0.36	2.64
crMILK	1.31	0.31	1.13	0.259	0.82	2.10
crJUICE	1.09	0.29	0.30	0.761	0.64	1.84
crSSB	1.13	0.29	0.47	0.640	0.68	1.87
crWHLGRAIN	0.91	0.25	-0.34	0.736	0.54	1.55
crDAIRY	0.50	0.17	-2.07	**0.038**	0.26	0.96
crVEG	1.41	0.41	1.18	0.240	0.80	2.50
crFRUIT	0.96	0.35	-0.12	0.904	0.47	1.95
crTREATS	0.69	0.18	-1.38	0.168	0.41	1.17
crBODYSAT	3.34	0.78	5.13	**<0.001**	2.11	5.29
crkuSLEEPrec	0.22	0.16	-2.10	**0.036**	0.05	0.90
crACTIVTR	1.41	0.33	1.46	0.143	0.89	2.24
__cons	0.01	0.01	-6.68	**<0.001**	0.00	0.04

The exploratory multivariable models are significant and achieve acceptable discrimination, but due to high levels of missing data the sample size is reduced, and models are over fitted for the number of obesity events kept in the model. Not all candidate variables make a significant contribution to the model; some could be eliminated.

Five dichotomous variables (Puberty/armpit hair, Mother's overweight, Dairy intake, Child's body satisfaction, Sleep recommendations), two categorical variables (Dairy intake, Vegetable intake) and one continuous variable (Dairy intake) are individually significant at $p \leq 0.05$ within the different versions of the model. Food intakes as continuous variables retain more information than the alternative categorical or dichotomous food intake variables but have O.R.s close to 1, indicating neither increased nor decreased risk of future obesity. The categorical and dichotomous variables are more useful for differentiating between higher and lower predicted risk or likelihood of an obesity outcome in a future risk score.

9.3.5 Single variable models

Results of fitting UNIVARIABLE logistic regression models for each dichotomous candidate variable are summarised in Table 9-4. Each row shows the results from a model with only that candidate variable as a single predictor.

In several cases the significance test for the odds ratio (P> |z|) and the overall test for the model (Prob < chi2) do not match. This is because the z statistic uses a Wald chi-square test and the test of the overall model uses a likelihood chi-square test, which can vary especially in smaller samples. Single variable models that achieve significance at $p \leq 0.05$ with one or both tests include Ethnicity, Puberty/armpit hair, Mother has degree, Mother's overweight, Dairy intake, Vegetable intake, Child's body satisfaction and Active travel.

The AUROC value is close to 0.50 for most dichotomous candidate variables, suggesting that they have little discrimination or predictive ability on their own. However, two dichotomous variables, Mother's overweight and Child's body satisfaction, reach 0.62; they could be useful predictors of future obesity.

Table 9-4 UNIVARIABLE logistic regression models - dichotomous variables

	Obs.	O.R.	Std. Error	z	P>|z|	Lower 95% CI for OR	Upper 95% CI for OR	Prob >chi2	AUROC
crEthnicity	3,752	2.25	0.74	2.49	**0.013**	1.19	4.27	0.24	0.52
crArmpit	4,114	2.45	0.51	4.27	**<0.001**	1.62	3.69	**<0.001**	0.55
crMumDegree	3,811	1.62	0.42	1.85	0.065	0.97	2.71	**0.05**	0.53
crONLYCHILD	3,639	1.14	0.30	0.50	0.619	0.68	1.92	0.62	0.51
crSMOKERS	3,594	1.25	0.23	1.23	0.217	0.88	1.79	0.22	0.52
crMumoverw	3,890	2.61	0.44	5.68	**<0.001**	1.87	3.63	**<0.001**	0.62
cr3Meals	3,339	1.33	0.26	1.48	0.139	0.91	1.95	0.15	0.53
crSnacking	4,114	1.30	0.21	1.63	0.102	0.95	1.79	0.10	0.53
crFastFood	3,393	1.03	0.41	0.07	0.947	0.47	2.23	0.94	0.50
crMILK	4,114	1.13	0.18	0.75	0.456	0.82	1.54	0.46	0.51
crJUICE	4,114	0.86	0.14	-0.95	0.343	0.62	1.18	0.35	0.52
crSSB	4,114	1.32	0.21	1.71	0.087	0.96	1.81	0.08	0.53
crWHLGRAIN	4,114	1.04	0.19	0.23	0.818	0.74	1.48	0.82	0.50
crDAIRY	4,114	0.61	0.14	-2.20	**0.028**	0.39	0.95	**0.04**	0.53
crVEG	4,114	1.48	0.30	1.95	0.051	1.00	2.19	**0.04**	0.53
crFRUIT	4,114	1.33	0.35	1.06	0.288	0.79	2.24	0.27	0.51
crTREATS	4,114	0.75	0.13	-1.67	0.095	0.53	1.05	0.10	0.53
crBODYSAT	3,319	2.89	0.52	5.88	**<0.001**	2.03	4.12	**<0.001**	0.62
crkuSLEEPrec	3,638	1.07	0.30	0.24	0.812	0.62	1.85	0.81	0.50
crACTIVTR	4,114	1.50	0.24	2.54	**0.011**	1.10	2.06	**0.01**	0.55

NOTE: Each row shows the results from a model with only that variable.

Results of fitting univariable logistic regression models for each continuous or categorical candidate variable (alternative food intake variables) are summarised in Table 9-5. Again, each row shows the results from a model with only that candidate variable as a single predictor.

Examination of the significance level of the likelihood test (Prob > chi2) in Table 9-5 shows that continuous and categorical intakes of vegetable and of fruit were significant at $p < 0.05$. Continuous and categorical intakes of dairy foods and treats were significant at $p < 0.25$. Two categories of intake of Vegetables and of Fruit (compared to the reference category of zero intake) were significant at $p < 0.05$ based on the Wald chi-square test for odds ratio ($P>|z|$). One category of intake of Treats and Energy dense snacks was also significant at $p < 0.05$ based on the Wald chi-square test.

In every case the categorical food intake variable was a better single predictor than the dichotomous food intake variable, based on a post-estimation results of the AUROC value, but whole grain intakes did not predict obesity outcomes.

Table 9-5 UNIVARIABLE logistic regression models - continuous and categorical variables

	Obs.	O.R.	Std. Error	z	P>|z|	Lower 95% CI for OR	Upper 95% C for ORI	Prob >chi2	AUROC
Continuous food intake variables									
crfd10WHLGR	4,114	1.00	0.00	-0.37	0.713	0.99	1.00	0.71	0.51
crfd10DAIRY	4,114	1.00	0.00	1.77	0.076	1.00	1.00	0.09	0.53
crfd10VEG	4,114	1.00	0.00	-2.06	**0.039**	0.99	1.00	**0.03**	0.55
crfd10FRUIT	4,114	1.00	0.00	-2.23	**0.025**	0.99	1.00	**0.02**	0.56
crfd10TREATS	4,114	1.00	0.00	-1.06	0.287	0.99	1.00	0.28	0.53
Categorical food intake variables									
crWHLGRcat	4,114							0.54	0.52
< 1 serving/day		0.82	0.15	-1.08	0.279	0.56	1.18		
≥ 1 serving/day		0.88	0.17	-0.66	0.510	0.60	1.29		
crDAIRcat	4,114							0.07	0.54
< 1 serving/day		0.84	0.16	-0.92	0.357	0.59	1.21		
≥ 1 serving/day		1.45	0.37	1.45	0.146	0.88	2.40		
crVEGcat	4,114							**<0.001**	0.57
< 2 servings/day		0.51	0.10	-3.51	**<0.001**	0.35	0.74		
≥ 2 servings/day		0.41	0.10	-3.72	**<0.001**	0.26	0.66		
crFRUITcat	4,114							**0.004**	0.56
< 2 servings/day		0.58	0.10	-3.20	**0.001**	0.42	0.81		
≥ 2 servings/day		0.53	0.15	-2.22	**0.026**	0.30	0.93		
crTREATScat	4,114							0.09	0.53
< 2 servings/day		0.40	0.22	-1.64	0.100	0.14	1.19		
≥ 2 servings/day		0.32	0.17	-2.14	**0.033**	0.11	0.91		

NOTE: Each row shows the results from a model with only that variable.

9.3.6 Multivariable models by subgroup

Results of fitting logistic regression models for each SUBGROUP of candidate variables (Child, family and socio-demographic variables or "Non-food" variables, Eating habits, Drinks frequency, Categorical Food intakes and Dichotomous Food intakes) are presented in Table 9-6, Table 9-7, Table 9-8, Table 9-9 and Table 9-10.

Examination of the significance level of the likelihood test (Prob > chi2) finds that models with non-food variables or with food intake variables were significant at $p \leq 0.05$. Their AUROC values indicate some useful predictive ability for these models. Subgroup models for eating habits and drinks frequency were not significant.

Table 9-6 SUBGROUP model with all 9 non-food variables, using 9 d.f.

Log likelihood at final iteration	-376.41
Number of observations	2,615
LR chi2 (9)	70.12
Prob > chi2	**<0.001**
Pseudo R2	0.0852
AUROC	0.73

crTF2obese	Odds Ratio	Std. Error	z	P>|z|	Lower 95% CI for OR	Upper 95% CI for OR
crEthnicity	1.30	0.63	0.55	0.585	0.50	3.38
crArmpit	2.81	0.72	4.04	**<0.001**	1.70	4.65
crMumDegree	1.42	0.46	1.09	0.274	0.76	2.67
crONLYCHILD	1.28	0.40	0.78	0.435	0.69	2.37
crSMOKERS	0.88	0.21	-0.54	0.590	0.55	1.41
crMumoverw	2.53	0.54	4.34	**<0.001**	1.66	3.85
crBODYSAT	2.77	0.59	4.78	**<0.001**	1.82	4.20
crkuSLEEPrec	0.65	0.28	-1.00	0.317	0.28	1.52
crACTIVTR	1.58	0.34	2.14	**0.033**	1.04	2.40
_cons	0.01	0.00	-13.32	**<0.001**	0.00	0.02

Table 9-7 SUBGROUP model with all 3 eating habit variables, using 3 d.f.

Log likelihood at final iteration	-509.65
Number of observations	3,298
LR chi2 (3)	4.97
Prob > chi2	0.1741
Pseudo R2	0.0049
AUROC	0.56

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI for OR**	**Upper 95% CI for OR**
cr3Meals	1.24	0.25	1.05	0.292	0.83	1.84
crSnacking	1.40	0.28	1.71	0.088	0.95	2.06
crFastFood	0.93	0.37	-0.18	0.859	0.42	2.05
_cons	0.03	0.00	-22.66	**<0.001**	0.02	0.04

Table 9-8 SUBGROUP model with all 3 drinks frequency variables, using 3 d.f.

Log likelihood at final iteration	-699.41
Number of observations	4,114
LR chi2 (3)	4.81
Prob > chi2	0.1860
Pseudo R2	0.0034
AUROC	0.55

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI for OR**	**Upper 95% CI for OR**
crMILK	1.12	0.18	0.69	0.489	0.82	1.53
crJUICE	0.83	0.14	-1.15	0.250	0.60	1.14
crSSB	1.35	0.22	1.83	0.067	0.98	1.85
_cons	0.04	0.01	-19.13	**<0.001**	0.03	0.05

Table 9-9 SUBGROUP model with 5 categorical food intake variables, using 10 d.f.

Log likelihood at final iteration	-686.97
Number of observations	4114
LR chi2 (10)	29.69
Prob > chi2	**0.0010**
Pseudo R2	0.0212
AUROC	0.62

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI for OR**	**Upper 95% CI for OR**
crWHLGRcat < 1	0.95	0.18	-0.29	0.773	0.65	1.38
crWHLGRcat ≥ 1	0.98	0.19	-0.09	0.929	0.67	1.45
crDAIRcat < 1	1.01	0.19	0.05	0.962	0.70	1.46
crDAIRcat ≥ 1	1.63	0.42	1.87	0.062	0.98	2.71
crVEGcat < 2	0.59	0.12	-2.70	**0.007**	0.40	0.86
crVEGcat ≥ 2	0.48	0.12	-2.93	**0.003**	0.30	0.79
crFRUITcat < 2	0.67	0.12	-2.25	**0.024**	0.47	0.95
crFRUITcat ≥ 2	0.61	0.18	-1.66	0.097	0.35	1.09
crTREATScat < 2	0.49	0.28	-1.27	0.203	0.16	1.47
crTREATScat ≥ 2	0.39	0.21	-1.73	0.084	0.13	1.14
_cons	0.21	0.12	-2.80	**0.005**	0.07	0.63

Table 9-10 SUBGROUP model with 5 dichotomous food intake variables, using 5 d.f.

Log likelihood at final iteration	-695.70
Number of observations	4114
LR chi2 (5)	12.24
Prob > chi2	**0.0316**
Pseudo R2	0.0087
AUROC	0.58

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI for OR**	**Upper 95% CI for OR**
crWHLGRAIN	1.05	0.19	0.29	0.772	0.74	1.50
crDAIRY	0.60	0.13	-2.27	**0.023**	0.39	0.93
crVEG	1.46	0.29	1.88	0.061	0.98	2.16
crFRUIT	1.30	0.35	0.98	0.326	0.77	2.21
crTREATS	0.75	0.13	-1.62	0.104	0.53	1.06
_cons	0.05	0.02	-8.14	**<0.001**	0.02	0.10

9.3.7 Stepwise removal/addition of candidate variables

Results from stepwise addition were the same as the results from stepwise removal, with identical sets of candidate variables added or kept at $p \leq 0.05$, and the same sets of candidate variables added or kept at $p \leq 0.25$. (A smaller significance level can miss potentially important predictors, picked up when a larger significance level is applied, although using a larger significance level may also keep less useful predictors in the model. A balance must be found.)

In the **multivariable model with categorical food variables**, the 5 retained candidate variables at $p \leq 0.05$ were Child's body satisfaction, Puberty/armpit hair, Mother's overweight, Categorical Vegetable intake and Meets sleep recommendation. At $p \leq 0.25$ a further 5 candidate variables were retained; Categorical Dairy intake, Only child, Categorical Treat intake, Active travel and Milk as a drink frequency.

In the **multivariable models with either continuous or dichotomous food variables**, the 5 retained candidate variables at $p \leq 0.05$ were Child's body satisfaction, Puberty/armpit hair, Mother's overweight, Continuous/Dichotomous Dairy intake and Meets sleep recommendation. At $p \leq 0.25$ a further 5 candidate variables were retained; Continuous/Dichotomous Vegetable intake, Snacking between meals, Only child, Active travel and Milk as a drink frequency.

In the **non-food sub-group model,** the 4 retained candidate variables at $p \leq 0.05$ and at $p \leq 0.25$ were Mother's overweight, Puberty/armpit hair, Active travel and Child's body satisfaction.

In the **eating habits sub-group model** all 3 variables were removed at $p \leq 0.05$ and only Snacking between meals was retained at $p \leq 0.25$.

In the **drinks frequency sub-group model** all 3 variables were removed at $p \leq 0.05$ but both Juice frequency and SSB frequency were retained at $p \leq 0.25$.

In all the **food intake sub-group models** Whole grain intake was removed at $p \leq 0.05$ and $p \leq 0.25$. At least one of the alternative variables (continuous, categorical or dichotomous) for Dairy intake, Vegetable intake, Fruit intake and Energy Dense Treats intake was kept in the corresponding sub-group model.

9.3.8 Missing observations

Several pre-specified candidate variables had missing observations, but none had more than 20% missing after imputation. See Chapter 8, Table 6.

Four dichotomous candidate variables had more missing observations than the frequency observed in one of their categories.

- Child's ethnic background (135 Non-white but 365 missing)
- Only child (398 Only children but 475 missing)
- Fast Food once a week or more (187 Yes but 721 missing)
- Meets sleep recommendation (363 No but 476 missing)

The first three were not kept in any model after stepwise removal or stepwise addition. Meets sleep recommendation was retained in the full models after stepwise removal or stepwise addition at $p \leq 0.05$ yet was not kept in the non-food sub-group model. All four variables were eliminated from further model testing in the ALSPAC sample to avoid introducing bias with spurious associations/predictions. Note that if missing had not been an issue these variables would have been tested further – they were not ruled out altogether.

Two dichotomous variables retained in the full models with stepwise removal or stepwise addition also had missing observations, but the number of missing observations did not exceed the frequency in the smallest category.

- Mother's overweight (224 missing but 1,323 overweight)
- Child's body satisfaction (795 missing but 914 dissatisfied)

Alternative variables with missing as an extra category were generated. Children with missing values for Mother's overweight had significant and higher O.R. for future obesity than children with Mothers who were overweight. Children with missing values for Child's body satisfaction were at a reduced risk of future obesity. This suggests that observations were not missing at random. Although the variables with missing as a category kept observations in the models, they used extra d.f. and reduced the model's overall predictive performance. The variables with missing as a category were not employed in further model testing.

9.3.9 Best predictors with reduced and interim models

Logistic regression model results for the 16 candidate variables that did not have unacceptable levels of missing are summarised in Table 9-11. The "best predictors" are highlighted and include four non-food variables (Puberty/armpit hair, Mother's overweight, Child's body satisfaction and Active travel) and two food intake variables (Categorical Vegetable intake and Dichotomous Dairy intake). Each variable achieved significance at p = 0.05 as a single predictor of future obesity and was kept in the sub-group model as well as in the full models with stepwise removal/addition. These six "best predictors" were included in the REDUCED model. (See Table 9-12.) The REDUCED model has acceptable discrimination with an AUROC value of 0.74.

Five candidate variables were removed from full models and their sub-group models with stepwise removal/addition: Mother has degree, Smokers in household, Eats 3 meals or more a day, Whole grain intakes and the Dichotomous Fruit intake variable. Although one of these variables, Mother has degree, was significant as a single predictor of future obesity it did not have much predictive value in the multivariable models. All five variables were eliminated from further model testing.

The predictive value of other candidate variables was less clear cut. Snacks between meals was not significant as a single predictor of future obesity and was removed from full models, only retained in the Eating habits sub-group model at p = 0.25. Similarly Drinks frequency variables did not predict obesity outcomes on their own, yet Milk as a drink frequency was a weak yet significant predictor in the full models and both Juice frequency and SSB frequency were kept in their sub-group model at p = 0.25. Models suggested that Fruit intake and Energy dense treats intake could have some predictive value. Unlike Dichotomous Fruit intake, Categorical Fruit intake was significant as a single predictor of future obesity at p = 0.05, but was kept only in the sub-group model, not in the full model. Energy dense treats intake, whilst not significant as a single predictor, made a small predictive contribution in multivariable models, being kept in sub-group models *and* the full model with categorical food intakes at p = 0.25, so was preferred to the Eating habit Snacks between meals.

Table 9-11 Summary of each candidate variable's performance in each type of model

MODEL		Single predictor			Sub-group		Full, cont. food		Full, categ. food		Full, dichot. food	
Variable type	Candidate variable	OR	Sig at 0.05?	c-stat	Kept at 0.05?	Kept at 0.25?	Kept at 0.05?	Kept at 0.25?	Kept at 0.05?	Kept at 0.25?	Kept at 0.05?	Kept at 0.25?
NON FOODS	crArmpit	2.44	Sig	0.55	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	crMumDegree	1.62	Sig	0.53								
	crSMOKERS	1.25	NS	0.52								
	crMumoverw	2.61	Sig	0.62	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	crBODYSAT	1.06	Sig	0.62	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	crACTIVTR	0.41	Sig	0.55	Yes	Yes		Yes		Yes		Yes
EATING HABITS	cr3Meals	1.33	NS	0.53								
	crSnacking	1.30	NS	0.53		Yes						
DRINKS	crMILK	1.13	NS	0.51				Yes		Yes		Yes
FREQUENCY	crJUICE	0.86	NS	0.52		Yes						
	crSSB	1.32	NS	0.53		Yes						
CONTINUOUS	crfd10WHLGRAIN	1.00	NS	0.51								
FOOD INTAKE	crfd10DAIRY	1.00	NS	0.53	Yes	Yes	Yes	Yes				
VARIABLES	crfd10VEG	1.00	Sig	0.55		Yes		Yes				
	crfd10FRUIT	1.00	Sig	0.56	Yes	Yes						
	crfd10TREATS	1.00	NS	0.53								
CATEGORICAL	crWHLGRcat	0.82 & 0.88	NS	0.52								
FOOD INTAKE	crDAIRcat	0.84 &1.45	NS	0.54		Yes				Yes		
VARIABLES	crVEGcat	0.51 & 0.41	Sig	0.57	Yes	Yes			Yes	Yes		
	crFRUITcat	0.58 & 0.53	Sig	0.56	Yes	Yes						
	crTREATcat	0.40 & 0.32	NS	0.53		Yes				Yes		
DICHOTOMOUS	crWHLGRAIN	1.04	NS	0.50								
FOOD INTAKE	crDAIRY	0.61	Sig	0.53	Yes	Yes					Yes	Yes
VARIABLES	crVEG	1.48	NS	0.53	Yes	Yes						Yes
	crFRUIT	1.33	NS	0.51								
	crTREAT	0.75	NS	0.53		Yes						

Table 9-12 REDUCED model with 6 "best predictors", using 7 d.f.

Log likelihood at final iteration	-468.01
Number of obs. (obesity outcomes)	3,171 (121)
LR chi2 (7)	91.68
Prob > chi2	**<0.001**
Pseudo R2	0.0892
AUROC	0.7381

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI for OR**	**Upper 95% CI for OR**
crArmpit Yes armpit hair	3.01	0.70	4.76	**<0.001**	1.91	4.75
crMumoverw Mum overweight	2.37	0.45	4.53	**<0.001**	1.63	3.45
crBODYSAT Unsatisfied	2.69	0.51	5.20	**<0.001**	1.85	3.90
crACTIVTR No active travel	1.58	0.30	2.41	**0.016**	1.09	2.30
crVEGcat < 2 servings Veg/day	0.58	0.14	-2.34	**0.019**	0.36	0.91
crVEGcat ≥ 2 servings Veg/day	0.43	0.13	-2.89	**0.004**	0.24	0.76
crDAIRY <1 serving Dairy/day	0.53	0.14	-2.41	**0.016**	0.31	0.89
_cons	0.04	0.01	-9.48	**<0.001**	0.02	0.07

Candidate variables were added to the REDUCED model singly and in combination. A model with eleven predictors (Puberty/armpit hair, Mother's overweight, Child's body satisfaction, Active travel, Categorical Vegetable intake, Categorical Fruit intake, Dichotomous Dairy food intake, Dichotomous Energy dense treats intake and Milk, Juice and SSB frequency) achieved the best predictive performance, but was over fitted for the 121 obesity outcomes in the 3,171 observations kept by the model, using 13 d.f. The variables causing list wise deletion of missing observations were Mother's overweight and Child's body satisfaction, two of the "best predictors". Using alternative variables with missing as a category kept all observations but reduced the model's predictive performance. Total milk intake (servings/day, measured by the 3 day diet diary) was a stronger single predictor of future obesity than Milk as a drink frequency and contributed more to the model's overall predictive performance, so was used in preference. Removing Fruit intake (correlated with Vegetable intake) and Juice frequency made marginal reductions to the model's LR chi2 or Pseudo R2, so both variables were eliminated. This gave an INTERIM model with nine predictors, presented in Table 9-13, which has a larger LR chi2 and Pseudo R2 and slightly better discrimination than the REDUCED model. The model is not over fitted, based on a calculated shrinkage value of 0.91.

Table 9-13 INTERIM model with 9 predictors, using 10 d.f.

Log likelihood at final iteration	-460.76
Number of obs. (obesity outcomes)	3,171 (121)
LR chi2 (10)	106.17
Prob > chi2	**<0.001**
Pseudo R2	0.1033

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI for OR**	**Upper 95% CI for OR**
crArmpit Yes armpit hair	2.98	0.70	4.67	**<0.001**	1.88	4.71
crMumoverw Mum overweight	2.36	0.45	4.45	**<0.001**	1.62	3.44
crBODYSAT Unsatisfied	2.69	0.51	5.16	**<0.001**	1.85	3.91
crACTIVTR No active travel	1.60	0.31	2.44	**0.015**	1.10	2.33
crVEGcat < 2 servings Veg/day	0.63	0.15	-1.95	0.051	0.39	1.00
crVEGcat ≥ 2 servings Veg/day	0.47	0.14	-2.54	**0.011**	0.26	0.84
crDAIRY <1 serving Dairy/day	0.55	0.15	-2.23	**0.026**	0.33	0.93
crMILKSERVE Milk intake is zero	2.16	0.50	3.32	**0.001**	1.37	3.41
crTREATS ≥ 2 servings/day	0.70	0.15	-1.67	0.094	0.46	1.06
crSSB ≥ 1 serving/day	1.32	0.27	1.36	0.174	0.88	1.97
_cons	0.03	0.01	-8.45	**<0.001**	0.01	0.07

9.3.10 Purposefully selected variables

The purposeful selection strategy identified the same set of six "best" predictors in a reduced model as the previous approach, and the same variables were removed altogether (Smokers in household, Eats 3 meals or more a day, Whole grain intakes and Fruit intakes).

The results of each purposeful selection step are described below, and the models used in Steps 2 to 5 of the purposeful selection of variables are shown in Appendix H.

Step 1

Univariable logistic regression analyses of 20 pre-specified candidate variables and obesity outcomes at age 13.5 years were run in the derivation sample. (See Table 9-4 and Table 9-5 again) Six UNIVARIABLE models had Wald statistics with a p-value > 0.25 for their single variable. These variables were Only child, Fast food, Milk as a drink frequency, Juice frequency, Whole grain intake and Meets sleep recommendation.

Examination of contingency tables for each of the 20 categorical candidate variables showed that none had zero frequency in a category (which can cause a failure of the logistic regression software to converge and generate an odds ratio). However, 4 variables had levels of missing that exceeded the frequency in one category, so were not tested further (Ethnic background, Only child, Fast food, Meets sleep recommendation). Milk as a drink frequency, Juice frequency and Whole grain intake were set aside until Step 4.

Step 2

The thirteen remaining variables were candidates for an initial multivariable model. If applicable, missing observations for these variables and the set aside variables were dropped from the derivation sample. This left 2,496 observations and 88 obesity "events" in the dataset, used for each subsequently fitted model, thereby allowing a fair comparison of each model's predictive performance.

In the fitted MULTIVARIABLE purposeful selection model seven variables had p values >0.05 for their Wald statistic, so made little contribution to the initial model (Mother has degree, Smokers, 3 meals a day, Snacking, SSB frequency, Fruit intake, Treats intake). These seven variables were set aside until Step 4.

Step 3

The six remaining "best" predictors (Puberty/armpit hair, Mother's overweight, Dairy intake, Vegetable intake, Child's body satisfaction and Active travel) were significant at or very close to $p \leq 0.05$, so were fitted in a REDUCED purposeful selection model, which had a LR of 89.17 using 8 d.f. The AUROC value was 0.771 indicating acceptable discrimination. The values of

their estimated O.R.s in the reduced model were similar to their values in the multivariable model, suggesting that the set aside variables did not greatly adjust the effect of the variables in the smaller model.

Step 4

The ten set aside candidate variables were added to the reduced model one at a time. None were significant at 0.05 based on their Wald statistic p-value, but four variables increased the LR to above 90, suggesting that they contributed to the model. Treats intake contributed the most, followed by Milk as a beverage frequency, then Snacking and SSB frequency. With only 88 obesity events and 8 d.f. already used in the reduced model, only the dichotomous variable for Treats intake was added to avoid overfitting. This gave a PRELIMINARY MAIN EFFECTS purposeful selection model.

Step 5

As one category of categorical Dairy intake was not significant in the reduced model compared to the reference category, the dichotomous alternative was used instead. Adding dichotomous Treats intake and Milk as a drink frequency produced the best MAIN EFFECTS purposeful selection model within the available d.f. and gave a modest improvement in discrimination (AUROC = 0.774). The smaller sample size and fewer degrees of freedom restricted the total number of variables in the main effects model to eight and left no scope to add interaction terms.

9.3.11 Food intake variables with changed reference categories

The direction of influence of candidate variables on obesity outcomes was as hypothesised, with two exceptions, as seen in the INTERIM model. Higher not lower intakes of Dairy food (not including milk) and lower not higher intakes of Energy dense treats predicted an increased likelihood of future obesity in the derivation sample. Dichotomous variables were not re-coded, but reference categories were changed prior to running further logistic regression models.

Both variables were checked and the Dairy food variables crfd10DAIRY, crDAIRYcat and crDAIRY were amended as follows.

Dairy food intake (not including milk) is based on total intake of Yoghurt + Cheese + Milk based sauces in g/day, representative foods from the 3 day diet diary at 10 years 6 months. As it was not known if the Puddings and ice-creams intake variable in the ALSPAC dataset was dairy based or non-dairy (most likely both) this measure was not included with Dairy food. As a consequence, the Dairy food variables may not fully capture dairy intake. A cut-off of 1 serving a day, based on a 125g serving of yoghurt, was used for categorical Dairy food intake variables. Only 406 children (<10%) of the 4,114 children in the derivation sample consumed 1 serving/day or more of Dairy foods, although intake ranged from 0g/day to 450g/day. Mean intakes of yoghurt, cheese and milk based sauces in the derivation sample were 32g/day, 12g/day and 2g/day respectively, giving a mean Dairy food intake of 46g/day, or just over one third of a 125g serving. We concluded that a uniform serving size of 125g was too high. Instead new Dairy food variables were created, based on intake in servings/day, using serving sizes of 125g, 30g and 60g for yoghurt, cheese and milk based sauces respectively. The recalibrated variables were named crfd10DAIRYSERV, crDAIRYSERVEcat and crDAIRYSERVE. Serving sizes for yoghurt and cheese are taken from the Eat well plate, the serving size for milk based sauces is based on standard recipes, assuming a modest serving. This redistributed the frequency in each category. Now 1,146 (28%) of the 4,114 children in the derivation sample consumed 1 serving/day or more of the included Dairy foods, which seems more realistic. Only 262 of the 4,114 children (6.4%) consumed 2 servings/day or more, so a cut-off of 1 serving a day was used. (See Table 9-14.) Higher intakes still predicted an increased likelihood of future obesity, compared with Dairy food servings < 1 /day.

Table 9-14 Revised Dairy food variables, in servings/day

Variable	Scale	Categories (frequency) / Range, mean and SD in derivation sample	d.f.	No. obs in derivation sample (missing)
Dairy food servings, not including milk 1 serving cheese = 30g 1 serving yogurt = 125g 1 serving milk based sauces = 60g	Continuous	0 to 5.7 servings/day, mean 0.70 servings/day, SD 0.74 servings/day	1	4,114 (0%)
	Categorical	Zero (1,041) < 1 serving a day (1,927) ≥ 1 serving a day (1,146)	2	4,114 (0%)
	Dichotomous	< 1 serving a day ≥ 1 serving a day	1	4,114 (0%)

9.3.12 Interaction terms

The INTERIM model had nine predictors (Puberty/armpit hair, Mother's overweight, Child's body satisfaction, Active travel, Categorical Vegetable intake, Dichotomous Dairy servings, Dichotomous Milk servings, Dichotomous Energy dense treats intake and SSB frequency) (See Table 9-13.) Some pairs of these predictors were already shown to be correlated (See section 9.3.3). In case any variable modified the effect of a second variable on obesity outcomes, interactions between pairs of candidate variables in the INTERIM model were investigated. Only three interactions were significant at $p \leq 0.05$ (95% CIs for their O.R.s did not cross 1) when added to the INTERIM model:

- Mother's overweight and SSB frequency, $p>|z| = 0.03$
- Mother's overweight and Vegetable intake >= 2 servings/day, $p>|z| = 0.01$
- Vegetable intake categories and SSB frequency, $p>|z| = 0.003$ and 0.03

The percentage of obesity outcomes in the derivation sample for each sub-category of these interacting pairs of variables are shown in Table 9-15. They illustrate how the effect of one variable on the outcome may shift with changes in the second variable of the interacting pair. Note that numbers in some sub-categories are small and not all interactions for each pair were significant when tried in the multivariable model.

Mother's overweight x SSB frequency.

Independently, having a mother who was overweight, or consuming ≥ 1 serving/day of sugar sweetened beverage (cola, fizzy drinks and sugar sweetened fruit drinks) a day, increased a child's risk of future obesity. In a simple logistic regression model with only these two interacting variables as predictors, the interaction was significant at $p \leq 0.05$. Some of the protective effect of a child's mother not being overweight was lost if the child consumed 1 or more servings/day of SSBs.

Mother's overweight x Vegetable intake category.

Having a mother who was overweight or consuming no vegetables (on the days surveyed), increased a child's risk of future obesity. In a simple logistic regression model, the interactions between the two were not significant at $p \leq 0.05$ and 95% CIs for the O.R.s crossed 1. Even if the child's mother was overweight, children who ate vegetables seem to have a reduced risk of future obesity, but in the interim model the interaction was only significant in the highest vegetable intake category.

Vegetable intake categories x SSB frequency.

Consuming no vegetables, or consuming ≥1 serving/day of sugar sweetened beverages, increased a child's risk of future obesity. In a simple logistic regression model, the interactions were significant at $p \leq 0.05$. Some of the protective effect of eating vegetables seemed to be lost if the child had 1 or more servings/day of SSBs.

Adding all three interaction terms to the model exceeds the available d.f. creating an over fitted and complex model that is hard to understand. Only the interaction which most improved model performance, Vegetable intake categories and SSB frequency, was added to the model, using another 2 degrees of freedom.

Table 9-15 Obesity outcomes by sub-category of interacting candidate variables

Sub-categories	Total obs.	Not obese at TF2	Obese at TF2	% Obese at TF2
Mother's overweight x SSB frequency: 3,890 observations in 4 sub-categories				
Mum not o/w x SSB < 1 serving/day	1,251	1,230	21	1.7%
Mum not o/w x SSB ≥ 1 serving/day	1,316	1,272	44	3.3%
Mum o/w x SSB < 1 serving/day	531	494	37	7.0%
Mum o/w x SSB ≥ 1 serving/day	792	745	47	5.9%
Mother's overweight x Veg. intake category: 3,890 observations in 6 sub-categories				
Mum not o/w x zero Veg	336	324	12	3.6%
Mum not o/w x Veg < 2 servings/day	1,533	1,496	37	2.4%
Mum not o/w x Veg ≥ 2 servings/day	698	682	16	2.3%
Mum o/w x zero Veg	232	206	26	11.2%
Mum o/w x Veg < 2 servings/day	792	745	47	5.9%
Mum o/w x Veg ≥ 2 servings/day	299	288	11	3.7%
Veg. intake categories x SSB frequency: 4,114 observations in 6 sub-categories				
Zero Veg x SSB < 1 serving/day	228	207	21	9.2%
Zero Veg x SSB ≥ 1 serving/day	376	354	22	5.9%
Veg < 2 serv./day x SSB < 1 serv./day	1,108	1,074	34	3.1%
Veg < 2 serv./day x SSB ≥ 1 serv./day	1,352	1,293	59	4.4%
Veg ≥ 2 serv./day x SSB < serv./day	546	535	11	2.0%
Veg ≥ 2 serv./day x SSB ≥ 1 serv./day	504	483	21	4.2%

9.3.13 The final model

The best performing model in the derivation sample includes nine predictors (Puberty/armpit hair, Mother's overweight, Child's body satisfaction, Active travel, Categorical Vegetable intake, Dichotomous Dairy servings, Dichotomous Milk servings, Dichotomous Energy dense treats intake and SSB frequency) with 1 interaction term (Categorical Vegetable intake and SSB frequency).

This FINAL model is presented with the reference categories for each predictor in Table 9-16. It keeps 3,171 observations from 4,114 in the derivation sample and has 121 obesity outcomes at follow-up. The FINAL model offers only a marginal improvement in discrimination (AUROC = 0.755) but has a larger LR chi2 and Pseudo R2 than the INTERIM model.

In the INTERIM model, eating vegetables reduces the risk of future obesity (O.R.s < 1) compared with eating no vegetables, and consuming one or more SSB a day increases the risk (O.R. > 1) compared with consuming less than one SSB a day. In the FINAL model with an interaction term between these two variables, at first glance it seems that not only does eating vegetables reduce the risk but consuming one or more SSB a day reduces the risk too. However, the interaction terms must also be considered. There we see that consuming one or more SSB a day *even if a child eats vegetables* increases the risk of future obesity, with the highest Odds Ratios of any predictors in the model.

Based on the exact shrinkage value of 0.895, the FINAL model is over fitted for the number of obesity outcomes in the derivation sample. However, if Dichotomous Energy dense treats (the candidate variable with the highest p value for the z statistic, not significant at $p < 0.05$) was removed so that the model was no longer over fitted, the comparative measures of predictive performance were reduced (LR chi2 = 111.82, Pseudo R2 = 0.1088, AUROC = 0.7528). This implies that Dichotomous Energy dense treats variable made a small but important predictive contribution and should be kept in the FINAL model.

Table 9-16 FINAL model with 9 predictors and 1 interaction term, using 12 d.f., in the DERIVATION sample

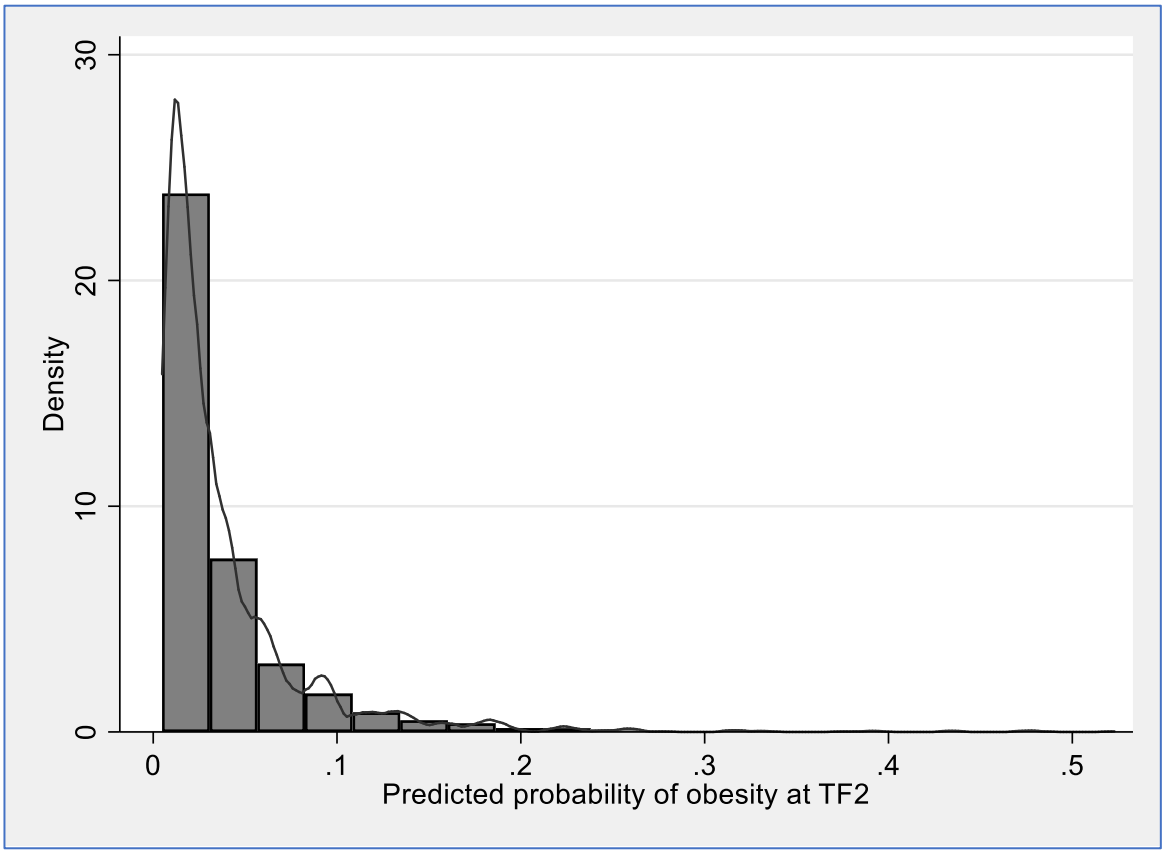
Log likelihood at final iteration	-456.24
Number of obs. (obesity outcomes)	3,171 (121)
LR chi2 (12))	115.22
Prob > chi2	**<0.001**
Pseudo R2	0.1121
AUROC	0.7549

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI for OR**	**Upper 95% CI for OR**
crArmpit No armpit hair	Ref					
crArmpit Armpit hair	2.81	0.66	4.40	**<0.001**	1.78	4.46
crMumoverw Mum not overweight	Ref					
crMumoverw Mum overweight	2.31	0.45	4.34	**<0.001**	1.58	3.37
crBODYSAT Child satisfied	Ref					
crBODYSAT Child unsatisfied	2.63	0.50	5.04	**<0.001**	1.81	3.83
crACTIVTR Active travel to school	Ref					
crACTIVTR No Active travel	1.58	0.30	2.38	**0.017**	1.08	2.30
crVEGcat Veg intake = 0/day	Ref					
crVEGcat < 2 servings Veg/day	0.24	0.09	-3.71	**<0.001**	0.11	0.51
crVEGcat ≥ 2 servings Veg/day	0.21	0.10	-3.35	**0.001**	0.08	0.52
crSSB < 1 serving SSB/day	Ref					
crSSB ≥ 1 serving SSB/day	0.44	0.18	-1.99	**0.047**	0.20	0.99
crVEGcat#crSSB < 2 servings Veg/day#≥1 serving SSB/day	4.47	2.24	2.99	**0.003**	1.67	11.92
crVEGcat##SSB ≥ 2 servings Veg/day#≥1 serving SSB/day	3.71	2.27	2.14	**0.032**	1.12	12.30
crDAIRYSERVE < 1 Dairy serving/day	Ref					
crDAIRYSERVE ≥ 1 Dairy serving/day	1.55	0.31	2.18	**0.029**	1.05	2.30
crMILKSERVE Milk servings > 0/day	Ref					
crMILKSERVE Milk servings = 0/day	2.19	0.51	3.35	**0.001**	1.38	3.46
crTREATS ≥ 2 Treats/day	Ref					
crTREATS < 2 Treats/day	1.50	0.32	1.89	0.059	0.99	2.27
_cons	0.02	0.01	-10.1	**<0.001**	0.01	0.05

9.3.14 Model predictions in the derivation sample

Post-estimation commands were used to generate the predicted probability that a child would be classified as obese at ~ 13.5 years, based on the final model with one interaction term. This fitted model used 3,171 observations from the derivation sample. Probabilities ranged between 0.005 and 0.523, Mean 0.038, SD 0.045. The median was 0.024, indicating that distribution was highly skewed (skewness 3.9) towards lower probabilities in this sample, as shown in Figure 9-1.

Figure 9-1 Distribution of predicted probability of obesity at TF2



As the predicted probability of obesity reduced, reassuringly the percentage of observed obesity outcomes also reduced, demonstrating that the model has some predictive value. (See Table 9-17.)

In total 168 children in the derivation sample of 4,114 were classified as obese by TF2. Due to missing observations for Mother's overweight and Child's body satisfaction, 943 children were omitted from the model. Hence, they had missing values for the predicted probability of future obesity. It was observed that 47 (5%) of them were obese by TF2, a larger proportion than among those children kept in the model (3.8%) or in the derivation sample (4.1%).

Table 9-17 Predicted probability of obesity and observed outcomes at TF2

Predicted probability of obesity at TF2	No. children	Observed obesity outcomes at TF2	Approx. %
>=0.5	1	1	100%
>=0.4	6	4	67%
>=0.3	14	7	50%
>=0.2	38	13	34%
>=0.1	215	32	15%
>=0.05	700	69	10%
<0.038 (MEAN)	2,190	44	2%
<0.024 (MEDIAN)	1,582	21	1%
Derivation sample	4,114	168	4.1%
Kept in model	3,171	121	3.8%
Missing	943	47	5.0%

9.3.15 Model performance

The FINAL model with one interaction term was run in the derivation sample (See Table 9-16), the validation sample (See Table 9-18) and in the combined cohort (model not shown).
In the smaller validation sample there are far fewer obesity outcomes or events for the model to predict. The whole model achieves significance at $p \leq 0.05$ and the AUROC touches 0.70, indicating acceptable discrimination, but LR chi2 and Pseudo R2 values are smaller. Most predictors in the model have the same direction of influence in the validation set as in the derivation set, except for the interaction term, Categorical Vegetable intake x SSB frequency. Only two predictors (Mother's overweight and Child's body satisfaction) were significant at $p < 0.05$ based on the Wald chi-square test for their odds ratio (P>|z|). The other predictors in the model have O.R.s with 95% CI that cross 1.

Measures of the model's performance in each sample and the combined cohort are summarised in Table 9-19 . Sample sizes and the number of obesity outcomes in each dataset and kept in the model are also presented for comparison.

Shrinkage

The prediction model was on the cusp of being over fitted in the derivation sample, with a shrinkage factor of 0.90. In the smaller validation sample, with fewer obesity outcomes, the shrinkage factor was 0.61. In the combined cohort with more observations the shrinkage factor rose to 0.91, so acceptable but still indicative of noise or error that may make the model less stable in a different context.

Brier score

This measure of overall performance was 0.03 in the derivation sample, far less than 0.25 or no better than chance, indicating that the fitted model is informative. In the validation sample the model's Brier score was 0.04, which shows a slightly higher level of disagreement between the predictions and the observed outcomes, but it is still informative.

Table 9-18 FINAL model with 9 predictors and 1 interaction term, using 12 d.f., in the VALIDATION sample

Log likelihood at final iteration	-190.02
Number of obs. (obesity outcomes)	1,077 (51)
LR chi2 (12))	30.62
Prob > chi2	**0.0023**
Pseudo R2	0.0746
AUROC	0.696

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI**	**Upper 95% CI**
crArmpit No armpit hair	Ref					
crArmpit Armpit hair	1.49	0.69	0.87	0.387	0.60	3.70
crMumoverw Mum not overweight	Ref					
crMumoverw Mum overweight	2.02	0.60	2.38	**0.017**	1.13	3.61
crBODYSAT Child satisfied	Ref					
crBODYSAT Child unsatisfied	3.19	0.95	3.90	**<0.001**	1.78	5.73
crACTIVTR Active travel to school	Ref					
crACTIVTR No Active travel	1.26	0.37	0.78	0.435	0.71	2.24
crVEGcat Veg intake = 0/day	Ref					
crVEGcat < 2 servings Veg/day	1.23	0.97	0.26	0.795	0.26	5.77
crVEGcat ≥ 2 servings Veg/day	1.54	1.29	0.52	0.603	0.30	7.90
crSSB < 1 serving SSB/day	Ref					
crSSB ≥ 1 serving SSB/day	1.22	1.06	0.23	0.817	0.22	6.69
crVEGcat#crSSB < 2 servings Veg/day#≥1 serving SSB/day	1.01	0.96	0.01	0.989	0.16	6.45
crVEGcat##SSB ≥ 2 servings Veg/day#≥1 serving SSB/day	0.39	0.44	-0.84	0.400	0.04	3.52
crDAIRYSERVE < 1 Dairy serving/day	Ref					
crDAIRYSERVE ≥ 1 Dairy serving/day	1.03	0.34	0.09	0.927	0.54	1.96
crMILKSERVE Milk servings > 0/day	Ref					
crMILKSERVE Milk servings = 0/day	1.86	0.73	1.57	0.116	0.86	4.03
crTREATS ≥ 2 Treats/day	Ref					
crTREATS < 2 Treats/day	1.63	0.51	1.56	0.118	0.88	3.01
_cons	0.01	0.01	-5.55	**<0.001**	0.00	0.06

Table 9-19 Summary of FINAL model performance in each sample

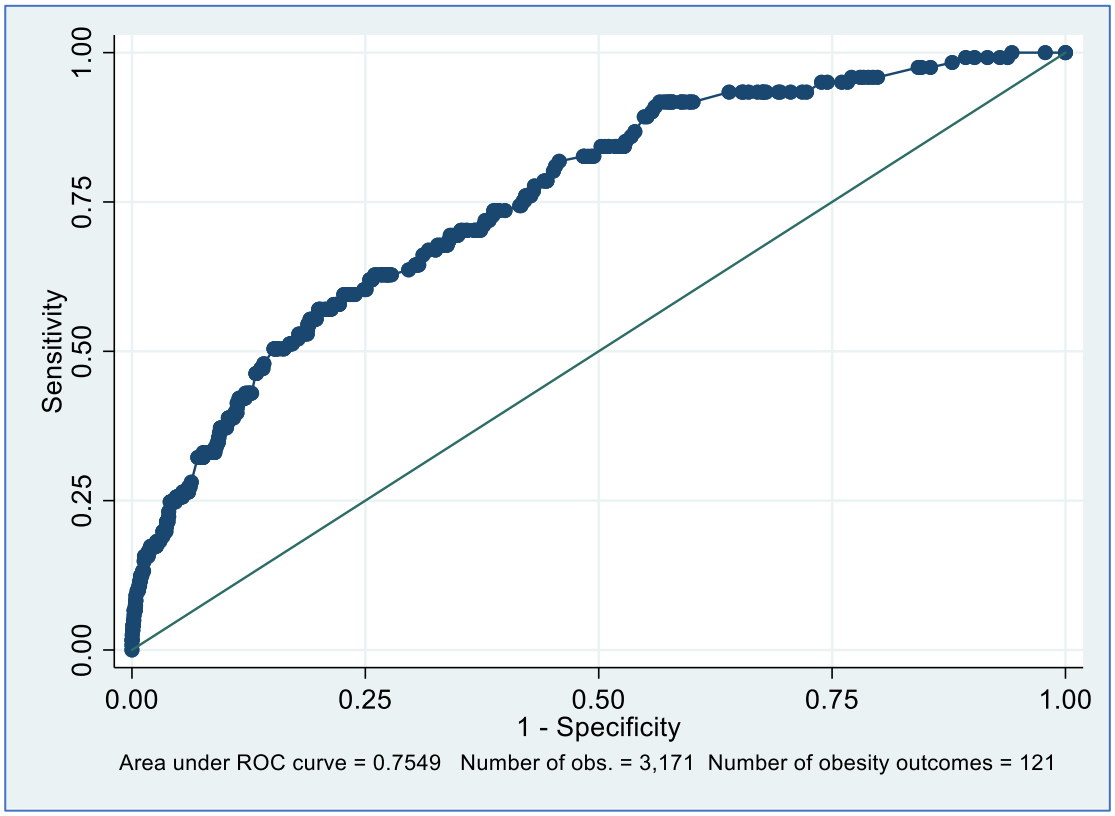
SAMPLE	Derivation	Validation	Combined
No. obs. in sample	4,114	1,372	5,486
Obesity outcomes at TF2	168	71	239
% obesity outcomes at TF2	4.1%	5.2%	4.4%
MODEL SUMMARY			
No. obs. dropped by model	943	295	1,238
No. obs. kept by model	**3,171**	**1,077**	**4,248**
Obesity outcomes kept in model	121	51	172
% obesity outcomes kept in model	3.8%	4.7%	4.0%
LR chi2 (12 d.f.)	115.22	30.62	134.54
Prob >chi2	<0.001	0.0023	<0.001
Pseudo R2	0.1121	0.0746	0.0934
MODEL PERFORMANCE			
Shrinkage	0.90	0.61	0.91
Brier Score	0.0345	0.0431	0.369
AUROC or concordance statistic	0.755	0.696	0.730
Calibration slope	1.00	1.00	1.00
Hosmer-Lemeshow chi2 (prob.)	9.41 (0.31)	3.22 (0.92)	13.93 (0.08)
Mean predicted probability in model	0.038	0.047	0.040
Sensitivity at 0.04	62.81%	64.71%	64.53%
Specificity at 0.04	72.92%	63.16%	68.67%
Correctly classified at 0.04?	73%	63%	69%
PPV at 0.04	8.43%	8.03%	8.00%
NPV at 0.04	98.02%	97.30%	97.87%
Sensitivity at 0.05	57.02%	54.90%	52.91%
Specificity at 0.05	79.31%	72.32%	77.94%
Correctly classified at 0.05?	78%	71%	77%
PPV at 0.05	9.86%	8.97%	9.19%
NPV at 0.05	97.90%	96.99%	97.51%

Discrimination

The discrimination plot for the model in the derivation sample is shown in Figure 9-2. The area under the receiver operating characteristic (AUROC) curve shows how well the receiver (the model) detects the existence of a signal (obesity outcome) in the presence of noise (errors). The plot lies well above the diagonal reference line of 0.5 (no better than chance) indicating that the predictive model can differentiate between children with the outcome and those without to some extent but does not make the correct prediction in every case.

The AUROC values used to classify discriminatory ability are: >0.5 to <0.7 poor, ≥ 0.7 to < 0.8 acceptable, ≥ 0.8 to < 0.9 excellent, ≥ 0.9 outstanding (Hosmer, 2013). Based on an AUROC value equal to or above 0.7 the model's discrimination in the derivation, validation and combined cohort samples was "acceptable".

Figure 9-2 Discrimination plot of the FINAL model run in the DERIVATION sample, for obesity outcomes at 13.5 years old



Calibration

Calibration plots are given for the final model in the derivation and validation samples (Figure 9-3 and Figure 9-4), showing the level of agreement between predicted obesity outcomes on the x axis and observed obesity outcomes on the y axis. The spike plot underneath shows observed outcome events (obese = 1) and non-events (not obese = 0) at the different predicted probabilities. The dotted reference line has a slope of 1 and an intercept of 0, indicating perfect agreement. The children in each sample were categorised into 10 risk groups of predicted probability of obesity by 13.5 years, with each risk group plotted on the graph as a small circle. (This is a visual depiction of the Hosmer-Lemeshow goodness-of-fit test with 10 "deciles of risk", see below.) Most groups have low predicted probability, so cluster to the left hand side of the graph. The slope of the plotted line is close to 1 with an intercept close to 0, showing that the model is well calibrated in the derivation and validation samples. In both graphs, the smoothed LOWESS (locally weighted scatterplot smoothing) line rises steeply where there are fewer observations, which suggests that the model may exaggerate risk at the very highest predicted probabilities.

The Pearson goodness-of-fit (g-o-f) test (not shown) found 428 different covariate patterns in the derivation sample, while in the validation sample there were 283 different patterns. In logistic regression, fitted values are calculated for each covariate pattern, so a large number of patterns is not a concern during model development but becomes an issue during assessment of the model (Hosmer, 2013), as we found. Pearson g-o-f was not a helpful measure of calibration here.

Results of the Hosmer-Lemeshow goodness-of-fit test with 10 groups, using 8 d.f., in the derivation and validation samples are shown in Table 9-20 and Table 9-21. In each sample the number of predicted obesity outcomes for each risk group broadly agrees with the number of observed obesity outcomes in that risk group. In the derivation sample the FINAL model has a Hosmer-Lemeshow chi squared value (HL $\chi 2$) of 9.41, with a p-value calculated from the chi-square distribution with 8 d.f. of 0.309. The model is less well calibrated in the validation sample based on the p-value, HL $\chi 2 = 3.22$, $p = 0.920$. Close examination of the calibration plots confirms this, as the LOWESS curve starts to deviate from the reference line sooner (at lower predicted probabilities) in the validation sample.

Figure 9-3 Calibration plot of the FINAL model run in the DERIVATION sample

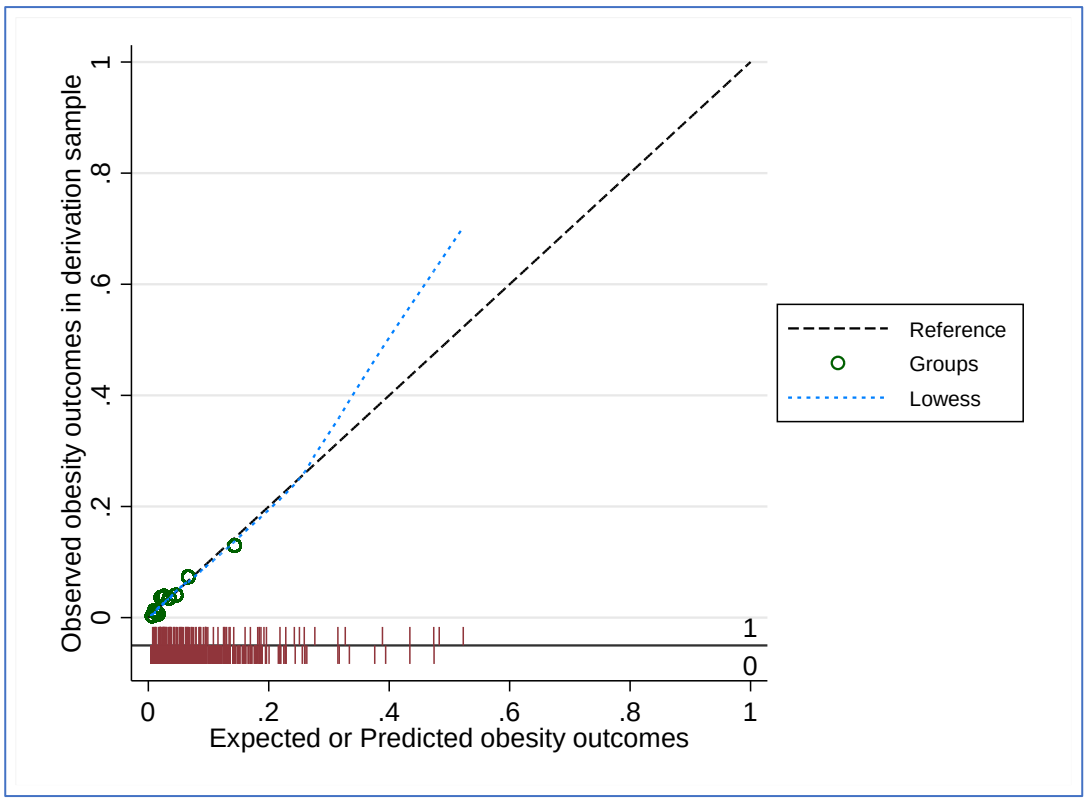


Figure 9-4 Calibration plot of the FINAL model run in the VALIDATION sample

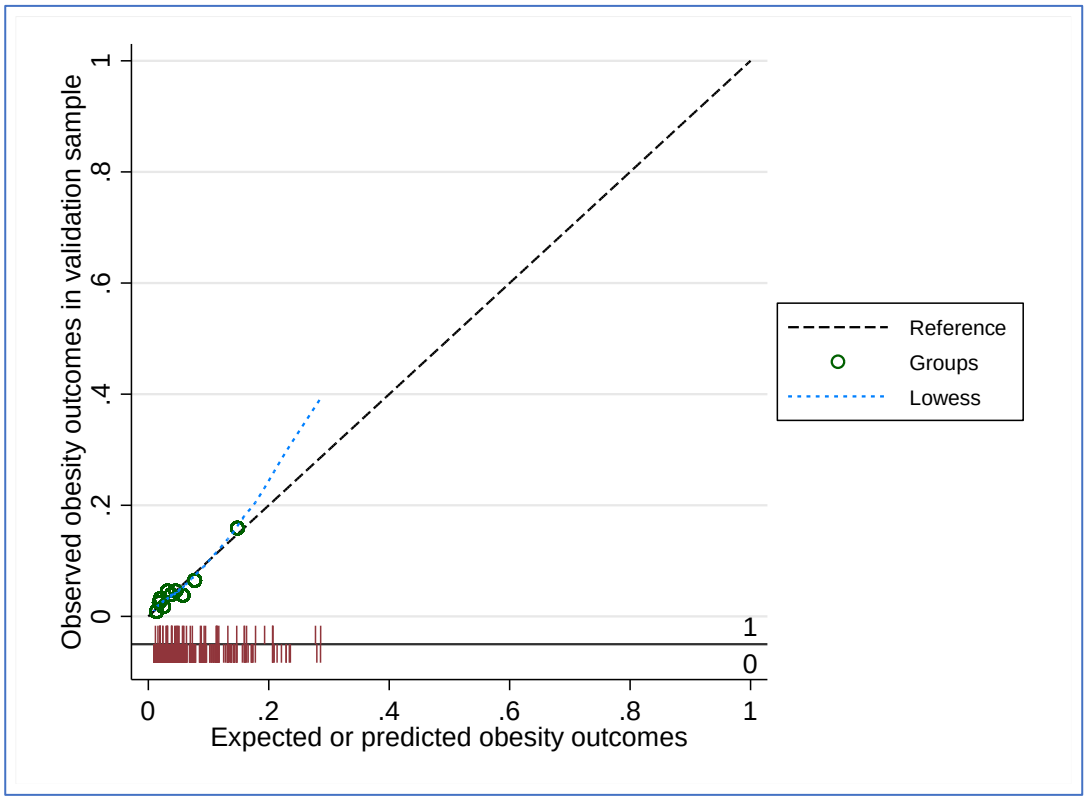


Table 9-20 Hosmer-Lemeshow goodness of fit test of the FINAL model in the DERIVATION sample

Group	Probability	Observed Obese at TF2	Predicted Obese at TF2	Observed Not obese at TF2	Predicted Not obese at TF2	Total
1	0.0082	1	2.1	326	324.9	327
2	0.0112	4	3.2	318	318.8	322
3	0.0144	3	4.3	334	332.7	337
4	0.0181	2	5.2	305	301.8	307
5	0.0241	11	6.3	289	293.7	300
6	0.0294	12	8.2	299	302.8	311
7	0.0384	12	11.9	331	331.1	343
8	0.0546	12	13.6	283	281.4	295
9	0.0863	23	20.8	290	292.2	313
10	0.5229	41	45.3	275	270.7	316
Observations = 3,171 Groups = 10, H-L chi2 (8) = 9.41, Prob > chi2 = 0.3093						

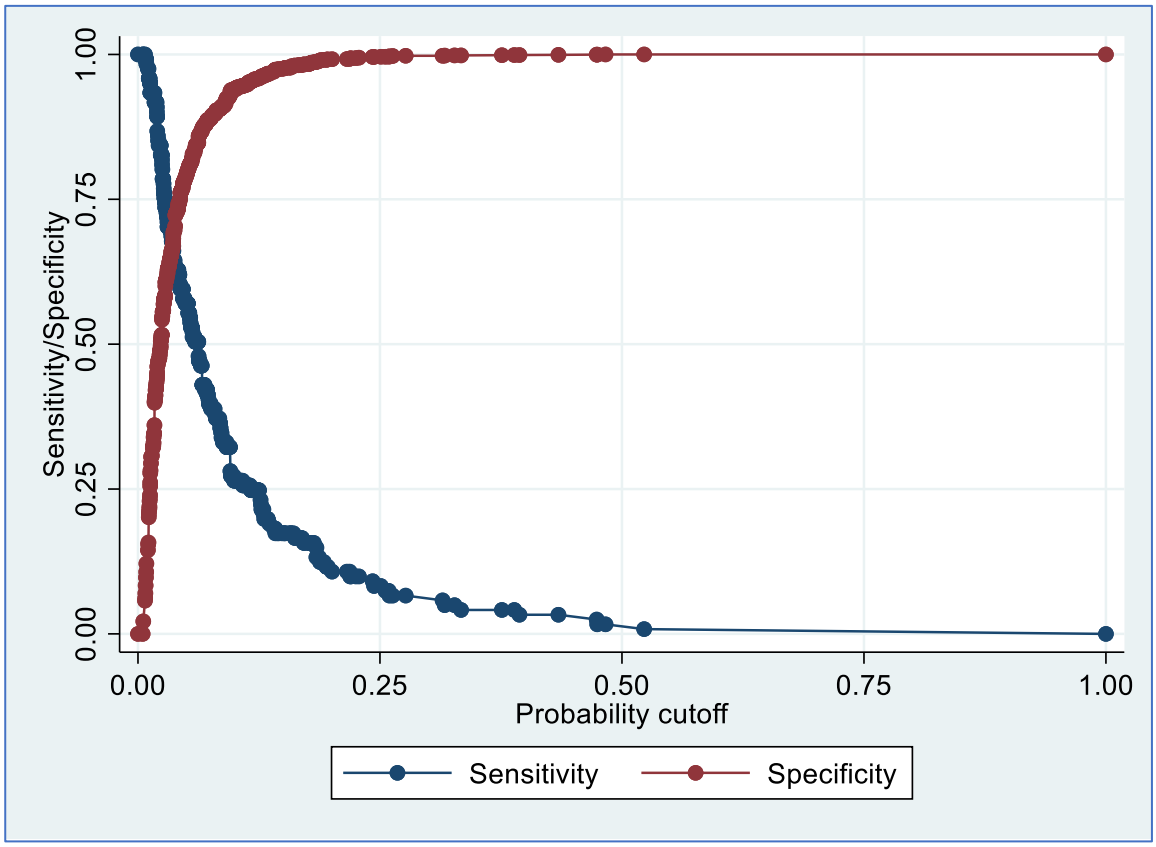
Table 9-21 Hosmer-Lemeshow goodness of fit test of the FINAL model in the VALIDATION sample

Group	Probability	Observed Obese at TF2	Predicted Obese at TF2	Observed Not obese at TF2	Predicted Not obese at TF2	Total
1	0.0156	1	1.5	111	110.5	112
2	0.0195	3	2.2	116	116.8	119
3	0.0235	3	1.9	90	91.1	93
4	0.0289	2	2.9	111	110.1	113
5	0.0352	5	3.5	104	105.5	109
6	0.0393	4	3.9	98	98.1	102
7	0.0496	5	5	104	104	109
8	0.0616	4	6	101	99	105
9	0.0966	7	8.3	101	99.7	108
10	0.2861	17	15.8	90	91.2	107
Observations = 1,077, Groups = 10, H-L chi2 (8) = 3.22, Prob > chi2 = 0.9198						

Probability cut-offs

The trade-off between sensitivity (true positive success rate) and specificity (true negative success rate) is clearly seen in Figure 9-5 which plots them both at different probability cut-offs, for the FINAL model in the derivation sample. The optimal balance between the two is where the two plotted lines cross, at a probability cut-off of approximately 0.05, or 1 in 20 (5%). The same plot in the validation sample gives similar results, with an optimal cut-off at ~ 0.05 (graph not shown).

Figure 9-5 Plot of sensitivity and specificity of the FINAL model at different probability cut-offs in the DERIVATION sample



Classification tables

A classification table for the FINAL model in the derivation sample, using 0.05 as the probability cut-off or decision threshold is shown (See Table 9-22), with worked examples of the calculations for sensitivity, specificity, PPV and NPV.

Table 9-22 Classification table with a cut-off 0.05 for the FINAL model in the DERIVATON sample

	Observed outcome		
Predicted outcome	**Obese**	**Not obese**	**Total**
Obese	69	631	700
Not obese	52	2,419	2,471
Total	121	3,050	3,171

Sensitivity	69/121	57.0%
Specificity	2,419/3,050	79.3%
Positive predictive value	69/(69+631)	9.9%
Negative predictive value	2,419/(52+2,419)	97.9%
Correctly classified	(69+2,419)/3,171	78.5%

In the derivation sample, 3,171 children were kept in the model of whom 121 were children with obesity at follow-up. Applying the model with a probability cut-off of 0.05 in the derivation sample correctly classifies 78.5% of the children. 700 children are predicted to have future obesity so are “at risk” and may benefit from an intervention. If the intervention was 100% successful, 69 cases of future obesity could be prevented (PPV = 9.9%), but 631 children are misclassified. The model identifies 2,471 children as “not at risk” of future obesity. 2,419 were correctly classified (NPV = 97.9%), but 52 children were misclassified and do experience obesity.

Classification tables for the derivation sample and the validation sample gave different sensitivity, specificity, and PPV and NPV rates at different probability cut-offs, as summarised in Table 9-23 and Table 9-24. The optimal decision threshold is a trade-off between the benefit of a true positive classification (the purpose of the model) and any harm or distress that may be caused by a false positive classification. At cut-offs above 0.05 the predictive model correctly classified more outcomes overall, but was less sensitive, so did not correctly predict as many of the observed (true positive) obesity outcomes. At cut-offs below 0.05 the model

became more sensitive but less specific, so even more true negative outcomes (not obese) were wrongly predicted to be positive outcomes (obese) and the number of correctly predicted outcomes fell. For the same cut-offs, the model performs less well in the validation sample than in the derivation sample, with fewer correctly classified overall, demonstrating that the FINAL model is "optimistic".

Table 9-23 Sensitivity, specificity, PPV and NPV at different cut-offs for the FINAL model in the DERIVATION sample

Probability cut-off	Sensitivity	Specificity	PPV	NPV	Correctly classified
0.5	<1%	100%	100%	96.2%	96.2%
0.1	26.5%	94.0%	14.9%	97.0%	91.4%
0.06	50.4%	84.5%	11.4%	97.7%	83.2%
0.05	57.0%	79.3%	9.9%	97.0%	78.5%
0.04	62.8%	72.9%	8.4%	98.0%	72.5%

Table 9-24 Sensitivity, specificity, PPV and NPV at different cut-offs for the FINAL model in the VALIDATION sample

Probability cut-off	Sensitivity	Specificity	PPV	NPV	Correctly classified
0.5	0%	100%	n/a	95.3%	95.3%
0.1	33.3%	91.2%	15.9%	96.5%	88.5%
0.06	47.1%	79.6%	10.3%	96.8%	78.1%
0.05	54.9%	72.3%	9.0%	97.0%	71.5%
0.04	64.7%	63.2%	8.0%	97.3%	63.2%

9.3.16 Risk score

When the amount of shrinkage is more than expected (with an observed shrinkage factor in the validation dataset smaller than the one estimated in the training or derivation sample, as we found) one option is to adjust the model for optimism by multiplying the regression coefficients in the models by the observed shrinkage factor (Van Houwelingen and Le Cessie, 1990). However, this step was not necessary for calculating a risk score. Regression coefficients were used as the basis of the risk score, assigning integer scores to the variable categories to reflect their relative strengths as predictors (Weng, S. F. et al., 2013).

The FINAL model, with and without the interaction term, was run in the combined cohort with the output as logit coefficients (See Table 9-25 and Table 9-26). Note that reference categories were amended for the following predictor variables, so that coefficients for the individual predictors were >0, indicating an increased risk of future obesity compared with the reference category.

Categorical Vegetable intake at 10.5 years

- Vegetable intake is zero
- < 2servings Veg/day
- ≥ 2 servings Veg/day REFERENCE

Dichotomous Dairy servings at 10.5 years

- Dairy foods (not milk) ≥ 1 serving/day
- Dairy foods (not milk) < 1 serving/day REFERENCE

Dichotomous Energy dense treats intake at 10.5 years

- Treats < 2 servings/day
- Treats ≥ 2 servings/day REFERENCE

Table 9-25 FINAL model with 9 predictors and 1 INTERACTION TERM in the COMBINED cohort, shown with logit coefficients

Log likelihood at final iteration	-652.76
Number of obs. (obesity outcomes)	4,248 (172)
LR chi2 (12))	134.54
Prob > chi2	**<0.001**
Pseudo R2	0.0934
AUROC	0.7302

crTF2obese	**Co-efficient**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI for coef**	**Upper 95% CI for coef**
crArmpit No armpit hair	Ref					
crArmpit Armpit hair	0.89	0.21	4.32	**<0.001**	0.49	1.30
crMumoverw Mum not overweight	Ref					
crMumoverw Mum overweight	0.82	0.16	5.09	**<0.001**	0.50	1.13
crBODYSAT Child satisfied	Ref					
crBODYSAT Child unsatisfied	1.01	0.16	6.32	**<0.001**	0.70	1.33
crACTIVTR Active travel to school	Ref					
crACTIVTR No Active travel	0.38	0.16	2.39	**0.02**	0.07	0.70
crVEGcat ≥ 2 servings Veg/day	Ref					
crVEGcat Veg intake = 0/day	1.04	0.38	2.74	**0.01**	0.30	1.79
crVEGcat < 2 servings Veg/day	0.0047	0.33	0.01	0.99	-0.64	0.65
crSSB < 1 serving SSB/day	Ref					
crSSB ≥ 1 serving SSB/day	0.12	0.36	0.32	0.75	-0.60	0.83
crVEGcat#crSSB < 2 servings Veg/day#≥1 serving SSB/day	-0.74	0.51	-1.45	0.15	-1.75	0.26
crVEGcat##SSB ≥ 2 servings Veg/day#≥1 serving SSB/day	0.40	0.43	0.93	0.35	-0.44	1.23
crDAIRYSERVE < 1 Dairy serving/day	Ref					
crDAIRYSERVE ≥ 1 Dairy serving/day	0.34	0.17	1.99	**0.05**	0.00	0.67
crMILKSERVE Milk servings > 0/day	Ref					
crMILKSERVE Milk servings = 0/day	0.74	0.20	3.73	**<0.001**	0.35	1.13
crTREATS ≥ 2 Treats/day	Ref					
crTREATS < 2 Treats/day	0.44	0.17	2.52	**0.01**	0.10	0.78
_cons	-4.87	0.32	-15.44	**<0.001**	-5.49	-4.25

Table 9-26 FINAL model with 9 predictors and NO INTERACTIONS in the COMBINED cohort, shown with logit coefficients

Log likelihood at final iteration	-656.28
Number of obs. (obesity outcomes)	4,248 (172)
LR chi2 (10))	127.49
Prob > chi2	**<0.001**
Pseudo R2	0.0885
AUROC	0.7304

crTF2obese	**Co-efficient**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI for coef**	**Upper 95% CI for coef**
crArmpit No armpit hair	Ref					
crArmpit Armpit hair	0.90	0.21	4.38	**<0.001**	0.50	1.31
crMumoverw Mum not overweight	Ref					
crMumoverw Mum overweight	0.83	0.16	5.15	**<0.001**	0.51	1.14
crBODYSAT Child satisfied	Ref					
crBODYSAT Child unsatisfied	1.02	0.16	6.38	**<0.001**	0.71	1.34
crACTIVTR Active travel to school	Ref					
crACTIVTR No Active travel	0.39	0.16	2.45	**0.01**	0.08	0.71
crVEGcat ≥ 2 servings Veg/day	Ref					
crVEGcat Veg intake = 0/day	0.56	0.26	2.20	**0.03**	0.06	1.07
crVEGcat < 2 servings Veg/day	0.26	0.21	1.27	0.21	-0.14	0.67
crSSB < 1 serving SSB/day	Ref					
crSSB ≥ 1 serving SSB/day	0.1956	0.17	1.16	0.25	-0.14	0.53
crDAIRYSERVE < 1 Dairy serving/day	Ref					
crDAIRYSERVE ≥ 1 Dairy serving/day	0.33	0.17	1.95	**0.05**	0.00	0.66
crMILKSERVE Milk servings > 0/day	Ref					
crMILKSERVE Milk servings = 0/day	0.76	0.20	3.80	**<0.001**	0.37	1.15
crTREATS ≥ 2 Treats/day	Ref					
crTREATS < 2 Treats/day	0.43	0.17	2.45	**0.01**	0.09	0.77
_cons	-4.93	0.26	-18.61	**<0.001**	-5.45	-4.41

The final model with no interaction term can also be presented as an equation, with the constant (intercept) and logit coefficients for the 10 covariates:

Log (p/1-p) =	-4.93 + (0.90 x Has armpit hair) + (0.83 x Mother is overweight) + (1.02 x Child unsatisfied with body) + (0.39 x No active travel to/from school) + (0.56 x Vegetable intake is zero) + (0.26 x Vegetable intake is < 2 servings/day) + (0.20 x SSB frequency ≥ 1 serving/day) + (0.33 x Dairy intake is ≥ 1 serving/day) + (0.76 x Milk intake is zero) + (0.43 x Treats intake is < 2 servings/day)

In the derivation sample the FINAL model with 1 interaction term was on the cusp of being over fitted, with a shrinkage factor of 0.90. In the larger combined cohort, the shrinkage factor of the FINAL model with 1 interaction term using 12 d.f. improved to 0.91. In the model with no interaction term using only 10 d.f. shrinkage equalled 0.92. AUROC values for both versions of the model are similar at 0.73 so still indicate acceptable discrimination.

Predicted probabilities that a child would experience obesity at ~ 13.5 years, based on the final model with one interaction term in the combined cohort (4,248 observations), were Mean 0.04 SD 0.04 Range 0.01 to 0.47. The median was 0.027. Distribution was skewed (skewness 3.3) towards lower probabilities as before.

Apart from the interacting variables, Vegetable intake and SSB frequency, the coefficients for individual predictors were similar in both versions of the model run in the combined cohort. As one interaction (Vegetable intake < 2 servings/day # SSB ≥1 serving/day) had a negative coefficient, and for simplicity in calculating scores based on answers to the questionnaire, the interactions were not assigned scores. Instead predictors were put in order of the coefficients (highest to lowest) in the model with interactions, for both models. Coefficients were divided by the value of the smallest positive coefficient in that model, to generate a standardised score (Weng, S. F. et al., 2013). The standardised score was rounded to the nearest whole number, and reduced further if necessary, to create a trial risk score for each version of the model. (See Table 9-27 and Table 9-28) Reference categories for each predictor were scored 0.

Table 9-27 Trial risk score for FINAL model with 1 INTERACTION TERM in the COMBINED cohort

Predictor	Coefficients in model with interaction	Standardised score (Coefficient ÷ 0.0047)	Rounded integer	% of total	Allocated score
Veg = zero	1.04	222.3	222	18%	3
Body dissatisfaction	1.01	215.6	216	17%	3
Armpit hair	0.89	189.6	190	15%	3
Mum overweight	0.82	174.2	174	14%	2
Milk intake is zero	0.74	158.3	158	13%	2
Treats < 2 servings/day	0.44	93.6	94	8%	1
Veg>=2 #SSB>=1	0.40	Interaction – no score	n/a	n/a	n/a
No active travel	0.38	81.4	81	7%	1
Dairy >= 1 serving/day	0.34	72.0	72	6%	1
SSB >= 1 serving/day	0.12	24.5	24	2%	1
Veg < 2 servings/day	0.0047	1	1	<1%	1
Veg<2 #SSB>=1	-0.74	Interaction – no score	n/a	n/a	n/a
Total score			1,232	100	**18**

Smallest coefficient is for Veg < 2 servings/day = 0.0047

Table 9-28 Trial risk score for FINAL model with NO INTERACTION TERM in the COMBINED cohort

Predictor	Coefficients in model with no interaction	Standardised score (Coefficient ÷ 0.1956)	Rounded integer
Veg = zero	0.56	2.89	3
Body dissatisfaction	1.02	5.22	5
Armpit hair	0.90	4.61	5
Mum overweight	0.83	4.23	4
Milk intake is zero	0.76	3.87	4
Treats < 2 servings/day	0.43	2.18	2
No active travel	0.39	2.00	2
Dairy >= 1 serving/day	0.33	1.70	2
SSB >= 1 serving/day	0.1956	1.00	1
Veg < 2 servings/day	0.26	1.35	1
Total score			**29**

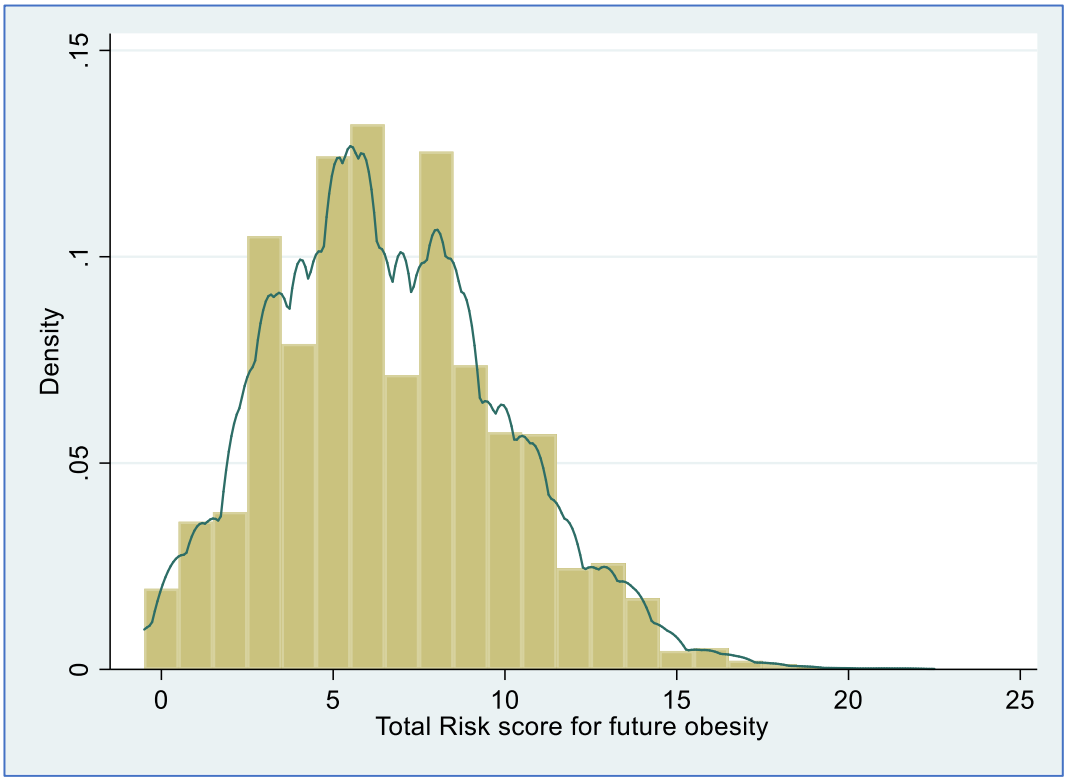
Smallest coefficient is for SSB>= 1 serving/day = 0.1956

In the model with the interaction term, Vegetable intake = zero and SSB frequency >= 1 serving/day became more important predictors of future obesity than in the model with no interaction term. The two trial risk scores were combined, adjusting the weights of the interacting variables to reflect this importance, while keeping all the predictors in the same relative proportions and high to low order of their coefficients as the model with interactions. (See Table 9-29) The resulting risk score had a maximum total of 25. Among the 4,248 children in the combined cohort with a risk score, 172 were obese by the age of 13.5 years. The median score was 6 and the mean score was 6.6 (SD 3.4) ranging between 0 and 22. No-one had a full score and only 60 children had risk scores at or above 15 out of 25. (See Figure 9-6.) As a single predictor of obesity at TF2, the risk score had an AUROC value of 0.72, lower than the full model but still in the acceptable range for discrimination.

Table 9-29 Risk score for the FINAL model in the COMBINED cohort

Predictor	Coefficients in model with interaction	Trial risk score, model with interaction	Trial risk score, model with NO interaction	Adjusted risk score, maximum 25
Veg = zero	1.04	3	3	**4**
Body dissatisfaction	1.01	3	5	3
Armpit hair	0.89	3	5	3
Mum overweight	0.82	2	4	3
Milk intake is zero	0.74	2	4	3
Treats & snacks < 2 servings/day	0.44	1	2	2
No active travel	0.38	1	2	2
Dairy >= 1 serving/day	0.34	1	2	2
SSB >= 1 serving/day	0.12	1	1	**2**
Veg < 2 servings/day	0.005	1	1	1
Total score	n/a	**18**	**29**	**25**

Figure 9-6 Distribution of risk scores in the COMBINED cohort, n = 4,248



The distribution of risk scores was divided into 4 quantiles. As a single predictor of obesity at TF2, the quantiles of risk had an AUROC value of 0.71, still (just) in the acceptable range for discrimination. The percentage of obesity outcomes approximately doubled with each additional quantile of risk. (See Table 9-30)

Table 9-30 Obesity outcomes by quantile of risk score in the COMBINED cohort, n = 4,248

Quantile (4 groups)	Risk score	Mean predicted probability	Not obese at TF2	Obese at TF2	Total	Obese at TF2 %
VERY LOW RISK	0 to 4	0.012	1,164	14	1,178	1.2%
LOW RISK	5 or 6	0.023	1,063	26	1,089	2.4%
Median	6	0.027	n/a	n/a	n/a	n/a
MEDIUM RISK	7 to 9	0.043	1,100	49	1,149	4.3%
HIGH RISK	≥10	0.100	749	83	832	10%
Total	n/a	n/a	**4,076**	**172**	**4,248**	**4.0%**

Children who were obese at baseline were excluded from model development, but 14.5% (793) of the 5,486 children in the combined cohort were overweight at baseline, of whom 584 had a risk score.

The risk score was not strongly correlated with BMIz at baseline (corr = 0.1447). However, children with overweight at baseline tended to have higher risk scores for future obesity (Mean 7.9 SD 3.6 range 0 to 22) than children who were not overweight at baseline (Mean 6.4 SD 3.3 range 0 to 19) and both groups included children in the high risk quantile for future obesity. (See Table 31)

Table 31 Baseline overweight status and obesity outcomes in the COMBINED cohort, n = 4,248

Baseline overweight status	Quantile (4 groups)	Not obese at TF2	Obese at TF2	Total	Obese at TF2 %
Overweight	V. LOW RISK	94	8	102	7.8%
Overweight	LOW RISK	103	19	122	15.6%
Overweight	MEDIUM RISK	145	34	179	19.0%
Overweight	HIGH RISK	115	66	181	36.0%
Overweight	**Total**	**457**	**127**	**584**	**21.7%**
Not o/w	V. LOW RISK	1,070	6	1,076	0.5%
Not o/w	LOW RISK	960	7	967	0.7%
Not o/w	MEDIUM RISK	955	15	970	1.5%
Not o/w	HIGH RISK	634	17	651	2.6%
Not o/w	**Total**	**3,619**	**45**	**3,664**	**1.2%**

Three quarters of the observed obesity outcomes occurred among children with overweight at baseline. There were fewer obesity outcomes among children who were not overweight at baseline, but in both groups the percentage of observed obesity outcomes increased with each additional quantile of risk.

9.3.17 Questionnaire scoring

The prototype questionnaire asked about exposure to 24 putative predictors of obesity. Nine predictors were included as covariates in the FINAL model. Risk scores allocated to each category for these variables are shown in Table 9-31.

Table 9-31 Risk score allocated to predictors in the FINAL model

Predictor	Risk score
Child has armpit hair at 10 years old – Yes	3
Child has armpit hair at 10 years old – No	0
Mother is overweight. Do you think one or both of your parents is overweight? – Yes	3
Mother is not overweight. Do you think one or both of your parents is overweight? – No	0
Milk intake is zero – Never or rarely	3
Milk intake is > zero – Sometimes, Once day, More than once a day	0
Sugar sweetened beverage frequency is ≥ 1 serving/day – Once day, More than once a day	2
Sugar sweetened beverage frequency is < 1 serving/day – Never or rarely, Sometimes	0
Dairy food (yogurt, cheese, sauces) intake is ≥ 1 serving/day – Two or more times a day	2
Dairy food intake is < 1 serving/day – Never or rarely, Sometimes, Once a day	0
Vegetable intake is zero – Never or rarely	4
Vegetable intake is < 2 servings/day – Sometimes, Once a day	1
Vegetable intake is ≥ 2 servings/day – Two or more times a day	0
Treats & snack intake is < 2 servings/day – Never or rarely, Sometimes, Once a day	2
Treats & snack intake is ≥ 2 servings/day – Two or more times a day	0
Chid unsatisfied with body. Happy with body shape – No	3
Child satisfied with body shape. Happy with body shape – Yes	0
No active travel to/from school. Car Taxi. Bus. Train. Other.	2
Active travel to/from school. Walk. Bike.	0

9.3.17.1 Extending the risk score

Four potential predictors were unmatched to candidate variables in the ALSPAC dataset:

- Breakfast frequency
- Family meals
- Diet drinks
- Dieting/fussy eating

Four potential predictors had more missing observations after imputation than there were observations in a category:

- Child's ethnic background
- Only child
- Fast Food once a week or more
- Meets sleep recommendation

These eight predictors could not be tested in the model, but they may have some predictive value. Evidence from papers in the Systematic Review that investigated multiple predictors, and from a published childhood obesity prediction model, suggests that their strength as predictors of obesity was modest in comparison with the predictors that were tested.
No comparative evidence was found for Diet drinks. As the association of Diet drinks with future obesity was uncertain, it was decided to remove the question about Diet drinks from the questionnaire. Estimated risk scores were allocated to dichotomous categories of the other 7 variables for future consideration, creating an extended risk score. (See Table 9-32.)

Table 9-32 Estimated risk scores allocated to untested predictors

Predictor	Estimated risk score	Source of comparative evidence
Child's ethnicity – Non white	1 est.	(Rehkopf et al., 2011) (Reilly et al., 2005)
Child's ethnicity – White	0	
Only child – Yes	1 est.	(Rehkopf et al., 2011; Reilly et al., 2005)
Only child – No, siblings at home	0	
Breakfast – Never or rarely	1 est.	(Rehkopf et al., 2011) (Quick et al., 2013)
Breakfast – Sometimes 1 to 3 times a week or Often 4 times a week or more	0	
Family meals – Never or rarely	1 est.	(Rehkopf et al., 2011)
Family meals – Sometimes 1 to 3 times a week or Often 4 times a week or more	0	
Fast food 1 or 2 times a week, or more	1 est.	(Rehkopf et al., 2011; Quick et al., 2013)
Fast food Less than once a week	0	
Diet drinks frequency is ≥ 1 serving/day, Once a day or More than once a day	n/a	No comparative evidence found.
Diet drinks frequency is < 1 serving/day, Never/rarely or Sometimes	n/a	
Dieting/fussy eating – Yes	1 est.	(Rehkopf et al., 2011) (Quick et al., 2013)
Dieting/fussy eating – No	0	
Meets sleep rec. No – asleep after 10pm	1 est.	(Reilly et al., 2005)
Meets sleep rec. Yes – asleep by 10pm	0	

9.3.17.2 Revised Questionnaire

Seven potential predictors proved to have little predictive value in the ALSPAC cohort. They made marginal contributions to model predictive performance and were dropped from models with step-wise removal/addition:

- Mother has degree
- Smokers in household
- Eats 3 meals or more a day
- Eats between meals (snacking)
- Juice frequency
- Whole grain intakes
- Fruit intakes

As the seven variables did not have allocated risk scores, questions about them were removed from the questionnaire.

The prototype Children's Obesity Risk Assessment (CORA) questionnaire was amended to reflect changes made during model fitting, notably for the questions about Milk and Dairy foods. As the questionnaire is intended for children in Year 6 of primary school, aged around 10.5 years old, a question about current age is presented at the start, but age was not used as a predictor and so date of birth is not required. The question about the child's sex was also removed as the questionnaire is not designed to be sex specific, and sex was not used as a predictor. The revised CORA questionnaire, containing 18 sequentially numbered questions, is presented with scores from the risk model and estimated scores in Appendix I. The maximum total score is 32, from 16 prognostic/risk factors.

Potentially sensitive questions about the child or their family have been moved to the end of the questionnaire, with a "Prefer not to say" tick box option, and response boxes are now set out vertically, as recommended for paper-based surveys (Fanning, 2005).

9.4 Discussion and Conclusion

Childhood obesity continues to be a problem in England. The latest results from the NCMP (NHSDigital, 2019) show that in 2018/19 obesity prevalence among pupils in Year 6 of state primary schools (mainly children aged 10 to 11 years) was not much changed at 20.2%, tending to be higher among boys (22.5%) than among girls (17.8%), highest among Black children (28.9%) and lowest in children of White or Chinese ethnicity (~ 18%). Children living in the most deprived areas had the highest obesity prevalence overall (26.9%), over twice that of children living in the least deprived areas (11.4%). It is evident that these children and their families need interventions to help the children reach a healthier weight.

However, 80% of Year 6 children in the NCMP were not classified as obese. Almost two-thirds of Year 6 children were of a healthy weight in 2018/19, with 14.1% who had overweight. The second public health challenge is to help such children maintain or achieve a healthy weight, and to prevent them from becoming obese. Although existing overweight and living in a more deprived area increase its likelihood, future obesity is not a given. Using an algorithm to further identify children/populations most at risk could help Public health planners and health professionals to target prevention interventions more effectively.

This chapter has described an obesity risk prediction model, developed using non-clinical prognostic/risk factors that are not routinely measured but could be surveyed at the same time as the NCMP Year 6 measurements. The predictive model identifies children aged approximately 10.5 years old (without obesity) who are at risk of obesity three years later, in their early teens. Predictions are based on nine prognostic/risk factors, including five that are related to diet: Early puberty, Mother's overweight, Child's body satisfaction, Active travel (as a measure of routine physical activity), Vegetable frequency, Dairy food frequency, Energy Dense treats frequency, Milk frequency and Sugar sweetened beverage frequency. Predictors in the model were translated into risk scores for use with the CORA questionnaire.

Other childhood and adolescent obesity risk prediction models have been developed (Ziauddeen et al., 2018, Canfell et al., 2018) but none considered the child's diet beyond breast feeding and weaning. Most tools considered prognostic/risk factors that were established between pregnancy and infancy/early childhood rather than in later childhood. Perhaps as a consequence of a longer follow-up period, only one model that predicted adolescent obesity reported a strong predictive performance (Morandi et al., 2012), relying on "traditional risk factors" at birth (Parental BMI, birthweight, gestational weight gain, number in household, mothers professional status, smoking) which were recorded by the Northern Finland Birth cohort. Arguably the child's birthweight, a precursor of child's weight used to calculate BMI

and thence define overweight and obesity using BMI cut-offs, should not be used as a predictor of overweight/obesity as the two are not independent.

The FINAL model was informative in the derivation and the validation samples from the ALSPAC cohort, with Brier scores of 0.03 and 0.04 respectively. These scores compare favourably with the Brier scores reported for men (0.09) and women (0.02) in the Cardiovascular Disease Population Risk Tool (Manuel et al., 2018), which was developed in a much larger cohort of 104,219 respondents.

The fitted model is well-calibrated, with a Hosmer-Lemeshow goodness-of-fit p-value = 0.3, equivalent to p-values reported by obesity risk tools for new-born children (Morandi et al., 2012, Steur et al., 2011) and to the p-values in the Obesity Population Risk Tool for adults (Males p = 0.649, Females p = 0.104) described as "acceptable" calibration (Lebenbaum et al., 2018).

The model also achieved acceptable discrimination (AUROC = 0.76) in the derivation sample, correctly classifying 79% of observed outcomes at the optimum decision threshold (5% risk). In the two Systematic reviews of obesity risk tools the best discriminative performance (AUROC =0.86+, "excellent") was reported for a mobile phone application to predict an infant's risk of childhood obesity, developed in the Born in Bradford cohort (Santorelli et al., 2013). Other childhood obesity risk models reported "acceptable" discrimination, with one, the CORE tool (Manios et al., 2013), developed with retrospectively collected predictors, reporting "poor" discrimination (AUROC = 0.64). Predictive models used in clinical settings use objective clinical measures as well as self-reported predictors, and necessarily have "excellent" levels of discrimination. For example, QRISK2, a score which predicts the 10 year risk of cardiovascular disease in the UK has a reported AUROC value of 0.83 (Collins and Altman, 2012).

Overall, the model's predictive performance proved "optimistic". In the smaller validation sample, with fewer obesity events to find, the model was less well calibrated at higher predicted probabilities, so tended to exaggerate risk. Discrimination was acceptable (AUROC = 0.70) but a lower percentage of observed outcomes (72%) were correctly classified at the 5% cut-off.

The level of accuracy seen in the PPV (the probability that someone predicted to have the outcome really does have the outcome) is less than 10%. This is far lower than the PPV reported for a risk algorithm to predict childhood overweight from predictors in infancy (PPV = 37%) (Weng, S. F. et al., 2013), and means that 9 out of 10 predictions of an obesity outcome are wrong, which may cause unwarranted distress if used at an individual level. This level of

false positives is not ideal, but at a population level the model may still be useful for preventive purposes. When the model was reduced to a risk score, discrimination dropped a little, but the percentage of observed obesity outcomes increased with each additional quantile of risk for children with overweight and for children with healthy weight. This suggests that the risk score may have some predictive value in addition to baseline weight status. It tells us something new.

Other childhood and adolescent obesity risk prediction models had between 4 and 10 predictors, with 6 on average (Ziauddeen et al., 2018). Indisputably, the model (with nine predictors and one interaction term) is on the cusp of being over fitted to the available information and this will have added error, explaining why the model's predictive performance was reduced in the validation sample.

The five dietary predictors included in the final model as frequency questions are based on robust measures of intake at baseline, with carefully considered category cut-offs. Four of the five (Vegetable frequency, Dairy food frequency, Energy Dense treats frequency and Milk frequency) are derived from quantified intakes from a validated 3 day food diary, with no imputed values. The fifth dietary predictor (Sugar sweetened beverage frequency), based upon a drinks frequency questionnaire, was validated by comparison with quantified SSB intakes from the 3 day food diary, which was also used to impute SSB frequency for the 20% that were missing. While categorisation of continuous food intake variables lost some information, the pre-specified categorical variables were useful predictors that translated readily to the questionnaire.

Individually, higher vegetable frequencies and drinking at least some milk predicted lower risk of future obesity, while higher SSB frequency predicted higher risk, as expected. In this cohort higher dairy food frequency (yoghurt, cheese, milk based sauces) was not beneficial, instead predicting higher risk of future obesity. This may be because larger quantities of sugary yoghurts and high fat cheese contribute to an energy dense dietary pattern, which invites further investigation. Conversely children having energy Dense treats (Sweet and savoury biscuits, Buns, cakes, pastries and fruit pies, Crisps, Sweets and Chocolate) less often were at higher risk of future obesity. This could be explained by reverse causality if children who already have concerns about their weight cut treats and snacks from their diet, or it may be because some foods in the composite variable reflect helpful eating habits. For example, teenagers who ate meals but did not snack much had higher intakes of Buns, cakes, pastries and fruit pies than teenagers who snacked. As the weakest dietary predictor in the model, the composite variable Energy Dense treats needs reconsideration and recalibration.

Three of the four non-dietary covariates, Mother's overweight, Early puberty and Child's body satisfaction, are among the strongest predictors in the final model. Mother's overweight is based on self-report so will contain measurement error and is likely biased towards under-reporting of overweight. Early puberty is based on the mother's report, not clinical measures, but as the question about armpit hair was repeated, this measure is more reliable. Child's body satisfaction is derived from two subjective measures and conceivably children's responses might vary day-to-day. The fourth non-dietary predictor, Active travel to/from school, was used as a proxy for physical activity. Active travel was the weakest non-dietary predictor in the model. Oher measures of physical activity may be stronger predictors. Maternal weight or BMI was a predictor often included in other childhood and adolescent obesity risk prediction models, but few models extended to a time when pubertal stage was a potential predictor, and no psychological or physical activity measures were considered, which is a limitation.

Not all potential predictors/candidate variables proved useful. Whole grain intakes were comparatively low in the ALSPAC cohort - possibly too few children had high enough intakes to predict differences in future obesity risk. Fruit intake was correlated with vegetable intake, but as it was the weaker predictor, added little beyond extra measurement error. Non-diet variables often used in childhood and adolescent obesity risk prediction models, such as Mother's education level and Smoking, were not important predictors of future obesity among ALSPAC children. Mother's education level and Smoking were both dummy variables derived from categorical data, so may have lost important information.

Some potential predictors could not be tested, either because no suitable candidate variable was available in the dataset, or due to high levels of missing observations which risked spurious associations and misleading predictions. Although estimated risk scores were allocated to these predictors (Child's ethnic background, Only child, Breakfast frequency, Family meals, Fast Food, Diet drinks, Dieting/fussy eating, Meets sleep recommendation) they are untested and so less robust than scores derived from the logistic regression model. Testing in a different dataset is recommended.

One strength of this study is that all the candidate predictors tested in the model were based on a systematic review, which provided evidence about the dietary patterns, eating habits, and food and drinks significantly associated with childhood overweight or obesity outcomes, and also identified predictors of weight gain, overweight or obesity, some of which were non-food variables. Published obesity risk prediction models provided corroborating evidence of the predictive value of some of the non-food factors that were identified, including parental BMI,

parental education level, ethnicity, smoking and number of siblings. The analysis plan was written in a study protocol, assumptions about casual relationships were set out in a DAG and all candidate predictors were pre-specified before examining any relationships with obesity outcomes in the ALSPAC derivation sample, to avoid data driven selection of variables. However, the selection process was not "blind". The associations of some candidate predictors with future overweight/obesity in the ALSPAC cohort were reported by papers in the Systematic review, including energy dense dietary patterns (Ambrosini et al., 2012) , Milk (Noel et al., 2011) and Flavoured milk (Noel et al., 2013) and Dairy foods with milk (Bigornia et al., 2014). Sleep was also identified as an early life predictor for childhood obesity in the ALSPAC cohort (Reilly et al., 2005).

One limitation of this study is the age of the ALSPAC data. Children in the cohort, whose mothers were recruited during pregnancy in 1990-1992, were 10 year olds between 1999 and 2003, and ~15% of them were classified as obese. Twenty years on, the situation has changed. Adult smoking rates have continued to fall, and more people are educated to degree level. Low fat dairy food is more widely available and the recent introduction of a sugar levy on soft drinks has reportedly decreased sugar sweetened beverage sales, yet obesity prevalence among 10 and 11 years olds in England now is 20%. It is unlikely that 10 year olds in 2020 will be exposed to an identical accumulation of the same risk factors for future obesity as their contemporaries at the turn of the last century.
An additional difficulty is the low incidence of obesity (5% or less) in the cohort, which gave so few new cases of obesity for the model to predict.

There were also some limitations in the analysis. Missing observations for candidate predictors and the variables needed to derive obesity outcomes were singly imputed to keep as many children in the models as possible, but the whole sample size available for analysis was still reduced, from 7,462 who took part in the Focus at 10+ clinic to 5,486, of whom only 4,248 had observations for all the predictors used in the final model. While the split sample used for internal validation was effective, it was not the most efficient use of data and restricted the d.f. available for model building and assessment of model performance. More sophisticated multiple imputation methods, which predict values for missing data (assuming missing at random – which was not always the case in the ALSPAC cohort) while accounting for the uncertainty due to imputed data, or resampling methods such as *k*-fold cross-validation and boot strapping might have improved efficiency (Hosmer, 2013).

A strength of the analysis is that two strategies were applied to select covariates for the model, an approach based on a series of models with stepwise removal/addition at $p<0.05$ and $p<0.25$, and the "purposeful selection" method (Hosmer, 2013) which used a smaller sample. Reassuringly, both methods produced the same reduced model, and similar final models within the d.f. available.

Every effort was made to internally validate the final model and to report all aspects of model performance (discrimination, calibration and clinical usefulness) as recommended by the ABCD framework (Steyerberg and Vergouwe, 2014).

Further work is needed to establish whether the risk algorithm and extended risk score are transportable to different settings. As a first step it is proposed that a paper version of the revised CORA questionnaire (without the scoring system on display) is trialled by children in the right age bracket. The risk model/risk score and revised questionnaire with (or without) the extended risk score can then be externally validated in a different population or dataset. Objectively, some predictors may need to be pruned or recalibrated to improve model stability in a new context.

Chapter 10 Conclusions and recommendations for future research

10.1 Conclusions

As set out in Chapter 1 the aim of this thesis was to develop a dietary assessment tool to identify populations of children who are at risk of obesity during early adolescence. The work had two major sections:

- A systematic review of children's dietary exposure and adiposity outcomes.
- The development and internal validation of a predictive model of obesity risk, based on evidence from the systematic review.

The research question that was the primary objective of the systematic review asked,

"To what extent does diet during childhood and adolescence influence future indicators of overweight or obesity?"

Longitudinal evidence from the systematic review was wide ranging. It showed that some aspects of children's diet *do* influence their future adiposity. Quantified intakes of specific foods or drinks, some dietary patterns and eating habits (see Chapters 5 and 6) were significantly associated with markers of adiposity, such as waist circumference, body fat percentage, Body Mass Index or BMI z score, or with higher risk of overweight or obesity. Foods, drinks and eating habits that were more likely to be associated with adverse adiposity outcomes included energy dense snacks and convenience foods, sugar sweetened beverages and eating fast food, all of which contributed to low fibre, high fat, high sugar dietary patterns. Foods and drinks which seemed beneficial included whole grains, dairy foods/milk and vegetables, as did more regular eating habits and the avoidance of dieting. Diet's beneficial or adverse associations with future adiposity were not always certain, as reported associations were not significant in every cohort that investigated a specific dietary exposure or were only significant in a population sub-group.

Most studies adjusted for total energy intake in their final model, treating it as a confounder that influences both dietary exposure and adiposity outcomes, but associations were often attenuated (made smaller) or no longer significant after adjustment for energy intake. For some foods and drinks, particularly those which had a sizeable share of energy intake (E.g. SSBs, 5%+ of TEI on average in some cohorts), the attenuation may have been due to over adjustment. A few studies argued that a food or drink's contribution to TEI is part of the mechanism whereby high intakes may lead to future overweight, electing not to adjust for energy intake at all.

The DONALD study adjusted for residual energy. i.e. Energy intake from all sources other than the food or drink exposure under investigation, which is a logical alternative.

Reported effect sizes were generally modest, but the *extent* to which diet influences future indicators of overweight or obesity could not be determined. Although included studies measured and quantified dietary intakes, specific exposures were examined in only a few cohorts, yielding insufficient quantitative data for meta-analysis. Even when meta-analysis was feasible (see Chapter 7, SSB intakes and adiposity outcomes) the methodological heterogeneity of studies made a dose-response meta-analysis impossible.

The two papers from Project EAT and the NGHS that considered dietary exposures alongside other potential predictors of future overweight or obesity (see Chapter 6), demonstrated that childhood/adolescent diet is not the only, or even the most important predictor of later adiposity outcomes. Socio-economic, familial and psychological factors also play a part, as do health behaviours such as physical/sedentary activity. Potential predictors of children's future obesity were included in a simple questionnaire.

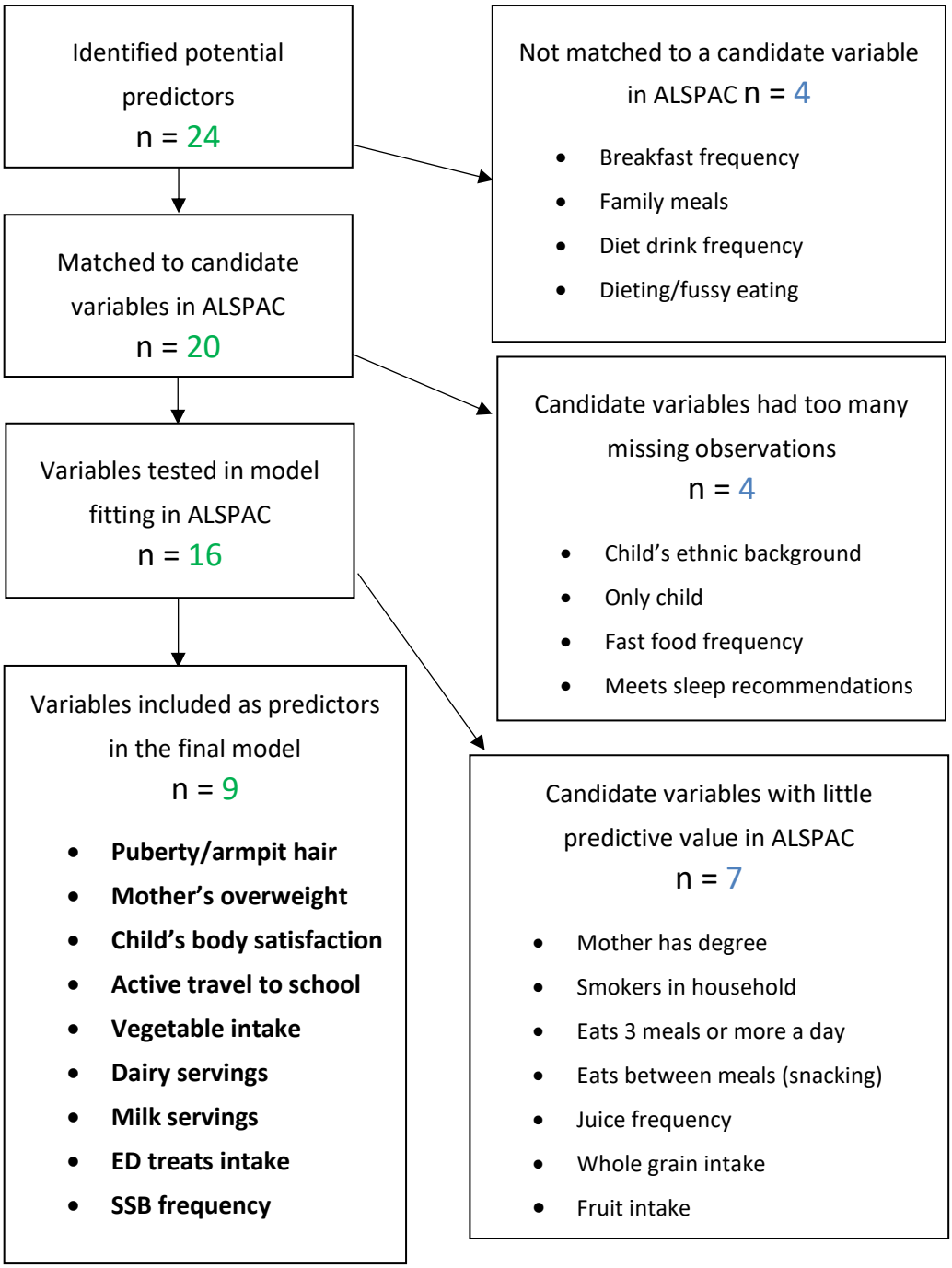
The secondary objective of the systematic review was to identify the best quality cohorts that had measured children's diet and adiposity outcomes, with the intention of acquiring enough data (from a single cohort or combined cohorts) to develop the predictive model. Systematic review screening showed that there are numerous childhood cohorts, but many of them are not of healthy children from the general population or have not measured and quantified children's dietary intake (see Chapter 3). Most of the included childhood cohort studies that had measured diet and adiposity outcomes were judged to be of moderate quality, based on their modified Newcastle-Ottawa quality score. Medium sized cohorts (> 500 to 1,000s) and studies that followed up participants annually or after only a few years had less loss to follow-up than larger cohorts (10,000+) or studies that followed up participants after 5 or 10 years or more. Small cohorts and studies with small sample sizes in analyses (low 100s) sometimes lacked the power to find an effect and had too few participants or outcomes for model development. Large cohorts, with enrolled participants numbering over 10,000+, typically used FFQs and relied upon self-reported height and weight as well as having attrition rates > 30%, all of which added bias and uncertainty about their findings – these studies received the lowest quality scores. The exception was the large ALSPAC cohort (14,000+) which received the highest possible quality score (8/9), alongside the medium sized IDEA & ECHO and NGHS cohorts. Data from the UK ALSPAC cohort were requested for model development on this basis.

The predictive model described and discussed in Chapter 9, shows that an evidence based dietary assessment tool to predict children's future obesity risk is feasible. The model aims to

identify children aged approximately 10.5 years old (without obesity) who are at risk of obesity in their early teens, three years later. The developed model is unique due to the inclusion of dietary predictors (quantified food and drink intakes, eating habits), which other childhood obesity risk models have not considered.

We identified 24 potential predictors of children's future obesity, of which 20 could be matched to candidate variables in the ALSPAC dataset, measured at or close to when the children were 10+ years old. See the flow chart in Figure 10-1 below.

Figure 10-1 Flow chart of potential predictors



Four potential predictors could not be matched, and another four potential predictors could not be tested in the ALSPAC data due to high levels of missing observations. The eight untested variables may have some predictive value which could be tested in other populations. The remaining sixteen potential predictors were tested in the ALSPAC data.

The final fitted model contained nine predictive variables and one interaction term: Puberty/armpit hair, Mother's overweight, Child's body satisfaction, Active travel, Vegetable intake, Dairy servings, Milk servings, Energy dense treats intake and SSB frequency, with an interaction between Vegetable intake and SSB frequency.
Seven potential predictors had relatively little predictive value in the ALSPAC dataset, so were eliminated: Mother has degree, Smokers in household, Eats 3 meals or more a day, Eats between meals (snacking), Juice frequency, Whole grain intake and Fruit intake.

The internal validation of the final predictive model found that it was informative in both the derivation and the validation sample. The model was well-calibrated and achieved acceptable discrimination. Its performance in this regard compared equally well with published childhood obesity risk models although the model was "optimistic" and tended to exaggerate risk at higher predicted probabilities.

The predictive model is not sensitive enough to use with individual children (too many false positives which may cause distress), and so should not be applied in this way. However, that was not the objective. Instead the tool is intended to identify *populations* of children who are at risk of obesity. It may prove useful for this purpose, as the questionnaire for Children's Obesity Risk Assessment (CORA), with risk scores derived from the model, has some predictive value beyond baseline weight status, telling us something new about future obesity risk.

As set out in Chapter 1, a decade ago Levine, Dahly and Rudolf developed a prototype obesity risk tool to predict a baby's obesity risk (Levine et al., 2012). They also concluded that their developed tool did not have acceptable levels of specificity and sensitivity for use at an individual level. Another ethical issue that they addressed was that of follow-up.
This is an important practical and ethical matter at the population level too. After external validation and feasibility testing it will only be appropriate to use a dietary risk assessment such as the CORA risk score if there is a firm intention to provide obesity prevention interventions for the populations found to be at risk, or to use the evidence to guide public health policies on childhood obesity prevention.

10.2 Limitations and challenges

The diet and health of children is a huge research field and it was hard to develop a literature search strategy which would find relevant records, while minimising the retrieval of irrelevant ones. The literature searches were run in April 2015. If time had allowed, updating the searches to find additional published evidence might have been helpful.

As discussed in Chapter 3, employing two search strategies in five databases was ambitious. The screening of over 6,500 de-duplicated records by two reviewers was time-consuming and laborious. The process was hindered by ambiguous or poorly written abstracts and full texts. Reaching a consensus about which records to include was not straightforward. Strictly, two included papers should have been excluded, based on the criteria:

- Dieting in GUTS was measured over a 3 year period but the association of dieting and change in BMI z scores was only investigated over one year (Field et al., 2003a).
- The first published paper about meal frequency in NGHS girls (Franko et al., 2008) only presented a cross-sectional analysis, but was kept as it was the basis of a later follow-up paper and longitudinal analysis (Ritchie, 2012).

There was also debate about whether to include the German DONALD cohort. Other observational cohorts were "closed" birth cohorts or studies that recruited similar aged children/adolescents at one point in time. DONALD is an "open" cohort that recruits new infant participants each year. As new children are recruited annually, and it not possible to separate out the children who were in the study from the beginning from those who were recruited later, the study method is a hybrid between a repeated cross-sectional and cohort analysis. In analyses, data about children of the same age are combined, even though individual children reached that age at different times. DONALD papers reported "concurrent" change (the annualised change in exposure vs. the annualised change in outcome) and adjusted for residual energy in final analyses. These methodological differences meant that evidence from DONALD was not directly comparable with evidence from other cohorts. However, there were relatively few cohorts from Europe, so the decision was made to keep records from DONALD.

In total 14 childhood cohorts were included, but they were not representative of children worldwide. Studies from the USA predominated (9 cohorts) with three cohorts from Northern European countries (Denmark, Germany, UK), one from Australia and one from Colombia. Undoubtedly the food environment in the USA is unlike anywhere else. Most of the cohorts were established in the 1980s, 1990s or 2000s. Inevitably there is a time-lag between

establishing a cohort, making longitudinal observations, analysing the data and publication of results, but as a result these cohorts are not fully representative of children today.

Half of the included papers came from just three USA cohorts, the large but poorer quality GUTS and Project EAT cohorts, and the higher quality, but single-sex NGHS (girls only) cohort. Other higher and moderate quality cohorts were far less prolific in terms of publication about children's diet and adiposity outcomes. Eight cohorts contributed only one paper to this systematic review.

All included papers employed quantitative or semi-quantitative methods of dietary assessment (see Chapter 4). Although most cohorts employed techniques to improve measurement accuracy, their estimations of dietary intake may still contain substantial error which will cloud any relationship with future adiposity, overweight or obesity.

No cohort validated their chosen DAT against an "ideal" reference measure such as doubly labelled water for energy intake or biomarkers for protein intake, but several used a different DAT (3 day FD or 24-HDR) as a "non-ideal" reference measure to validate an FFQ. There was no reference to any validation study for the FFQ used in the GUTS II cohort. No study described using the findings of validation studies to correct observed results for measurement error, as later recommended by STROBE-nut (Lachat et al., 2016). However, several cohorts assessed mis-reporting by comparing reported energy intake with objectively measured energy expenditure; in some cohorts identified under/over reporters were excluded from subsequent analyses.

Papers examined a breadth of dietary exposures, but few were examined by multiple studies, so the retrieved evidence lacked depth. An additional problem was the lack of common definitions, such as which drinks are included in the term "sugar sweetened beverage" or which foods constitute "snacks". Even when more than one cohort looked at a specific food or drink, they were not direct equivalents.

Papers often used non-dietary family and child factors as confounders in their regression models, while two papers specifically examined such factors as potential predictors of future overweight or obesity. Some non-dietary factors such as early puberty or parental overweight are heritable traits influenced by genetics but included papers did not consider the child's genotype directly. However, studies of identical (monozygotic) twins show that gene-environment interactions influence obesity outcomes. For example, in the FinnTwin16 study several eating behaviour patterns proved moderately heritable and a frequent snacking pattern partly mediated genetic susceptibility to obesity, mainly due to shared genetic factors (Masip et al., 2020). Cohorts of twins were not explicitly excluded from the systematic review

but the few retrieved records about twin studies had not quantified whole diet, which demonstrates a gap in the evidence. As mentioned in Chapter 1, genetic variables (in the form of scores derived from selected single nucleotide polymorphisms or SNPs associated with obesity) have been tested in several childhood obesity risk tools in attempts to enhance predictive accuracy (Morandi et al., 2012; Seyednasrollah et al., 2017). However, data about SNPs can only be obtained via genetic testing of blood samples. Such clinical measures are not suitable for use in population based risk tools, which are designed to use routine and/or easily obtained data.

The included papers used a range of measures to assess children's adiposity outcomes (see Chapter 4). Most studies used height and weight to calculate BMI or change in BMI, which was a useful comparison within the study (if children were the same sex and of a similar age) but not for comparing outcomes between studies. BMI z scores or categorical outcomes (overweight/obesity or not) based on age and sex growth references were more helpful in this regard. One advantage of BMI as a measure of obesity is that height and weight are easily obtained; measured values are more reliable than self-report. However, as BMI is based on excess mass, rather than excess fat, BMI is not a direct measure of adiposity, only an estimate (Simmonds et al., 2015).

The methods used to obtain waist circumference or more direct measures of adiposity, such as body fat percentage, were well described but showed that measurement procedures varied from cohort to cohort. There were also methodological differences in the way that such measurements were processed, so that assessments of children's body fat from (2, 3 or 4) skin fold thicknesses or from DXA were not directly comparable between cohorts. Body fat percentage is difficult to measure accurately even with dual x-ray absorptiometry, while skin fold thickness measurements may contain intra- and inter-observer variability, introducing measurement error. Body fat percentage cut-offs for overweight and obesity in children have not been universally agreed (Freedman et al., 2004), as mentioned in Chapter 4.

Most included papers used regression models but took different approaches to longitudinal analyses. The simplest models looked at baseline dietary exposure and outcome at follow-up. Others factored in change in dietary exposure and/or adiposity outcome. Sometimes the analytical approach was not immediately clear. Better reporting, or the adoption of standard methods with agreed terminology would be helpful.

All the papers included in the systematic review were from observational cohorts, many of which reported significant associations between aspects of diet and adiposity outcomes. However, causality can only be inferred from observational evidence – cause and effect are not certain. A perennial problem with observational studies is that of reverse causality -

although an exposure X and an outcome Y are shown to be associated, instead of X causing a change in Y as expected, Y may be causing a change in X. As presented in Chapters 5 and 6, the possibility of reverse causality was recognised by included papers. For example, higher intakes of diet (low calorie) beverages and fewer eating episodes at baseline were each associated with higher risk of overweight at follow-up, but participants with concerns about their weight status at baseline (not necessarily limited to those who were overweight) might have already altered their diet or eating behaviour in an attempt to control their weight. To address this difficulty, Project EAT participants classified as overweight at baseline were excluded from analyses, although some reverse causality may remain. Other studies acknowledged that reverse causality was still an issue even after adjusting for baseline weight status in analyses.

As a result of the diversity of studied dietary exposures and adiposity outcomes, combined with methodological heterogeneity, the narrative synthesis was complex (see Chapters 5 and 6) and there was frustratingly little data for quantitative synthesis. Only a small meta-analysis of SSB intake and change in BMI was possible (see Chapter 7). However, the systematic review did provide enough evidence to hypothesize which foods and drinks and eating habits might be adverse or beneficial for future adiposity outcomes, and to establish theoretical quantitative cut-offs for categorical intakes of foods and drinks (see Chapter 8). Dietary patterns also provided useful insights, but it was not practical to integrate complex questions about dietary patterns into the questionnaire.

Many of the limitations of the ALSPAC data and the predictive model have already been discussed in Chapters 8 and 9. ALSPAC was the only available dataset with all the measures that were wanted for model development. The ALSPAC birth cohort, established in 1990-1992, is broadly representative of the UK population at that time, but ethnic minorities are under-represented, and ongoing loss-to-follow-up has resulted in an over representation of more affluent and better educated mothers. It is unlikely to be representative of UK 10 year old children and their diets today.

The predictive model's baseline age of 10.5 years ties in with the National Child Measurement Programme, which takes place in England when Year 6 primary school children are aged between 10 and 11 years old. Fortunately, measures of dietary intake were available at age 10.5 years (baseline) for approximately half the children in the ALSPAC cohort used for model development.

The child's BMI z score was selected as the indicator of obesity for the model, using classifications based on the UK90 reference data for age and sex, to match classifications used by the NCMP. Although ALSPAC researchers made assessments of children's body fat percentage using bio-impedance or DXA scans, these direct measures of adiposity contained

more missing observations than height and weight measures, from which obesity status at baseline and follow-up could be ascertained. In hindsight, available measures of body fat percentage could have been employed to validate BMI classifications, as arguably BMI is not the most reliable indicator of overall body fatness (sometimes greater body mass is explained by greater bone density or muscle mass rather than greater fat mass), while adolescent skin fold thickness has been shown to be better than adolescent BMI at predicting adult fatness (Nooyens et al., 2007). Such a validation could be done in future. However, BMI z score is well suited to assessing childhood adiposity on a single occasion (Cole et al., 2005) and age and sex adjusted BMI is recommended by NICE as a practical method of estimating overweight in children and adolescents (NICE., 2006).

Approximately half the cohort had measures of dietary intake at age 10.5 years (baseline) with measures of height and weight available to ascertain obesity status at baseline. More height and weight observations were available for ALSPAC children at age 13.5 years than at age 15.5 years, so a 3 year follow-up period was chosen for model development. Three years is a short time for obesity to manifest and there were relatively few new obesity events in the cohort. This likely reduced the final model's predictive performance. A 5 year follow-up with a higher proportion of new obesity events might have been the better option for model development but the reduced sample size and increased number of missing observations would have introduced a higher risk of bias into the model. Using overweight as the predicted outcome was also considered (more children became overweight, so more events to find) but that would have made it necessary to exclude children with overweight at baseline (in addition to excluding children with obesity), further reducing the sample size available for model development.

Between the ages of 10 and 13 years old children often begin puberty, heralding the adolescent developmental stage typified by rapid increases in body size and changes in body composition (Adair, 2008). Adolescence is also a time of change in lifestyle and behaviour as young people become more independent of their parents/carers and start to exercise more autonomy. The many physical and behavioural changes in adolescence make it difficult to research, but they also present an opportunity for obesity prevention interventions, especially if at risk populations can be identified, which is the purpose of the predictive model. The updated Cochrane review of interventions for preventing childhood obesity (Brown et al., 2019) cited in Chapter 1, found that combined diet and physical activity interventions may be effective in older children and adolescents.

A practical limitation of the predictive model is that it can only consider diet and other factors at one time (baseline) and obesity at one later time (follow up). Whilst diet/eating habits may

lead to changes in adiposity/body mass, in turn adiposity/body mass may lead to changes in diet/eating habits, and so on, in a continuing loop through time. It was a challenge to unpick this ongoing "causal spiral", even with the help of a directed acyclic graph (DAG). The ensuing model is necessarily a simplification. Being predictive, the model cannot account for changes between baseline and follow-up. It must be recognised that, although predictors such as early puberty (or not) are fixed at baseline, other predictors such as dietary and behavioural factors will change, and those changes may influence the outcome so that the baseline prediction proves wrong. Additionally, adiposity fluctuates naturally during growth and some adolescents might experience obesity at earlier time points, even if they are not classified as obese (or predicted to have obesity) by follow-up.

10.3 Strengths

This is the first predictive model of childhood obesity to include detailed measures of diet. Furthermore, the predictive model and resulting CORA risk score is evidence based. All the potential predictors tested in the model were chosen based on reports from a wide-ranging systematic review, enhanced by evidence from other published childhood obesity risk prediction models and corroborated by more focussed systematic reviews and meta-analyses in the published research literature.

A protocol for the systematic review was written and published beforehand, and deviations from the protocol are reported, as recommended by PRISMA (Moher et al., 2009). The systematic review had a comprehensive literature search strategy, developed using a PICO-S framework and applied in five bibliographic databases to reduce bias. All stages of the screening process were carried out in duplicate by two independent reviewers. For consistency, piloted screening questionnaires and a data extraction form based on a Cochrane template were used. Disagreements were resolved by a third independent reviewer.

PRISMA reporting guidelines were followed throughout and the numbers of records retrieved, screened and included in the review, with reasons for exclusions at each stage, were summarised with a flow diagram. The literature searches retrieved a broad range of evidence about children's dietary exposures (food and drink intakes, eating habits and dietary patterns) and future adiposity as intended. Different statistical methods were applied for different purposes, which was challenging for data extraction, but also a valuable learning opportunity.

Unexpectedly, some papers that met the inclusion criteria considered both dietary and non-diet predictors of BMI change, future overweight or future obesity, using logistic regression or regression tree analysis. This proved fortuitous, as the additional evidence justified inclusion of non-diet predictors as well as dietary predictors of future obesity in CORA. Non-diet

variables undoubtedly strengthened the predictive model's performance and elevated the tool to more than just a diet quality score.

A major strength of the model development process is that an analysis plan was written, and all putative predictors were pre-specified and matched to candidate variables before exploring exposure-outcome associations or attempting any model fitting. The informal study protocol was improved by the inclusion of a directed acyclic graph which set out assumptions about causal relationships (presumed to be predictive). A formal protocol was not registered or published, although this practice is encouraged by TRIPOD. Other TRIPOD guidelines were followed during model development and internal validation (Collinset al., 2015).

The ALSPAC dataset used for model development is from a good quality UK birth cohort of 14,755 live born children, which measured and quantified children's diet well. Data was pre-cleaned by the University of Bristol. 7,462 children completed a 3 day food diary at age 10+ years, with parental help. Based on the ratio of reported Energy intake: Estimated Energy requirement, 42% of 10 year old children had plausible dietary intakes (Noel et al., 2011). Children's heights and weights were measured by trained personnel at clinics at age 10+ years and 13+ years. Loss to follow-up was low, but participants that were lost were often children with the hypothesised prognostic/risk factors for future obesity that were the putative predictors of interest.

A further strength is that imputation was used to keep as many children in the sample as possible, as not all children had all the potential predictors. Single imputation methods were used to fill the gaps intelligently, by interpolating an estimated value from other observations about the same individual, or by applying unconditional mean imputation (replacing missing values with the estimated mean from available cases). Single imputation is conceptually simple, but underestimates standard error and can add bias, the extent of which depends how much data is missing, the other information in the dataset and the reason for the missing data, as discussed in Chapter 8.

The single imputation methods required some exploration of the data in order to make assumptions. Hence the model development process was not completely "blind". In addition, papers from ALSPAC were included in the systematic review and ALSPAC data had been used elsewhere to develop obesity risk tools, so the model developer had fore-knowledge of variables that might prove to be useful predictors in the ALSPAC dataset.

Obesity status could be derived for most children at 10+ years but some children were missing height and weight measures needed to calculate BMI z-scores and obesity status at 13+ years. Where possible missing values were imputed from previous and/or later measures of height

and weight, using unconditional means, to generate "best estimates" of BMI z-scores and obesity outcomes. Imputation achieved the aim of keeping more children in the model but added some uncertainty to some of the obesity outcomes sought by the model.

After imputation there were 5,486 children in the ALSPAC sample, enough for a split-sample for internal validation, albeit a slightly smaller number than expected. Children were randomly allocated 3:1 into a derivation sample and a validation sample. This simple approach saved time but restricted the size of the sample available for model development. Alternative methods such as k-fold cross-validation or boot-strapping would have used the data more efficiently.

A further methodological forte is that two strategies were used to select covariates for the predictive model. The first used a series of models using stepwise removal/addition. The second "purposeful selection" strategy, run in a smaller sample, confirmed the selection.

The careful consideration of potential predictors using a DAG highlighted the inappropriateness of using sex or age as predictors, as both variables were used to derive the BMI classification which was the outcome. Although the predictive model was created with the Year 6 age group in mind, calendar age was not used as a predictor and it may be possible to recalibrate the model for use at other ages.

Elsewhere, biological age, rather than calendar age, has been shown to be the better predictor of obesity risk in children. Girls who experienced menarche before age 11 years were more than twice as likely to be overweight in young adulthood (Adair and Gordon-Larsen, 2001), while early puberty in boys predicted greater central adiposity in young adult males (Kindblom et al., 2006). Early puberty made an important predictive contribution to our model but may simply be a proxy for higher baseline BMI (even though not obese), as a higher pre-pubertal BMI has been linked to earlier puberty in girls (Davison et al., 2003). It is not clear whether early puberty acts independently of pre-pubertal BMI as a risk factor for future obesity (Must et al., 2005).

A novel attribute of the predictive model is the use of an interaction term, as explained in Chapter 9. The interaction showed that, in terms of a child's future obesity risk, some of the benefit of eating vegetables seemed to be lost if the child consumed 1 or more servings/day of sugary drinks such as cola, fizzy drinks and sweetened fruit drinks.

As recommended by the ABCD framework (Steyerberg and Vergouwe, 2014), all aspects of the final model's performance (discrimination, calibration, clinical usefulness) in the derivation, validation and combined samples, are presented for the internal validation of the model. Model performance was equal to published childhood obesity risk models and the model has a

level of accuracy that is adequate for its intended purpose of screening populations, but not for use with individuals.

10.4 Future research

The Children's Obesity Risk Assessment requires further work before it can be put to its intended use as a population screening tool. As a first step a paper version of the questionnaire could be piloted in 10 year old children, to assess their comprehension and whether the current questionnaire and scoring system is practical or overlong. The questionnaire could later be converted to an on-line or digital version, with automatically calculated scoring.

Although there were relatively few new obesity events in the derivation sample used for model development in ALSPAC (168 out of 4,114, or 4%) more girls than boys (102 vs. 66) who were not classified as obese at age 10+ years, went on to have obesity by 13+ years. With more events to find, the model may be better at predicting future obesity in girls rather than boys, which could be checked by a sensitivity analysis. CORA has been developed for use with both sexes, but it may need to be calibrated differently for girls and boys.

From the National Child Measurement programme, we know that 10 year old children living in the most deprived areas of England are more likely to experience overweight/obesity than children living in the least deprived areas. The ALSPAC dataset contained an Index of Multiple Deprivation (IMD) variable about the area deprivation level for each child's home, but IMD was not considered as a candidate predictor in the model, as it is a complex multi-faceted variable, containing elements of other socio-economic variables that were tried instead. It would be interesting to test whether a CORA risk score makes better predictions of obesity outcomes than IMD alone.

Following this, the risk model/score and questionnaire requires external validation. The performance of a developed model should be evaluated in different participants, by using the published regression formula (risk algorithm) to predict outcomes for each individual and checking the prediction against the observed outcome. The participant data may come from the same original source, but from a different time (temporal validation) which may be feasible with the ALSPAC data, as diet was measured again at 13+ years and teenagers' heights and weights were measured at later clinics.

Preferably the external validation should be done with similar participants but in a different setting, population or dataset. If the model performs poorly it can be adjusted or recalibrated to better suit the validation dataset. It will be a challenge to find a cohort that has measures of all the predictors in CORA but one with more new obesity "events" might be helpful.

After external validation and feasibility-testing, wide scale implementation of CORA as a population screening tool will only be worthwhile if the risk score leads to better policy decisions that reduce the prevalence of obesity outcomes or make interventions more cost-effective. Ideally a comparative study or impact assessment should be run, comparing decision making and obesity outcomes with and without the risk score.

Taking a wider view, this thesis has encountered several areas where more research and improved practices are needed, to better understand the relationship between children's diets and their future weight status, as well as other health outcomes. Observational studies and cross-sectional surveys indicate that many children have mean intakes of fat that exceed WHO recommendations. There is evidence that children and adolescents in the USA, Australia and Northern European countries who follow energy dense dietary patterns, with high intakes of sugar sweetened drinks or high fat foods, and eating habits that reinforce those high intakes, are more susceptible to obesity or overweight. Obesity in childhood is linked to musculoskeletal problems, an increased risk of asthma, high blood pressure, dyslipidaemia and insulin resistance which may lead to poor cardiovascular health in adulthood. Childhood or adolescent obesity/overweight and the habits that promote it often track into adulthood.

Most of the evidence we found about children's diets and future adiposity originated from cohorts located in the USA and Northern European countries. Some cohorts included in the systematic review diet only contributed one or two papers, suggesting that there is a wealth of data that could (and should) be exploited further. Possibly the research is done, but papers are still to be written, submitted and published.

Based on the published literature there are surprisingly few longitudinal, observational cohorts of healthy children that have quantified children's diet. The limited evidence about children's diets in the USA and Northern Europe is unlikely to be representative of children's diets elsewhere in the world. To widen the evidence, more studies are needed from other countries. Moderate to higher quality cohorts that assessed children annually or over several years and had low attrition rates generated the most convincing evidence in the systematic review. As resources allow, well run medium sized (>500) cohorts should be established in preference to larger but more poorly executed cohorts. Additionally, with amendments to the study protocol and with participants consent, existing longitudinal children's cohorts including twin studies could add quantitative dietary assessment to their repertoire of measures.

It is hard to measure diet well. The traditional paper-based DATS used by included cohorts contained measurement error, which likely obscured some of the associations of dietary intake with adiposity outcomes, but this was not always acknowledged. Based on comparisons of reported energy intake vs energy requirements/expenditure, extensive mis-reporting of

dietary intake was commonplace even in studies which had taken practical steps to reduce it. Better dietary assessment methods are needed. New technologies may improve dietary assessment, by easing the respondent and researcher burden, and possibly reducing error. However, it is likely that some biases will remain with subjective self-reports of diet, so there will still be a need for validation studies. Comparing the chosen DAT against an "ideal" standard such as doubly labelled water or biomarkers is recommended, making more use of the findings to correct observations for measurement error. This might increase the validity of measures of diet quantity and quality.

Diet can be quantitatively measured and assessed at the level of micro and macronutrients, as specific foods or drinks, or as a dietary pattern. Eating habits are more usually based on frequency, rather than quantity of intake. All these approaches to assessing diet are helpful and were used by included papers in the systematic review. However, coverage of specific exposures was often limited to only one or two studies. More investigation of specific dietary exposures and adiposity outcomes in children and adolescents is required. Further investigations of juice and of fruit (separately from vegetables) and dairy products (separately from milk) are needed. (Updating the systematic review literature search may show that these research gaps have been addressed.)

In academic research, researchers are often encouraged to produce and publish novel findings or use novel methods., but from the perspective of a systematic reviewer, researchers should not shy away from replicating others research. Following the same methods in a different population will facilitate quantitative synthesis and add to the evidence.

One of the difficulties encountered when attempting meta-analysis was the multiplicity of methods to measure adiposity in children. It would help if nutrition researchers adopted common protocols for measuring waist circumference or agreed standard methods for assessing body fat from skin fold thicknesses. The best comparative measure of adiposity in children of different ages is undoubtedly the BMI z-score, ideally using the IOTF growth reference to allow international comparisons. With modern statistical packages BMI can readily be converted to BMIz based on a suitable age and sex specific growth reference. BMIz and change in BMIz should be used more often when investigating children's adiposity.

Again, from the perspective of a systematic reviewer, poor reporting of methods in abstracts and full texts was a problem. The recently issued STROBE-nut guidelines may help improve reporting. Authors should refer to these guidelines and consider the use of supplementary material if publication space is limited. Greater care in describing exactly what was measured and the statistical methods used to explore any association would remove ambiguity. Is the

exposure at baseline only, or is it change in exposure over time? Is the outcome only at follow-up, or does the study look at change in outcomes?

In nutrition research it is usual to adjust for Total Energy intake, but often this attenuates the effect size, sometimes making an apparent association no longer significant. There is an argument for not adjusting for Total Energy intake when Energy intake is assumed to be part of the mechanism contributing to future overweight. However, I believe it is better practice to make no á priori assumptions, instead testing models with *and* without adjustment for TEI. Results of both models can then be presented in the interests of transparency, and for ease of comparison and data synthesis. Adjusting models for residual Energy is a sensible alternative, but unless this technique is more widely adopted, it will be less helpful for comparative purposes and meta-analysis. In either case the share of TEI made by the dietary exposure under investigation is an informative measure for comparing differences between cohorts. The reporting of the share of TEI made by a specific dietary exposure should be encouraged.

The predictive model described in this thesis was developed and internally validated with secondary data from the UK ALSPAC cohort using single imputation methods, a split sample for validation and traditional step-wise statistical modelling techniques. One benefit of applying these simple methods is that they required learning the key principles of predictive modelling and a good understanding of the dataset. The model's predictive performance was acceptable, but a combination of better quality data and more sophisticated modelling techniques offer routes to improvement in the future.

The dietary data used for the model (collected 20 years ago) may not be representative of children's diets in other countries, or indeed in the UK now or in the future. As discussed in Chapter 8, the ALSPAC data contains different kinds of bias including loss to follow-up and mis-reporting of food and drink intake. Such biases restrict the model's predictive performance. In future, an improved predictive model could be developed with more up-to-date data using measures of diet quantity and quality which have greater validity. Better dietary data might allow the exploration of non-linear relationships between food/drink intakes and adiposity outcomes to find optimal intakes, rather than estimating a quantitative or frequency threshold for greater adiposity risk based on an assumed linear relationship.

The chosen outcome variable for prediction was a classification of obesity based on a BMI z score, which gives an estimate of adiposity. In future, it may be more usual for children's body fat percentage to be measured directly and more accurately by BIA or DXA which will increase the validity of the adiposity outcome measure. More certain measures of adiposity could also improve the model's predictive performance.

As described in Chapter 9, multiple imputation methods and resampling techniques such as k-fold cross-validation or boot strapping for validation are more efficient ways of using the available data and should be considered for future predictive modelling research. A possible alternative to traditional statistical modelling techniques is to employ a branch of artificial intelligence known as machine learning. Machine learning is based on the premise that systems can learn to identify patterns in the data and automatically improve the ensuing computer algorithms with little human intervention. Machine learning has already been employed in clinical prediction models and shown promise, significantly improving the accuracy of cardiovascular risk prediction (Weng et al., 2017). It may have a place in developing predictive algorithms for use in population settings too. However, a recent article in the Lancet cautioned that the utility and performance of any machine learning algorithm (whether for diagnosis or prognosis) is "highly dependent on the quality and relevance of the data on which it is trained" and called for collaboration between "traditional methodologists and experts in machine learning" to avoid wasted research effort (Wilkinson et al., 2020).

10.5 Public health implications

Obesity is a social, environmental, economic and political issue. Preventing obesity is a major public health challenge, but obesity is a complex condition, with many underlying causes. It has recently been argued that some of the risks associated with non-communicable diseases and obesity outcomes have gone unchecked "in the name of economic growth and free trade" (Oni et al., 2019).

A first step towards preventing obesity is to identify risk factors. By synthesising published longitudinal research about childhood and adolescent diet and overweight/obesity outcomes, the systematic review has given a better understanding of some of the determinants of obesity, adding evidence to the IOTF framework as intended.

The systematic review provided evidence that energy dense dietary patterns in childhood and adolescence are associated with an increased likelihood of overweight or obesity. This finding was embedded in the dietary assessment questionnaire, CORA. The resulting predictive model included dietary variables that contributed to energy dense dietary patterns such as sugary drinks and foods high in sugar and/or fat such as snacks and dairy foods, but also included vegetables and milk, which are less energy dense. The model demonstrated an interaction between vegetable intake and sugary drinks frequency, which indicates that in terms of a child's future obesity risk, some of the benefit of eating vegetables is lost if the child also has a sugary drink every day. This may be an important public health message to promote to children and their parents/carers, but the mechanism is uncertain and merits further

investigation. Elsewhere it has been postulated that sugary drinks add energy to the diet without satisfying the appetite, resulting in little compensatory reduction in energy intake from other sources (Malik et al., 2006).

A second step towards preventing obesity is to identify those at risk. This thesis has demonstrated a prognostic tool to identify at risk groups of children who might benefit from obesity prevention measures, adding further evidence to the IOTF framework. Once externally validated and proved reliable, the CORA questionnaire and risk score could be used to identify high risk populations of children and to intervene *before* they experience obesity/overweight. Additionally, the CORA questionnaire could be used to evaluate the effectiveness of a targeted public health message or intervention, by comparing risk scores before and afterwards. Although the model on which the CORA risk score is based is relatively simple, with only nine predictors and one interaction term (using 12 degrees of freedom), the Pearson goodness-of-fit test found 428 different covariate patterns in children kept in the model in the derivation sample (n = 3,171) and 283 different covariate patterns in children kept in the model in the validation sample (n = 1,077). The model only explains some of the risk. This illustrates the difficulties encountered when trying to make predictions. Each child's weight status comes about through a unique combination of contributing factors (known and unknown). Obesity prediction will likely remain probabilistic, as we do not know exactly which risk factors should be measured, or exactly when. Although scientists may find ways to improve predictions, including machine learning, the complexity of obesity will limit the predictive capability of childhood obesity risk tools.

The next step towards preventing obesity is to act. Approaches to obesity prevention include "downstream" individual intervention measures, such as the diet and physical activity interventions featured in the updated 2019 Cochrane systematic review of interventions for preventing obesity in children (Brown et al., 2019) mentioned in Chapter 1. Other methods for obesity prevention include "upstream" intervention measures, such as the Soft Drinks Industry Levy (SDIL) introduced by the UK government in 2018 as part of their Childhood obesity strategy. Soft drink manufacturers were given two years' advance warning that they would be taxed 24p a litre on drinks containing 8g of sugar per 100ml and 18p a litre on those with 5-8g of sugar per 100ml. The aim was to reduce sugar consumption by encouraging manufacturers to reformulate their high sugar soft drinks and avoid paying the levy. A recent study shows that the SDIL did incentivise many manufacturers to cut the sugar content of soft drinks, but some of the cost of the levy on high sugar drinks was passed on to consumers, directly or indirectly (Scarborough et al., 2020). Although this may have the intended effect of reducing

population exposure to sugary drinks, it is too soon to tell whether the SDIL intervention will have an impact on UK childhood obesity levels.

No single action will make a difference. In order to reduce the current high prevalence of obesity, Governments and public health professionals may need to adopt a strategy akin "marginal gains" approach used by British Cycling's performance director Dave Brailsford, who attributed the success of the British track cycling team in the 2012 London Olympics to an accumulation of small improvements, which became "significant" only when they were all put together (Slater, 2012). One option is the "small-changes approach" advocated by James Hill and the Joint Task Force of the American Society for Nutrition, the Institute of Food Technologists and the International Food Information Council (Hill, 2009). They pointed out that obesity rates continue to rise in most countries, driven by gradual weight gain across populations. Initiatives to stop the upward trend have failed thus far. Hill suggests that this is due to a focus on weight loss, requiring large and permanent changes in lifestyle that are very hard for individuals who are already overweight or obese to maintain. Instead Hill and colleagues proposed that countries switch to promoting small, sustainable changes in diet and physical activity, such as simple food substitutions or walking more steps a day, to prevent further gradual weight gain in populations and individuals, including adults with overweight or of normal weight. Initially, this could help to stabilise obesity rates. Then, by helping people to keep making conscious small changes, and by working with the public and private sectors to reduce environmental factors that contribute to excess energy intake and reduced physical activity, obesity rates might gradually be reduced. The small-changes approach is intended to be a "unifying platform" which gives all parties credit for making positive changes, rather than blaming existing forces. Such a campaign would undoubtedly require long-term government input and policy changes, reinforced by educational support and social marketing, perhaps over a period of decades.

"Prediction is very difficult, especially about the future."

Attributed to Niels Henrik David Bohr (7 October 1885 – 18 November 1962), Danish physicist, possibly based on an old Danish proverb.

Appendix A Protocol registered with PROSPERO

UNIVERSITY of York
Centre for Reviews and Dissemination

NHS
National Institute for Health Research

PROSPERO International prospective register of systematic reviews

Review title and timescale

Review title
Give the working title of the review. This must be in English. Ideally it should state succinctly the interventions or exposures being reviewed and the associated health or social problem being addressed in the review.
Childhood and adolescent cohorts which measure whole diet and subsequent adiposity: a systematic review and meta-analysis

Original language title
For reviews in languages other than English, this field should be used to enter the title in the language of the review. This will be displayed together with the English language title.

Anticipated or actual start date
Give the date when the systematic review commenced, or is expected to commence.
07/04/2015

Anticipated completion date
Give the date by which the review is expected to be completed.
31/03/2016

Stage of review at time of this submission
Indicate the stage of progress of the review by ticking the relevant boxes. Reviews that have progressed beyond the point of completing data extraction at the time of initial registration are not eligible for inclusion in PROSPERO. This field should be updated when any amendments are made to a published record.

The review has not yet started ×

Review stage	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	Yes	No
Data extraction	No	No
Risk of bias (quality) assessment	No	No
Data analysis	No	No

Provide any other relevant information about the stage of the review here.

Review team details

Named contact
The named contact acts as the guarantor for the accuracy of the information presented in the register record.
Catherine Rycroft

Named contact email
Enter the electronic mail address of the named contact.
fs11cr@leeds.ac.uk

Named contact address
Enter the full postal address for the named contact.
Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds, Leeds, LS2 9JT.

Named contact phone number
Enter the telephone number for the named contact, including international dialing code.
0113 3439581

Organisational affiliation of the review
Full title of the organisational affiliations for this review, and website address if available. This field may be completed as 'None' if the review is not affiliated to any organisation.
Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds
Website address:

Review team members and their organisational affiliations

Give the title, first name and last name of all members of the team working directly on the review. Give the organisational affiliations of each member of the review team.

Title	First name	Last name	Affiliation
Mrs	Catherine	Rycroft	Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds
Dr	Charlotte	Evans	Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds
Professor	Janet	Cade	Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds

Funding sources/sponsors

Give details of the individuals, organizations, groups or other legal entities who take responsibility for initiating, managing, sponsoring and/or financing the review. Any unique identification numbers assigned to the review by the individuals or bodies listed should be included.

No funding other than an Emma and Leslie Reid studentship used to fund a doctoral programme at the University of Leeds

Conflicts of interest

List any conditions that could lead to actual or perceived undue influence on judgements concerning the main topic investigated in the review.

Are there any actual or potential conflicts of interest?

None known

Collaborators

Give the name, affiliation and role of any individuals or organisations who are working on the review but who are not listed as review team members.

Title	First name	Last name	Organisation details
Miss	Marion	Hery	National School of Agricultural Sciences in Bordeaux
Mr	Benjamin	Clapinson	School of Food Science and Nutrition, University of Leeds

Review methods

Review question(s)

State the question(s) to be addressed / review objectives. Please complete a separate box for each question.

To what extent does diet during childhood or adolescence influence future indicators of overweight or obesity?

Searches

Give details of the sources to be searched, and any restrictions (e.g. language or publication period). The full search strategy is not required, but may be supplied as a link or attachment.

The following databases will be searched (in the order shown) for relevant articles about human studies published from January 1990, with no limit on language. - Ovid Medline - Embase - Cochrane central register of controlled trials, CENTRAL (only for trials) - Scopus - Web of Science (ISI web of Knowledge) Reference lists of relevant reviews and papers will be cross checked for relevant articles not found by the database searches.

URL to search strategy

If you have one, give the link to your search strategy here. Alternatively you can e-mail this to PROSPERO and we will store and link to it.

I give permission for this file to be made publicly available

Yes

Condition or domain being studied

Give a short description of the disease, condition or healthcare domain being studied. This could include health and wellbeing outcomes.

The condition being studied is difference in body fatness or adiposity across populations or groups

Participants/population

Give summary criteria for the participants or populations being studied by the review. The preferred format includes details of both inclusion and exclusion criteria.

Inclusion: Studies which include healthy children or adolescents aged from 8 to 19 years at baseline which have followed up participants for a minimum of two years. Mixed-age group studies will be included if half or more of the participants are aged from 8 to 19 years at baseline. Single and mixed sex studies will be included. Participants may be underweight, normal weight, overweight or obese at baseline, reflecting the range of weights that exist within populations. Exclusion: Studies where more than half of the participants are aged less than 8 years at baseline. Studies where more than half of participants are adults aged over 19 years at baseline. Studies of specific groups which are not representative of the population. E.g. Children born pre-term, children recruited due to overweight or obesity at baseline, children with diabetes or other long-term health conditions, vegans, small ethnic minority groups.

Intervention(s), exposure(s)

Give full and clear descriptions of the nature of the interventions or the exposures to be reviewed

The primary exposure being studied is whole diet, or the usual intake of food and nutrients, which may be characterised as a dietary pattern using collected dietary information. Inclusion: Studies which have used an objective measure of whole dietary intake, derived from either a weighed or un-weighed diet diary, 24 hour diet recall or a quantitative or semi-quantitative food frequency questionnaire, from which partial or total intake of foods and drink can be quantified. Exclusion: Studies which have not measured whole diet or dietary pattern. Studies specifically of breast feeding. Studies which have used a diet history method to gauge usual rather than actual intake. Studies which have used a tool in which consumption of food and drink is not quantified (e.g. Simple food frequency questionnaire, Diet Quality index or dietary habits such as breakfast eating)

Comparator(s)/control
Where relevant, give details of the alternatives against which the main subject/topic of the review will be compared (e.g. another intervention or a non-exposed control group).
Any

Types of study to be included initially
Give details of the study designs to be included in the review. If there are no restrictions on the types of study design eligible for inclusion, this should be stated.
This review will include: Observational, longitudinal or cohort studies or clinical trials that have measured diet and anthropometry, measuring anthropometry at a later point in time (at least two years later) from diet. Studies that have been published from 1990 (cohort studies) or 2010 (clinical trials). This review will exclude: Studies where only the abstract is available. Studies where the abstract or the full paper is not available in English. Intervention studies with no untreated or placebo control group. E.g. Studies including promotion of healthy eating, physical activity or weight loss. Cohort studies or clinical trials of less than two years duration.

Context
Give summary details of the setting and other relevant characteristics which help define the inclusion or exclusion criteria.
Any

Primary outcome(s)
Give the most important outcomes.
The primary outcome of interest is adiposity, reported as underweight, normal weight, overweight or obesity (categorical variables), body mass index or body mass index z-score, waist circumference or height: waist circumference ratio (continuous variables). Inclusion: Studies which have used anthropometric measurements from which an assessment of adiposity or overweight/obesity status can be made. As a minimum this will include self-reported height and weight. Exclusion: Studies which have not used any type of anthropometric measurement. Studies which have not measured anthropometry at least two years after measuring diet.
Give information on timing and effect measures, as appropriate.
Studies must have measured anthropometry at least two years after measuring diet

Secondary outcomes
List any additional outcomes that will be addressed. If there are no secondary outcomes enter None.
The secondary outcomes of interest are indicators of heart health or metabolic syndrome, such as blood pressure, blood lipids or insulin resistance, if they have been measured.

Give information on timing and effect measures, as appropriate.
If indicators of heart health or metabolic syndrome have been measured, this must be at least two years after measuring diet

Data extraction, (selection and coding)
Give the procedure for selecting studies for the review and extracting data, including the number of researchers involved and how discrepancies will be resolved. List the data to be extracted.
Bibliographic details of all records found using the search strategy will be imported into an Endnote library. Where duplicate records occur, the first imported record will be kept. All papers will be sorted by publication date; any that pre-date 1990 will be excluded. Keyword searches in the title field will be used to identify clearly irrelevant papers, which will also be excluded. The title and abstract of all remaining records will be independently screened by two trained reviewers. Where there are differences between reviewers' decisions a third reviewer will be consulted to resolve matters. A copy of the full article will be sought for each potentially relevant record. If articles are unavailable at the University of Leeds and cannot be obtained elsewhere, authors will be contacted electronically. It is anticipated that potentially relevant studies may have multiple papers; any related papers will be grouped by their common study. All groups of full text articles will be independently screened by two trained reviewers. Any differences will be resolved by discussion and consultation with a third reviewer. For each of the studies which meet the inclusion criteria the following characteristics will be extracted (if reported) in duplicate using a modified Cochrane data extraction template: Study/cohort name, study type, study aims, linked papers, decades studied (when established and times/length of follow up), country, setting of study, sampling frame, sample size and attrition rates, sample population age, gender, ethnicity, socio-economic measures, exposure (food or nutrient intake or dietary pattern), dietary assessment method (validated or not) with timings/ages at assessment, measurements of potential confounders (physical activity, sedentary behaviours, risk behaviours including diet related behaviours) and primary and secondary outcomes of interest with methods (validated or not) and timings/ages at assessment. Extracted data will be entered into Review Manager 5 software.

Risk of bias (quality) assessment
State whether and how risk of bias will be assessed, how the quality of individual studies will be assessed, and whether and how this will influence the planned synthesis.
A modified NICE Quality appraisal checklist will be used to assess the internal validity of each study.

Strategy for data synthesis
Give the planned general approach to be used, for example whether the data to be used will be aggregate or at the level of individual participants, and whether a quantitative or narrative (descriptive) synthesis is planned. Where appropriate a brief outline of analytic approach should be given.
The opportunity to carry out a statistical synthesis of the study results will be explored for different dietary factors. If appropriate a meta-analysis will be carried out for each dietary factor with results presented as forest plots for all eligible studies in the review. Heterogeneity across studies will be checked using the I-squared test, with an I2 of 25-50% indicating moderate heterogeneity, an I2 of 50-75% indicating substantial heterogeneity and an I2 above 75% indicating considerable heterogeneity. It is expected that heterogeneity will be above 50% based on previous research. Meta-regression will be carried out using factors potentially having an impact on heterogeneity such as age, dietary assessment method and weight at baseline in order to attempt to explain some of the heterogeneity. If a quantitative synthesis is not feasible, a narrative synthesis including a discussion of study quality will be produced.

Analysis of subgroups or subsets
Give any planned exploration of subgroups or subsets within the review. 'None planned' is a valid response if no subgroup analyses are planned.
No subgroup analysis is planned

Review general information

Type of review
Select the type of review from the drop down list.
Epidemiologic

Language
Select the language(s) in which the review is being written and will be made available, from the drop down list. Use the control key to select more than one language.
English
Will a summary/abstract be made available in English?
Yes

Country
Select the country in which the review is being carried out from the drop down list. For multi-national collaborations select all the countries involved. Use the control key to select more than one country.
England

Other registration details
Give the name of any organisation where the systematic review title or protocol is registered together with any unique identification number assigned. If extracted data will be stored and made available through a repository such as the Systematic Review Data Repository (SRDR), details and a link should be included here.

Reference and/or URL for published protocol
Give the citation for the published protocol, if there is one.
Protocol not yet published
Give the link to the published protocol, if there is one. This may be to an external site or to a protocol deposited with CRD in pdf format.

I give permission for this file to be made publicly available
Yes

Dissemination plans
Give brief details of plans for communicating essential messages from the review to the appropriate audiences.
Do you intend to publish the review on completion?
Yes

Keywords
Give words or phrases that best describe the review. (One word per box, create a new box for each term)
child
adolecent
cohort
dietary intake
adiposity

Details of any existing review of the same topic by the same authors
Give details of earlier versions of the systematic review if an update of an existing review is being registered, including full bibliographic reference if possible.

Current review status
Review status should be updated when the review is completed and when it is published.
Ongoing

Any additional information
Provide any further information the review team consider relevant to the registration of the review.

Details of final report/publication(s)
This field should be left empty until details of the completed review are available.
Give the full citation for the final report or publication of the systematic review.
Give the URL where available.

Appendix B Database searches

B.1 Search in Cochrane Library

Search Name: COCHRANE Refocused Combined Strategy 10th April 2015

Date Run: 10/04/15 10:48:44.292

Description: 10th April 2015

ID Search Hits

#1 MeSH descriptor: [Child] explode all trees 156

#2 "child*":ti,ab,kw (Word variations have been searched) 80796

#3 MeSH descriptor: [Adolescent] explode all trees 77199

#4 "adolescen*":ti,ab,kw (Word variations have been searched) 95786

#5 "teen*":ti,ab,kw (Word variations have been searched) 970

#6 #1 or #2 or #3 or #4 or #5 145136

#7 MeSH descriptor: [Diet] explode all trees 12400

#8 "diet*":ti,ab,kw (Word variations have been searched) 41495

#9 MeSH descriptor: [Diet Records] explode all trees 511

#10 "food diar*":ti,ab,kw (Word variations have been searched) 194

#11 "food record*":ti,ab,kw (Word variations have been searched) 409

#12 "food frequency questionnaire*" or "ffq*":ti,ab,kw (Word variations have been searched) 579

#13 "24 hour recall*" or "twenty four hour recall*" or "24 hr recall*":ti,ab,kw (Word variations have been searched) 122

#14 MeSH descriptor: [Energy Intake] explode all trees 3590

#15 "energy intake*" or "food intake*":ti,ab,kw 6974

#16 MeSH descriptor: [Nutrition Assessment] explode all trees 422

#17 "nutrition*-assessment*":ti,ab,kw 778

#18 "energy dens*":ti,ab,kw 403

#19 #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #16 or #17 or #18 44299

#20 MeSH descriptor: [Anthropometry] explode all trees 15709

#21 "anthropometr*":ti,ab,kw 4613

#22 MeSH descriptor: [Body Composition] explode all trees 3242

#23 "body composition*" or "body fat*" or "adipos*" or "fat percentage*" or "fat mass":ti,ab,kw (Word variations have been searched) 8358

#24 MeSH descriptor: [Body Height] explode all trees 1226

#25 "height*":ti,ab,kw 8111

#26 MeSH descriptor: [Body Weight] explode all trees 16541

#27 "body-weight*":ti,ab,kw (Word variations have been searched) 5

#28 (#24 or #25) and (#26 or #27) 1205

#29 MeSH descriptor: [Body Mass Index] explode all trees 6162

#30 "body mass index" or "bmi":ti,ab,kw (Word variations have been searched) 18572

#31 "z score*":ti,ab,kw (Word variations have been searched) 1219

#32 MeSH descriptor: [Waist Circumference] explode all trees 429

#33 MeSH descriptor: [Waist-Hip Ratio] explode all trees 191

#34 "waist circumference*" or "waist to hip ratio*":ti,ab,kw (Word variations have been searched) 2649

#35 (#20-#23 or #28-#34) 34217

#36 MeSH descriptor: [Follow-Up Studies] explode all trees 44179

#37 "follow up":ti,ab,kw (Word variations have been searched) 112030

#38 #36 or #37 112030

#39 #6 and #19 and #35 and #38 556

All Results (556)

Cochrane Reviews (4)

Other Reviews (4)

Trials (547) from Cochrane central register of controlled trials (CENTRAL) – sorted by date, exported records 1 to **266**, March 2015 to 2010 inclusive.

Method Studies (0)

Technology Assessments (0)

Economic Evaluations (1)

Cochrane Groups (0)

B.2 Searches in Medline

B.2.1 MED Refocused combined strategies for cohort with limits

Database(s): **Ovid MEDLINE(R)** 1946 to April Week 1 2015

Search Strategy: Run on 9th April 2015

#	Searches	Results
1	exp Child/	1574433
2	child$.mp.	1880547
3	1 or 2	1880547
4	exp Adolescent/	1648687
5	adolescen$.mp.	1673202
6	teen$.mp.	21044
7	4 or 5 or 6	1675746
8	3 or 7	2732527
9	exp Diet/	201626
10	diet$.tw.	372632
11	exp Diet Records/	4231
12	(food adj (diary or diaries)).tw.	829
13	"food-record$".tw.	1980
14	"food-frequency-questionnaire$".tw.	6678
15	"ffq$".tw.	1758
16	"24-hour-recall$".tw.	1041
17	"twenty-four-hour-recall$".tw.	24
18	"24-hr-recall$".tw.	79
19	exp Energy Intake/	36093
20	"energy-intake$".tw.	14392
21	"food-intake$".tw.	32635
22	exp Nutrition Assessment/	8337
23	"nutrition$-assessment$".tw.	3050
24	(energy adj1 dens$).tw.	3941
25	9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24	492131

26	exp Anthropometry/	401168
27	anthropometr$.tw.	29815
28	exp Body Composition/	39210
29	"body-composition$".tw.	21143
30	"body-fat$".tw.	21711
31	adipos$.tw.	63367
32	"fat-percentage$".tw.	2515
33	"fat-mass".tw.	11897
34	exp Body Mass Index/	85404
35	"body-mass-index".tw.	98531
36	bmi.tw.	72029
37	"z-score".tw.	7144
38	exp Waist Circumference/	5327
39	"waist-circumference$".tw.	13747
40	"waist-to-hip-ratio".tw.	6182
41	exp Body Height/ or height$.tw.	131563
42	exp Body Weight/ or "body-weight$".tw.	446580
43	41 and 42	34232
44	26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 43	547144
45	exp Cohort Studies/ or cohort$.mp.	1516519
46	exp Longitudinal Studies/ or (longitudinal adj5 (stud$ or trial$ or design$)).mp.	121008
47	exp Prospective Studies/ or (prospectiv$ adj5 (stud$ or trial$ or design$)).mp.	488284
48	exp Observational Study/ or (observation$ adj5 (stud$ or trial$ or design$)).mp.	87370
49	45 or 46 or 47 or 48	1647296
50	exp Intervention Studies/	7523
51	49 not 50	1644824
52	8 and 25 and 44 and 51	4057
53	limit 52 to yr="1990 -Current"	3780
54	limit 53 to humans	3772

B.2.2 MED Refocused combined strategy for clinical trials with follow up with limits

Database(s): **Ovid MEDLINE(R)** 1946 to April Week 1 2015

Search Strategy: Run on 9th April 2015

#	Searches	Results
1	exp Child/	1574433
2	child$.mp.	1880547
3	1 or 2	1880547
4	exp Adolescent/	1648687
5	adolescen$.mp.	1673202
6	teen$.mp.	21044
7	4 or 5 or 6	1675746
8	3 or 7	2732527
9	exp Diet/	201626
10	diet$.tw.	372632
11	exp Diet Records/	4231
12	(food adj (diary or diaries)).tw.	829
13	"food-record$".tw.	1980
14	"food-frequency-questionnaire$".tw.	6678
15	"ffq$".tw.	1758
16	"24-hour-recall$".tw.	1041
17	"twenty-four-hour-recall$".tw.	24
18	"24-hr-recall$".tw.	79
19	exp Energy Intake/	36093
20	"energy-intake$".tw.	14392
21	"food-intake$".tw.	32635
22	exp Nutrition Assessment/	8337
23	"nutrition$-assessment$".tw.	3050
24	(energy adj1 dens$).tw.	3941
25	9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24	492131
26	exp Anthropometry/	401168

27	anthropometr$.tw.	29815
28	exp Body Composition/	39210
29	"body-composition$".tw.	21143
30	"body-fat$".tw.	21711
31	adipos$.tw.	63367
32	"fat-percentage$".tw.	2515
33	"fat-mass".tw.	11897
34	exp Body Mass Index/	85404
35	"body-mass-index".tw.	98531
36	bmi.tw.	72029
37	"z-score".tw.	7144
38	exp Waist Circumference/	5327
39	"waist-circumference$".tw.	13747
40	"waist-to-hip-ratio".tw.	6182
41	exp Body Height/ or height$.tw.	131563
42	exp Body Weight/ or "body-weight$".tw.	446580
43	41 and 42	34232
44	26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 43	547144
45	exp Clinical Trial/	802258
46	exp Follow-Up Studies/	513147
47	"follow-up".mp.	898451
48	45 and (46 or 47)	140545
49	8 and 25 and 44 and 48	531
50	limit 49 to yr="2010 -Current"	261
51	limit 50 to humans	261

B.3 Searches in EMBASE

B.3.1 EMB Refocused combined strategies for cohort with limits

Database(s): **Embase Classic+Embase** 1947 to 2015 April 08

Search Strategy: Run on 9th April 2015

#	Searches	Results
1	child/	1463907
2	child$.mp.	2299754
3	adolescent/	1279444
4	adolescen$.mp.	1351402
5	teen$.mp.	30125
6	1 or 2 or 3 or 4 or 5	2922945
7	exp diet/	256020
8	diet$.tw.	544213
9	"diet-record$".mp.	972
10	"food-diar$".mp.	1387
11	"food-record$".mp.	2780
12	"food-frequency-questionnaire$".mp.	10278
13	ffq$.mp.	2505
14	"24-hour-recall$".mp.	1629
15	"twenty-four-hour-recall$".mp.	37
16	"24-hr-recall$".mp.	198
17	"energy-intake$".mp.	19453
18	exp food intake/	241980
19	"food-intake$".tw.	45151
20	exp nutritional assessment/	16354
21	"nutrition$-assessment$".tw.	4863
22	(energy adj1 dens$).mp.	8040
23	7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22	823303
24	exp anthropometry/ or anthropometr$.tw.	67392
25	exp body composition/ or "body-composition$".tw.	74793

26	(body-fat$ or adipos$ or fat-percentage$ or fat-mass).tw.	127745
27	(body-mass-index or bmi).mp.	237299
28	"z-score".mp.	12518
29	exp waist circumference/ or "waist-circumference$".tw.	32390
30	"waist-to-hip-ratio".mp.	12040
31	exp body height/ or height$.tw.	196435
32	exp body weight/ or "body-weight$".tw.	562245
33	31 and 32	46873
34	24 or 25 or 26 or 27 or 28 or 29 or 30 or 33	453054
35	"Cohort-Stud$".mp.	135338
36	exp longitudinal study/	76226
37	(longitudinal adj5 (stud* or trial* or design*)).tw.	91483
38	exp prospective study/	284620
39	(prospectiv* adj5 (stud* or trial* or design*)).tw.	444547
40	exp observational study/	69076
41	(observation* adj5 (stud* or trial* or design*)).tw.	136330
42	35 or 36 or 37 or 38 or 39 or 40 or 41	825523
43	exp intervention study/	23205
44	42 not 43	821046
45	6 and 23 and 34 and 44	2681
46	limit 45 to yr="1990 -Current"	2627
47	limit 46 to human	2526

B.3.2 EMB Refocused combined strategy for clinical trials with follow up with limits

Database(s): **Embase Classic+Embase** 1947 to 2015 April 08

Search Strategy: Run on 9th April 2015

#	Searches	Results
1	child/	1463907
2	child$.mp.	2299754
3	adolescent/	1279444
4	adolescen$.mp.	1351402

5	teen$.mp.	30125
6	1 or 2 or 3 or 4 or 5	2922945
7	exp diet/	256020
8	diet$.tw.	544213
9	"diet-record$".mp.	972
10	"food-diar$".mp.	1387
11	"food-record$".mp.	2780
12	"food-frequency-questionnaire$".mp.	10278
13	ffq$.mp.	2505
14	"24-hour-recall$".mp.	1629
15	"twenty-four-hour-recall$".mp.	37
16	"24-hr-recall$".mp.	198
17	"energy-intake$".mp.	19453
18	exp food intake/	241980
19	"food-intake$".tw.	45151
20	exp nutritional assessment/	16354
21	"nutrition$-assessment$".tw.	4863
22	(energy adj1 dens$).mp.	8040
23	7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22	823303
24	exp anthropometry/ or anthropometr$.tw.	67392
25	exp body composition/ or "body-composition$".tw.	74793
26	(body-fat$ or adipos$ or fat-percentage$ or fat-mass).tw.	127745
27	(body-mass-index or bmi).mp.	237299
28	"z-score".mp.	12518
29	exp waist circumference/ or "waist-circumference$".tw.	32390
30	"waist-to-hip-ratio".mp.	12040
31	exp body height/ or height$.tw.	196435
32	exp body weight/ or "body-weight$".tw.	562245
33	31 and 32	46873
34	24 or 25 or 26 or 27 or 28 or 29 or 30 or 33	453054
35	exp clinical trial/	1023786

36 exp follow up/ or "follow-up".tw.	1245654
37 35 and 36	139806
38 6 and 23 and 34 and 37	365
39 limit 38 to yr="2010 -Current"	221
40 limit 39 to human	213

B.4 Searches in Scopus

B.4.1 Saved in Scopus as Search #24

Database(s): Scopus

Search Strategy: Run on 10th April 2015, found **2557** results

((TITLE-ABS-KEY (child*) OR TITLE-ABS-KEY (adolescen*) OR TITLE-ABS-KEY (teen*))) AND ((TITLE-ABS-KEY (diet*) OR TITLE-ABS-KEY ("Diet record*") OR TITLE-ABS-KEY ("Food diar*") OR TITLE-ABS-KEY ("food record*") OR TITLE-ABS-KEY ("Food frequency questionnaire*") OR TITLE-ABS-KEY ("ffq*") OR TITLE-ABS-KEY ("twenty four hour recall*") OR TITLE-ABS-KEY ("24 hr recall") OR TITLE-ABS-KEY ("24 hour recall") OR TITLE-ABS-KEY ("energy intake*") OR TITLE-ABS-KEY ("food intake*") OR TITLE-ABS-KEY ("energy dens*") OR TITLE-ABS-KEY ("nutrition* assessment*"))) AND ((TITLE-ABS-KEY (anthropometr*) OR TITLE-ABS-KEY ("body composition") OR TITLE-ABS-KEY ("body fat*") OR TITLE-ABS-KEY (adipos*) OR TITLE-ABS-KEY ("fat percentage*") OR TITLE-ABS-KEY ("fat mass") OR TITLE-ABS-KEY ("body mass index") OR TITLE-ABS-KEY (bmi) OR TITLE-ABS-KEY ("z-score") OR TITLE-ABS-KEY ("waist circumference*") OR TITLE-ABS-KEY ("waist to hip ratio") OR TITLE-ABS-KEY ("body weight*" AND height*))) AND ((TITLE-ABS-KEY ("Cohort*") OR TITLE-ABS-KEY ("Longitudinal stud*") OR TITLE-ABS-KEY ("Prospectiv* stud*") OR TITLE-ABS-KEY ("Observationa* stud*") AND NOT TITLE-ABS-KEY ("Intervention")) AND PUBYEAR > 1989) AND (LIMIT-TO (EXACTKEYWORD , "Human")) AND (EXCLUDE (EXACTKEYWORD , "Middle Aged"))

B.4.2 Saved in Scopus as Search #25

Database(s): Scopus

Search Strategy: Run on 10th April 2015, found **135** results

(TITLE-ABS-KEY (child*) OR TITLE-ABS-KEY (adolescen*) OR TITLE-ABS-KEY (teen*))) AND ((TITLE-ABS-KEY (diet*) OR TITLE-ABS-KEY ("Diet record*") OR TITLE-ABS-KEY ("Food diar*") OR TITLE-ABS-KEY ("food record*") OR TITLE-ABS-KEY ("Food frequency questionnaire*") OR TITLE-ABS-KEY ("ffq*") OR TITLE-ABS-KEY ("twenty four hour recall*") OR TITLE-ABS-KEY ("24 hr recall") OR TITLE-ABS-KEY ("24 hour recall") OR TITLE-ABS-KEY (

"energy intake*") OR TITLE-ABS-KEY ("food intake*") OR TITLE-ABS-KEY ("energy dens*") OR TITLE-ABS-KEY ("nutrition* assessment*"))) AND ((TITLE-ABS-KEY (anthropometr*) OR TITLE-ABS-KEY ("body composition") OR TITLE-ABS-KEY ("body fat*") OR TITLE-ABS-KEY (adipos*) OR TITLE-ABS-KEY ("fat percentage*") OR TITLE-ABS-KEY ("fat mass") OR TITLE-ABS-KEY ("body mass index") OR TITLE-ABS-KEY (bmi) OR TITLE-ABS-KEY ("z-score") OR TITLE-ABS-KEY ("waist circumference*") OR TITLE-ABS-KEY ("waist to hip ratio") OR TITLE-ABS-KEY ("body weight*" AND height*))) AND ((TITLE-ABS-KEY ("Clinical trial*") AND TITLE-ABS-KEY ("follow up")) AND PUBYEAR > 2009) AND (LIMIT-TO (EXACTKEYWORD , "Human")) AND (EXCLUDE (EXACTKEYWORD , "Middle Aged"))

B.5 Searches in Web of Science

B.5.1 RefocusComboCohorts

Database(s): Web of Science Core Collection, SCI EXPANDED 1990 to 2015

Search Strategy: Run on 10th April 2015, found **1851**results

# 7	**1,851**	(#5 AND #4 AND #3 AND #2) not #1 *Indexes=SCI-EXPANDED Timespan=1990-2015*
# 6	**1,897**	#5 AND #4 AND #3 AND #2 *Indexes=SCI-EXPANDED Timespan=1990-2015*
# 5	**1,528,392**	TS = (Cohort OR Longitudinal OR Prospective OR Observation*) NOT TS = (Intervention) *Indexes=SCI-EXPANDED Timespan=1990-2015*
# 4	**300,106**	TS = (Anthropometr* OR "Body composition" OR "Body fat" OR "Adipos*"OR "Fat percentage*" OR "fat mass" OR (*Weight AND *Height) OR "Body Mass Index" OR bmi OR "z-score" OR "Waist circumference*" OR "waist to hip ratio") *Indexes=SCI-EXPANDED Timespan=1990-2015*
# 3	**543,578**	TS = (Diet* OR "Diet record*" OR "Food diar*" OR "Food record*" OR " Food frequency questionnaire*" OR "ffq" OR "twenty four hour recall*" OR "24 hr recall*" OR "energy intake*" OR "food intake*" OR "energy dens*" OR "nutrition* assessment*") *Indexes=SCI-EXPANDED Timespan=1990-2015*
# 2	**879,278**	TS = (Child* OR Adolescen* OR Teen*) *Indexes=SCI-EXPANDED Timespan=1990-2015*

1 **563,452** TS = ("Animal Experiment" OR "Veterinary medicine" OR (Animals NOT (Animals AND Humans)))

Indexes=SCI-EXPANDED Timespan=1990-2015

B.5.2 RefocusComboClinical

Database(s): Web of Science Core Collection, SCI EXPANDED 2010 to 2015

Search Strategy: Run on 10th April 2015, found **23** results

7 **23** (#5 AND #4 AND #3 AND #2) NOT #1

Indexes=SCI-EXPANDED Timespan=2010-2015

6 **23** #5 AND #4 AND #3 AND #2

Indexes=SCI-EXPANDED Timespan=2010-2015

5 **7,892** TS = ("Clinical Trial" AND "Follow up")

Indexes=SCI-EXPANDED Timespan=2010-2015

4 **132,504** TS = (Anthropometr* OR "Body composition" OR "Body fat" OR "Adipos*"OR "Fat percentage*" OR "fat mass" OR (*Weight AND *Height) OR "Body Mass Index" OR bmi OR "z-score" OR "Waist circumference*" OR " waist to hip ratio")

Indexes=SCI-EXPANDED Timespan=2010-2015

3 **180,742** TS = (Diet* OR "Diet record*" OR "Food diar*" OR "Food record*" OR " Food frequency questionnaire*" OR "ffq" OR "twenty four hour recall*" OR "24 hr recall*" OR "energy intake*" OR "food intake*" OR "energy dens*" OR "nutrition* assessment*")

Indexes=SCI-EXPANDED Timespan=2010-2015

2 **320,658** TS = (Child* OR Adolescen* OR Teen*)

Indexes=SCI-EXPANDED Timespan=2010-2015

1 **168,885** TS = ("Animal Experiment" OR "Veterinary medicine" OR (Animals NOT (Animals AND Humans)))

Indexes=SCI-EXPANDED Timespan=2010-2015

Appendix C Endnote Libraries used for systematic review record management

Systematic Review stage	**Endnote Library name**	**Created**	**Number of records**
Database searches	Clinical trials and follow up *plus MASTER COPY*	April 2015	898
Database searches	Cohorts only *plus MASTER COPY*	April 2015	10706
All searches plus additional sources	Systematic Review	April 2015	11714
De-duplication	De-duplicated Clinical trials and follow up	April 2015	551
De-duplication	De-duplicated Cohorts only	April 2015	6205
Duplicates kept	Duplicates removed from Cohorts only	May 2015	4511
Duplicates kept	Duplicates removed from Clinical trials and follow up	June 2015	447
Combined library	All de-duplicated records and additional sources	June 2015	6589
First pass #1 screening	First pass screening all records	June 2015	6589
Second pass #2 screening	Second pass screening by Cath (*Reviewer 1*)	July 2015	3431
Second pass #2 screening	Second pass screening by Marion (*Reviewer 2*)	July 2015	3431
Resolving #2 differences	Second pass screening for 3rd reviewer to decide	Sept 2015	838
Resolving #2 differences	Second pass screening with 3rd reviewer decisions	Sept 2015	671

Appendix C

Systematic Review stage	**Endnote Library name**	**Created**	**Number of records**
Third pass #3 screening	Third pass screening papers to be found	Sept 2015	390
Third pass #3 screening	Third pass screening by Cath (*Reviewer 1)*	Oct 2015	390
Third pass #3 screening	Third pass screening by Ben (and others*, Reviewer 2*)	Oct 2015	390
Resolving #3 differences	Third pass screening for 3rd reviewer to decide	Feb 2016	158
Included studies for data extraction	Included in Systematic Review	March 2016	61

These Endnote libraries were originally saved in University of Leeds shared drive:

N: MAPS/Research/PRC/NEG/NEG036 Obesity Risk in Young People/D. Literature/ Systematic Review of childhood and adolescent cohorts

Later transferred to:

N: Faculty of EPS/Research/PRC/NEG/NEG036 Obesity Risk in Young People/D. Literature/ Systematic Review of childhood and adolescent cohorts

Appendix D Systematic Review screening

D.1 First pass screening keywords used for searches in the title

Rats	Gene	Bacter	Gestational
Mice	Supplementation	Serum	Hydro
Mouse	Laparoscopic	Lung	Obese “and” Adult
Drug	Leukemia/Leukaemia	Teeth/Tooth	Cell
Animal	Treatment	Hermia	Whale
Vet/Veterinary	Disorder	Acid	Monkey
Pregnan	Diarrhoea	Atresia	Neuro
Foetal/Fetal	Balloon	Renal	Mercury
Native	Epilepsy/Epileptic	Organic	Fertile
Hormon	Pulmonary	Failure	Menstruation
Transplant	Allele	Intestine/Intestinal	Menstrual
Anorexi	Insulin	Chronic	Obese “and”
Experiment	Eczema	Pancreas/Pancreatic	Adolescents
Bariatric	Elderly	Microbiota	Reduction
Coeliac/Celiac	FTO	Baby/Babies	Sex
IBD	Dyskinesia	Skeletal	Outpatient
Diabetes/Diabetic	Gastric/Gastro	Immune/Immunology	Stature
Inflammat	Palsy	Deficiency	Haemoglobin
Disease	Fluor	Oestrogen/Estrogen	Spina
Syndrome	Vitamin	Injury/Injuries	Hospital
Surgery/Surgical	Osteoporosis	Facial	Arrhythmia
Ghrelin	Tuberculosis	Urol	Judo
Leptin	Dialysis	Tube	Ball
Liver	Arthritis	Menopause/Menopausal	Stunt
Bone	Infection	Gymnast	Malnutrition
Cancer	Hepatic	Urinal	Congenital
Cystic	Cirrho	Allergic	Psoriasis
Sleep	Virus	Muscular	Endoscopic
Malaria	Tumor/Tumour	Thymus	Ulcer
Zince	Brain	Dementia	Insecticide
Smoke/Smoking	Oncology	Hyper	Autism
Enteral	Malnourished	Primates	Narcolepsy
Lactation/Lactatin	Snoring	Phenyl	Malignant/Malignancy
Crohn	Asthma	Steroid	Mammo
Peptide	Striae	Ovarian	Colonic
Underweight	Atherosclerosis	Loss	Ventilation
HIV	Anaemia/Anemia	“omy”	Gyne
Therapy	Function	Subject	Laser
Bowel	Gynaecomastia	Morbidity/Morbidities	
Phenylketonuria	Bacteria	Obese “and” Children	
Natal		Fracture	
Bulimia		Swim	

D.2 Second stage screening questionnaire for screening on title and abstract

Language:

Is the abstract in English?

Exclude if the abstract is written in a language other than English.

Study Design:

Is this a cohort (observational) study or a controlled trial *with follow-up*?

Exclude if it is clearly an intervention study (E.g. to promote healthy eating, physical activity or weight loss) with no untreated control group or if the trial is of less than 2 years duration.

Exclude if it is clearly a review or systematic review.

Participants:

Are some participants aged between 8 and 19 years old during this study?

Exclude if it is clear that participants are not representative of the general population, even if participants are aged between 8 and 19 years old. E.g. If they were selected because they were obese or overweight at the start, were born pre-term, have diabetes or some other long-term health condition, are vegan or belong to a small ethnic minority group.

Exposure:

Was diet measured or assessed, in part or in total, with a diet diary, 24 hour recall, or a food frequency questionnaire?

Exclude if it is clear that diet was not measured.

Exclude if a diet history was used to gauge usual intake.

Exclude if diet was not quantified, in part or in total. E.g. If only a simple food frequency questionnaire or diet quality index was used to assess diet quality.

Outcome:

Was body fatness measured *at a later point in time from diet*?

E.g. did the study report height and weight, or BMI, z-score, waist circumference, hip to waist ratio, fat percentage or fat mass?

Exclude if it is clear that there were no body measurements.

Exclude if it is clear that body measurements were at baseline but not at follow-up.

Screening on Title and Abstract – Instructions to reviewers

Screening on title and abstract will be done in duplicate, with at least two independent reviewers.

For each record in the Endnote library ask the key questions in turn.

If an answer to any question is NO that record can be excluded. (No further questions need to be answered.)

If the answer is YES or UNCLEAR, proceed to the next question.

Some records may not give enough detail in the title or abstract for a sensible decision to be made – it may be unclear whether the record should be included in the review or not.

After screening the title and abstract enter your decision in the research notes field for that record, save it and then file the record in the appropriate one of these groups within the Endnote library:

- Non-English title/abstract **NE**
- Exclude on title/abstract: Study Type **XST**
- Exclude on title/abstract: Participants **XP**
- Exclude on title/abstract: Exposure **XE**
- Exclude on title/abstract: Outcome **XO**
- Include on title/abstract (all answers are YES) **I**
- Title/abstract unclear (some/all answers are UNCLEAR but none are NO) **IU**

NOTE: Records in the latter two groups will proceed to full text screening.

Differences in decisions between the two reviewers (exclude vs include /unclear) will be resolved by a third reviewer.

D.3 Third stage screening questionnaire for screening on full text

Screening on Full Text – Instructions to reviewers

Screening on full text will be done in duplicate, with two independent reviewers.

Records which report research from the same study should be considered together rather than separately.

For each study/record in the Endnote library ask the key questions shown below in turn. Answers will be found mainly in the Methods section.

If an answer to any question is NO that study/record can be excluded. No further questions need to be answered.

If the answer is YES or UNCLEAR then proceed to the next question.

Differences in decisions between the two reviewers (exclude vs include) will be resolved by a third reviewer.

Key Questions

Full Text

We want to include full text papers– abstracts are unlikely to include enough detail.

Is the full text available?

- Exclude if the full text is not available E.g. Abstract only

Language:

We want to include studies published in English.

Is the full text in English?

- Exclude if the full text is written in a language other than English.

Study Type:

We want to include observational longitudinal or cohort studies or clinical trials that have measured diet and body anthropometry, measuring body anthropometry at a later point in time (at least two years later) from diet

Is this a prospective cohort (observational) study *or* a controlled trial?

- Exclude if it is clearly an intervention study (E.g. to promote healthy eating, physical activity or weight loss) *with no untreated control or placebo group*. Exclude if it is clearly a review or systematic review.

Does this study span two years or more?

- Exclude if the study is of less than two years duration *or* if the trial has a follow-up period of less than two years.

Participants:

We want to include studies of healthy children or adolescents aged from 8 to 19 years, which have followed up participants for two years or more. Studies where fewer than 50% of participants are below or above these age cut-offs (E.g. younger than 8 years or older than 19 years) at baseline will be included.

Both single sex and mixed sex studies will be considered.

Participants may be underweight, normal weight, overweight or obese at baseline, reflecting the range of weights that exist within populations.

Are participants representative of the general population?

- Exclude if participants are from specific, unrepresentative groups. E.g. If they were recruited because they were obese or overweight at the start, were born pre-term, have diabetes or some other long-term health condition, are vegan or belong to a small ethnic minority group.

Are more than 50% of participants aged 8 years old or more at baseline?

- Exclude if more than half of participants are babies, infants or children aged less than 8 years at baseline.

Are more than 50% of participants aged less than 19 years old at baseline?

- Exclude if more than half of participants are aged 19 years or older at baseline.

Exposure:

We want to include studies which have used an objective measure of dietary intake derived from either a weighed or un-weighed diet diary, 24 hour diet recall or a quantitative or semi-quantitative food frequency questionnaire, from which partial or total intake of foods and drink can be quantified.

Was diet or dietary pattern measured or assessed, in part or in total?

- Exclude if it is clear that diet was not measured at all.

Was diet measured or assessed, in part or in total, with a diet diary, 24 hour recall, or a quantitative or semi-quantitative food frequency questionnaire?

- Exclude if a diet history was used to gauge usual (rather than actual) intake.
- Exclude if the study only used a tool in which consumption of food and drink is not quantified. E.g. the study used either a simple food frequency questionnaire, a diet quality index to assess diet quality or a questionnaire about dietary habits.
- Exclude if the study was specifically about breast feeding.

Later amendment: ***Was diet or dietary pattern reported as an exposure?***

- Exclude if diet or dietary pattern was not reported as an exposure.

Outcome:

We want to include studies which have used anthropometric measurements from which an assessment of overweight/obesity status can be made, measured at least two years after measuring diet. Such measures may be reported as height *and* weight, body mass index (BMI) or BMI z-score, waist circumference, hip to waist ratio, fat percentage or fat mass.

Was anthropometry measured or body fatness assessed?

- Exclude if it is clear that no anthropometric measurements were taken.

Was body anthropometry measured at least two years later than measuring diet?

- Exclude if it is clear that anthropometric measurements were only made at baseline but not at follow-up.
- Exclude if anthropometric measurements were made less than two years after measuring diet i.e. the follow up is too short.

Later amendment: ***Was anthropometry or body fatness reported as an outcome?***

- Exclude if anthropometry or body fatness was not reported as an outcome.

Coding

After screening the full text(s) enter the reason for your decision in the Endnote library at the start of the research notes field for that study/record, using the codes shown below in bold. Save the record, and then file it in the similarly coded decision group.

- Abstract only: **AO**
- Non-English: **NE**
- Exclude on Study Type: **XST**
- Exclude Study too short: **SHORT**
- Exclude Participants not representative: **MINORITY**
- Exclude More than half of participants too young: **YOUNG**
- Exclude More than half of participants too old: **OLD**
- Exclude on exposure, Dietary intake not measured: **XE**
- Exclude Used Diet history: **DHIST**
- Exclude Diet not quantified: **XQUANT**
- Exclude on outcome, Anthropometry not measured: **XO**
- Exclude Anthropometry at baseline only: **XOB**
- Exclude Anthropometry follow up too short: **XFU**
- *Later amendment:* Exclude if association between diet and body fatness not reported: **AXR**
- Unclear (No answers are NO but some are UNCLEAR): **UNCLEAR ***
- Include in Review (All answers are YES): **INCLUDE**

* It may be necessary to contact the authors for clarification.

D.4 Third stage screening form for screening on full text

Review title	**A systematic review of childhood and adolescent cohorts which measure diet and subsequent adiposity**
Study ID *(surname of first author and year first full report of study was published e.g. Smith 2001)*	
Report ID - use the Endnote # number	
Study Name (and abbreviation) if given	
Report ID of other reports of this study including errata or retractions	
Notes	

General Information

Date form completed *(dd/mm/yyyy)*	
Name of person extracting data	Cath Rycroft / Ben Clapinson/ Charlotte Evans
Reference citation	
Study author contact details	
Publication type *(e.g. full report, abstract, letter)*	
Full Text available?	
Full Text in English? (If not, which language?)	
Notes:	

Study eligibility

Study Characteristics	Eligibility criteria/screening question *(Insert inclusion criteria as defined in the Protocol)*	Are eligibility criteria met? Yes No ?	Location in text or source *(page & figure, table, paragraph, other)*
Type of study	Is this a prospective cohort study?	☐ ☐ ☐	
	Is this a controlled trial with an untreated control group?	☐ ☐ ☐	
	Does this study span two years or more?	☐ ☐ ☐	
Participants	Are participants representative of the general population?	☐ ☐ ☐	
	Are more than 50% aged over 8 years old?	☐ ☐ ☐	
	Are more than 50% aged less than 19 years old?	☐ ☐ ☐	
Exposure, Diet	Was diet measured at all?	☐ ☐ ☐	
	Was diet measured in part or in total with a diet diary (weighed or un-weighed), 24 hour recall or a quantitative or semi-quantitative FFQ?	☐ ☐ ☐	
Outcome, Body fatness	Was anthropometry measured or body fatness assessed?	☐ ☐ ☐	
	Was body anthropometry measured at least two years later than measuring diet?	☐ ☐ ☐	
INCLUDE ☐		EXCLUDE ☐	
Reason for exclusion	Code:		
Notes:			

DO NOT PROCEED IF STUDY EXCLUDED FROM REVIEW

D.5 The customised Newcastle-Ottawa quality assessment scale for cohort studies

Selection

1. **Representativeness of the exposed cohort**
    - ❖ Star given if cohort was truly or somewhat representative of the average food and drink consuming child or adolescent in the community.
2. **Selection of the non-exposed cohort**
    - ❖ Star given if non-exposed persons were drawn from exactly the same population or community as the exposed persons.
3. **Ascertainment of exposure (whole diet and dietary factors)**
    - ❖ Star given if diet was assessed by using (weighed or un-weighed) food record or food diary OR multiple 24 hour recalls.
4. **Demonstration that outcome of interest (body fatness or adiposity) was not present at start of study**
    - ❖ No star given as included studies were of cohorts reflecting the full distribution of body fatness; in each cohort some children or adolescents were overweight or obese at the start of the study.

Comparability

1. **Comparability of cohorts on the basis of the design or analysis**
    - ❖ Star given if age was controlled for in analysis.
    - ❖ Star given if total energy intake was controlled for in analysis.

Outcome

1. **Assessment of outcome**
    - ❖ Star given if body fatness or adiposity outcome was obtained from measures done by trained research staff.
2. **Was follow-up long enough for outcomes to occur**
    - ❖ Star given if follow-up period was 2 years or longer.
3. **Adequacy of follow up of cohorts**
    - ❖ Star given if complete follow-up or if the loss to follow-up was less than 25% or a description was provided of those lost.

D.6 Newcastle-Ottawa quality appraisal form

Review title	A systematic review of childhood and adolescent cohorts which measure diet and subsequent body fatness
Study ID *(# number author year)*	
Date form completed *(dd/mm/yyyy)*	
Name of quality appraiser	
Criteria used	Newcastle-Ottawa Quality Assessment Scale – Cohort studies
Notes:	

Type	**Appraisal question**	**Selection options**	**Outcome**
Selection	1. Representativeness of exposed cohort	a) truly representative of the average ***child or adolescent*** in the community	□ *
		b) somewhat representative of the average ***child or adolescent*** in the community	□ *
		c) selected group of users e.g. nurses, volunteers *or convenience sampling*	□
		d) no description of the derivation of the cohort	□
	2. Selection of the non-exposed cohort	a) drawn from the same community as the exposed cohort	□*
		b) drawn from a different source	□
		c) no description of the derivation of the non-exposed cohort	□
	3. Ascertainment of exposure (***whole diet and dietary factors***)	a) secure record (e.g. surgical records)	□*
		b) structured interview	□*
		c) written self-report	□
		d) no description	□
	4. Outcome of interest (***body fatness or***	a) yes	□*

Type	Appraisal question	Selection options	Outcome
	adiposity) was not present at start of study *(no star)*	b) no	□
Comparability	1. Comparability of cohorts on the basis of the design or analysis	a) Study controls for ***age***	□*****
		b) Study controls for ***Total Energy intake***	□*****
Outcome	1. Assessment of outcome	a) Independent blind assessment	□*****
		b) Record linkage	□*****
		c) Self-report	□
		d) *No description*	□
	2. Was follow-up long enough for outcomes to occur	a) yes (***minimum 2 years*** follow up period for outcome of interest)	□*****
		b) No	□
	3. Adequacy of follow up of cohorts	a) Complete follow up - all subjects accounted for	□*****
		b) Subjects lost to follow up unlikely to introduce bias - small number lost. **> or = *75* %** follow up, or description provided of those lost	□*****
		c) Follow up rate **< *75* %** and no description of those lost	□
		d) No statement	□
Star total *(maximum of 9):*			**/9**

Appendix E Included cohorts

E.1 ALSPAC Avon Longitudinal Study of Parents and Children

http://www.bristol.ac.uk/alspac/researchers/cohort-profile/

Country: United Kingdom

Study dates: Established 1990 - 1992, ongoing.

To date more than 2,000 research papers from the ALSPAC study have been published.

Study design and setting: The Avon Longitudinal Study of Parents and Children, also known as "Children of the 90s", is a transgenerational observational birth cohort study set up to investigate how genetics and the environment influence health and development across the life-course. The study is centred on the city of Bristol, which straddles the River Avon in the southwest of England, and its surrounding towns, villages and farms. In 1970 this area had a population of parents and children broadly similar to Great Britain as a whole. Assessments started with pregnant mothers and their offspring as babies and have continued through infancy, childhood and adolescence into young adulthood. Additional funding has allowed follow-up assessments of the mothers and more recently fathers, siblings and the next generation - Children of the Children of the 90s, COCO90s.

Recruitment and participants: Recruitment to ALSPAC was "opportunistic", using local media, antenatal and maternity health services to promote the study. Pregnant mothers with a due delivery date between 1 April 1991 and 31 December 1992, from three health districts in the South-West Regional Health Authority (later Bristol and District Health Authority) and the children resulting from those pregnancies were eligible for the study. 14,541 mothers were recruited during pregnancy, resulting in 14,062 known live births and 13,988 children alive at one year old. Post-natal recruitment at 7 years and at 8 years added a further 456 and 257 children respectively, increasing the total to 14,701 children who were alive at one year of age. Compared to the 1991 census of Great Britain and all potentially eligible mothers in Avon, recruited mothers were more likely to be married, live in owner-occupied homes and have use of a car. Only 2.2% of recruited mothers were non-white, compared with 4.1% in Avon and 7.6% in Great Britain as a whole. Attrition rates in the child cohort were highest during infancy and increased again during the transition to adulthood. Over 3,000 families have responded to every assessment and 5,777 responded to three quarters of them or more. During the "adolescence" phase, the average response rate was 48%, with 75% of the 12,776 individuals still eligible responding to at least one survey.

Measures: The ALSPAC cohort has been assessed by a series of questionnaires and clinics during pregnancy and at 68 time points between the child's birth and 18 years of age. Mothers have completed questionnaires about demographics and the child's health, psychological, physical and social development and the environment. Food frequency questionnaires within these self-reported surveys were used to measure children's diet. Later questionnaires were answered by the children directly and education questionnaires and tests were administered in schools. Clinical assessments of the children included cognitive and psychological measures as well as physiological measures such as anthropometry, blood pressure and dual-emission X-ray absorptiometry scans. Children's diet was quantified using 3 day dietary diaries. Physical activity was measured using activity accelerometers. ALSPAC also collected genetic and biological samples (placenta, blood, urine, hair, nails, teeth, saliva). Records have been linked to ONS deaths and cancer registrations, the National Pupil database and the General Practice Research database.

Funding and declared interests: Core funding for ALSPAC has been provided by the UK Medical Research Council (MRC), the Wellcome Trust and the University of Bristol. Funding for specific projects has come from many sources including the UK Department of Health, Department of Transport, the Deputy Prime Minister's Office, Department of Education and DEFRA, the Wellcome Trust, joint UK Research Councils, the National Lottery, the British Heart Foundation, the US National Institute of Health and the World Cancer Research Fund. Individual named researchers within ALSPAC have received Wellcome Trust grants, MRC research fellowships or work in centres that receive funding from the MRC.

E.2 BDPP Bienestar Diabetes Prevention Program

Country: United States of America.

Study dates: 2001 to 2004

Study design and setting: The BDPP longitudinal cohort originated from the control arm of an intervention programme. The primary study, Bienestar: a school-based type 2 diabetes prevention program (Trevino, 2005), took place in 27 inner city elementary schools in San Antonio, Texas. (Bienestar means "well-being" in Spanish.) 13 schools received a health program to reduce the risk of type 2 diabetes mellitus in high risk Mexican-American children and 14 schools were randomised as control schools. 1,024 children were assigned to the control group and 706 children with parental consent signed an assent form. The control group was followed up after almost two and three academic years. Only the 625 control group children with complete baseline data were included in the analysis of the association between

frequency of ready to eat cereal consumption and body mass index (Balvin Frantzen et al., 2013).

Recruitment and participants: Fourth grade children (mean age 9.13 years, SD 0.46 years) were recruited at the start of the 2001/2002 school year and were followed up at the end of fifth grade and again at the end of sixth grade, 2003/4 school year. The children's families provided demographic information at baseline. 78% of the children were Hispanic, 49% were male and 62% were from low-income households.

Measures: At each data collection point, trained interviewers took three 24 hour dietary recalls from the children using a multiple-pass method. Dietary intake was analysed using the Nutrition Data System for Research (NDS-R) version 2006. Anthropometric measurements were taken between August 2001 and May 2004. Children, barefoot, had their height measured to the nearest 0.1cm using a wall mounted stadiometer and their weight measured to the nearest 0.1kg using a combined bioelectric impedance/weight scale. Body Mass Index was calculated for each child and BMI values were converted to BMI percentiles using appropriate sex and age specific Centers for Disease Control and Prevention reference charts.

Funding and declared interests: Funding for the intervention program was provided in part by a grant from the National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Disease and by the University Health System, San Antonio. Funding for the secondary analysis also came from the National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases. The first author, Dr Frantzen, was employed by Dairy MAX, a regional dairy council established by American dairy farmers.

E.3 BHS Bogalusa Heart Study

Country: United States of America

Study dates: Established 1973, ongoing

Study design and setting: The Bogalusa Heart Study (Berenson and Bogalusa Heart Study, 2001) is a longitudinal cohort study that initially set out to investigate cardiovascular risk factors in children. Subjects were residents of Ward 4 of Washington Parish (Louisiana), a district which includes Bogalusa, a semi-rural, single industry town typical of many other towns in the southern United States. With continued funding 16,000 subjects have been examined in seven observational cross-sectional surveys of school children at approximately five year intervals, followed by post high school surveys extending to adulthood (up to 45 years old to date). Additional sub-studies (Blood Pressure Study, Lipid Study, and Diabetes Study) have also been conducted.

Recruitment and participants: Researchers requested support from school boards, head teachers, teachers and parents to recruit children to the bi-racial study (European American and African American). The initial participants included 440 infants, 800 pre-school children and 4,000 school-aged children and adolescents. Approximately two-thirds of the initial cohort were white and the remainder black.

Measures: Standard protocols were used throughout. Questionnaires were used to assess lifestyle factors such as dietary intake, tobacco and alcohol use and physical activity. Clinical measures/observations included height, body weight, waist circumference, subscapular and triceps skinfold thickness, Tanner stage, dental caries, blood pressure, low and high density lipoprotein-cholesterol, triglycerides, insulin, glucose and C-reactive protein.

During cross-sectional surveys conducted in 1973-74, in 1976-77 and in 1978-79, dietary intake and physiological measures of cardiovascular risk of children aged 10 years were assessed (O'Neil et al., 2015). Follow-up data from 247 participants were collected in 1989-91 by which time respondents were adults aged 19 to 28 years (mean age 23 years). A follow-up of a further set of 222 respondents was conducted in 1995-96 (mean age 29 years). At baseline and the first follow-up survey diet was measured by trained interviewers using a single 24 hour recall and quantified using the Moore Extended Nutrients (MENu) computerised database. During the second follow-up survey a 131 food item semi-quantitative food frequency questionnaire - the youth/adolescent questionnaire (YAQ) was employed to measure diet.

At baseline and both these follow-ups cardiovascular risk factors were assessed using the standard protocols. Body Mass Index was calculated from height and weight. For children at baseline overweight was defined as at or above the 85^{th} percentile and obesity was defined as at or above the 95^{th} percentile, with reference to the Centers for Disease Control and Prevention 2000 Growth charts for the appropriate sex and age. For young adults at follow-up, overweight was defined as a BMI from 24.9 to 29.9 kgm^{-2} and obesity was defined as a BMI ≥ 30 $kgm^{-2.}$

Funding and declared interests: The Bogalusa Heart Study was originally funded by the National Heart Lung and Blood Institute, with further funding from the National Institute of Child Health and Human Development, the American Heart Association, the National Institute of Ageing and, for the O'Neil et al study of Candy consumption, the National Confectioners Association.

E.4 BSCC Bogotá School Children cohort

Country: Colombia

Study dates: 2006 to 2008

Study design and setting: The Bogotá School Children cohort was a longitudinal study of the nutrition and health of school children in Bogotá, Columbia. It was originally established to investigate whether a mid-morning food ration given to children enrolled in public primary schools in the city of Bogotá was improving children's nutritional status (Arsenault et al., 2009).

Recruitment and participants: A sample of 4,000 children aged between five and 12 years old was chosen from all those enrolled in the 361 public primary schools in Bogotá. From a sampling frame of 8,500 classes, 166 classes were randomly selected to reach the target sample size. After seeking consent from parents 3,202 children from 3,032 households were enrolled, of whom 51.3% were girls. Most children were from low and middle income families, reflecting the socio-economic groups who typically enrolled in the public system. The mean age at baseline in February 2006 was 8.6 years (SD 1.7 years).

Measures: After enrolment parents from 2,466 households completed a survey of background information (parent's age, education, occupation, parity, anthropometry and indicators of household socio-economic status) and reported on the child's physical activity (television watching, outdoor play). Between May and June 2006, trained dietitians used a 38 item semi-quantitative food frequency questionnaire in the school setting to ask a random sample of 1,027 mothers about their child's usual dietary intake (Isanaka et al., 2007) . Principal component analysis was used to identify dietary patterns from this information (Shroff et al., 2014). Trained research staff visited schools in February 2006 to collect anthropometric measurements from the children and to obtain fasting blood samples from which to assess micronutrient status. If children in were not in school on the day of assessment they were later visited at home. Height was measured to the nearest 1mm using SECA 202 stadiometers. Weight was measured to the nearest 0.1kg with Tanita HS301 electronic scales. Body Mass Index was calculated from height and weight. Children's BMI for age and height for age z scores were derived using the World Health Organization's 2007 sex-specific growth references for children aged 5 to 19 years. Skinfold thicknesses (triceps and subscapular) were measured with Slimguide Skinfold callipers. Repeat anthropometric measurements were taken at follow-ups in June and November 2006 and again during 2007 and 2008. In the latter two follow-ups waist circumference was measured to the nearest 1mm using a non-extendible measuring tape. The median follow-up period was 2.5 years.

Funding and declared interests: The Bogotá School Children cohort was set up with support from the Secretary of Education of Bogotá, the David Rockefeller Center for Latin American Studies at Harvard University, the National University of Colombia and the National Institute of Health of Colombia. Later research sponsorship came from the ASISA Research Fund at the University of Michigan.

E.5 DONALD Dortmund Nutritional and Anthropometric Longitudinally Designed Study

https://www.ernaehrungsepidemiologie.uni-bonn.de/forschung/donald-1

Country: Germany

Study dates: Established 1985, ongoing

Study design and setting: The DOrtmund Nutritional and Anthropometric Longitudinally Designed Study is a longitudinal open cohort, initially set up to investigate and describe trends in dietary intake and behaviours, obtain metabolic reference data from healthy children and analyse the links between nutrition and growth in order to determine the nutritional needs of children and adolescents. The scope of the DONALD study has broadened over time to include the relationships between food consumption and nutritional behaviour, growth, development, metabolism and health, following subjects from infancy and childhood to adolescence and young adulthood. The study is based in the city of Dortmund and its surrounding communities (Kroke et al., 2004). Since 2012 the study has been conducted through the University of Bonn's Department of Nutritional Epidemiology.

Recruitment and participants: The DONALD study uses convenience sampling to recruit via personal contacts, maternity wards and paediatric practices. Healthy German babies aged 3 to 9 months are eligible if at least one parent speaks German and is willing to take part in the long-term study. Parents of the DONALD study children tend to have a higher socio-economic status and education level than the general population. The starting sample in 1985 also included older children between 2 and 18 years old, from cross-sectional studies in kindergartens and schools, who were added to the DONALD study to boost numbers - their data from infancy is missing. Since then 35 to 40 infants have been recruited each year and by 2010 approximately 1,400 children had joined the cohort (Buyken et al., 2012). Currently there are over 1,500 participants. Dropout rates are highest during puberty, but overall are described as low.

Measures: Regular examinations have taken place in infancy at 3, 6, 9, 12, 18 and 24 months and annually until adulthood (girls until 20 years, boys until 23 years old). Since 2005

participants have been invited to follow-up examinations every 5 years. Assessments include anthropometry, dietary intake and physical activity, 24 hour urine sampling, a medical examination and parental interviews. Child anthropometric measures taken by trained nurses include body mass measured to the nearest 100g with an electronic scale, height with no shoes and abdominal circumference to the nearest 0.1cm and skinfold thicknesses (biceps, triceps, scapular, iliac) to the nearest 0.1mm. Diet is scheduled to be measured each year by a 3 day weighed dietary record and coded records are linked to the LEBTAB database, developed during the DONALD study. LEBTAB is based on nutrient data from German food tables, supplemented by information from other national food tables, particularly from the UK, USA and the Netherlands. From the age of 3 or 4 years children provide a 24 hour urine sample. Since 2005, participants invited for follow-up during adulthood have also provided fasting blood samples.

Funding and declared interests: The DONALD Study is supported by the Ministry of Science and Research of North Rhine-Westphalia, Dusseldorf. Other funding acknowledgments include a research grant from the Ministry of the Environment, Nature Conservation, Agriculture and Consumer Protection of North Rhine-Westphalia (Alexy et al., 2011), a grant from the DANONE Institute (Cheng et al., 2009) and support from the German Federal Ministry of Food, Agriculture and Consumer Protection (Libuda et al., 2008).

E.6 EYHS European Youth Heart Study (Danish part)

Country: Denmark

Study dates: 1997 - 2009

Study design and setting: The European Youth Heart Study was originally a cross-sectional survey of children in four European locations, Odense (Denmark), Tartu (Estonia), Oslo (Norway) and Madeira (Portugal) which was set up with the aim of measuring cardiovascular disease risk factors and their associated influences in children (Riddoch et al., 2005). A follow-up was planned for six years later, along with the recruitment of a new cohort of 9 year old children. In Denmark children were followed up with repeat measures after six and twelve years (Zheng et al., 2014).

Recruitment and participants: In the four locations 5,664 children were invited to take part and 4,072 children were recruited. At least 1,000 girls and boys at the ages of 9 years and 15 years were recruited from each country. In each location a list of schools was used from which to randomly select schools and each school's register was used to randomly sample children of

appropriate ages using random number tables. Testing was done throughout the school year to minimise seasonal effects, using standard procedures that were validated for this age group.

In the Danish section of the EYHS cohort, 590 children aged 9 years took part in the baseline interview in 1997. They were recruited from the 25 of the 28 invited schools in Odense who agreed to participate. At baseline 56.1% of Danish participants were female, 53.4% were of high socio-economic status, based on parental education, occupation and family income data and 54.6% were classed as active, reporting regular physical exercise in a computer-based questionnaire.

After six years, in 2003, 384 Danish children, now aged 15 years, took part in the first follow-up. Participants and non-participants had similar socio-economic status, dietary intake, physical activity levels, anthropometry and pubertal status (Zheng et al., 2015). After 12 years, in 2009, a second follow-up took place, with 237 participants who were now young adults aged 21 years. In total 187 Danish children (98 girls and 89 boys) had measures at all three time points (Zheng et al., 2014).

Measures: Measures included height and weight, waist circumference, sum of skinfold thicknesses (biceps, triceps, subscapular, supra-iliac and medial calf sites), pubertal status assessed by Tanner stages and systolic and diastolic blood pressure. Fasting blood samples were tested for total cholesterol, high density lipoprotein cholesterol, triglycerides, glucose, and insulin. Dietary intake was measured by a face-to-face 24 hour recall interview using pictures of portion sizes of common foods and supported by a parent-assisted food record. Physical activity was measured over 4 consecutive days (2 weekdays and 2 weekend days) using an MTI accelerometer but a large number of participants did not complete the accelerometer measurements. Cardio-respiratory fitness was assessed by a graded maximal exercise test on a cycle ergometer, with greater workloads for the older children. A computer based questionnaire was employed to ask children about their smoking habits, alcohol intake, diet preferences and physical activity. A separate questionnaire asked parents about parental occupation and education, family income, health status (self-reported), ethnicity, CVD risk and family history of CVD, child's birth weight and breastfeeding history.

Funding: Funding came from grants from the Danish Heart Foundation, Danish Medical Research Council, Health Foundation, Danish Council for Sports Research, Foundation of 17-12-1981, Foundation in memory of Asta Florida Bolding and the Faculty of Health Sciences, University of Southern Denmark (all Denmark), the Estonian Science Foundation (Estonia), Norwegian Council of Cardiovascular Diseases and the Eckbo Legacy (Norway), and the European Social Fund (Portugal)

E.7 FAMS Female Adolescent Maturation Study (girls only)

Country: Unites States of America

Study dates: 2000/2001 to 2002/2003

Study design and setting: The Female Adolescent Maturation study was a longitudinal study of two years duration run by the University of Hawaii. The aim of the study was to compare differences in body size and fat distribution between Asian and white adolescent girls in Hawaii and to investigate the influence of diet and physical activity on their body composition (Novotny et al., 2006).

Recruitment and participants: Girls aged between 9 and 14 years old were recruited from the Kaiser Permanente membership database for the Island of Oahu, Hawaii. Only healthy adolescent girls, non-smokers, without chronic diseases, not using asthma or epilepsy medication (Lee et al., 2007), who identified as Asian (Japanese, Korean, Chinese, Filipino, Indian, Thai, and Vietnamese), white or as mixed Asian/white ethnicities were eligible. Kaiser Permanente had 1,106 female patients in the correct age range; 349 girls met the health and ethnicity criteria. With consent from their parents, they agreed to take part in Exam 1 at Kaiser Permanente's Honolulu clinic. From the original cohort, 160 girls (46%) took part in Exam 2 at the Kapiolani Clinical Research Centre two years (± 2 months) later.

Measures: One week before each exam girls were asked to complete a three day food record (Thursday, Friday, Saturday) and a questionnaire about dietary supplements, with help from their parent/guardian. This information was analysed by the University of Hawaii Cancer Research Center using the Shared Nutrition Food Composition database (version 1999) (St-Jules et al., 2014). Girls also answered a validated adolescent physical activity questionnaire which they brought to the exam to be checked.

At each exam parents provided information about the girl's age, ethnicity and menstrual status and about the mother's and father's weight, height and education level. Anthropometric measures of the girls at both examinations included weight in kg measured using a digital scale (SECA), height measured in cm using a height stadiometer (Measurement Concepts) and waist circumference measured in cm with an inextensible measuring tape (Hoechst). Height and weight measures were used to calculate BMI which were translated into z scores for age and sex based on the references from the Centers for Disease Control and Prevention. Skinfold thicknesses (subscapular, triceps, biceps, iliac) in mm were measured using Lange skinfold callipers. Puberty status was assessed by a nurse practitioner using Tanner staging criteria for breast and pubic hair development. During Exam 2 the girls' total, trunk and peripheral fat mass was assessed by dual energy x-ray absorptiometry (Lunar Prodigy DXA scanner).

Funding and declared interests: The Female Adolescent Maturation study was funded by the U.S. Department of Agriculture grant 9900700 and National Institutes of Health/National Center for Research Resources/Research Centers in Minority Institutions award P20 RR11091.

E.8 FCS Framingham Children's Study

Country: United States of America

Study dates: 1987 to 1999

Study design and setting: The Framingham Children's study was a longitudinal cohort of children and parents, set up to investigate diet and physical activity habits during childhood. Parents and children (one per family) were examined annually for 12 years. Recruited families lived within 64 km of Framingham, near Boston, Massachusetts.

Recruitment and participants: Families in the cohort were third and fourth generation descendants of subjects in the Framingham Heart Study established in 1948 (Hasnain et al., 2014). In total 106 white, non-Hispanic, two-parent families with a child aged 3 to 5 years old were enrolled to the Framingham Children's study. Approximately 60% of the children were boys (Moore, L.L. et al., 2006).

Measures: At yearly visits children were interviewed about their diet, activity and beliefs. Usual hours of television watching, and video time were evaluated by questionnaire. Parents completed questionnaires and interviews about their child's diet and physical activity, as well as about their own health, dietary habits, physical activity, risky behaviours and attitudes.

The child's dietary intake was measured using 3 day diet records, collected four times in the first year and once or twice each year thereafter. The Nutrition Data System (NDS) of the University of Minnesota was employed to calculate mean macronutrient and micronutrient intakes. Average daily intakes of foods and beverages were estimated by combining data from food records with the Food Guide Pyramid serving database, via the technical files of the U.S. Department of Agriculture's Continuing Survey of Food Intakes by Individuals.

Children's physical activity was assessed using a Caltrac accelerometer; worn for three to five consecutive days on one to four occasions each year.

At an annual clinic exam, parents and children were weighed without shoes and in light clothing to the nearest pound, using a standard counterbalance scale. Height was measured with a measuring bar to the nearest 0.25 inch. These measures were used to calculate BMI kgm^{-2}. Waist circumference to the nearest mm was measured with a cloth tape. Skin fold

thicknesses (triceps, subscapular, suprailiac, abdominal) were measured to the nearest mm using Lange callipers. At the final follow-up percent body fat was assessed by dual energy x-ray absorptiometry (Lunar DXA scanner).

Funding and declared interests: Work was supported by grant HL35653 from the National Heart, Lung and Blood Institute and by the National Dairy Council.

E.9 GUTS Growing Up Today Study

https://gutsweb.org/

Country: United States of America

Study dates: GUTS established 1996, ongoing

Study design and setting: The original Growing Up Today study is a longitudinal cohort set up to study how diet and exercise influence weight changes throughout the life course. Researchers from Brigham and Women's Hospital and Harvard School of Public Health recruited children of the Nurses' Health Study II participants from 50 states across the U.S.A. in 1996 (Taveras et al., 2005) Now that the children are adults their health can be compared with that of their mothers.

Recruitment and participants: Using information from the Nurses' Health Study II, mothers with children aged 9 to 14 years old in 1996 were identified. There were 53,000 children in this age range. Letters about GUTS were sent to the mothers, seeking permission to invite their child to the study. With the mothers' consent, invitation letters and questionnaires were sent to 13,261 girls and 13,504 boys. Completed baseline questionnaires were returned by 9,039 girls (68%) and 7,843 boys (58%) which was taken as assent to take part in the cohort. Non respondents were not re-contacted (Field et al., 2003b). At baseline girls had a mean age of 12 years (SD 1.6) and boys had a mean age of 11.8 years (SD 1.5). Children were mostly white (approximately 95%) (Berkey et al., 2005)

Measures: Maternal measures were obtained from questionnaires that mothers completed as part of the Nurses' Health Study II. Child measures taken annually in the autumn between 1996 and 1999 were obtained by mailed out, self-administered questionnaires (Field et al., 2004).

Dietary intake over the past year was measured by a 131 item semi-quantitative food frequency questionnaire, the Youth/Adolescent Questionnaire. The YAQ was validated by comparing intakes with those measured by three non-consecutive 24 hour recalls. Dieting was assessed with an adapted question from the Youth Risk Behaviour Surveillance System questionnaire.

Physical activity was estimated by questions about the number of hours per week engaged in 18 specific activities in each season. Inactivity was assessed by questions about the number of hours per week spent watching television, doing homework or playing video games.

The questionnaire gave detailed instructions about measuring height and weight, suggesting that the respondent asked someone to help. Self-reported height and weight information was used to calculate BMI and BMIz scores were calculated based on age and gender specific CDC growth chart references.

Drawings of the five Tanner stages of pubic hair development were used to estimate the child's pubertal development.

Funding and declared interests: Research and analysis has been supported by grants from the National Institutes of Health DK46834, DK59570 and HL68041, a special interest project grant U48-CCU115807 from the Centers for Disease Control and Prevention, the Boston Nutrition Research Center DK46200, Harvard Medical School and the Harvard Pilgrim Health Care Foundation and the Kellogg company. Dr Taveras was partly supported by the Minority Medical Faculty Development Program of the Robert Wood Johnson Foundation

E.10 GUTS II Growing Up Today Study II

Country: United States of America

Study dates: GUT II established 2004, ongoing

Study design and setting: The Growing Up Today Study II is a separate longitudinal cohort study, expanding on the work of GUTS with the enrolment of a second group of children. This second cohort was established to investigate diet and physical activity in adolescents and to assess their relationships with height velocity and weight gain (Field et al., 2014).

Recruitment and participants: Letters were sent to 20,700 women in the Nurses' Health Study II who had children aged 9 to 15 years old in 2004. With the mothers' consent, invitation letters and questionnaires were then sent to 8,826 girls and 8,454 boys. Completed baseline questionnaires were returned by 6,002 girls (68%) and 4,917 boys (58%) which was taken as assent to take part in the cohort.

Measures: Child measures were obtained by further questionnaires sent in the fall (autumn) of 2006 and 2008 and in winter 2011 (Field et al., 2014).

Dietary intake was measured with the same semi-quantitative food frequency questionnaire as before, the Youth/Adolescent Questionnaire. Physical activity/inactivity was estimated by

asking participants to recall activities each season and to report the number of hours per day spent watching television.

Again, self-reported height and weight information was used to calculate BMI. However, this time International Obesity Task Force cut-offs were used to classify obesity in children and adolescents under 18 years old. For those participants who were 18 years or older overweight was defined as a BMI between 25 and 29.9 and obesity was classified as a BMI >30.

Funding and declared interests: The GUTS II study was funded by the Breast Cancer Research Foundation and by a National Institutes of Health grant (R01-DK084001). Study researchers were funded by grants from the General Mills Company, the National Institutes for Health and the Maternal and Child Health Bureau.

E.11 IDEA & ECHO, Identifying Determinants of Eating and Activity combined with Etiology of Childhood Obesity

Country: United States of America

Study dates: IDEA 2006//2007 until 2008/2009, ECHO 2007/2008 until 2009/2010

Study design and setting: The IDEA study and the ECHO study were longitudinal cohort studies of young people and a significant adult in their life, which used a social ecological framework to investigate the etiology of childhood obesity. Both studies took place in a 7 county metropolitan area of Minneapolis-St Paul, Minnesota. Study protocols were approved by the University of Minneapolis and Ohio State University.

Recruitment and participants: The IDEA study recruited 6th to 11th grade school students, aged 10 to 16 years at baseline, regardless of weight status, plus their parent, guardian or adult carer. Recruits came from an existing cohort, the Minnesota Adolescent Community Cohort Tobacco study (Widome et al., 2007), from a convenience sample in the local community and from a Minnesota Department motor vehicles list limited to the 7 county area (Lytle, 2009). Similarly the ECHO study recruited 6th to 11th grade students and one parent or guardian/carer, but drawn from the membership of Health Partners, a health organisation in Minnesota, as this cohort was designed to be more ethnically diverse (Laska et al., 2012). The two cohorts were combined to create a larger sample for analysis. In the combined cohort there were 723 adolescents, with a mean age of 14.6 years at baseline, 327 males (86.5% white, 78.3% with a parent who was a college graduate) and 339 females (83.8% white, 72.9% with a parent who was a college graduate).

Measures: Identical measurement protocols were used by both studies. Each student/adult pair was invited to clinic visits at baseline and follow-up. Each participant's height without

shoes was measured to the nearest 0.1cm with a Shore height board. Weight to the nearest 0.1kg and body fat percentage to the nearest 0.1% was measured with a Tanita TBF-300A Body Composition Analyzer. BMI (kg/m^2) was calculated from these height and weight measures. Student blood pressure was measured using Dinamap machines and pubertal status was established using a 7-item self-report puberty scale. All participants completed surveys about behaviours related to energy balance such as breakfast eating, beverage intake, fast food intake, sleep patterns, weight perceptions, sedentary and screen time. The surveys included questions about the home environment and family structure, depression, smoking and alcohol use and family health history. A three day physical activity record (3D-PAR) was completed by students whereas physical activity of the adults was assessed using the International Physical Activity Questionnaire (IPAQ).

Physical activity of the students was also measured over 7 days using the ActiGraph accelerometer at baseline and two year follow-up. Adolescent diet was measured at baseline and two year follow-up, using three telephone-administered 24 hour dietary recalls (2 weekdays, 1 weekend day). Interactive dietary recall interviews were conducted by trained, certified staff using the Nutrition Data System Research (NDS-R) to allow direct data entry linked to a nutrient database. Adult diet was assessed using the National Cancer Institute Diet History Questionnaire (NCI-DHQ 2007) food frequency instrument, completed after the clinic visit. Willing students gave fasting blood samples (glucose, insulin, triglycerides, total cholesterol, low and high density lipoprotein and inflammatory markers) at baseline and follow-up.

Additional measures pertaining to the eating and physical activity environment at home, at school and in the local neighbourhood were also taken.

Funding and declared interests: Research was funded by a grant from the National Cancer Institute Transdisciplinary research in Energetics and Cancer Initiative and by the National Heart, Lung and Blood Institute. Salary support was provided by the National Cancer Institute. Participating families in IDEA received up to $150 per measurement year. Students who took part in the blood sampling could receive a further $100.

E.12 NGHS National Heart, Lung and Blood Institute's Growth & Health Study (girls only)

Country: United States of America

Study dates: 1987 to 1997

Study design and setting: The National Heart, Lung and Blood Institute's Growth & Health Study was set up as a 5 year prospective bi-racial cohort study, designed to investigate factors linked to the onset of obesity in black and white, non-Hispanic preadolescent girls and to assess obesity's effects on cardiovascular disease risk (NHLBI, 1992). The cohorts were eventually followed for 10 years, from age 9 to age 19 years old.

Recruitment and participants: Between January 1987 and May 1988 1,213 black and 1,166 white 9 or 10 year old girls were recruited by three clinical research centres. The University of California at Berkeley recruited girls from public and parochial schools in the Richmond Unified Schools district, an area with little income and occupational disparity between black and white families. The University of Cincinnati/Cincinnati Children's Hospital Medical Centre chose girls from schools in greater Cincinnati, selected to be racially and socioeconomically representative of the inner city, urban and suburban areas of Hamilton County. Westat Inc. selected participants from a membership list of eligible families with girls aged 9 or 10 years old, enrolled in a prepaid medical programme of Group Health Association in Washington D.C. White girls in this last sample were boosted by recruiting Girl Scouts from the same area. All subjects assented and parental consent to their participation was obtained. At baseline, black girls had a mean age of 10.1 years, whereas white girls were slightly younger with a mean age of 10.0 years. Over 90% of girls lived with their biological mother, but only 67.4 % of white girls and 41.9% of black girls lived with both natural parents. Black households had, on average, lower family incomes, lower education levels and less home ownership than white households, but a wide range of incomes and education levels was represented in each race group (Albertson et al., 2007). Over 90% of recruited girls had repeat measures at clinic visits 2, 3 and 4. Participation rates fell to 82% by visit 7, but at visit 10 (final measures) 89% of recruits took part (Striegel-Moore et al., 2006).

Measures: Parental data collected at baseline in clinics included demographics (self-reported race, education, income, family composition) diet and physical activity patterns and beliefs, medical history, anthropometry, blood pressure and serum lipids. Parent follow-up questionnaires were completed in Year 3.

Ten examinations of the girls were conducted at approximately one year intervals between 1987 and 1997 by centrally trained examiners. In Berkeley and Cincinnati examinations were

carried out in schools. In Washington D.C. girls were examined at Group Health Association clinics. Wearing hospital gowns or large T-shirts girls were weighed using a Health-o-meter electronic scale and their height in socks was measured using portable stadiometers. Height and weight were used to calculate BMI and BMIz scores were derived from the age and gender specific Centers for Disease Control and Prevention references. Upper thigh and upper arm circumferences were measured using a fiberglass tape. Skinfold thicknesses (triceps, suprailiac and subscapular) were measured using Holtain callipers. Blood pressure was measured using a standardised protocol. Puberty was assessed based on pubic hair distribution and areolar development stage. Twelve hour fasting blood samples (total cholesterol, HDL cholesterol, LDL cholesterol, apolipoprotein A1, apolipoprotein B) were collected. As a quality control procedure, repeat anthropometry and blood pressure measures were taken for 10% of the sample.

Girls' food intakes were measured with three day food records (2 weekdays, 1 weekend day) collected for visits 1, 2, 3, 4, 5 and visits 7, 8 and 10. Dietitians reviewed the food records, using standard questions to clarify incomplete responses. Food records were analysed using the Food Table version 19 of the Nutrition Coordinating Center nutrient database (Affenito et al., 2005).

Funding and declared interests: Research was supported by the National Heart, Lung and Blood Institute contract number HC-55023-26 and co-operative agreement number U01-HL-48941-44. Additional funding came from NHLBI grant HL/DK71122, General Mills Inc., the Bell Institute of Health and Nutrition, Minneapolis and grant 5R21DK075068 from the National Institute of Diabetes and Digestive and Kidney Diseases.

E.13 Project EAT Eating and Activity in Teens

http://www.sphresearch.umn.edu/epi/project-eat/

Country: United States of America

Study dates: Established 1998/1999, ongoing

Study design and setting: Project EAT is a longitudinal cohort study of nutrition, physical activity and weight status in young people from Minnesota of diverse ethnic and socioeconomic backgrounds. The baseline survey at Time 1 (Project EAT-I) took place during the 1998/99 school year, with follow-ups at approximately five year intervals, Time 2 (Project EAT-II) between April 2003 and June 2004 and Time 3 (Project EAT-III) between November 2008 and October 2009. A fifteen year follow-up began in 2015 at Time 4 (Project EAT-IV).

Recruitment and participants: At baseline Project EAT-I recruited 4,746 school students in the age range 11 to 18 years, from 31 public schools in the Minneapolis, St Paul and Osseo metropolitan school districts of Minnesota. Approximately 34% of students attended middle school, 66% attended high school. The mean age at baseline was 14.9 years and 50% were female. 49% were White, 19% African American, 19% Asian American, 6% Hispanic, 3.5% Native American and 4% other/mixed ethnicity. Participants' socio-economic status, based on parental education and employment status plus poverty indicators, was 18% lower, 19% lower-middle, 27% middle, 23% upper-middle and 13% upper SES (Fulkerson et al., 2008).

By Time 2 1,074 (23%) of the original participants were lost to follow-up due to missing contact information at Time 1 or no address at Time 2. Researchers were able to contact 3,672 participants of whom 2,516 (53% of the original cohort) completed valid EAT II five year follow-up surveys (Fulkerson et al., 2008). At this point 55% of respondents were female and 61% were white. Two-thirds of them were young adults who had left high school. At Time 3, 2,287 participants (all now adult) completed the EAT-III 10 year follow-up survey. These respondents included 1,030 men and 1,257 women (55%), aged between 20 and 31 years (Berge et al., 2015).

Measures: Baseline measures were conducted in schools during the 1998/99 school year. Trained research staff administered the youth form of the 2007 Willet semi-quantitative food frequency questionnaire, the 149 item Youth/adolescent Questionnaire. Students also completed the Project EAT-I survey, a 221 item survey which included questions about demographic characteristics, eating habits, physical activity (Leisure Time questionnaire) and sedentary behaviours as well as questions on self-reported height and weight. After completing the surveys students also had their height and weight measured by trained staff using a standardised protocol.

At follow-ups the food frequency questionnaires and repeat surveys, including questions on self-reported height and weight, were sent by mail. The Project EAT-I survey was revised for use at Time 2, with one version for participants still at high school and a second version for post high school adults, and re-revised for use at Time 3. The youth or adult form of the 2007 Willett semi-quantitative food frequency questionnaire was applied as appropriate.

At baseline high correlations (*r* 0.85 for girls, *r* 0.89 for boys) were observed between self-reported and measured height and weight (Cutler et al., 2012) so researchers relied upon self-reported height and weight to calculate BMI. At baseline and during adolescence, overweight was defined as a BMI at or above the 85th percentile, based on age and gender specific reference data from the Centers for Disease Control and Prevention. Overweight status at follow-up for adults was defined as a BMI ≥ 25 kgm^{-2} (Quick et al., 2013).

Funding and declared interests: Project EAT- I and -II received support from the Maternal and Child Health Program, Health Resources and Services Administration and the Department of Health and Human Services (grants MCJ-270834 and R40 MC 00319). Project EAT-III was funded by the National Heart, Lung, and Blood Institute National Institutes of Health (grant R01HL084064). The ongoing Project EAT-IV is also supported by the NHLBI NIH (grant R01HL116892).

E.14 RAINE Western Australian Pregnancy Cohort Study

Country: Australia

Study dates: Established 1989/1991, ongoing

Study design and setting: The Raine Study is a pregnancy/offspring cohort based around Perth, Western Australia.

Recruitment and participants: Pregnant women with a gestational age between 16 and 20 weeks, who attended antenatal clinics at King Edward Memorial Hospital, Perth or nearby private practices between May 1989 and November 1991, were invited to enrol. Conditions of enrolment included an expectation to deliver at King Edward Memorial Hospital, the intention to live in Western Australia (thus allowing childhood follow-up) and enough understanding of English to perceive the implications of taking part in the study (Newnham et al., 1993). The study began with 2,900 pregnant women and 2,804 of those women had 2,868 live births. By the 14 year follow-up, 2,337 subjects (81.5%) were still in the cohort, of whom 49% were female.

Measures: Children were followed up at birth and throughout childhood at 1, 2, 3, 5, 8, 10 years old (Ambrosini, Gina L et al., 2009) with follow-ups in adolescence during 2003-2006 (at 14 years) and 2006-2009 (at 17 years) (Ambrosini et al., 2013). During adolescent follow-up the primary caregiver gave sociodemographic information including mother's highest level of education, family income, single-parent household status and caregiver's smoking status. The primary caregiver also completed the 12 question General Functioning Scale, about family communication, problem solving, responses and behaviour control.

The adolescent's usual dietary intake over the previous year was measured using a modified version of a semi-quantitative food frequency questionnaire developed by the Commonwealth Scientific and Industrial Research Organization (CSIRO). The modified FFQ was compared with a 3 day food diary in the same cohort and was found to correctly rank most nutrient intakes. Adolescents completed the FFQ with help from their parent or primary caregiver and responses were checked by a research nurse at the clinic visit.

Adolescent physical activity was assessed by self-report of how many times they did exercise that caused them to become sweaty or out of breath. Their sedentary behaviour was assessed by self-report of television and video watching.

During study clinic visits, research nurses measured height and weight with standard calibrated equipment. Body Mass Index was calculated from these measures to assess weight status using age and gender specific BMI references from the US Centers for Disease Control and Prevention. Adolescents with a BMI below the 5th percentile were classified as underweight, those between the 5th and 85th percentile were considered to be a healthy weight, those between the 85th and 95th percentile were considered at risk of overweight and those above the 98th percentile were deemed overweight.

Funding and declared interests: The RAINE study was supported by the Faculty of Medicine, Dentistry and Health Sciences and the Raine Research Foundation of the University of Western Australia, the National Health and Medical Research Council of Australia (NHMRC) grants 353514 and 403981 and the King Edward Memorial Hospital Research Foundation. Later funding came from the Telstra Research Foundation of Australia, the Australian Rotary Health Research Fund, the Western Australian Health Promotion Research Foundation (Healthway), the National Heart Foundation of Australia, Beyond Blue Cardiovascular Disease and Depression Strategic Research Program (ID G08P4036), the UK Medical Research Council grant U105960389, the Telethon Institute for Child Health Research; the Women's and Infants Research Foundation and Curtin University.

Appendix F Prototype CORA questionnaire with notes

INTRODUCTION

Hello.

We'd like to ask you (*OR your parent or carer*) some questions about yourself and your family, what you eat and drink and some of the things that you do. It is O.K. to ask someone if you need help to fill in your answers.

Thank you.

ABOUT TODAY

1. What is today's date?

DAY	MONTH	YEAR

ABOUT YOU[1]

1. What is your date of birth? (*OR how old are you now? When is your next birthday?*)

DAY	MONTH	YEAR

2. Are you a boy or a girl? Please tick one box.

BOY	GIRL

3. What is your ethnicity?[2] Please tick one box

WHITE	NON-WHITE	PREFER NOT TO SAY

[1] Models often adjusted for age, sex, ethnicity and pubertal status. These variables are considered to be confounders of the relationship between diet and obesity.

[2] Race (White or black girl) was a predictor of BMI percentile change, onset of overweight and onset of obesity

4. Have you noticed any changes to your body as you begin to grow up, such as pubic hair or hair in your armpits?[3] Please tick one box

YES	NO

ABOUT YOUR HOME AND FAMILY

1. What is your home postcode (*for IMD*) /average household income/mother's highest education level?[4]

2. How many brothers and sisters do you have, living at home with you?[5]

BROTHERS	SISTERS

3. Do you think one or both of your parents is overweight?[6] Please tick one box.

YES	NO

4. Do any of the adults living at home with you, smoke while at home?[7] Please tick one box.

YES	NO

[3] These are variables we have as measures of pubertal stage in the ALSPAC dataset.

[4] Models often adjusted for SES. Only one measure is needed. Mother's highest education level is frequently used as a measure of SES in ALSPAC papers.

[5] Number of siblings in the household was a predictor of BMI percentile change and onset of overweight.

[6] Females who perceived that biological parent(s) were overweight were at risk of o/w.

[7] Models often adjusted for parental smoking.

ABOUT MEALS and EATING

1. How often do you eat breakfast in the morning?

Never or rarely	1 to 3 times a week	4 times a week, or more

2. How often do you eat three meals a day? (This may include breakfast)

Never or rarely	1 to 3 times a week	4 times a week, or more

3. How often do you have a meal at home with grown-ups and children eating together?

Never or rarely	1 to 3 times a week	4 times a week, or more

4. How many times a day do you have something to eat in-between meals?[8]

Never or rarely	1 to 3 times a day	4 times a day, or more

5. How often do you eat a take-away meal or eat a meal in a fast food restaurant?[9]
   Examples of take-away and fast food meals are fish and chips, pizza, fried chicken, Chinese, Indian, burgers or kebabs. (By meal we mean more than just a drink or a portion of chips.)

Less than once a week	1 or 2 times a week, or more.

[8] i.e. "Snacking" eating behaviour – want to ask about eating any kind of food between meals, not just "snack food". Will capture energy dense snack type foods in a later Q.

[9] Take-away and fast food examples are modelled on questions in the NDNS survey.

ABOUT DRINKS[10]

1. How often do you drink milk, plain or flavoured? (Not counting milk on breakfast cereal)

Never or rarely	Sometimes	Once a day	More than once a day.

2. How often do you drink 100% fruit juice?

Never or rarely	Sometimes	Once a day	More than once a day.

3. How often do you drink sugary drinks?

   Examples of sugary drinks are non-diet, fizzy soft drinks such as lemonade, Fanta, Coca-Cola, Pepsi, energy drinks, sports drinks, sweetened fruit drinks, fruit squash such as Robinsons and fruit cordial such as Ribena.

Never or rarely	Sometimes	Once a day	More than once a day.

4. How often do you drink diet drinks?[11]

   Examples of diet drinks are low calorie or low sugar soft drinks such as Diet Coke, Coke Zero, Diet Pepsi and Pepsi Max.

Never or rarely	Sometimes	Once a day	More than once a day.

[10] Drinks FFQs in ALSPAC use 5 frequency categories: Never or rarely, once in 2 weeks, 1 to 3 times a week, 4 to 7 times a week, more than once a day.

[11] The list of examples of diet drinks does not include “no added sugar” squash/cordial as they still contain free sugars from the fruit. Children may not know if they consume added sugar squash or no added sugar squash, especially with manufacturers’ recent reformulations in response to the UK sugar levy on SSBs.

ABOUT FOOD

1. How often do you eat wholegrain foods?[12]
   Examples of wholegrain foods are whole wheat breakfast cereal[13] such as Weetabix, Shreddies or Shredded Wheat, bran flakes, porridge oats, wholemeal or granary bread, brown rice and brown pasta.

Never or rarely	Sometimes	Once a day	Two or more times a day

2. How often do you eat dairy foods? Examples of dairy foods are cheese, cottage cheese, yogurt and milk added to breakfast cereal.

Never or rarely	Sometimes	Once a day	Two or more times a day

3. How often do you eat vegetables including salad (not including potatoes or beans)?
   Vegetables can be fresh, frozen, dried or canned, raw or cooked. Include vegetables that are in cooked dishes too, such as onions or carrots in a casserole or stew.

Never or rarely	Sometimes	Once a day	Two or more times a day

4. How often do you eat fruit (not including fruit juice)?
   Fruit can be fresh, frozen, dried or canned, raw or cooked.

Never or rarely	Sometimes	Once a day	Two or more times a day

5. How often do you eat food such as cakes, buns, Danish pastries, biscuits, chocolate, sweets, ice-cream or crisps, either as part of a meal or in-between meals?

Never or rarely	Sometimes	Once a day	Two or more times a day

[12] Categories are designed to capture > 1 serving/day or > 2 servings/day cut off.

[13] Examples of whole wheat breakfast cereal taken from https://www.bda.uk.com/foodfacts/wholegrains.pdf

OTHER QUESTIONS

1. Have you ever tried to diet?[14] (*OR do you think that you are a fussy eater?*) Please tick one box

YES	NO

2. Are you happy with your body shape?[15] Please tick one box?

YES	NO

3. When you have school the next day, by what time in the evening are you usually asleep?[16]

By 8 pm	By 9 pm	By 10 pm	By 11 pm	After 11 pm

4. How do you usually travel to and from school?[17] Please tick one box.

Walk	Bicycle	Car	Taxi	Bus	Train	Other

[14] Some girls in NGHS had dieting concerns even at 11 years old. In GUTS dieters gained more weight than non-dieters.

[15] Baseline body satisfaction was associated with lower risk of overweight at follow-up in Project EAT females. Body dissatisfaction predicted change in BMI percentile in NGHS girls.

[16] Sleep affects energy balance via hormonal influences on metabolism and appetite, so confounds the relationship between diet and obesity outcomes. In ALSPAC, short sleep duration at 3 years predicted a higher risk of obesity at 7 years. The NHS advises 10 hours sleep for 9 year olds. If most children get up between 7am and 8am for a 9am school start, this means going to sleep between 9pm and 10pm.

[17] Physical activity mediates energy balance so confounds the relationship between diet and obesity outcomes. Physical activity was not an important predictor of BMI change in NGHS girls but increased physical activity reduced the risk of overweight in females in Project EAT. Active travel (walk or bicycle) to and from school is a simple measure of routine physical activity.

Appendix G Preparation of variables and imputation of missing observations.

Variables in the ALSPAC dataset were identified that would help determine which children in the cohort should be included or excluded at baseline, each child's obesity status at baseline and follow-up and the potential predictors of obesity to which they were exposed. (See Chapter 8 Table 5 and Table 6). This appendix sets out how each variable was prepared for model development and, if applicable, how missing observations were imputed.
New variables were generated, based on existing variables. Existing variables are shown with their original name, as given in the ALSPAC dataset. New variable names begin with the initials of the post graduate researcher who first generated that variable (zy = Ziyi Li, cr = Catherine Rycroft). Dummy candidate variables were coded 1 for the category thought most likely to predict future obesity and 0 for the opposite category.

G.1 Inclusion and exclusion variables

Attended the Focus at 10+ Clinic

7,557 children attended the F10+ clinic (In_f10). Children who did not attend F10 were excluded from analyses (7,888 observations).

Returned a 3 day food diary at 10 years

7,462 children in F10+ returned a 3 day food diary in enough detail to have Energy intake in kilojoules as a daily average (fd10kj). 95 children did not have this information so were excluded from analyses.

Twins

Six children in F10 are missing pregnancy size (mz010), so it was not clear if they are singletons (1) or twins (2). Their status was checked by listing pregnancy ID (cidB2798) and birth order within pregnancy (qlet).

- Singletons had a unique pregnancy ID and birth order = A.

If in F10+, missing pregnancy size (mz010) was recoded as singleton (1).

- Twins had a duplicated pregnancy ID, first birth order = A, second birth order = B.

If in F10, missing pregnancy size (mz010) was recoded as twin (2).

There were 202 twins (101 pairs) in F10. They were excluded from analyses.

G.2 Obesity status variables

Sex, age, height and weight were needed to calculate each child's BMIz score at the Focus at 10+, Teen Focus 2 and Teen Focus 3 clinics and to classify weight status as obese or not obese. Missing observations for sex, age, height and weight were imputed as follows:

Sex

Six children in F10 are missing Sex (kz021). Sex was imputed based on responses to sex-specific puberty questions about development of pubic hair.

- If a child had a non-missing response to a question about development stage of pubic hair (male) at different ages (pub355, pub455, pub555, pub655, pub755 or pub955), missing sex (kz021) was recoded as male (1).
- If a child had a non-missing response to a question about development stage of pubic hair (female) at different ages (pub335, pub435, pub535, pub635, pub735, or pub935), missing sex (kz021) was recoded as female (2).

Age at F10

The same six children in F10 are missing age in days (fd003a) and age in weeks (fd003b). Five of them have ages in weeks at other times, reported in child based questionnaires completed by mother, KU, KW, TA and TB. All five are close to the mean age.

- Missing age in days at F10 (fd003a) was imputed using the integer of mean age at F10, 3902 days.
- Missing age in weeks at F10 (fd003b) was imputed using the integer of mean age at F10, 557 weeks.

Age at TF2

1,857 children in F10 are missing age in weeks at TF2 (fg0011b). The difference between mean age at F10 and mean age at TF2 is (723 - 557) = 166 weeks.

- If in F10, missing age in weeks at TF2 (fg0011b) was imputed using observed age in weeks at F10 (fd003b) + 166 weeks

Age at TF3

2,553 children in F10 are missing age in weeks at TF3 (fh0011b). The difference between mean age at F10 and mean age at TF3 is (807 - 557) = 250 weeks.

- If in F10, missing age in weeks at TF3 (fh0011b) was imputed using observed age in weeks at F10 (fd003b) + 250 weeks

Height at F10, TF2 and TF3 for boys and girls

73 children in F10 are missing height in cm at F10 (fdms010).

1,865 children in F10 are missing height in cm at TF2 (fg3100).

2,594 children in F10 are missing height in cm at TF3 (fh3000).

Measured heights in cm at each consecutive clinic F10, F11, TF1, TF2, TF3 and TF4, at ages 10+, 11+, 12+, 13+, 15+ and 17+ years, were summarised and the differences in mean heights between each clinic were calculated. Missing observations were imputed by subtracting the difference from known height at the next clinic(s) or by adding the difference to known height at the previous clinic. Imputations were run separately for boys and girls as they had different growth trajectories.

Differences in mean heights between clinics for boys:

F10 to F11	+ 6.22cm
F11 to TF1	+ 6.88cm
TF1 to TF2	+ 7.92cm
TF2 to TF3	+ 9.42cm
TF3 to TF4	+ 4.44cm

- Missing height for boys at F10 (fdms010) was imputed using height at F11 (fems010) - 6.22. No height variable was available from the previous clinic.

If still missing, height for boys at F10 (fdms010) was imputed using height at TF1 (ff2000) -13.1

- Missing height for boys at TF2 (fg3100) was imputed using height at TF3 (fh3000) - 9.42

If still missing, height for boys at TF2 (fg3100) was imputed using height at TF1 (ff2000) + 7.92

- Missing height for boys at TF3 (fh3000) was imputed using height at TF4 (FJMR020) - 4.44

If still missing, height for boys at TF3 (fh3000) was imputed using height at TF2 (fg3100) + 9.42

Differences in mean heights between clinics for girls:

F10 to F11	+ 7.28cm
F11 to TF1	+ 6.14cm
TF1 to TF2	+ 4.32cm
TF2 to TF3	+ 2.71cm
TF3 to TF4	+ 0.57cm

- Missing height for girls at F10 (fdms010) was imputed using height at F11 (fems010) – 7.28. No height variable was available from the previous clinic.

If still missing, height for goys at F10 (fdms010) was imputed using height at TF1 (ff2000) - 13.42

- Missing height for girls at TF2 (fg3100) was imputed using height at TF3 (fh3000) – 2.71

If still missing, height for girls at TF2 (fg3100) was imputed using height at TF1 (ff2000) + 4.32

- Missing height for girls at TF3 (fh3000) was imputed using height at TF4 (FJMR020) – 0.57

If still missing, height for girls at TF3 (fh3000) was imputed using height at TF2 (fg3100) + 2.71

After these imputations, all but 16 children in F10 have observations for height at F10. 906 are still missing height at TF2 and 799 are missing height at TF3.

Weight at F10, TF2 and TF3 for boys and girls

46 children in F10 are missing weight in kg at F10 (fdms026),

1,873 children in F10 are missing weight in kg at TF2 (fg3130).

2,603 children in F10 are missing weight in kg at TF3 (fh3010).

Measured weights in kg at each consecutive clinic F10, F11, TF1, TF2, TF3 and TF4, at ages 10+, 11+, 12+, 13+, 15+ and 17+ years, were summarised and the differences in mean weights between each clinic were calculated. Missing observations were imputed either by subtracting the difference from known weight at the next clinic(s) or adding the difference to known weight at the previous clinic. Imputations were run separately for boys and girls as they had different growth trajectories.

Differences in mean weights between clinics for boys:

F10 to F11	+ 5.11 kg
F11 to TF1	+ 5.71 kg
TF1 to TF2	+ 6.19 kg
TF2 to TF3	+ 9.51 kg
TF3 to TF4	+ 8.23 kg

- Missing weight in kg for boys at F10 (fdms026) was imputed using weight at F11 (fems026) - 5.11. No weight variable was available from the previous clinic.

If still missing, weight for boys at F10 (fdms026) was imputed using weight at TF1 (ff2030) - 10.82 (- 5.11 - 5.71)

- Missing weight for boys at TF2 (fg3130) was imputed using weight at TF3 (fh3010) - 9.51

If still missing, weight for boys at TF2 (fg3130) was imputed using weight at TF1 (ff2030) + 6.19

- Missing weight for boys at TF3 (fh3010) was imputed using weight at TF4 (FJMR022) - 8.23

If still missing, weight for boys at TF3 (fh3010) was imputed using weight at TF2 (fg3130) + 9.51

Differences in mean weights between clinics for girls:

F10 to F11	+ 6.14 kg
F11 to TF1	+ 5.36 kg
TF1 to TF2	+ 4.42 kg
TF2 to TF3	+ 4.69 kg
TF3 to TF4	+ 3.76 kg

- Missing weight in kg for girls at F10 (fdms026) was imputed using weight at F11 (fems026) – 6.14. No weight variable was available from the previous clinic.

If still missing, weight for girls at F10 (fdms026) was imputed using weight at TF1 (ff2030) -11.5

- Missing weight for girls at TF2 (fg3130) was imputed using weight at TF3 (fh3010) – 4.69

If still missing, weight for girls at TF2 (fg3130) was imputed using weight at TF1 (ff2030) + 4.42

- Missing weight for girls at TF3 (fh3010) was imputed using weight at TF4 (FJMR022) – 3.76

If still missing, weight for girls at TF3 (fh3010) was imputed using weight at TF2 (fg3130) + 4.69

After these imputations, all but 14 children in F10 have observations for weight at F10. 919 are missing weight at TF2 and 805 are missing weight at TF3.

BMI at F10, TF2 and TF3

Using observed and imputed height and weight, NEW variables for BMI at F10, TF2 and TF3 clinics were generated using the formula: weight in kg/(height in m)^2.

- BMI at F10 (zyF10bmi) mean BMI 18.3 kg/m^2
- BMI at TF2 (zyTF2bmi) mean BMI 20.5 kg/m^2
- BMI at TF3 (zyTF3bmi) mean BMI 21.6 kg/m^2

BMIz at F10, TF2 and TF3

BMI was converted to age and sex adjusted BMIz scores using the STATA command "zanthro" (Vidmar et al., 2013) to generate NEW variables for BMIz, based on UK 1990 age and sex specific growth references, at F10, TF2 and TF3 clinics.

- BMIz at F10 (crF10bmizuk)
- BMIz at TF2 (crTF2bmizuk)
- BMIz at TF3 (crTF3bmizuk)

Obesity status at F10, TF2 and TF3

Obesity status was based on BMIz at F10, TF2 and TF3, using ≥ 1.64 (95th percentile) as the cut-off. NEW dummy variables for obesity status were coded obese (1) or not obese (0). These derived variables had some missing observations for the 7,557 children who attended the Focus at 10+ clinic.

- Obesity at F10 (crF10obese) 26 missing. 7,531 obs. (**14.5% obese**).
- Obesity at TF2 (crTF2obese) 919 missing. 6,638 obs. (**13.3% obese**).
- Obesity at TF3 (crTF3obese) 805 missing. 6,752 obs. (**13.0% obese**).

Children who were already obese at baseline (F10) or whose obesity status was unknown were excluded from analyses. Obesity at TF2 (crTF2obese) = 1 was the obesity "event" that the model aimed to predict. There were 251 new obesity outcomes at TF2, which reduced to 239 after exclusion criteria were applied.

G.3 Candidate variables (Potential predictors)

G.3.1 Non-diet predictors

Child's ethnic background

Child's ethnic background (c804) is a dichotomous variable, based on the child's mother's ethnic group and the mother's partner's ethnic group. In F10, 6,524 children have a white ethnic background (mother and partner are white) and 270 children have a non-white ethnic background (one or both of mother and partner are non-white). The remaining 763 children (10%) are missing this variable. *There were no suitable variables for imputation*. The variable was recoded as a NEW dummy variable (crc804, shown as crEthnicity in the model tables) as Non-white = 1, White = 0.

Puberty status: armpit hair at age 10 years 8 months

In F10, 2,568 children in F10 are missing a response for whether "hair has started to grow in respondent's armpits" in the Puberty Questionnaire completed by the child's mother/ carer at 10 years 8m (pub370). Responses to the same question at older ages, 11+, 13+, 14+, 15+, 16+ and 17+ years are available in the dataset (pub470, pub570, pub670, pub770, pub870, pub970). It is reasonable to assume that children without armpit hair at older ages did not have armpit hair at 10 years 8 months.

- If the response to "hair has started to grow in respondent's armpits" was No (2) at an older age, missing responses for armpit hair at age 10+ years (pub370) if in F10 were recoded as No (2).

Responses to the same question at earlier ages are not available in the dataset. Less than half the children in F10 who had armpit hair at age 11+ years had armpit hair at 10+ years; we cannot assume that those who responded yes at 11 years also had armpit hair at the earlier age. However, 10 years 8 months is early for puberty onset, so it is reasonable to assume that most children with missing responses do not have armpit hair.

- If still missing, responses for armpit hair at age 10+ years (pub370) if in F10 were recoded as No (2).

After imputations, all 7,557 children in F10 have observations. The variable was recoded as a NEW dummy variable (crpub370, shown as crArmpit in the model tables) as Armpit hair = 1, No armpit hair = 0.

Socio-economic status (SES) - Mother's highest education level

Several SES variables were available in the ALSPAC dataset, including quintile of index of multiple deprivation (IMD) when the child was ~10 yrs. 6 months old. IMD is known to be associated with childhood obesity but was not chosen as a predictor as it is a composite measure which may introduce unexpected confounding. Mother's highest education level was preferred as it was used in several ALSPAC papers as a confounder of associations between an exposure and obesity/overweight outcomes (Bigornia et al., 2014; Noel et al., 2011) and parental education was a predictor in several childhood obesity predictive models documented elsewhere (Pei et al., 2013; Manios et al., 2013; Classen and Hokayem, 2005). During pregnancy, mothers were asked about their highest education level (c645a), certificate of secondary education (C.S.E.), vocational, O level, A level and degree. In F10, 1,094 children (16% of those with observations) had mothers who were educated to degree level. 661 children (9%) are missing this variable. *There were no suitable variables for imputation*. A NEW dummy variable (crMumDegree) was generated, coded as No degree = 1, Mother has degree = 0.

Number of siblings – Number of children in household

When the child was approximately 10 years and 2 months old, mothers were asked how many children, under the age of 16 years, were living in the household (q3002). In F10, 791 children (12%) were in households with only one child, assumed to be only children with no siblings. However, 1,114 children (15%) are missing this variable. *There were no suitable variables for imputation*. A NEW dummy variable (crONLYCHILD) was generated, coded as Only child = 1, Not an only child = 0.

Perception of parental overweight – based on mother's BMI

Children in ALSPAC were not asked directly whether they thought one (or both) of their parents were overweight. However, mothers self-reported their own weight in kg (p1290) and height in cm (p1291) when their child was approximately 9 years and 2 months old. Mothers height and weight were used to derive a NEW variable for mother's BMI (crMumsBMI), from which to establish mother's overweight status. In F10, 25% were missing mother's self-reported height and 28% were missing mother's self-reported weight. As a consequence, 29% were missing mother's BMI.

Mothers also self-reported pre-pregnancy height and weight in a questionnaire at approximately 12 weeks gestation, from which pre-pregnancy BMI (dw042) had been generated. For mothers of children in F10, the difference between mother's mean BMI when the child was 9 years and 2 months and mother's mean pre-pregnancy BMI was (24.74 – 22.89) = 1.85 kg/m^2.

- Missing mother's BMI (crMumsBMI) in F10 was imputed by adding the difference (1.85) to known pre-pregnancy BMI.

This assumes that all mothers experienced BMI gains by the time their child was 9 years old, with the same absolute gain regardless of baseline BMI, taking no account of later pregnancies (unknown). After imputation, 503 observations (6.6%) were still missing for mother's BMI.

Using a cut-off of BMI ≥ 25 for adult overweight, the variable was recoded as a NEW dummy variable for Mother's overweight (crMumoverw) coded Mum overweight = 1, Mum not overweight = 0.

Smokers in household

Mothers reported the number of smokers in the household, when their child was approximately 10 years and 2 months old (q3031). In F10, 70% reported no smokers in the household, but 17% of observations were missing.
Mothers/carers were also asked about their own frequency of smoking in the past two weeks (r6020), one year later when their child was approximately 11 years and 2 months old. In F10, 70% of respondents did not answer this question, but 106 mothers with a missing observation for the number of smokers in the household (q3031) did smoke. It is reasonable to assume that those who smoked when their child was 11 years old were also smokers when their child was 10 years old.

- Missing (.) number of smokers in household (q3031) was recoded as 1 (1 smoker in household) if the mother/carer smoked at any frequency other than none (0) or infrequently (97) just one year later (r6020).

Cross tabulation of q3031 and r6020 showed that in supposedly non-smoking households, 101 mothers did smoke one year later.

- For consistency, reported no (0) smokers in household (q3031) was also recoded as 1 (1 smoker in household) if the mother/carer smoked at any frequency other than none (0) or infrequently (97) just one year later (r6020).

In F10, after imputation and recoding, 67.5% have no smokers in the household, but 15.5% of observations are still missing. There are no other suitable variables for imputation. The

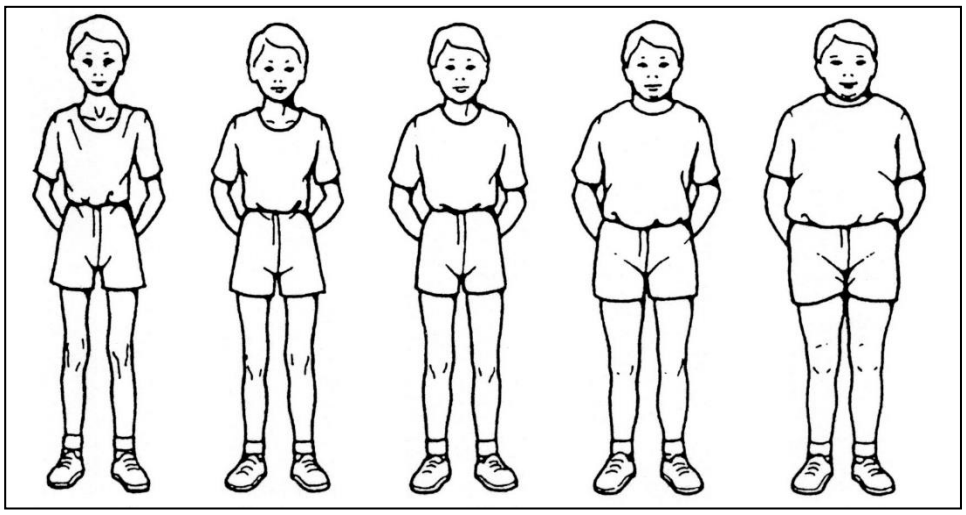
variable was recoded as a NEW dummy variable for Smokers in household (crSMOKERS) coded Smokers in household = 1, No smokers = 0.

Body satisfaction – child's perceived body shape versus child's desired body shape

Children answered questions about their body shape when they were approximately 10 years and 8 months old, selecting one of five drawings (of girls or boys as appropriate) that was most like themselves (cch200). From the same set of drawings, children then chose the one that they would most like to be (cch201). Drawings were coded Very thin = 1, Thin = 2, Average = 3, Fat = 4, Very fat = 5. A NEW variable, Difference between child's perceived and desired body shape was generated (crBShbDiff) by subtracting the coded values one from the other. Matching answers, with a difference of 0, were interpreted as body satisfaction, mismatched answers with any other value, were interpreted as body dissatisfaction.

The drawings used in ALSPAC (See Figure G-1, representing boys) were not referenced, but they are similar to a pictorial instrument developed to examine perceptions of body figure in a cross-sectional study of preadolescent children (Collins, M.E., 1991).

Figure G-1 Pictures of 5 different boys from ALSPAC questionnaire CCH



In F10 1,644 children (22%) did not identify their perceived body shape and/or their desired body shape. *There were no suitable variables for imputation*.

The variable was recoded as a NEW dummy variable for Child's Body satisfaction (crBODYSAT), coded Unsatisfied = 1, Satisfied = 0.

Sleep duration

Mothers were asked about the time their child usually wakes up, in hours (ku340a) and minutes (ku340b), on school days and the time their child usually goes to sleep, in hours (ku341a) and minutes (ku341b) on school days, when the child was aged approximately 9 years and 7 months. A new variable, ku school days Total sleep in minutes (crkuSLEEPmn) was generated from these four variables. Mean sleep duration was 626.8 minutes or 10 hours and 27 minutes.
Approximately 15% of children in F10 were missing waking and sleeping time variables, so sleep duration at 9+ years could not be generated for 1,115 children.

Mothers were asked the same questions when the child was aged approximately 11 years and 8 months old (kw4060a, kw4060b, kw4061a, kw4061b). A new variable, kw school days Total sleep in minutes (crkwSLEEPmn) was generated from these four variables. Mean sleep duration at this older age was 588.9 minutes or 9 hours and 49minutes, 37.9 minutes less than before. The difference in sleep duration between the two means was used to impute missing observations.

- If in F10, missing ku school days Total sleep in minutes (crkuSLEEPmn) was imputed using kw school days Total sleep in minutes (crkwSLEEPmn) + 37.9 minutes.

This added 450 observations, but 665 or 9 % of children in F10 were still missing sleep duration at 9+ years. There were no alternative variables for imputation.

The Millpond Children's Sleep Clinic, cited by the NHS, recommends that children aged 10 years should have 9 hours 45 minutes sleep a night, with children aged 11 years needing slightly less at 9 hours 30 mins (NHSUK, 2017). Based on Total sleep in minutes (crkuSLEEPmn) and Age of study child in weeks (ku991b) when the sleep questions (ku340a, ku340b, ku341a and ku341b) were asked, a NEW dummy variable for Child met NHS sleep recommendations for age next birthday (crkuSLEEPrec) was generated, coded Does not meet sleep recommendation = 1, Meets sleep recommendation = 0.

Physical activity - Active travel

Moderate to vigorous physical activity (MVPA) was objectively assessed by the child wearing an activity monitor at age 11 years 6 months as part of the Focus at 11+ clinic, but this variable did not readily translate into a question.

Instead active transport (walking, bicycling) to/from school when the child was approximately 11 years and 8 months old was used as a proxy for physical activity at an earlier age. Mothers/carers were asked to tick their child's usual mode of transport between home and

school. Options were walking, wheelchair, public transport, school bus, car, bicycle and other. It is likely that some children were attending secondary school when the question was asked. Approximately 3,500 or 46% of children in F10 travelled to school most days or on some days by walking (kw7010) or by bicycle (kw7015). Slightly more, 3,600 or 47% of children in F10 travelled home most days or on some days by walking (kw7020) or by bicycle (kw7025). Those missing active transport to and/or from school, were assumed to have used an inactive mode of travel, such as wheelchair, public transport, school bus or car. No imputation was required.

A NEW dummy variable for Child walks or cycles to and/or from school (crACTIVTR) was generated, coded No active travel = 1, Active travel = 0

G.3.2 Eating habits

Mothers/carers answered questions about their teenager's eating habits, when their child was approximately 13 years and 1 month old. (These variables are used as proxies for eating habits at an earlier age.) There were no ALSPAC questions that readily matched Eats breakfast or Eats family meals, although these eating habits may be captured to some extent by the question about Three meals (or more) a day. Breakfast cereal intake from the 3 day food diary at F10 was considered as an indicator of breakfast eating, but high fibre breakfast cereal was already used as a component of whole grain intake.

Three meals a day - Number of real meals a day the teenager has now (ta8004).
In F10, 70% of children with observations had three or more meals/ day as teenagers.
25% of children in F10 are missing observations, but there are no alternative variables suitable for imputation. The variable was recoded as a NEW dummy variable for Teenager has three (or more) meals a day (cr3Meals) coded No, not 3 meals = 1, Yes, three meals or more = 0.

Eats in-between meals - Teenager snacks all day or has meals – school days (ta8000) or at weekend (ta8002).
Responses to questions about snacking and/or meals were consistent with responses to how many real meals a day the teenager has now. Snacking was more prevalent at weekends; more than two thirds of teenagers surveyed had meals but did not snack much on school days whereas not quite half had meals but did not snack much at weekends.
Cross tabulation showed that 46% of teenagers with observations who had taken part in F10 did not snack much on either school days or at weekends. Approximately 25% of teenagers snacked and also had meals on school days and at the weekend. 22% of teenagers from F10 had meals and didn't snack much on school days but did snack at weekends. Far fewer, 2%, had meals and snacked on school days but did not snack much at weekends. A further 2% of

teenagers from F10 had days when they snacked all day but had no real meals, while the remaining 3% were "other" on school days and/or weekends.

25% of children in F10 are missing observations for these two variables. Mean intakes of foods typically consumed as snacks, measured by a 3-day food diary at 13 years old, were explored by snacking frequency to find a suitable imputation variable. Among children who had taken part in F10, mean intakes of sweet biscuits (fg13bisc) and chocolate confectionery (fg13choc) were only slightly higher in teenagers who snacked compared with those who did not snack much. Conversely, mean intakes of buns, cakes, pastries and fruit pies (fg13bun) and items of fruit such as apples and pears (fg13sppl) and bananas (fg13bana) were marginally lower in teenagers who snacked compared with those who did not snack much but had meals. Mean intakes of savoury biscuits (fg13sabi) and nuts (fg13nuts) were too low to see any difference by snacking category.

However mean intakes of crisps and savoury snacks (fg13snck) were almost 50% higher in teenagers who snacked (~19g/day or ~ 2/3 of a small packet of crisps) than in teenagers who did not snack much (~13g/day or ~ 1/2 of a small packet of crisps). Similarly mean intakes of confectionery (fg13conf) were higher among teenagers who snacked (9.3g/day) than teenagers who did not snack much (6.5g/day). As more teenagers consumed crisps and savoury snacks than consumed confectionery (which had a more skewed distribution and wider variance) the former was chosen for imputation, using the mean intake of teenagers who did not snack much on either school days or at weekends (12.7g/day) as the cut point between those who did not snack much and those who snacked all day.

- If in F10, missing Teenager snacks all day or has meals – school days (ta8000) was recoded as doesn't snack much (3) if crisps and savoury snacks weight (g)DD at 13 years (fg13snck) was less than or equal to 12.7g/day
- If in F10, missing Teenager snacks all day or has meals – school days (ta8000) was recoded as snacks all day but has meals (2) if crisps and savoury snacks weight (g)DD at 13 years (fg13snck) was more than 12.7g/day
- If in F10, missing Teenager snacks all day or has meals – weekends (ta8002) was recoded as doesn't snack much (3) if crisps and savoury snacks weight (g)DD at 13 years (fg13snck) was less than or equal to 12.7g/day
- If in F10, missing Teenager snacks all day or has meals – weekends (ta8002) was recoded as snacks all day but has meals (2) if crisps and savoury snacks weight (g)DD at 13 years (fg13snck) was more than 12.7g/day

After imputations, all 7,557children in F10 have observations but a larger proportion are categorised as having snacks. Now only 40% (2,997) of teenagers in F10 did not snack much on either school days or at weekends while 37% (2,798) snacked and had meals on school days and at the weekend. The variable was recoded as a NEW dummy variable for Teenager snacks or eats between meals (crSnacking) coded Snacks between meals = 1, Does not snack = 0.

Takeaway or fast-food frequency– Number of times teenager eats in fast food restaurant per week (ta8230)

In F10, only 6% of children with observations ate in a fast food restaurant once a week or more frequently as teenagers.

24% of children in F10 are missing observations. Mean intakes of foods typically consumed at fast food restaurants, such as burgers and kebabs (fg13kebb) and fried/roast potatoes and chips (fg13frpo) measured by a 3 day food diary at 13 years old, did not vary in line with fast food frequency. There are no other variables suitable for imputation. The variable was recoded as a NEW dummy variable, Teenager eats in Fast food restaurant one or more times a week (crFastFood) coded Fast Food once a week or more = 1, Fast Food less than once a week = 0.

G.3.3 Drinks frequency

Children completed a small drinks frequency questionnaire, as part of a larger questionnaire "Teeth and things", when they were approximately 10 years and 8 months old. Drinks included cola, other fizzy drinks such as flavoured fizzy water and lemonade, plain water, plain fizzy water, pure fruit juice, sweetened fruit drinks, drinks with added water (cordial, squash), flavoured milk, plain milk, tea, coffee and other. Frequency options were:

- does not drink
- drinks on special occasions only
- drinks at mealtimes only
- drinks at any time of day

The frequency distributions in F10 only are similar to distributions for all responses. After comparing reported drinks frequency with drink intakes measured by the 3 day food diary when children were approximately 10 years and 6 months old, three drinks frequency variables were selected for the predictive model. Between 18% and 30% of children in F10 were missing observations for drinks frequency, depending on the drink. Missing drinks frequencies for children who took part in F10 were imputed based upon the child's corresponding drinks intakes measured by the 3 day food diary, relating the quantified intake

to an appropriate drinks frequency category, with intake cut-offs derived from the exploration of the data.

Milk – Frequency child drinks plain milk (cch709)

In F10, 33% of children with observations did not drink plain milk and tended not to drink flavoured milk either. This frequency category had the highest proportion (24%) of children with zero total milk intakes (whole milk + semi-skimmed milk + skimmed milk) in the 3 day food diary, but many children did consume some milk, albeit in low quantities.
A further 7% of children reported that they only drank plain milk on special occasions, of whom 13% consumed no milk on the days measured by the 3 day food diary.
Once non-consumers were excluded, the interquartile ranges of total milk intake for "does not drink plain milk" and "drinks plain milk on special occasions only" overlapped considerably, indicating little difference between the two categories.

60% of children reported that they drank plain milk at mealtimes or at any time. Those who drank plain milk at any time tended to drink flavoured milk at any time too. Never-the-less 6% of these children consumed no milk on the days measured by the 3 day food diary. Once non-consumers were excluded, the interquartile ranges of total milk intake for "drinks plain milk at mealtimes" and "drinks plain milk at any time", overlapped almost completely, indicating poor discrimination between these two categories.

30% of children in F10 are missing observations for plain milk frequency. Their mean intake of total milk was slightly higher than that of children with observations (217g/day vs 204g/day), and greater than the mean intake of milk consumers in the "drinks milk on special occasions" category (186g/day), suggesting that more of them drank milk at meal times or any time.

Missing observations for plain milk frequency (cch709) were imputed using a NEW variable, Total milk weight (g) DD at 10 years (crfd10totMILK), which was generated by addition of whole milk, semi-skimmed milk and skimmed milk intakes from the 3 day food diary. Intake cut-offs for frequency categories were based on the serving size (1 serving = 244g) previously used for plain milk in the ALSPAC cohort (REF Noel et al 2011).

- If in F10, missing plain milk frequency (cch709) was recoded as does not drink plain milk (1) if Total milk weight(g) DD at 10 years (crfd10totMILK) was less than or equal to 122g/day or 0.5 servings/day. This allows for some milk intake not as a beverage – such as on cereal. On average about 100ml to 125ml of milk is added to a 30 to 40g serving of cereal.
- If in F10, missing plain milk frequency (cch709) was recoded as drinks plain milk on special occasions (2) if Total milk weight (g) DD at 10 years (crfd10totMILK) was greater

than 122g/day or 0.5 servings/day and less than or equal to 143g/day or 0.6 servings/day. (This is the 25th percentile of intake for the next intake category, based on those with milk intake>0.)

- If in F10, missing plain milk frequency (cch709) was recoded as drinks plain milk at mealtimes (3) if Total milk weight(g) DD at 10 years (crfd10totMILK) was greater than 143g/day or 0.6 servings/day and less than or equal to 244g/day or 1 serving/day.
- If in F10, missing plain milk frequency (cch709) was recoded as drinks plain milk at any time (4) if Total milk weight (g) DD at 10 years (crfd10totMILK) was greater than 244g/day or 1 serving/day.

After imputation, all 7,557 children in F10 have observations, but now 61% (up from 60%) of children drank plain milk at mealtimes or at any time. The variable was recoded as a NEW dummy variable for Beverage Milk (crMILK) coded Milk < 1 serving a day = 1 (Plain milk: does not drink, special occasions only), Milk >= 1 serving a day = 0 (Plain milk: mealtimes, any time).

Fruit juice – Frequency child drinks pure fruit juices (cch704)

In F10, 11% of children with observations did not drink pure fruit juice. This frequency category had the highest proportion (57%) of children with fruit juice intakes of 0g/day in the 3 day food diary. 14.5% only drank juice on special occasions, of whom almost half consumed no juice on the days measured by the 3 day food diary. Once non-consumers were excluded, mean intakes and interquartile ranges of fruit juice intake for “does not drink pure fruit juice” and “drinks pure fruit juice on special occasions only” were almost identical, indicating little difference between these two categories.
Almost three quarters of children reported that they drank juice at meals times only or at any time. Despite this over 25% of them did not have any recorded fruit juice intake on the 3 days surveyed by the food diary. Again, once non-consumers were excluded, the interquartile ranges of total milk intake for “drinks pure fruit juice at mealtimes” and “drinks pure fruit juice at any time”, overlapped almost completely, indicating poor discrimination between these two categories

20% of children in F10 are missing observations for pure fruit juice frequency. Their mean intake of fruit juice was lower than that of children with observations (102g/day vs 120g/day) and below the median intake of juice consumers in the “drinks pure fruit juice on special occasions” category (107g/day) so after imputation we might expect a greater proportion to be in this category. UK guidelines advise no more than 1 serving/day or 150ml/day of fruit juice, but individual cartons of fruit juice often contain 200g, so this was used as the serving size.

Missing observations for pure fruit juice frequency (cch704) were imputed using Fruit juice weight (g) DD at 10 years (fd10frju) from the 3 day food diary. Chosen intake cut-offs for frequency categories were based on the mean intakes of juice for juice consumers (excluding non-consumers) in each category.

- If in F10, missing pure fruit juice frequency (cch704) was recoded as does not drink pure fruit juice (1) if Fruit juice weight (g) DD at 10 years (fd10frju) was equal to 0g/day. (This assumes that those with intakes of 0g/day are non-consumers of fruit juice.)
- If in F10, missing pure fruit juice frequency (cch704) was recoded as drinks pure fruit juice on special occasions (2) if Fruit juice weight (g) DD at 10 years (fd10frju) was less than or equal to 154g/day.
- If in F10, missing pure fruit juice frequency (cch704) was recoded as drinks pure fruit juice at mealtimes (3) if Fruit juice weight (g) DD at 10 years (fd10frju) was greater than 154g/day and less than or equal to 204g/day.
- If in F10, missing pure fruit juice frequency (cch704) was recoded as drinks pure fruit juice at any time (4) if Fruit juice weight (g) DD at 10 years (fd10frju) was greater than 204g/day or 1 individual carton/day.

After imputation, all 7,557children in F10 have observations, but now only 65% of children drank pure fruit juice at mealtimes or at any time. The variable was recoded as a NEW dummy variable for Beverage Juice (crJUICE) coded Juice ≥1 serving a day = 1 (Juice: mealtimes, any time), Juice < 1 serving a day = 0 (Juice: does not drink, special occasions only).

Sugary drinks - Frequency child drinks sugary drinks (crcchSSBs).

Sugary drinks (crcchSSBs) is a NEW variable generated from combined responses to frequency child drinks cola (cch700), other fizzy drinks (cch701) and sweetened fruit drinks (cch705).

Almost all (99%) of children reported drinking some kind of sugary drink, at least on special occasions. In F10 only 6% of children with observations did not drink added water drinks (diluted squash or cordial), 9% did not drink other fizzy drinks, 12% did not drink cola and 48% did not drink sweetened fruit drinks such as Sunny Delight. At this age few children in ALSPAC drank tea (sugar sweetened or otherwise) and even fewer drank coffee.

Although 87% of children in F10 drank added water drinks at mealtimes or any time, half of them did not drink other kinds of sugary drinks regularly and so tended to have moderate overall intakes of normal fizzy drinks and made up squash (fd10noma) as measured by the 3 day food diary. This suggests that frequency child drinks added water drinks (cch706) does not

discriminate clearly between moderate and high intakes of sugary drinks. For this reason, added water drinks (cch706) was not included in the new sugary drinks variable (crcchSSBs).

In F10, 3% of children with observations did not drink sugary drinks (other than added water drinks). This frequency category had the highest proportion (77%) of children with normal fizzy drinks and made up squash intakes of 0g/day in the 3 day food diary. A further 37% of children only drank sugary drinks (other than added water drinks) on special occasions, of whom over half (55%) consumed no normal fizzy drinks and made up squash on the days surveyed by the 3 day food diary.

The remaining 60% reported that they drank sugary drinks at meals times only or at any time, although 41% of them did not have any recorded normal fizzy drinks and made up squash juice intake on the 3 days surveyed by the food diary. Once non-consumers were excluded, mean and median intakes of normal fizzy drinks and made-up squash get larger across the four frequency categories. The interquartile ranges of normal fizzy drinks and made-up squash for "drinks at mealtimes" and "drinks at any time" overlap, yet again showing little difference between these two categories.

Approximately 19% of children in F10 were missing observations for cola (cch700), other fizzy drinks (cch701) or sweetened fruit drinks (cch705). In combination, 18% of children are missing observations for the new frequency child drinks sugary drinks (crcchSSBs). Their mean intake of normal fizzy drinks and made-up squash was higher than that of children with observations (115g/day vs 103g/day) but similar to the median intake in the "drinks sugary drinks on special occasions" category (113g/day) so after imputation we might reasonably expect a greater proportion to be in this category and a smaller proportion to drink sugary drinks at mealtimes or any time.

Missing observations for frequency child drinks sugary drinks (crcchSSBs) were imputed based on Normal fizzy drinks and made-up squash weight (g) DD at 10 years (fd10noma) from the 3 day food diary, using <100g/day as the cut off for drinking sugary drinks on special occasions and >150g/day as the cut-off for drinking sugary drinks at any time.

- If in F10, missing sugary drinks frequency (crcchSSBs) was recoded as does not drink sugary drinks, other than added water drinks (1) if Normal fizzy drinks and made up squash weight (g) DD at 10 years (fd10noma) was equal to 0g/day. (This assumes that those with intakes of 0g/day do not drink any sugary drinks - although some might occasionally.)
- If in F10, missing sugary drinks frequency (crcchSSBs) was recoded as drinks sugary drinks, other than added water drinks, on special occasions only (2) if Normal fizzy

drinks and made up squash weight (g) DD at 10 years (fd10noma) was less than or equal to 100g/day. (This allows for moderate intake of added water drinks.)

- If in F10, missing sugary drinks frequency (crcchSSBs) was recoded as drinks sugary drinks, other than added water drinks, at mealtimes (3) if Normal fizzy drinks and made up squash weight (g) DD at 10 years (fd10noma) was greater than 100g/day and less than or equal to 150g/day.
- If in F10, missing sugary drinks frequency (crcchSSBs) was recoded as drinks sugary drinks, other than added water drinks, at any time (4) if Normal fizzy drinks and made up squash weight (g) DD at 10 years (fd10noma) was greater than 150g/day.

After imputation, all 7,557 children in F10 have observations, but now 56% of children drank sugary drinks, other than added water drinks, at mealtimes or at any time. Using a serving size of 150g, the variable was recoded as a NEW dummy variable for Sugar Sweetened Beverages other than added water drinks (crSSB) coded SSB ≥1 serving a day = 1 (SSB: meal times, any time), SSB < 1 serving a day = 0 (SSB: does not drink, special occasions only).

G.3.4 Food intakes – continuous variables

Food intakes are based on the 3 day food diary that children completed with help from their parent or carer before the F10 clinic, when they were approximately 10 years and 6 months old. Food diaries were checked for completeness by trained nutrition field workers at the clinic visit. Researchers assigned food codes and weights to foods and beverages using DIDO (Diet In, Data Out) software developed by the Medical Research Council Human Nutrition Research Unit, Cambridge, UK. 7,462 children returned a food diary.
NEW continuous food intake variables are made up of totalled intakes of representative foods as g/day. They may not fully capture all the foods consumed in that food category.

Whole grain intake

Made up of High fibre breakfast cereals + Brown & granary bread + Soft grain white bread + Wholemeal bread weight (g) DD at 10 years

crfd10WHLGRAIN = (fd10bkhi + fd10bnbr + fd10hfbr + fd10wlbr)

NOTE: Pasta, rice, pizza etc. weight (g) DD at 10 years does not differentiate between white and brown/whole grain varieties. Reasonable to assume most is not wholegrain, so fd10rice is not included in Wholegrain intake.

Dairy intake (not including milk)

Made up of Yoghurt and fromage frais + Cheese + Milk-based sauces weight (g) DD at 10 years.

crfd10DAIRY = (fd10yog + fd10chse + fd10misa)

NOTE: The dataset provided gives full-fat and reduced-fat Dairy foods together, not separately. Puddings and ice-creams weight (g) DD at 10 years does not differentiate between dairy and non-dairy. Reasonable to assume it is a mix of both, so fd10pudd is not included in Dairy intake.

Milk intake (as an alternative to Milk as a beverage frequency)

Made up of whole milk, semi-skimmed milk and skimmed milk weight (g) DD at 10 years.

crfd10totMILK = (fd10whmk + fd10ssmk + fd10skmk)

NOTE: Goats and sheep's milks, soya milk and other plant-based milks had very low mean intakes and are not included.

Vegetable intake

Made up of Raw carrots + Cooked carrots + Green leafy vegetables + Peas + Green and runner beans + Cooked and canned tomatoes + Raw tomatoes + Other salad and raw vegetables + Other cooked vegetables + Vegetable dishes weight (g) DD at 10 years.

crfd10VEG = (fd10carr + fd10ckcr + fd10grlf + fd10peas + fd10rnnr + fd10otto + fd10toma + fd10rveg + fd10ckvg + fd10vgds)

Fruit intake

Made up of Fruit canned in syrup + Fruit canned in juice + Citrus fruit + Apples and pears + Bananas + Other fruit weight (g) DD at 10 years.

crfd10FRUIT = (fd10frsy + fd10caju + fd10citr + fd10appl + fd10bana + fd10otfr)

Energy dense treats and snacks intake

Made up of Sweet biscuits + Buns, cakes, pastries and fruit pies + Crisps and savoury snacks + Sugar confectionery + Chocolate confectionery + Nuts + Savoury biscuits and crackers weight (g) DD at 10 years

crfd10TREATS = (fd10bisc + fd10bun + fd10snck + fd10conf + fd10choc + fd10nuts + fd10sabi)

NOTE: Snack foods can be defined as foods which "tend to be Energy dense and of little nutritional value". Buns, cakes, pastries and fruit pies intake was higher among teenagers who ate meals but did not snack much, compared with teenagers who snacked all day, but they were still included as they are Energy dense foods which are optional "treats" or "non-core" foods.

G.3.5 Food intakes – categorical and dichotomous variables

NEW categorical and dichotomous food intake variables are based on the continuous food intakes, with intake cut-offs for frequency categories matched to the response options in the questionnaire. The response options themselves were guided by intakes and frequencies used by studies in the Systematic Review, average portion sizes at different ages in the UK National Diet and Nutrition Survey (NDNS) and UK dietary recommendations.

Whole grain

1 serving whole grain = 40g
Assuming 50% whole grain content, a 40g serving such as Weetabix gives 20g whole grain.

Whole grain categorical (crWHLGRcat) has 3 categories: 0, <1, >=1 serving/day

Whole grain dichotomous (crWHLGRAIN) is a DUMMY variable, coded "No < 1 serving/day" =1, "Yes >=1 serving/day" = 0

Dairy foods not including milk

1 serving dairy = 125g yoghurt*

Initially a cut-off of 2 servings/day was proposed, assuming 1 serving of dairy plus 1 serving of milk on cereal, plus a further serving of milk as a beverage to bring total intake of milk and dairy to the recommended 3 servings/day for this age group. (89% of children who consumed Dairy foods also consumed milk on the days surveyed.) However, we wanted to consider Dairy foods as an independent predictor of future obesity, separately from milk. As mean combined intakes of the three dairy foods, yoghurt, cheese and milk-based sauces were 46g/day SD 55g/day it was evident that few children had above 2 servings/day. Instead a 1 serving/day cut-off was employed.

Dairy foods categorical (crDAIRcat) has 3 categories: 0, <1, >=1 serving/day

Dairy foods dichotomous (crDAIRY) is a DUMMY variable, coded "No < 1 serving/day" =1, "Yes >=1 serving/day" = 0.

NOTE: * Serving sizes were later amended to 125g yoghurt, 30g cheese and 60g milk based sauces. (See Chapter 9 Table 14)

Milk

1 serving milk = 250g

Milk categorical (crMILKSERVEcat) has 3 categories: 0, <1, >=1 serving/day

Milk dichotomous (crMILKSERVE) is a DUMMY variable, coded "Zero servings/day" =1, "> zero servings/day" = 0.

Vegetables

1 child sized serving of vegetables = 50g, 2 servings = 100g, 3 servings 150g

Recommendations are that vegetables should make up at least 3 of “5 a day” servings of fruit and vegetables. As mean combined intakes of vegetables were 68g/day SD 59g/day it was evident that relatively few children had above 3 child sized servings/day. Instead a 2 servings/day cut-off was employed.

Vegetables categorical (crVEGcat) has 3 categories: 0, <2, >=2 servings/day

Vegetables dichotomous (crVEG) is a DUMMY variable, coded "No < 2 servings/day"=1, "Yes >=2 servings/day" =0

Fruit

1 serving fruit = 80g, 2 servings = 160g

Recommendations are that fruit should make up 2 of “5 a day” servings of fruit and vegetables. As mean combined intakes of fruit were 71g/day SD 77g/day it was evident that many children had close to 2 adult sized servings/day. A 2 servings/day cut-off was employed.

Fruit categorical (crFRUITcat) has 3 categories: 0, <2, >=2 servings/day

Fruit dichotomous (crFRUIT) is a DUMMY variable, coded "No < 2 servings/day"= 1, "Yes >=2 servings/day" =0,

Energy dense treats and snacks

1 serving = 30g, 2 servings = 60g

Mean combined intakes of Energy dense treats and snacks were 97g/day SD 48g/day. Most children had at least some of these foods. Only 44 children had zero intake.

Energy dense treats and snacks categorical (crTREATScat) has 3 categories: 0, <2, >=2 servings/day

Energy dense treats and snacks dichotomous (crTREATS) is a DUMMY variable, coded "No < 2 servings/day"=1, "Yes >=2 servings/day" =0

Appendix H Purposeful selection models

Table H-1 Purposeful selection Step 2 MULTIVARIABLE model with 13 variables in the DERIVATION sample

Log likelihood at final iteration	-333.25
Number of obs. (obesity outcomes)	2,496 (88)
LR chi2 (17)	95.10
Prob > chi2	<0.001
Pseudo R2	0.1249
AUROC	0.779

crTF2obese	Odds Ratio	Std. Error	z	P>|z|	Lower 95% CI	Upper 95% CI
crpub370	3.49	0.94	4.65	<0.001	2.06	5.90
crMumDegree	1.30	0.45	0.76	0.449	0.66	2.54
crSMOKERS	0.83	0.22	-0.70	0.482	0.50	1.39
crMumoverw	2.67	0.61	4.28	<0.001	1.70	4.18
cr3Meals	1.00	0.25	-0.01	0.993	0.61	1.64
crSnacking	1.28	0.31	1.02	0.309	0.80	2.05
crSSB	1.22	0.30	0.80	0.425	0.75	1.97
crDAIRcat < 1	1.20	0.35	0.64	0.524	0.68	2.13
crDAIRcat ≥ 1	2.53	0.96	2.46	0.014	1.21	5.31
crVEGcat < 2	0.58	0.17	-1.79	0.073	0.33	1.05
crVEGcat ≥ 2	0.50	0.18	-1.92	0.055	0.24	1.01
crFRUITcat < 2	0.92	0.25	-0.30	0.765	0.54	1.57
crFRUITcat ≥ 2	0.95	0.39	-0.12	0.905	0.43	2.13
crTREATScat < 2	0.40	0.34	-1.07	0.284	0.07	2.14
crTREATScat ≥ 2	0.30	0.25	-1.45	0.147	0.06	1.53
crBODYSAT	3.25	0.74	5.20	<0.001	2.08	5.06
crACTIVTR	1.57	0.35	1.99	0.047	1.01	2.44
_cons	0.03	0.03	-3.85	<0.001	0.00	0.17

Table H-2 Purposeful selection Step 3 REDUCED model with 6 "best predictors" in the DERIVATION sample

Log likelihood at final iteration	-336.22
Number of obs. (obesity outcomes)	2,496 (88)
LR chi2 (8)	89.17
Prob > chi2	<0.001
Pseudo R2	0.1171
AUROC	0.771

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI**	**Upper 95% CI**
crpub370 Yes armpit hair	3.49	0.93	4.72	<0.001	2.08	5.88
crMumoverw Mum overweight	2.83	0.64	4.61	<0.001	1.82	4.41
crDAIRcat < 1 serving a day	1.10	0.31	0.35	0.727	0.63	1.93
crDAIRcat ≥ 1 serving a day	2.38	0.89	2.33	0.020	1.15	4.94
crVEGcat < 2 servings Veg/day	0.55	0.16	-2.09	0.037	0.31	0.96
crVEGcat ≥ 2 servings Veg/day	0.45	0.15	-2.33	0.020	0.23	0.88
crBODYSAT Unsatisfied	3.30	0.74	5.32	<0.001	2.13	5.12
crACTIVTR No active travel	1.56	0.35	1.98	0.048	1.00	2.42
_cons	0.01	0.01	-11.14	<0.001	0.01	0.03

Table H-3 Purposeful selection Step 4 PRELIMINARY MAIN EFFECTS model with 7 predictors, in the DERIVATION sample

Log likelihood at final iteration	-335.43
Number of obs. (obesity outcomes)	2,496 (88)
LR chi2 (9))	90.75
Prob > chi2	<0.001
Pseudo R2	0.1192
AUROC	0.773

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI**	**Upper 95% CI**
crpub370 Yes armpit hair	3.45	0.92	4.67	<0.001	2.05	5.81
crMumoverw Mum overweight	2.85	0.64	4.63	<0.001	1.83	4.43
crDAIRcat < 1 serving a day	1.12	0.32	0.41	0.682	0.64	1.96
crDAIRcat ≥ 1 serving a day	2.38	0.89	2.32	0.020	1.14	4.94
crVEGcat < 2 servings Veg/day	0.55	0.16	-2.11	0.035	0.31	0.96
crVEGcat ≥ 2 servings Veg/day	0.44	0.15	-2.36	0.019	0.23	0.87
crBODYSAT Unsatisfied	3.29	0.74	5.30	<0.001	2.12	5.11
crACTIVTR No active travel	1.55	0.35	1.96	0.050	1.00	2.41
crTREATS < 2 Treats/day	0.73	0.18	-1.28	0.200	0.44	1.19
_cons	0.02	0.01	-9.53	<0.001	0.01	0.04

Table H-4 Purposeful selection Step 5 MAIN EFFECTS model with 8 predictors in the DERIVATION sample

Log likelihood at final iteration	-334.79
Number of obs. (obesity outcomes)	2,496 (88)
LR chi2 (9))	92.01
Prob > chi2	<0.001
Pseudo R2	0.1208
AUROC	0.774

crTF2obese	**Odds Ratio**	**Std. Error**	**z**	**P>|z|**	**Lower 95% CI**	**Upper 95% CI**
crpub370 Yes armpit hair	3.57	0.95	4.78	<0.001	2.12	6.03
crMumoverw Mum overweight	2.87	0.65	4.66	<0.001	1.84	4.47
crDAIRY < 1 serving a day	0.46	0.14	-2.52	0.012	0.25	0.84
crVEGcat < 2 servings Veg/day	0.55	0.16	-2.11	0.035	0.31	0.96
crVEGcat ≥ 2 servings Veg/day	0.44	0.15	-2.36	0.018	0.23	0.87
crBODYSAT Unsatisfied	3.29	0.74	5.31	<0.001	2.12	5.11
crACTIVTR No active travel	1.57	0.35	2.01	0.044	1.01	2.45
crTREATS ≥ 2 servings Treats/day	0.71	0.18	-1.34	0.179	0.44	1.17
crMILK < 1 serving a day	1.32	0.30	1.21	0.227	0.84	2.07
_cons	0.04	0.02	-7.22	<0.001	0.02	0.09

Appendix I Revised CORA questionnaire with scoring

YEAR 6 EATING HABITS SURVEY

Hello,

We'd like to ask you (*OR your parent or carer*) some questions about what you eat and drink, things that you do, and about you and your family. It is O.K. to ask someone if you need help to fill in the answers.

Your answers are private and afterwards, no-one can find out who you are.

Everyone's answers put together will show how often children your age usually eat different foods and drinks.

Thank you for your help.

ABOUT TODAY

1. What is today's date?

Day	Month	Year

YOUR AGE

2. How old are you now?

Age in years

ABOUT FOOD AND DRINKS

3. How often do you drink sugary drinks (not including fruit squash or cordial that has had water added)? Please tick one box.
   Examples of sugary drinks are non-diet, fizzy soft drinks such as lemonade, Fanta, Coca-Cola, Pepsi, energy drinks, sports drinks and sweetened fruit drinks.

Never or rarely		0
Sometimes		0
Once a day		2
More than once a day		2

4. How often do you drink milk (plain or flavoured) or have milk on breakfast cereal? Please tick one box.

Never or rarely		**3**
Sometimes		**0**
Once a day		**0**
More than once a day		**0**

5. How often do you eat dairy foods? Please tick one box.
   Examples of dairy foods are cheese, yogurt and milky sauces such as cheese sauce, parsley sauce or custard.

Never or rarely		**0**
Sometimes		**0**
Once a day		**0**
More than once a day		**2**

ABOUT FOOD AND DRINKS continued

6. How often do you eat vegetables (not including potatoes or beans)? Please tick one box.

   Vegetables can be fresh, frozen, dried or canned, raw or cooked. Include vegetables that are in cooked dishes too, such as onions or carrots in a casserole or stew.

Never or rarely		**4**
Sometimes		**1**
Once a day		**1**
More than once a day		**2**

7. How often do you eat cakes, buns, Danish pastries, biscuits, chocolate, sweets, ice-cream or crisps, either as part of a meal or in-between meals? Please tick one box.

Never or rarely		**2**
Sometimes		**2**
Once a day		**2**
More than once a day		**0**

ABOUT MEALS and EATING

8. How often do you eat breakfast in the morning? Please tick one box.

Never or rarely		**1 est.**
1 to 3 times a week		**0**
4 times a week, or more		**0**

9. How often do you have a meal at home with grown-ups and children eating together? Please tick one box.

Never or rarely		**1 est.**
1 to 3 times a week		**0**
4 times a week, or more		0

10. How often do you eat a take-away meal or eat a meal in a fast food restaurant? Please tick one box.
    Examples of take-away and fast food meals are fish and chips, pizza, fried chicken, Chinese, Indian, burgers or kebabs. (By meal we mean more than just a drink or a portion of chips.)

Less than once a week		**0**
1 or 2 times a week, or more		**1 est.**

11. Have you ever tried to diet? (*OR do you think that you are a fussy eater?*) Please tick one box.

Yes		**1 ext.**
No		**0**

ABOUT YOUR FAMILY

12. Do other children, babies or teenagers, including your brothers or sisters, live at home with you? Please tick one box.

Yes		**0**
No, no other children living at home		**1 est.**
Prefer not to say		n/a

13. Do you think one of your parents is overweight? Please tick one box.

Yes		**3**
No		**0**
Prefer not to say		n/a

ABOUT YOU

14. What is your ethnicity? Please tick one box.

White		**0**
Non-white		**1 ext.**
Prefer not to say		n/a

15. Have you noticed any changes to your body as you begin to grow up, such as hair under your armpits, or pubic hair? Please tick one box.

Yes		**3**
No		**0**
Prefer not to say		n/a

MORE ABOUT YOU

16. Are you happy with your body shape? Please tick one box.

Yes		**0**
No		**3**
Prefer not to say		n/a

17. When you have school the next day, by what time in the evening are you usually asleep? Please tick one box.

By 8 pm		**0**
By 9 pm		**0**
By 10 pm		**0**
By 11 pm		**1 est.**
After 11 pm		**1 est.**

18. How do you usually travel to and from school? You can tick more than one box

Walk		**0**
Bicycle		**0**
Car		**1 (Only if walk or bicycle not also ticked)**
Taxi		
Bus		
Train		
Other (please describe)		**0 if active, 1 if not**

You have finished.

Thank you for answering the questions.

Bibliography

Adair, L.S. 2008. Child and adolescent obesity: Epidemiology and developmental perspectives. *Physiology & Behavior.* **94**(1), pp.8-16.

Adair, L.S. and Gordon-Larsen, P. 2001. Maturational timing and overweight prevalence in US adolescent girls. *American journal of public health.* **91**(4), p642.

Affenito, S.G., Thompson, D.R., Barton, B.A., Franko, D.L., Daniels, S.R., Obarzanek, E., Schreiber, G.B. and Striegel-Moore, R.H. 2005. Breakfast consumption by African-American and white adolescent girls correlates positively with calcium and fiber intake and negatively with body mass index. *Journal of the American Dietetic Association.* **105**(6), pp.938-945.

Al-Khudairy, L., Loveman, E., Colquitt, J.L., Mead, E., Johnson, R.E., Fraser, H., Olajide, J., Murphy, M., Velho, R.M. and O'Malley, C. 2017. Diet, physical activity and behavioural interventions for the treatment of overweight or obese adolescents aged 12 to 17 years. *Cochrane database of systematic reviews.* (6).

Albar, S.A., Alwan, N.A., Evans, C.E., Greenwood, D.C. and Cade, J.E. 2016. Agreement between an online dietary assessment tool (myfood24) and an interviewer-administered 24-h dietary recall in British adolescents aged 11-18 years. *British Journal of Nutrition.* **115**(9), pp.1678-1686.

Albertson, A.M., Franko, D.L., Thompson, D., Eldridge, A.L., Holschuh, N., Affenito, S.G., Bauserman, R. and Striegel-Moore, R.H. 2007. Longitudinal patterns of breakfast eating in black and white adolescent girls. *Obesity.* **15**(9), pp.2282-2292.

Albertson, A.M., Thompson, D., Franko, D.L., Holschuh, N.M., Bauserman, R. and Barton, B.A. 2009. Prospective associations among cereal intake in childhood and adiposity, lipid levels, and physical activity during late adolescence. *Journal of the American Dietetic Association.* **109**(10), pp.1775-1780.

Alexy, U., Libuda, L., Mersmann, S. and Kersting, M. 2011. Convenience foods in children's diet and association with dietary quality and body weight status. *European Journal of Clinical Nutrition.* **65**(2), pp.160-166.

Alles, B., Samieri, C., Feart, C., Jutand, M.A., Laurin, D. and Barberger-Gateau, P. 2012. Dietary patterns: a novel approach to examine the link between nutrition and cognitive function in older individuals. *Nutrition Research Reviews.* **25**(2), pp.207-222.

Altman, D.G. 1991. *Practical Statistics for Medical Research.* London: Chapman & Hall.

Altman, D.G., Vergouwe, Y., Royston, P. and Moons, K.G. 2009. Prognosis and prognostic research: validating a prognostic model. *BMJ.* **338**, pb605.

Ambrosini, G.L. 2014. Childhood dietary patterns and later obesity: a review of the evidence. *Proceedings of the Nutrition Society.* **73**(1), pp.137-146.

Ambrosini, G.L., de Klerk, N.H., O'Sullivan, T.A., Beilin, L.J. and Oddy, W.H. 2009. The reliability of a food frequency questionnaire for use among adolescents. *European Journal of Clinical Nutrition.* **63**(10), pp.1251-1259.

Ambrosini, G.L., Emmett, P.M., Northstone, K., Howe, L.D., Tilling, K. and Jebb, S.A. 2012. Identification of a dietary pattern prospectively associated with increased adiposity during childhood and adolescence. *International Journal of Obesity.* **36**(10), pp.1299-1305.

Ambrosini, G.L., Oddy, W.H., Huang, R.C., Mori, T.A., Beilin, L.J. and Jebb, S.A. 2013. Prospective associations between sugar-sweetened beverage intakes and cardiometabolic risk factors in adolescents. *American Journal of Clinical Nutrition.* **98**(2), pp.327-334.

Ambrosini, G.L., Oddy, W.H., Robinson, M., O'Sullivan, T.A., Hands, B.P., De Klerk, N.H., Silburn, S.R., Zubrick, S.R., Kendall, G.E. and Stanley, F.J. 2009. Adolescent dietary patterns are associated with lifestyle and family psycho-social factors. *Public health nutrition.* **12**(10), pp.1807-1815.

Arnold, K.F., Berrie, L., Tennant, P.W.G. and Gilthorpe, M.S. 2020. A causal inference perspective on the analysis of compositional data. *International Journal of Epidemiology.*

Arsenault, J.E., Mora-Plazas, M., Forero, Y., Lopez-Arana, S., Marin, C., Baylin, A. and Villamor, E. 2009. Provision of a School Snack Is Associated with Vitamin B-12 Status, Linear Growth, and Morbidity in Children from Bogota, Colombia. *Journal of Nutrition.* **139**(9), pp.1744-1750.

Asayama, K., Hayashibe, H., Endo, A., Okada, T., Hara, M., Masuda, H. and Sugihara, S. 2005. Threshold values of visceral fat and waist girth in Japanese obese children. *Pediatrics International.* **47**(5), pp.498-504.

Balvin Frantzen, L., Trevino, R.P., Echon, R.M., Garcia-Dominic, O. and DiMarco, N. 2013. Association between frequency of ready-to-eat cereal consumption, nutrient intakes, and body mass index in fourth- to sixth-grade **low-income minority c**hildren. *Journal of the Academy of Nutrition & Dietetics.* **113**(4), pp.511-519.

Barba, G., Giacco, R., Clemente, G., Venezia, A., Russo, P., Grimaldi, C. and Siani, A. 2001. The BRAVO project: screening for childhood obesity in a primary school setting *Nutrition Metabolism & Cardiovascular Diseases.* **11**, pp.103-108.

Barlow, S.E. 2007. Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. *Pediatrics.* **120 Suppl 4**, pp.S164-192.

Barton, B.A., Eldridge, A.L., Thompson, D., Affenito, S.G., Striegel-Moore, R.H., Franko, D.L., Albertson, A.M. and Crockett, S.J. 2005. The relationship of breakfast and cereal consumption to nutrient intake and body mass index: the National Heart, Lung, and Blood Institute Growth and Health Study. *Journal of the American Dietetic Association.* **105**(9), pp.1383-1389.

Beato, G.C., Ravelli, M.N., Crisp, A.H. and de Oliveira, M.R.M. 2019. Agreement between body composition assessed by bioelectrical impedance analysis and doubly labeled water in obese women submitted to bariatric surgery. *Obesity surgery.* **29**(1), pp.183-189.

Berenson, G.S. and Bogalusa Heart Study, I. 2001. Bogalusa Heart Study: a long-term community study of a rural biracial (Black/White) population. *American Journal of the Medical Sciences.* **322**(5), pp.293-300.

Berge, J.M., Wall, M., Hsueh, T.F., Fulkerson, J.A., Larson, N. and Neumark-Sztainer, D. 2015. The Protective Role of Family Meals for Youth Obesity: 10-Year Longitudinal Associations. *Journal of Pediatrics.* **166**(2), pp.296-301.

Berkey, C.S., Rockett, H.R., Willett, W.C. and Colditz, G.A. 2005. Milk, dairy fat, dietary calcium, and weight gain: a longitudinal study of adolescents. *Archives of Pediatrics & Adolescent Medicine.* **159**(6), pp.543-550.

Berkey, C.S.S. and Colditz, G.A.M.D.D. 2006. Adiposity in Adolescents: Change in Actual BMI Works Better Than Change in BMI z Score for Longitudinal Studies. *Annals of Epidemiology.* **17**(1), pp.44-50.

Berz, J.P., Singer, M.R., Guo, X., Daniels, S.R. and Moore, L.L. 2011. Use of a DASH food group score to predict excess weight gain in adolescent girls in the National Growth and Health Study. *Archives of Pediatrics & Adolescent Medicine.* **165**(6), pp.540-546.

Bigornia, S.J., LaValley, M.P., Moore, L.L., Northstone, K., Emmett, P., Ness, A.R. and Newby, P.K. 2014. Dairy intakes at age 10 years do not adversely affect risk of excess adiposity at 13 years. *Journal of Nutrition.* **144**(7), pp.1081-1090.

Borenstein, M., Hedges, L.V., Higgins, J.P. and Rothstein, H.R. 2011. *Introduction to meta-analysis.* John Wiley & Sons.

Bouwmeester, W., Zuithoff, N.P., Mallett, S., Geerlings, M.I., Vergouwe, Y., Steyerberg, E.W., Altman, D.G. and Moons, K.G. 2012. Reporting and methods in clinical prediction research: a systematic review. *PLoS Medicine / Public Library of Science.* **9**(5), pp.1-12.

Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S. and 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**(1), pp.111-127.

Brambilla, P., Bedogni, G., Heo, M. and Pietrobelli, A. 2013. Waist circumference-to-height ratio predicts adiposity better than body mass index in children and adolescents. *International Journal of Obesity.* **37**(7), pp.943-946.

Brettschneider, A.K., Rosario, A.S. and Ellert, U. 2011. Validity and predictors of BMI derived from self-reported height and weight among 11- to 17-year-old German adolescents from the KiGGS study. *BMC Research Notes.* **4**, p414.

Brown, T., Moore, T.H.M., Hooper, L., Gao, Y., Zayegh, A., Ijaz, S., Elwenspoek, M., Foxen, S.C., Magee, L., O'Malley, C. and et al. 2019. Interventions for preventing obesity in children. *Cochrane Database of Systematic Reviews.* (7).

Bucher Della Torre, S., Keller, A., Laure Depeyre, J. and Kruseman, M. 2016. Sugar-Sweetened Beverages and Obesity Risk in Children and Adolescents: A Systematic Analysis on How Methodological Quality May Influence Conclusions. *Journal of the Academy of Nutrition & Dietetics.* **116**(4), pp.638-659.

Burkhard-Jagodzinska, K., Ladyga, M., Nazar, K., Starczewska-Czapowska, J., Borkowski, L., Malczewska, J. and Aniol-strzyzewska, K. 2001. Evaluation of nutrition, development and body composition of girls aged 11-15 training rowing at sports school. . *Biology of Sport.* **18**, pp.335-349.

Burley, V.J., Timmins, K., Cade, J., Greenwood, D., Husain, F., Gill, V., Vowden, K., Page, P., Lennox, A., Erens, B., Steer, T. and Hulme, C. 2015. *Making the best use of new technologies in the National Diet and Nutrition Survey: a review.* University of Leeds.

Buyken, A.E., Alexy, U., Kersting, M. and Remer, T. 2012. [The DONALD cohort. An updated overview on 25 years of research based on the Dortmund Nutritional and Anthropometric Longitudinally Designed study]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* **55**(6-7), pp.875-884.

Cade, J.E. 2017. Measuring diet in the 21st century: use of new technologies. *Proceedings of the Nutrition Society.* **76**(3), pp.276-282.

Canfell, O.J., Littlewood, R., Wright, O.R. and Walker, J.L. 2018. Clinical relevance and validity of tools to predict infant, childhood and adulthood obesity: a systematic review. *Public Health Nutrition.* pp.1-13.

Carter, M.C., Albar, S.A., Morris, M.A., Mulla, U.Z., Hancock, N., Evans, C.E., Alwan, N.A., Greenwood, D.C., Hardie, L.J., Frost, G.S., Wark, P.A., Cade, J.E. and myfood, C. 2015. Development of a UK Online 24-h Dietary Assessment Tool: myfood24. *Nutrients.* **7**(6), pp.4016-4032.

Carter, M.C., Hancock, N., Albar, S.A., Brown, H., Greenwood, D.C., Hardie, L.J., Frost, G.S., Wark, P.A. and Cade, J.E. 2016. Development of a New Branded UK Food Composition Database for an Online Dietary Assessment Tool. *Nutrients.* **8**(8), p05.

Cheng, G., Karaolis-Danckert, N., Libuda, L., Bolzenius, K., Remer, T. and Buyken, A.E. 2009. Relation of dietary glycemic index, glycemic load, and fiber and whole-grain intakes during puberty to the concurrent development of percent body fat and body mass index. *American Journal of Epidemiology.* **169**(6), pp.667-677.

Chowdhury, R., Warnakula, S., Kunutsor, S., Crowe, F., Ward, H.A., Johnson, L., Franco, O.H., Butterworth, A.S., Forouhi, N.G., Thompson, S.G., Khaw, K.T., Mozaffarian, D., Danesh, J. and Di Angelantonio, E. 2014. Association of dietary, circulating, and supplement fatty acids with coronary risk: a systematic review and meta-analysis. *Annals of Internal Medicine.* **160**(6), pp.398-406.

Classen, T. and Hokayem, C. 2005. Childhood influences on youth obesity. *Economics & Human Biology.* **3**(2), pp.165-187.

Cole, T.J. 1990. The LMS method for constructing normalized growth standards. *European Journal of Clinical Nutrition.* **44**(1), pp.45-60.

Cole, T.J., Bellizzi, M.C., Flegal, K.M. and Dietz, W.H. 2000. Establishing a standard definition for child overweight and obesity worldwide: international survey. *British Medical Journal.* **320**(7244), pp.1240-1243.

Cole, T.J., Faith, M.S., Pietrobelli, A. and Heo, M. 2005. What is the best measure of adiposity change in growing children: BMI, BMI %, BMI z-score or BMI centile? *EUROPEAN JOURNAL OF CLINICAL NUTRITION.* **59**(3), pp.419-425.

Cole, T.J., Freeman, J.V. and Preece, M.A. 1995. Body mass index reference curves for the UK, 1990. *Archives of Disease in Childhood.* **73**(1), pp.25-29.

Cole, T.J. and Green, P.J. 1992. Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine.* **11**(10), pp.1305-1319.

Collins, G.S. and Altman, D.G. 2012. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. *British Medical Journal.* **344**.

Collins, G.S., de Groot, J.A., Dutton, S., Omar, O., Shanyinde, M., Tajar, A., Voysey, M., Wharton, R., Yu, L.M., Moons, K.G. and Altman, D.G. 2014. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Medical Research Methodology.* **14**, p40.

Collins, G.S., Reitsma, J.B., Altman, D.G. and Moons, K.G. 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Medicine.* **13**, p1.

Collins, M.E. 1991. Body Figure perceptions and preferences among preadolescent children. *International Journal of Eating Disorders.* **10**(2), pp.199-208.

Costa, C.S., Del-Ponte, B., Assunção, M.C.F. and Santos, I.S. 2018. Consumption of ultra-processed foods and body fat during childhood and adolescence: a systematic review. *Public health nutrition.* **21**(1), pp.148-159.

Craigie, A., Lake, A.A., Kelly, S.A., Adamson, A.J. and Mathers, J.C. 2011. Tracking of obesity-related behaviours from childhood to adulthood: A systematic review. *Maturitas.* **70**(3), pp.266-284.

Crawford, P.B., Obarzanek, E., Morrison, J. and Sabry, Z.I. 1994. Comparative advantage of 3-day food records over 24-hour recall and 5-day food frequency validated by observation of 9- and 10-year-old girls. *Journal of the American Dietetic Association.* **94**(6), pp.626-630.

Cutler, G.J., Flood, A., Hannan, P. and Neumark-Sztainer, D. 2009. Major Patterns of Dietary Intake in Adolescents and Their Stability over Time. *Journal of Nutrition.* **139**(2), pp.323-328.

Cutler, G.J., Flood, A., Hannan, P.J., Slavin, J.L. and Neumark-Sztainer, D. 2012. Association between major patterns of dietary intake and weight status in adolescents. *British Journal of Nutrition.* **108**(2), pp.349-356.

Davison, K.K., Susman, E.J. and Birch, L.L. 2003. Percent body fat at age 5 predicts earlier pubertal development among girls at age 9. *Pediatrics.* **111**(4), pp.815-821.

de la Hunty, A., Gibson, S. and Ashwell, M. 2013. Does regular breakfast cereal consumption help children and adolescents stay slimmer? A systematic review and meta-analysis. *Obesity facts.* **6**(1), pp.70-85.

de Onis, M. 2004. The use of anthropometry in the prevention of childhood overweight and obesity. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity.* **28 Suppl 3**, pp.S81-85.

de Onis, M. and Lobstein, T. 2010. Defining obesity risk status in the general childhood population: Which cut-offs should we use? *International Journal of Pediatric Obesity.* **5**(6), pp.458-460.

de Onis, M., Onyango, A.W., Borghi, E., Siyam, A., Nishida, C. and Siekmann, J. 2007. Development of a WHO growth reference for school-aged children and adolescents/Mise au point d'une reference de croissance pour les enfants d'age scolaire et les adolescents/ Elaboracion de valores de referencia de la OMS para el crecimiento de escolares y adolescentes.(Research)(Author abstract). *Bulletin of the World Health Organization.* **85**(9), p660.

Deeks, J.J., Dinnes, J., D'Amico, R., Sowden, A.J., Sakarovitch, C., Song, F., Petticrew, M., Altman, D.G., International Stroke Trial Collaborative, G. and European Carotid Surgery Trial Collaborative, G. 2003. Evaluating non-randomised intervention studies. *Health technology assessment (Winchester, England).* **7**(27), pp.iii-x, 1-173.

Devlin, U.M., McNulty, B.A., Nugent, A.P. and Gibney, M.J. 2012. The use of cluster analysis to derive dietary patterns: methodological considerations, reproducibility, validity and the effect of energy mis-reporting. *Proceedings of the Nutrition Society.* **71**(4), pp.599-609.

Dietz, W.H. and Bellizzi, M.C. 1999. Introduction: the use of body mass index to assess obesity in children. *American Journal of Clinical Nutrition.* **70**(1), pp.123S-125S.

Dougkas, A., Barr, S., Reddy, S. and Summerbell, C.D. 2019. A critical review of the role of milk and other dairy products in the development of obesity in children and adolescents. *Nutrition research reviews.* **32**(1), pp.106-127.

Downs, S.H. and Black, N. 1998. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health.* **52**(6), pp.377-384.

Dror, D. 2014. Dairy consumption and pre-school, school-age and adolescent obesity in developed countries: a systematic review and meta-analysis. *Obesity reviews.* **15**(6), pp.516-527.

Druet, C., Stettler, N., Sharp, S., Simmons, R.K., Cooper, C., Smith, G.D., Ekelund, U., Levy-Marchal, C., Jarvelin, M.R., Kuh, D. and Ong, K.K. 2012. Prediction of childhood obesity by infancy weight gain: an individual-level meta-analysis. *Paediatric and Perinatal Epidemiology.* **26**(1), pp.19-26.

Egger, M., Davey Smith, G. and Schneider, M. 2001. Systematic reviews of observational studies. In: Egger, M., et al. eds. *Systematic Reviews in Health Care.* 2nd ed. London: BMJ Publishing Group, pp.213-215.

Egger, M., Smith, G.D., Schneider, M. and Minder, C. 1997. Bias in meta-analysis detected by a simple, graphical test. *Bmj.* **315**(7109), pp.629-634.

Elgar, F.J., Roberts, C., Tudor-Smith, C. and Moore, L. 2005. Validity of self-reported height and weight and predictors of bias in adolescents. *Journal of Adolescent Health.* **37**(5), pp.371-375.

Ensor, J., Snell, K. and Martin, E. 2018. *Investigation of prediction model performance with Stata: pmstats and pmcalplot*. Unpublished.

Fanning, E. 2005. Formatting a Paper-based Survey Questionnaire: Best Practices. *Practical Assessment Research & Evaluation.* [Online]. 12. **10**. [Accessed 19.01.2020.]. Available from: https://pdfs.semanticscholar.org/1c5c/5b7db0392c4e4b53e96647f4dd7918180704.pdf

Feskanich, D., Rimm, E.B., Giovannucci, E.L., Colditz, G.A., Stampfer, M.J., Litin, L.B. and Willett, W.C. 1993. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *Journal of the American Dietetic Association.* **93**(7), pp.790-796.

Field, A.E., Aneja, P. and Rosner, B. 2007. The validity of self-reported weight change among adolescents and young adults. *Obesity.* **15**(9), pp.2357-2364.

Field, A.E., Austin, S.B., Gillman, M.W., Rosner, B., Rockett, H.R. and Colditz, G.A. 2004. Snack food intake does not predict weight change among children and adolescents. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity.* **28**(10), pp.1210-1216.

Field, A.E., Austin, S.B., Taylor, C.B., Malspeis, S., Rosner, B., Rockett, H.R., Gillman, M.W. and Colditz, G.A. 2003a. Relation between dieting and weight change among preadolescents and adolescents. *Pediatrics.* **112**(4), pp.900-906.

Field, A.E., Gillman, M.W., Rosner, B., Rockett, H.R. and Colditz, G.A. 2003b. Association between fruit and vegetable intake and change in body mass index among a large sample of children and adolescents in the United States. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity.* **27**(7), pp.821-826.

Field, A.E., Sonneville, K.R., Falbe, J., Flint, A., Haines, J., Rosner, B. and Camargo, C.A. 2014. Association of sports drinks with weight gain among adolescents and young adults. *Obesity.* **22**(10), pp.2238-2243.

Flodmark, C.E., Lissau, I., Moreno, L.A., Pietrobelli, A. and Widhalm, K. 2004. New insights into the field of children and adolescents' obesity: the European perspective. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity.* **28**(10), pp.1189-1196.

Frank, G.C., Hollatz, A.T., Webber, L.S. and Berenson, G.S. 1984. Effect of interviewer recording practices on nutrient intake--Bogalusa Heart Study. *Journal of the American Dietetic Association.* **84**(12), pp.1432-1436, 1439.

Franko, D.L., Striegel-Moore, R.H., Thompson, D., Affenito, S.G., Schreiber, G.B., Daniels, S.R. and Crawford, P.B. 2008. The relationship between meal frequency and body mass index in black and white adolescent girls: more is less. *International Journal of Obesity.* **32**(1), pp.23-29.

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S.M. and Lawlor, D.A. 2013. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol.* **42**(1), pp.97-110.

Freedman, D.S., Khan, L.K., Serdula, M.K., Dietz, W.H., Srinivasan, S.R. and Berenson, G.S. 2004. Inter-relationships among childhood BMI, childhood height, and adult obesity: the Bogalusa Heart Study. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity.* **28**(1), pp.10-16.

Fulkerson, J.A., Neumark-Sztainer, D., Hannan, P.J. and Story, M. 2008. Family meal frequency and weight status among adolescents: cross-sectional and 5-year longitudinal associations. *Obesity.* **16**(11), pp.2529-2534.

Garner, D.M., Olmstead, M.P. and Polivy, J. 1983. Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders.* **2**(2), pp.15-34.

Gasser, C.E., Mensah, F.K., Russell, M., Dunn, S.E. and Wake, M. 2016. Confectionery consumption and overweight, obesity, and related outcomes in children and adolescents: a systematic review and meta-analysis. *The American journal of clinical nutrition.* **103**(5), pp.1344-1356.

Goldberg, G.R., Black, A.E., Jebb, S.A., Cole, T.J., Murgatroyd, P.R., Coward, W.A. and Prentice, A.M. 1991. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *European Journal of Clinical Nutrition.* **45**(12), pp.569-581.

Goodman, E., Hinden, B.R. and Khandelwal, S. 2000. Accuracy of teen and parental reports of obesity and body mass index. *Pediatrics.* **106**(1 Pt 1), pp.52-58.

Greenwood, D.C., Gilthorpe, M.S., Golding, C. and Cade, J.E. 2003. Stability over time of dietary patterns in the UK Women's cohort study. In: *Nutrition Society King's College, London*. Cambridge University Press, p.89A.

Harrell, F.E. 2001. *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis.* New York ;: Springer.

Hasnain, S.R., Singer, M.R., Bradlee, M.L. and Moore, L.L. 2014. Beverage Intake in Early Childhood and Change in Body Fat from Preschool to Adolescence. *Childhood Obesity.* **10**(1), pp.42-49.

Hatipoglu, N., Mazicioglu, M.M., Kurtoglu, S. and Kendirci, M. 2010. Neck circumference: an additional tool of screening overweight and obesity in childhood. *European Journal of Pediatrics.* **169**(6), pp.733-739.

Hemingway, H., Croft, P., Perel, P., Hayden, J.A., Abrams, K., Timmis, A., Briggs, A., Udumyan, R., Moons, K.G., Steyerberg, E.W., Roberts, I., Schroter, S., Altman, D.G., Riley, R.D. and Group, P. 2013. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ.* **346**, pe5595.

Higgins, J. and Wells, G. 2011. *Cochrane handbook for systematic reviews of interventions.* [Online]. [Accessed 1.11. 2015]. Available from: http://www.handbook.cochrane.org

Higgins, J.P., Thompson, S.G., Deeks, J.J. and Altman, D.G. 2003. Measuring inconsistency in meta-analyses. *Bmj.* **327**(7414), pp.557-560.

Higgins, J.P.T., Altman, D.G., Gotzsche, P.C., Juni, P., Moher, D., Oxman, A.D., Savovic, J., Schulz, K.F., Weeks, L., Sterne, J.A.C., Cochrane Bias Methods, G. and Cochrane Statistical Methods, G. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Clinical research ed.).* **343**, pp.d5928-d5928.

Hill, J.O. 2009. Can a small-changes approach help address the obesity epidemic? A report of the Joint Task Force of the American Society for Nutrition, Institute of Food Technologists, and International Food Information Council. *The American journal of clinical nutrition.* **89**(2), pp.477-484.

Himes, J.H., Hannan, P., Wall, M. and Neumark-Sztainer, D. 2005. Factors associated with errors in self-reports of stature, weight, and body mass index in Minnesota adolescents. *Annals of Epidemiology.* **15**(4), pp.272-278.

Hingorani, A.D., Windt, D.A., Riley, R.D., Abrams, K., Moons, K.G., Steyerberg, E.W., Schroter, S., Sauerbrei, W., Altman, D.G., Hemingway, H. and Group, P. 2013. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ.* **346**, pe5793.

Hoffmann, K., Schulze, M.B., Schienkiewitz, A., Nothlings, U. and Boeing, H. 2004. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *American Journal of Epidemiology.* **159**(10), pp.935-944.

Hosmer, D.W. 2013. *Applied logistic regression.* Third edition / David Hosmer,Stanley Lemeshaw, Rodney Sturdivant. ed. Hoboken: Wiley.

Hughes, A.R., Sherriff, A., Lawlor, D.A., Ness, A.R. and Reilly, J.J. 2011. Incidence of obesity during childhood and adolescence in a large contemporary cohort. *Preventive Medicine.* **52**(5), pp.300-304.

Isanaka, S., Mora-Plazas, M., Lopez-Arana, S., Baylin, A. and Villamor, E. 2007. Food insecurity is highly prevalent and predicts underweight but not overweight in adults and school children from Bogota, Colombia. *Journal of Nutrition.* **137**(12), pp.2747-2755.

Johnson, L., Mander, A.P., Jones, L.R., Emmett, P.M. and Jebb, S.A. 2008. A prospective analysis of dietary energy density at age 5 and 7 years and fatness at 9 years among UK children. *International Journal of Obesity.* **32**(4), pp.586-593.

Keast, D.R., Nicklas, T.A. and O'Neil, C.E. 2010. Snacking is associated with reduced risk of overweight and reduced abdominal obesity in adolescents: National Health and Nutrition Examination Survey (NHANES) 1999–2004. *The American Journal of Clinical Nutrition.* **92**(2), pp.428-435.

Keller, A. and Bucher Della Torre, S. 2015. Sugar-Sweetened Beverages and Obesity among Children and Adolescents: A Review of Systematic Literature Reviews. *Childhood Obesity.* **11**(4), pp.338-346.

Kimm, S.Y.S., Kwiterovich JR, P.O., Santanello, N.C., Obarzanek, E., Lakatos, E., Lauer, R.M., Van Horn, L., Barton, B.A., Stevens, V., Robson, A. and Greenlick, M.R. 1993. Dietary intervention study in children (DISC) with elevated low-density- lipoprotein cholesterol: Design and baseline characteristics. . *Annals of Epidemiology.* **3**, pp.393-402.

Kindblom, J.M., Lorentzon, M., Norjavaara, E., Lönn, L., Brandberg, J., Angelhed, J.-E., Hellqvist, Å., Nilsson, S. and Ohlsson, C. 2006. Pubertal timing is an independent predictor of central adiposity in young adult males: the Gothenburg osteoporosis and obesity determinants study. *Diabetes.* **55**(11), pp.3047-3052.

Kopelman, P. 2007. Health risks associated with overweight and obesity. *Obesity Reviews.* **8 Suppl 1**, pp.13-17.

Kroke, A., Manz, F., Kersting, M., Remer, T., Sichert-Hellert, W., Alexy, U. and Lentze, M.J. 2004. The DONALD Study. *European Journal of Nutrition.* **43**(1), pp.45-54.

Kromeyer-Hauschild, K., Wabitsch, M., Kunze, D., Geller, F., Geiß, H.C., Hesse, V., von Hippel, A., Jaeger, U., Johnsen, D., Korte, W., Menner, K., Müller, G., Müller, J.M., Niemann-Pilatus, A., Remer, T., Schaefer, F., Wittchen, H.U., Zabransky, S., Zellner, K., Ziegler, A. and Hebebrand, J. 2001. Perzentile für den Body-mass-Index für das Kindes- und Jugendalter unter Heranziehung verschiedener deutscher Stichproben. *Monatsschrift Kinderheilkunde.* **149**(8), pp.807-818.

Kuczmarski, R.J., Ogden, C.L., Guo, S.S., Grummer-Strawn, L.M., Flegal, K.M., Mei, Z., Wei, R., Curtin, L.R., Roche, A.F. and Johnson, C.L. 2002. 2000 CDC Growth Charts for the United States: methods and development. *Vital & Health Statistics - Series 11: Data From the National Health Survey.* (246), pp.1-190.

Kuhn, M. and Johnson, K. 2016. *Applied Predictive Modeling.* New York: Springer.

LaBerge, R.C., Vaccani, J.P., Gow, R.M., Gaboury, I., Hoey, L. and Katz, S.L. 2009. Inter- and intra-rater reliability of neck circumference measurements in children. *Pediatric Pulmonology.* **44**(1), pp.64-69.

Lachat, C., Hawwash, D., Ocke, M.C., Berg, C., Forsum, E., Hornell, A., Larsson, C., Sonestedt, E., Wirfalt, E., Akesson, A., Kolsteren, P., Byrnes, G., De Keyzer, W., Van Camp, J., Cade, J.E., Slimani, N., Cevallos, M., Egger, M. and Huybrechts, I. 2016. Strengthening the Reporting of Observational Studies in Epidemiology-Nutritional Epidemiology (STROBE-nut): An Extension of the STROBE Statement. *PLoS Medicine / Public Library of Science.* **13**(6), pe1002036.

Lambert, J., Agostoni, C., Elmadfa, I., Hulshof, K., Krause, E., Livingstone, B., Socha, P., Pannemans, D. and Samartín, S. 2004. Dietary intake and nutritional status of children and adolescents in Europe. *British Journal of Nutrition.* **92**(S2), pp.S147-S211.

Larson, N., Harnack, L. and Neumark-Sztainer, D. 2012. Assessing dietary intake during the transition to adulthood: a comparison of age-appropriate FFQ for youth/adolescents and adults. *Public Health Nutrition.* **15**(4), pp.627-634.

Laska, M.N., Murray, D.M., Lytle, L.A. and Harnack, L.J. 2012. Longitudinal associations between key dietary behaviors and weight gain over time: transitions through the adolescent years. *Obesity.* **20**(1), pp.118-125.

Lebenbaum, M., Espin-Garcia, O., Li, Y. and Rosella, L.C. 2018. Development and validation of a population based risk algorithm for obesity: The Obesity Population Risk Tool (OPoRT). *PLoS ONE [Electronic Resource].* **13**(1), pe0191169.

Ledoux, T., Hingle, M.D. and Baranowski, T. 2011. Relationship of fruit and vegetable intake with adiposity: a systematic review. *Obesity reviews.* **12**(5), pp.e143-e150.

Lee, S.K., Novotny, R., Daida, Y.G., Vijayadeva, V. and Gittelsohn, J. 2007. Dietary patterns of adolescent girls in Hawaii over a 2-year period. *Journal of the American Dietetic Association.* **107**(6), pp.956-961.

Levine, R.S., Dahly, D.L. and Rudolf, M.C.J. 2012. Identifying infants at risk of becoming obese: Can we and should we? *Public Health.* **126**(2), pp.123-128.

Levine, R.S. and Rudolf, M.C.J. 2011. Can we identify infants at risk of becoming obese, and if so, should we? *Maternal & child nutrition.* **7**(3), pp.328-329.

Levitan, E.B., Wolk, A. and Mittleman, M.A. 2009. Consistency with the dash diet and incidence of heart failure. *Archives of Internal Medicine.* **169**(9), pp.851-857.

Libuda, L., Alexy, U., Sichert-Hellert, W., Stehle, P., Karaolis-Danckert, N., Buyken, A.E. and Kersting, M. 2008. Pattern of beverage consumption and long-term association with body-weight status in German adolescents--results from the DONALD study. *British Journal of Nutrition.* **99**(6), pp.1370-1379.

Liu, Y.J., Zhan, J., Liu, X.L., Wang, Y., Ji, J. and He, Q.Q. 2014. Dietary flavonoids intake and risk of type 2 diabetes: a meta-analysis of prospective cohort studies. *Clinical Nutrition.* **33**(1), pp.59-63.

Lo, C.K., Mertz, D. and Loeb, M. 2014. Newcastle-Ottawa Scale: comparing reviewers' to authors' assessments. *BMC Medical Research Methodology.* **14**, p45.

Lobstein, T., Baur, L., Uauy, R. and TaskForce, I.I.O. 2004. Obesity in children and young people: a crisis in public health. *Obesity Reviews.* **5 Suppl 1**, pp.4-104.

Louie, J.C.Y., Flood, V.M., Hector, D.J., Rangan, A.M. and Gill, T.P. 2011. Dairy consumption and overweight and obesity: a systematic review of prospective cohort studies. *Obesity Reviews.* **12**(7), pp.E582-E592.

Lytle, L.A. 2009. Examining the Etiology of Childhood Obesity: The IDEA Study. *American journal of community psychology.* **44**(3-4), p338.

MacIntyre, U.E. 2009. Measuring Food Intake. In: Gibney, M.J., et al. eds. *Introduction to Human Nutrition (The Nutrition Society Textbook Series)* 2nd. ed. Hong Kong: Wiley-Blackwell, pp.238-275.

Malik, V.S., Pan, A., Willett, W.C. and Hu, F.B. 2013. Sugar-sweetened beverages and weight gain in children and adults: a systematic review and meta-analysis. *American Journal of Clinical Nutrition.* **98**(4), pp.1084-1102.

Malik, V.S., Schulze, M.B. and Hu, F.B. 2006. Intake of sugar-sweetened beverages and weight gain: a systematic review–. *The American journal of clinical nutrition.* **84**(2), pp.274-288.

Manios, Y., Birbilis, M., Moschonis, G., Birbilis, G., Mougios, V., Lionis, C., Chrousos, G.P. and Healthy Growth Study, g. 2013. Childhood Obesity Risk Evaluation based on perinatal factors and family sociodemographic characteristics: CORE index. *European Journal of Pediatrics.* **172**(4), pp.551-555.

Manios, Y., Vlachopapadopoulou, E., Moschonis, G., Karachaliou, F., Psaltopoulou, T., Koutsouki, D., Bogdanis, G., Carayanni, V., Hatzakis, A. and Michalacos, S. 2016. Utility and applicability of the "Childhood Obesity Risk Evaluation" (CORE)-index in predicting obesity in childhood and adolescence in Greece from early life: the "National Action Plan for Public Health". *EUROPEAN JOURNAL OF PEDIATRICS.* **175**(12), pp.1989-1996.

Mann, K.D., Pearce, M.S., McKevith, B., Thielecke, F. and Seal, C.J. 2015. Low whole grain intake in the UK: results from the National Diet and Nutrition Survey rolling programme 2008-11. *BRITISH JOURNAL OF NUTRITION.* **113**(10), pp.1643-1651.

Manuel, D.G., Rosella, L.C., Hennessy, D., Sanmartin, C. and Wilson, K. 2012. Predictive risk algorithms in a population setting: an overview. *J Epidemiol Community Health.* **66**(10), pp.859-865.

Manuel, D.G., Tuna, M., Bennett, C., Hennessy, D., Rosella, L., Sanmartin, C., Tu, J.V., Perez, R., Fisher, S. and Taljaard, M. 2018. Development and validation of a cardiovascular disease risk-prediction model using population health surveys: the Cardiovascular Disease Population Risk Tool (CVDPoRT). *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* **190**(29), pp.E871-E882.

Masip, G., Silventoinen, K., Keski-Rahkonen, A., Palviainen, T., Sipilä, P.N., Kaprio, J. and Bogl, L.H. 2020. The genetic architecture of the association between eating behaviors and obesity: combining genetic twin modeling and polygenic risk scores. *The American Journal of Clinical Nutrition.*

McCance, R.A. 2013. *McCance and Widdowson's the composition of foods.* Seventh summary edition / compiled by Institute of Food Research and Public Health England. ed. Cambridge: Royal Society of Chemistry.

McCarthy, H.D. and 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**(6), pp.988-992.

McDonald, C.M., Baylin, A., Arsenault, J.E., Mora-Plazas, M. and Villamor, E. 2009. Overweight is more prevalent than stunting and is associated with socioeconomic status, maternal obesity, and a snacking dietary pattern in school children from Bogota, Colombia. *Journal of Nutrition.* **139**(2), pp.370-376.

Mead, E., Brown, T., Rees, K., Azevedo, L.B., Whittaker, V., Jones, D., Olajide, J., Mainardi, G.M., Corpeleijn, E. and O'Malley, C. 2017. Diet, physical activity and behavioural interventions for the treatment of overweight or obese children from the age of 6 to 11 years. *Cochrane Database of Systematic Reviews.* (6).

MGRS. 2006. WHO Child Growth Standards based on length/height, weight and age. *Acta Paediatrica Supplement.* **450**, pp.76-85.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. and Grp, P. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Bmj-British Medical Journal.* **339**.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A. and Group, P.-P. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews.* **4**(1), pp.1-1.

Moons, K.G., Altman, D.G., Vergouwe, Y. and Royston, P. 2009a. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ.* **338**, pb606.

Moons, K.G., Royston, P., Vergouwe, Y., Grobbee, D.E. and Altman, D.G. 2009b. Prognosis and prognostic research: what, why, and how? *BMJ.* **338**, pb375.

Moons, K.G.M., Kengne, A.P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D.G. and Grobbee, D.E. 2012. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart.* **98**(9), pp.683-690.

Moore, J.B., Horti, A. and Fielding, B.A. 2018. Evaluation of the nutrient content of yogurts: a comprehensive survey of yogurt products in the major UK supermarkets. *BMJ open.* **8**(8), pe021387.

Moore, L.L., Bradlee, M.L., Gao, D. and Singer, M.R. 2006. Low dairy intake in early childhood predicts excess body fat gain. *Obesity.* **14**(6), pp.1010-1018.

Morandi, A., Meyre, D., Lobbens, S., Kleinman, K., Kaakinen, M., Rifas-Shiman, S.L., Vatin, V., Gaget, S., Pouta, A., Hartikainen, A.L., Laitinen, J., Ruokonen, A., Das, S., Khan, A.A., Elliott, P., Maffeis, C., Gillman, M.W., Jarvelin, M.R. and Froguel, P. 2012. Estimation of newborn risk for child or adolescent obesity: lessons from longitudinal birth cohorts. *PLoS ONE [Electronic Resource].* **7**(11), pe49919.

MRC. 2014. *Medical Research Council Diet, Anthropometry and Physical Activity (DAPA) Measurement Toolkit.* [Online]. [Accessed 2.3.2020.]. Available from: https://www.dapa-toolkit.mrc.ac.uk/

Must, A., Dallal, G.E. and Dietz, W.H. 1991. Reference data for obesity: 85th and 95th percentiles of body mass index (wt/ht2) and triceps skinfold thickness.[Erratum appears in Am J Clin Nutr 1991 Nov;54(5):773]. *American Journal of Clinical Nutrition.* **53**(4), pp.839-846.

Must, A., Naumova, E.N., Phillips, S.M., Blum, M., Dawson-Hughes, B. and Rand, W.M. 2005. Childhood overweight and maturational timing in the development of adult overweight and fatness: the Newton Girls Study and its follow-up. *Pediatrics.* **116**(3), pp.620-627.

Neeland, I.J., Ross, R., Despres, J.P., Matsuzawa, Y., Yamashita, S., Shai, I., Seidell, J., Magni, P., Santos, R.D., Arsenault, B., Cuevas, A., Hu, F.B., Griffin, B., Zambon, A., Barter, P., Fruchart, J.C., Eckel, R.H., International Atherosclerosis, S. and International Chair on Cardiometabolic Risk Working Group on Visceral, O. 2019. Visceral and ectopic fat, atherosclerosis, and cardiometabolic disease: a position statement. *The Lancet Diabetes & Endocrinology.* **7**(9), pp.715-725.

Neovius, M., Linne, Y., Barkeling, B. and Rossner, S. 2004. Discrepancies between classification systems of childhood obesity. *Obesity Reviews.* **5**(2), pp.105-114.

Newnham, J.P., Evans, S.F., Michael, C.A., Stanley, F.J. and Landau, L.I. 1993. Effects of frequent ultrasound during pregnancy: a randomised controlled trial. *The Lancet.* **342**(8876), pp.887-891.

NHLBI, G.a.H.S.R.G. 1992. Obesity and cardiovascular disease risk factors in black and white girls: the NHLBI Growth and Health Study. *American Journal of Public Health.* **82**(12), pp.1613-1620.

NHSDigital. 2018. *Health Survey for England 2017 [National Statistics].* [Online]. [Accessed 31.03.2020.]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2017

NHSDigital. 2019. *National Child Measurement Programme, England 2018/19* [Online]. [Accessed 15 Jan 2020]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/national-child-measurement-programme/2018-19-school-year

NHSUK. 2017. *How much sleep do children need?* [Online]. [Accessed 23.8.2018.]. Available from: https://www.nhs.uk/live-well/sleep-and-tiredness/how-much-sleep-do-kids-need/

NICE. 2006. *Obesity Prevention, Clinical Guideline 43.* [Online]. [Accessed 1st May 2020]. Available from: https://www.nice.org.uk/guidance/cg43

Nicklas, T.A. 1995. Dietary studies of children and young adults (1973-1988): the Bogalusa Heart Study. *American Journal of the Medical Sciences.* **310 Suppl 1**, pp.S101-108.

Noel, S.E., Mattocks, C., Emmett, P., Riddoch, C.J., Ness, A.R. and Newby, P.K. 2010. Use of accelerometer data in prediction equations for capturing implausible dietary intakes in adolescents. *American Journal of Clinical Nutrition.* **92**(6), pp.1436-1445.

Noel, S.E., Ness, A.R., Northstone, K., Emmett, P. and Newby, P.K. 2011. Milk intakes are not associated with percent body fat in children from ages 10 to 13 years. *Journal of Nutrition.* **141**(11), pp.2035-2041.

Noel, S.E., Ness, A.R., Northstone, K., Emmett, P. and Newby, P.K. 2013. Associations between flavored milk consumption and changes in weight and body composition over time: differences among normal and overweight children. *European Journal of Clinical Nutrition.* **67**(3), pp.295-300.

Nooyens, A.C., Koppes, L.L., Visscher, T.L., Twisk, J.W., Kemper, H.C., Schuit, A.J., van Mechelen, W. and Seidell, J.C. 2007. Adolescent skinfold thickness is a better predictor of high body fatness in adults than is body mass index: the Amsterdam Growth and Health Longitudinal Study. *The American journal of clinical nutrition.* **85**(6), pp.1533-1539.

Novotny, R., Daida, Y.G., Grove, J.S., Le Marchand, L. and Vijayadeva, V. 2006. Asian adolescents have a higher trunk:peripheral fat ratio than Whites. *Journal of Nutrition.* **136**(3), pp.642-647.

O'Neil, C.E., Nicklas, T.A., Liu, Y. and Berenson, G.S. 2015. Candy consumption in childhood is not predictive of weight, adiposity measures or cardiovascular risk factors in young adults: the Bogalusa Heart Study. *Journal of Human Nutrition and Dietetics.* **28**, pp.59-69.

O'Sullivan, T.A., Hafekost, K., Mitrou, F. and Lawrence, D. 2013. Food sources of saturated fat and the association with mortality: a meta-analysis. *American Journal of Public Health.* **103**(9), pp.e31-42.

Oni, T., Yudkin, J.S., Fonn, S., Adongo, P., Kaseje, M., Ajuwon, A., Doyal, L. and London, L. 2019. Global public health starts at home: upstream approaches to global health training. *The Lancet Global Health.* **7**(3), pp.e301-e302.

Patterson, R. and Pietinen, P. 2004. Assessment of Nutritional Status in Individuals and Populations. In: Gibney, M.J., et al. eds. *Public Health Nutrition (The Nutrition Society Textbook Series).* Tunbridge Wells, Kent: Blackwell Science Ltd., pp.66-82.

Peat, G., Riley, R.D., Croft, P., Morley, K.I., Kyzas, P.A., Moons, K.G., Perel, P., Steyerberg, E.W., Schroter, S., Altman, D.G., Hemingway, H. and Group, P. 2014. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. *PLoS Medicine / Public Library of Science.* **11**(7), pe1001671.

Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. and Feinstein, A.R. 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology.* **49**(12), pp.1373-1379.

Pei, Z., Flexeder, C., Fuertes, E., Thiering, E., Koletzko, B., Cramer, C., Berdel, D., Lehmann, I., Bauer, C.P., Heinrich, J., Giniplus and Group, L.I.S. 2013. Early life risk factors of being overweight at 10 years of age: results of the German birth cohorts GINIplus and LISAplus. *European Journal of Clinical Nutrition.* **67**(8), pp.855-862.

Peirson, L., Fitzpatrick-Lewis, D., Morrison, K., Ciliska, D., Kenny, M., Usman Ali, M. and Raina, P. 2015. Prevention of overweight and obesity in children and youth: a systematic review and meta-analysis. *CMAJ open.* **3**(1), pp.E23-33.

Pereira, M.A. and Ludwig, D.S. 2001. Dietary fiber and body-weight regulation: observations and mechanisms. *Pediatric Clinics of North America.* **48**(4), pp.969-980.

Pérez-Morales, E., Bacardí-Gascón, M. and Jiménez-Cruz, A. 2013. Sugar-sweetened beverage intake before 6 years of age and weight or BMI status among older children; systematic review of prospective studies. *Nutricion hospitalaria.* **28**(1), pp.47-51.

PHE. 2015a. *Economics of Obesity.* [Online]. Available from: https://www.noo.org.uk/NOO_about_obesity/economics

PHE. 2015b. *SACN Carbohydrates and Health Report*

The Scientific Advisory Committee on Nutrition recommendations on carbohydrates, including sugars and fibre. [Online]. [Accessed 31.3.2020.]. Available from: https://www.gov.uk/government/publications/sacn-carbohydrates-and-health-report

PHE. 2017. *Health Matters: Obesity and the environment.* [Online]. [Accessed 31.03.2020.]. Available from: https://www.gov.uk/government/publications/health-matters-obesity-and-the-food-environment/health-matters-obesity-and-the-food-environment--2

Pietrobelli, A., Faith, M.S., Allison, D.B., Gallagher, D., Chiumello, G. and Heymsfield, S.B. 1998. Body mass index as a measure of adiposity among children and adolescents: a validation study. **1**(2), pp.204-210.

Pingitore, R., Spring, B. and Garfieldt, D. 1997. Gender Differences in Body Satisfaction. *Obesity Research.* **5**(5), pp.402-409.

Quick, V., Wall, M., Larson, N., Haines, J. and Neumark-Sztainer, D. 2013. Personal, behavioral and socio-environmental predictors of overweight incidence in young adults: 10-yr longitudinal findings. *International Journal of Behavioral Nutrition & Physical Activity.* **10**, p37.

Rader, T., Mann, M., Stansfield, C., Cooper, C. and Sampsone, M. 2014. Methods for documenting systematic review searches: a discussion of common issues. *Research Synthesis Methods.* **5**(2), pp.98-115.

Rampersaud, G.C., Pereira, M.A., Girard, B.L., Adams, J. and Metzl, J.D. 2005. Breakfast habits, nutritional status, body weight, and academic performance in children and adolescents. *Journal of the American Dietetic Association.* **105**(5), pp.743-760.

Rauner, A., Mess, F. and Woll, A. 2013. The relationship between physical activity, physical fitness and overweight in adolescents: a systematic review of studies published in or after 2000. *Bmc Pediatrics.* **13**, p9.

Redsell, S.A., Rose, J., Weng, S., Ablewhite, J., Swift, J.A., Siriwardena, A.N., Nathan, D., Wharrad, H.J., Atkinson, P., Watson, V., McMaster, F., Lakshman, R. and Glazebrook, C. 2017. Digital technology to facilitate Proactive Assessment of Obesity Risk during Infancy (ProAsk): a feasibility study. *BMJ Open.* **7**(9).

Redsell, S.A., Weng, S., Swift, J.A., Nathan, D. and Glazebrook, C. 2016. Validation, Optimal Threshold Determination, and Clinical Utility of the Infant Risk of Overweight Checklist for Early Prevention of Child Overweight. *Childhood Obesity.* **12**(3), pp.202-209.

Rehkopf, D.H., Laraia, B.A., Segal, M., Braithwaite, D. and Epel, E. 2011. The relative importance of predictors of body mass index change, overweight and obesity in adolescent girls. *International Journal of Pediatric Obesity.* **6**(2-2), pp.e233-242.

Reilly, J.J. 2002. Assessment of childhood obesity: National reference data or international approach? *Obesity Research.* **10**(8), pp.838-840.

Reilly, J.J., Armstrong, J., Dorosty, A.R., Emmett, P.M., Ness, A., Rogers, I., Steer, C., Sherriff, A., Avon Longitudinal Study of, P. and Children Study, T. 2005. Early life risk factors for obesity in childhood: cohort study. *BMJ.* **330**(7504), p1357.

Reilly, J.J., Dorosty, A.R., Emmett, P.M., Avon Longitudinal Study of, P. and Childhood Study, T. 2000. Identification of the obese child: adequacy of the body mass index for clinical practice and epidemiology. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity.* **24**(12), pp.1623-1627.

Reilly, J.J. and Kelly, J. 2011. 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**(7), pp.891-898.

Reilly, J.J., Methven, E., McDowell, Z.C., Hacking, B., Alexander, D., Stewart, L. and Kelnar, C.J. 2003. Health consequences of obesity. *Archives of Disease in Childhood.* **88**(9), pp.748-752.

Riddoch, C., Edwards, D., Page, A., Froberg, K., Anderssen, S.A., Wedderkopp, N., Brage, S., Cooper, A.R., Sardinha, L.B., Harro, M., Klasson-Heggebø, L., Mechelen, W.v., Boreham, C., Ekelund, U., Andersen, L.B. and Team, T.E.Y.H.S. 2005. The European Youth Heart Study—Cardiovascular Disease Risk Factors in Children: Rationale, Aims, Study Design, and Validation of Methods. *Journal of Physical Activity and Health.* **2**(1), pp.115-129.

Riley, R.D., Hayden, J.A., Steyerberg, E.W., Moons, K.G., Abrams, K., Kyzas, P.A., Malats, N., Briggs, A., Schroter, S., Altman, D.G., Hemingway, H. and Group, P. 2013. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Medicine / Public Library of Science.* **10**(2), pe1001380.

Ritchie, L.D. 2012. Less frequent eating predicts greater BMI and waist circumference in female adolescents. *American Journal of Clinical Nutrition.* **95**(2), pp.290-296.

Ritchie, L.D., Spector, P., Stevens, M.J., Schmidt, M.M., Schreiber, G.B., Striegel-Moore, R.H., Wang, M.C. and Crawford, P.B. 2007. Dietary patterns in adolescence are related to adiposity in young adulthood in black and white females. *Journal of Nutrition.* **137**(2), pp.399-406.

Rockett, H.R., Breitenbach, M., Frazier, A.L., Witschi, J., Wolf, A.M., Field, A.E. and Colditz, G.A. 1997. Validation of a youth/adolescent food frequency questionnaire. *Preventive Medicine.* **26**(6), pp.808-816.

Rockett, H.R., Wolf, A.M. and Colditz, G.A. 1995. Development and reproducibility of a food frequency questionnaire to assess diets of older children and adolescents. *Journal of the American Dietetic Association.* **95**(3), pp.336-340.

Royston, P., Moons, K.G., Altman, D.G. and Vergouwe, Y. 2009. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* **338**, pb604.

Rudolf, M. 2011. Predicting babies risk of obesity. *Archives of Disease in Childhood.* **96**(11), p995.

SACN. 2011. *Dietary Reference values for Energy* [Online]. [Accessed 28.2.2020.]. Available from: https://www.gov.uk/government/publications/sacn-dietary-reference-values-for-energy

Sanderson, S., Tatt, I.D. and Higgins, J.P. 2007. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *International Journal of Epidemiology.* **36**(3), pp.666-676.

Santorelli, G., Petherick, E.S., Wright, J., Wilson, B., Samiei, H., Cameron, N. and Johnson, W. 2013. Developing prediction equations and a mobile phone application to identify infants at risk of obesity. *PLoS ONE [Electronic Resource].* **8**(8), pe71183.

Scarborough, P., Adhikari, V., Harrington, R.A., Elhussein, A., Briggs, A., Rayner, M., Adams, J., Cummins, S., Penney, T. and White, M. 2020. Impact of the announcement and implementation of the UK Soft Drinks Industry Levy on sugar content, price, product size and number of available soft drinks in the UK, 2015-19: A controlled interrupted time series analysis. *PLoS medicine.* **17**(2), pe1003025.

Schlichting, P., Hoilund-Carlsen, P.F. and Quaade, F. 1981. Comparison of self-reported height and weight with controlled height and weight in women and men. *International journal of obesity.* **5**(1), pp.67-76.

Schwingshackl, L., Hoffmann, G., Kalle-Uhlmann, T., Arregui, M., Buijsse, B. and Boeing, H. 2015. Fruit and Vegetable Consumption and Changes in Anthropometric Variables in Adult Populations: A Systematic Review and Meta-Analysis of Prospective Cohort Studies. *PLoS ONE [Electronic Resource].* **10**(10), pe0140846.

Seyednasrollah, F., Makela, J., Pitkanen, N., Juonala, M., Hutri-Kahonen, N., Lehtimaki, T., Viikari, J., Kelly, T., Li, C., Bazzano, L., Elo, L.L. and Raitakari, O.T. 2017. Prediction of Adulthood Obesity Using Genetic and Childhood Clinical Risk Factors in the Cardiovascular Risk in Young Finns Study. *Circulation. Cardiovascular Genetics.* **10**(3).

Shannon, B., Smiciklas-Wright, H. and Wang, M.Q. 1991. Inaccuracies in self-reported weights and heights of a sample of sixth-grade children. *Journal of the American Dietetic Association.* **91**(6), pp.675-678.

Shroff, M.R., Perng, W., Baylin, A., Mora-Plazas, M., Marin, C. and Villamor, E. 2014. Adherence to a snacking dietary pattern and soda intake are related to the development of adiposity: a prospective study in school-age children. *Public Health Nutrition.* **17**(7), pp.1507-1513.

Simmonds, M., Burch, J., Llewellyn, A., Griffiths, C., Yang, H., Owen, C., Duffy, S. and Woolacott, N. 2015. 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.* **19**(43), pp.1-1.

Singh, A.S., Mulder, C., Twisk, J.W.R., van Mechelen, W. and Chinapaw, M.J.M. 2008. Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obesity Reviews.* **9**(5), pp.474-488.

Siqueira, L.M. 2006. Management of preteen obesity could reduce risk of metabolic syndrome. . *International Pediatrics.* **21**, p60.

Slater, M. 2012. *Olympics cycling: Marginal gains underpin Team GB dominance.* [Online]. [Accessed 10.5.2020.]. Available from: https://www.bbc.co.uk/sport/olympics/19174302

Slaughter, M.H., Lohman, T.G., Boileau, R.A., Horswill, C.A., Stillman, R.J., Van Loan, M.D. and Bemben, D.A. 1988. Skinfold equations for estimation of body fatness in children and youth. *Human Biology.* **60**(5), pp.709-723.

St-Jules, D.E., Watters, C.A. and Novotny, R. 2014. Estimation of fish intake in Asian and white female adolescents, and association with 2-year changes in body fatness and body fat distribution: the female adolescent maturation study. *Journal of the Academy of Nutrition & Dietetics.* **114**(4), pp.543-551.

Stang, A. 2010. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *European Journal of Epidemiology.* **25**(9), pp.603-605.

STATA 2017. College Station, Texas ,USA: Stata Corporation.

Steur, M., Smit, H.A., Schipper, C.M., Scholtens, S., Kerkhof, M., de Jongste, J.C., Haveman-Nies, A., Brunekreef, B. and Wijga, A.H. 2011. Predicting the risk of newborn children to become overweight later in childhood: the PIAMA birth cohort study. *International Journal of Pediatric Obesity.* **6**(2-2), pp.e170-178.

Steyerberg, Moons, K.G., van der Windt, D.A., Hayden, J.A., Perel, P., Schroter, S., Riley, R.D., Hemingway, H., Altman, D.G. and Group, P. 2013. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med.* **10**(2), pe1001381.

Steyerberg and Vergouwe, Y. 2014. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European Heart Journal.* **35**(29), pp.1925-1931.

Steyerberg, E.W. 2009. *Clinical Prediction Models: a practical approach to development, validation and updating.* New York: Springer.

Strauss, R.S. 1999. Comparison of measured and self-reported weight and height in a cross-sectional sample of young adolescents. *International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity.* **23**(8), pp.904-908.

Striegel-Moore, R.H., Thompson, D., Affenito, S.G., Franko, D.L., Obarzanek, E., Barton, B.A., Schreiber, G.B., Daniels, S.R., Schmidt, M. and Crawford, P.B. 2006. Correlates of beverage intake in adolescent girls: the National Heart, Lung, and Blood Institute Growth and Health Study. *Journal of Pediatrics.* **148**(2), pp.183-187.

Szajewska, H. and Ruszczynski, M. 2010. Systematic Review Demonstrating that Breakfast Consumption Influences Body Weight Outcomes in Children and Adolescents in Europe. *Critical Reviews in Food Science and Nutrition.* **50**(2), pp.113-119.

Taljaard, M., Tuna, M., Bennett, C., Perez, R., Rosella, L., Tu, J.V., Sanmartin, C., Hennessy, D., Tanuseputro, P., Lebenbaum, M. and Manuel, D.G. 2014. Cardiovascular Disease Population Risk Tool (CVDPoRT): predictive algorithm for assessing CVD risk in the community setting. A study protocol. *BMJ OPEN.* **4**(10), pe006701.

Taveras, E.M., Berkey, C.S., Rifas-Shiman, S.L., Ludwig, D.S., Rockett, H.R., Field, A.E., Colditz, G.A. and Gillman, M.W. 2005. Association of consumption of fried food away from home with body mass index and diet quality in older children and adolescents. *Pediatrics.* **116**(4), pp.e518-524.

Te Morenga, L., Mallard, S. and Mann, J. 2013. Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. *BMJ.* **346**, pe7492.

Textor, J., van der Zander, B., Gilthorpe, M.S., Liśkiewicz, M. and Ellison, G.T. 2017. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology.* **45**(6), pp.1887-1894.

Threapleton, D.E., Greenwood, D.C., Evans, C.E., Cleghorn, C.L., Nykjaer, C., Woodhead, C., Cade, J.E., Gale, C.P. and Burley, V.J. 2013. Dietary fibre intake and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ.* **347**, pf6879.

Trevino. 2005. Impact of the Bienestar school-based diabetes mellitus prevention program on fasting capillary glucose levels: A randomized controlled trial (vol 158, pg 911, 2004). *Archives of Pediatrics & Adolescent Medicine.* **159**(4), pp.341-341.

UCLA, S.C.G. 2019. *Logistic regression analysis STATA annotated output.* [Online]. [Accessed 16.12.2019.]. Available from: https://stats.idre.ucla.edu/stata/output/logistic-regression-analysis/

Valdés, J., Rodríguez-Artalejo, F., Aguilar, L., Jaén-Casquero, M.B. and Royo-Bordonada, M.Á. 2013. Frequency of family meals and childhood overweight: a systematic review. *Pediatric Obesity.* **8**(1), pp.e1-e13.

Van Houwelingen, J.C. and Le Cessie, S. 1990. Predictive value of statistical models. *Statistics in Medicine.* **9**(11), pp.1303-1325.

Vandenbroucke, J.P., von Elm, E., Altman, D.G., Gotzsche, P.C., Mulrow, C.D., Pocock, S.J., Poole, C., Schlesselman, J.J., Egger, M. and Initiative, S. 2014. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *International Journal Of Surgery.* **12**(12), pp.1500-1524.

Vidmar, S.I., Cole, T.J. and Pan, H. 2013. Standardizing anthropometric measures in children and adolescents with functions for egen: Update. *Stata Journal.* **13**(2), pp.366-378.

Wark, P.A., Hardie, L.J., Frost, G.S., Alwan, N.A., Carter, M., Elliott, P., Ford, H.E., Hancock, N., Morris, M.A., Mulla, U.Z., Noorwali, E.A., Petropoulou, K., Murphy, D., Potter, G.D.M., Riboli, E., Greenwood, D.C. and Cade, J.E. 2018. Validity of an online 24-h recall tool (myfood24) for dietary assessment in population studies: comparison with biomarkers and standard interviews. *BMC Medicine.* **16**(1), p136.

Waters, E., de Silva-Sanigorski, A., Hall, B.J., Brown, T., Campbell, K.J., Gao, Y., Armstrong, R., Prosser, L. and Summerbell, C.D. 2011. Interventions for preventing obesity in children. *Cochrane Database of Systematic Reviews.* (12), pCD001871.

WCRF. 2017. *Are you making yourself attractive to cancer?* [Online]. [Accessed 25.11.2019.]. Available from: https://www.wcrf-uk.org/uk/cancer-health-check

Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M. and al, e. 2009. *The Newcastle-Ottawa Scale (NOS) for assessing the quality of Nonrandomised Studies in Meta-Analyses.* [Online]. [Accessed 25 November 2015]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Weng, S.F., Redsell, S.A., Nathan, D., Swift, J.A., Yang, M. and Glazebrook, C. 2013. Estimating overweight risk in childhood from predictors during infancy. *Pediatrics.* **132**(2), pp.e414-421.

Weng, S.F., Reps, J., Kai, J., Garibaldi, J.M. and Qureshi, N. 2017. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PloS one.* **12**(4), pe0174944.

WHO. 2015. *World Health Organisation Obesity and overweight factsheet.* [Online]. [Accessed 5th March 2015]. Available from: http://www.who.int/mediacentre/factsheets/fs311/en/

WHO. 2020. *World Health Organisation Obesity and overweight fact sheet.* [Online]. [Accessed 08.04.2020.]. Available from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

WHO/FAO. 2003. *Diet, nutrition and the prevention of chronic diseases.* [Online]. [Accessed 28.2.2020.]. Available from: http://www.who.int/nutrition/publications/obesity/WHO_TRS_916/en

Widome, R., Forster, J.L., Hannan, P.J. and Perry, C.L. 2007. Longitudinal patterns of youth access to cigarettes and smoking progression: Minnesota Adolescent Community Cohort (MACC) study (2000–2003). *Preventive Medicine.* **45**(6), pp.442-446.

Wilkinson, J., Arnold, K.F., Murray, E.J., van Smeden, M., Carr, K., Sippy, R., de Kamps, M., Beam, A., Konigorski, S., Lippert, C., Gilthorpe, M.S. and Tennant, P.W.G. 2020. Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health.*

Winpenny, E.M., Penney, T., Corder, K., White, M. and van Sluijs, E. 2017. Changes in consumption of added sugars from age 13 to 30 years: a systematic review and meta-analysis of longitudinal studies. *Obesity Reviews.* **18**(11), pp.1336-1349.

Ye, E.Q., Chacko, S.A., Chou, E.L., Kugizaki, M. and Liu, S. 2012. Greater Whole-Grain Intake Is Associated with Lower Risk of Type 2 Diabetes, Cardiovascular Disease, and Weight Gain. *The Journal of Nutrition.* **142**(7), pp.1304-1313.

Yoshii, A., Plaut, D.A., McGraw, K.A., Anderson, M.J. and Wellik, K.E. 2009. Analysis of the reporting of search strategies in Cochrane systematic reviews. *Journal of the Medical Library Association.* **97**(1), pp.21-29.

Zeng, X., Zhang, Y., Kwong, J.S., Zhang, C., Li, S., Sun, F., Niu, Y. and Du, L. 2015. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. *J Evid Based Med.* **8**(1), pp.2-10.

Zhang, Z., Xu, G., Ma, M., Yang, J. and Liu, X. 2013. Dietary fiber intake reduces risk for gastric cancer: a meta-analysis. *Gastroenterology.* **145**(1), pp.113-120.e113.

Zheng, M., Rangan, A., Olsen, N.J., Andersen, L.B., Wedderkopp, N., Kristensen, P., Grontved, A., Ried-Larsen, M., Lempert, S.M., Allman-Farinelli, M. and Heitmann, B.L. 2015. Substituting sugar-sweetened beverages with water or milk is inversely associated with body fatness development from childhood to adolescence. *Nutrition.* **31**(1), pp.38-44.

Zheng, M., Rangan, A., Olsen, N.J., Bo Andersen, L., Wedderkopp, N., Kristensen, P., Grontved, A., Ried-Larsen, M., Lempert, S.M., Allman-Farinelli, M. and Heitmann, B.L. 2014. Sugar-sweetened beverages consumption in relation to changes in body fatness over 6 and 12 years among 9-year-old children: the European Youth Heart Study. *European Journal of Clinical Nutrition.* **68**(1), pp.77-83.

Ziauddeen, N., Roderick, P.J., Macklon, N.S. and Alwan, N.A. 2018. Predicting childhood overweight and obesity using maternal and early life risk factors: a systematic review: Predicting childhood overweight. *Obesity Reviews.* **19**(3), pp.302-312.