The feasibility of screening for obesity-related co-morbidities in children attending community weight management services

Vishal Sharma

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

Leeds Institute of Clinical Trials Research

Faculty of Medicine and Health

September 2018

Declaration

The candidate confirms that the work submitted is his own and that appropriate credit has been given 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 Vishal Sharma to be identified as Author of this work has been asserted by Vishal Sharma in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

Firstly I would like to give thanks to my supervisors, Maria Bryant, Jane Nixon, Susanne Coleman, Linda Sharples, and Julian Hamilton-Shield, for providing me with opportunity to undertake this programme of work and for their support and guidance over the years.

I would also like to express my gratitude to my colleagues and desk buddies at the Leeds Institute for Clinical Trials Research for their support and motivation during the PhD.

I am also extremely grateful to the participants who took the time to take part in the consensus study. This includes all the service users and the panel of experts; Dr Martha Ford-Adams, Professor Robbie Foy, Professor Jason Halford, Professor Andy Hill, Dr Laura Johnson, Sarah Meanwell, Dr Talat Mushtaq, Dr Bernadette Moore, Dr Ken Ong, Dr Miranda Pallan, Professor Jonathan Pinkney, Dr Duncan Radley, Lindsey Reece, Professor Pinki Sahota, Dr Georgina Selby, Professor Julian Shield, Jenny Walker, Dr Christina Wei, Carolyn Wellings, and Professor John Wilding.

Finally, I would like to thank my parents, my sister for proof-reading drafts of the chapters and providing emotional support, and to my extended family and friends for their continued patience, understanding, and motivation in completing the PhD.

Abstract

1.1 Background

NICE public health guidance 47 recommended screening for obesity-associated co-morbidities in children attending community weight management services to identify those at risk, and provide support/treatment to minimise future complications.

1.2 Aim

Assess the feasibility of developing a screening programme for obesity-associated co-morbidities in children.

1.3 Methods

The programme of work consisted of three stages based on the National Screening Committee's criteria:

1. A systematic review and meta-analyses of observational studies to determine the prevalence ratios of co-morbidities in children who were overweight and obese, relative to those of a healthy weight using a random effects model.

2. Results were presented to a panel of health professionals and researchers, and a separate panel of service users to obtain consensus on co-morbidities and screening methods for the proposed screening programme.

3. Feasibility of the proposed screening programme was assessed via thematic analysis of the transcripts from consensus meetings with the health professionals and researchers.

1.4 Results

1. Twenty-six co-morbidities from 162 studies including 1,801,388 children were identified. Prevalence ratios ranged from 1.4 (diabetes) to 58.0 (metabolic syndrome) for children with obesity, and 1.2 to 15.8, respectively, for children who are overweight, relative to those of a healthy weight.

2. Consensus was achieved for the screening of five co-morbidities, hyperglycaemia, hypertension, obstructive sleep apnoea, depression and anxiety. However, consensus for screening methods was not achieved.

3. Thematic analysis identified concerns regarding the appropriateness of screening for the co-morbidities, relating to acceptability of screening methods and impact of the screening programme.

1.5 Conclusions

Although screening for five co-morbidities was deemed important, further work is required to explore costs and benefits of a screening programme in line with National Screening Committee criteria with regards to screening methods, treatment and the practicalities of implementing a screening programme.

Table of Contents

List of Tables

List of Figures

List of Boxes

Chapter 1: Thesis Overview

1.1 Introduction to Thesis

In 2013, NICE released Public Health Guidance 47 (Weight management: lifestyle services for overweight or obese children and young people)(NICE 2013). The guidance recommended screening children and young people, henceforth referred to as children, attending community weight management services for obesity-associated co-morbidities to identify those at risk, and provide support/treatment to minimise future complications.

This PhD was undertaken to assess the feasibility of developing a co-morbidity screening programme over three stages based on criteria developed by the National Screening Committee (NSC) which considers four areas: the condition, the test, the treatment, and the screening programme (Public Health England 2015a). The following outlines the stages of the PhD, and the work undertaken.

Chapter 2: Background to Obesity and Screening Programme Development

Chapter two provides the background to the prevalence and impact of obesity in children globally and in the UK. This includes definitions for weight categorisation and the health and socio-economic impacts to the individual. The chapter goes on to present the NSC's criteria and how this applies to the proposed screening programme. The chapter ends with the aims and objectives of the PhD.

Chapters 3 and 4: Systematic Review and Meta-Analyses of Obesity-Associated Co-morbidities

Chapters three and four detail the systematic review and meta-analysis undertaken to identify a comprehensive list of co-morbidities associated with childhood obesity and their prevalence relative to those of a healthy weight. Additionally screening methods for the co-morbidities were identified. This work addressed part of the first and second criterion put forward by the NSC, the condition and the test, when considering the appropriateness of implementing a screening programme. The chapters provide an account of the aims, methods, and results of the systematic review and meta-analyses along with a discussion of potential confounding factors, and implications of the results.

Chapter 5: Consensus on Co-morbidities and Screening Methods for the Proposed Screening Programme

Chapter five provides an overview and examination of four consensus methods (Delphi, Nominal Group Technique, RAND/UCLA Appropriateness Method, and Consensus Development Conference), before providing the rationale for selecting the RAND/UCLA Appropriateness Method for undertaking the consensus study. The chapter goes on to detail the methods and results of the consensus study, and their impact on the proposed screening programme.

Chapter 6: Expert Opinion on the Feasibility of the Proposed Screening Programme

Chapter six provides a thematic analysis of the face-to-face meetings conducted with the health professionals and researchers as part of the consensus study. The aim and methods are described, as well as the results, which assisted in explaining the results of Chapter 5, their implications and areas for future work.

Chapter 7: General Discussion and Conclusion

Chapter seven summaries the results obtained in the previous chapters and the implications of the results in line with literature on screening programme development. The chapter goes on to discuss the strengths and limitations of the programme of research, and the opportunities for future work.

Chapter 2: Background to Obesity and Screening Programme Development

2.1 Introduction

This chapter provides the background and rationale for the feasibility of developing a co-morbidities screening programme for children and young people with obesity, for the purposes of the thesis this is hereon referred to as children with obesity. Section 2.2 provides an overview of obesity and compares definitions of obesity in adults and children. The section then provides an overview of the prevalence and risk factors for obesity. Section 2.2.4 discusses the individual and societal consequences of childhood obesity. Section 2.3 presents the NICE recommendation which was the premise for this PhD and provides the rationale for screening children attending UK weight management services. Section 2.4.3 discusses literature associated with the development of screening programmes and the criteria developed by the National Screening Committee, which are presented in relation to childhood obesity and its associated co-morbidities. Section 2.5 provides an overview of the work and the aim of the PhD, followed by Section 2.6 which summarises the remaining chapters of the thesis and the steps undertaken to assess the feasibility of developing a co-morbidity screening programme for children with obesity attending UK weight management services.

2.2 Introduction to Overweight and Obesity

Before 1980, less than one in 10 people in OECD (Organisation for Economic Co-operation and Development) countries were overweight or obese (OECD 2014). Since then, worldwide obesity levels have doubled, and in some countries tripled (OECD 2014; WHO 2016) (Figure 1). In 2016, it was estimated that globally more than three in 10 adults (aged 18+) were overweight or obese (OECD 2014; World Health Organisation 2017). Obesity is considered a worldwide epidemic, and has emerged as one of the most serious public health concerns of the 21st century (James *et al.* 2001; Racette, Deusinger and Desusinger 2003; Güngör 2014). Based on current trends, it is estimated that by 2030, 38% of the world's adult population will be overweight and an additional 20% will be obese (Kelly *et al.* 2008; Ng *et al.* 2014). This will have major implications for individuals and society.
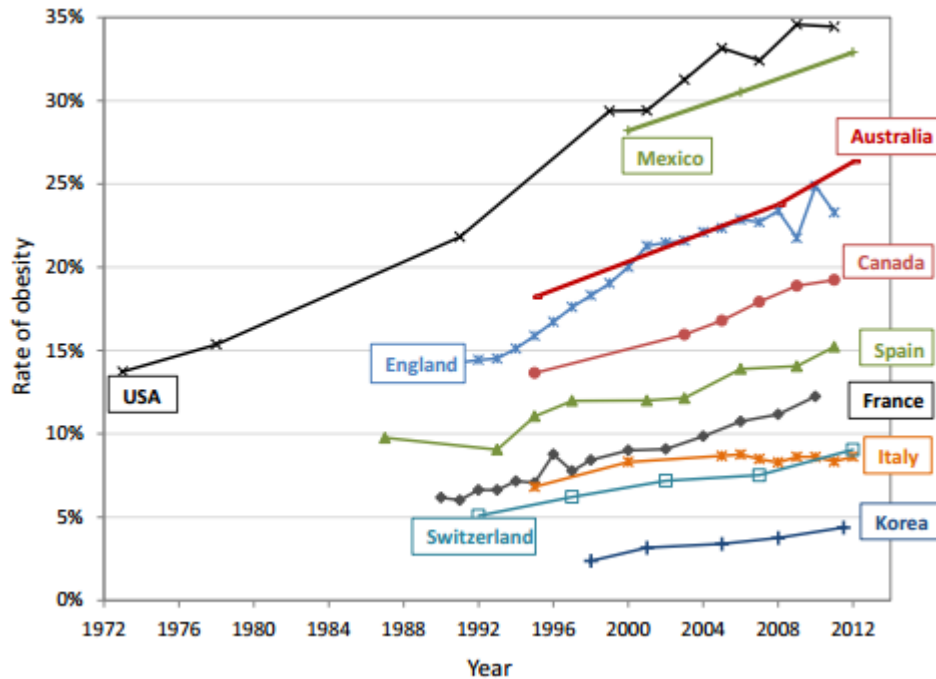
Figure 1: Worldwide obesity levels in adults
Graph taken from OECD (2014), which illustrates the rise in obesity prevalence in adults, by country, since the 1970s.

2.2.1 Rise in the Prevalence of Childhood Obesity

Over the past four decades there has also been a steady rise in the global prevalence of overweight and obesity in children (Han, Lawlor and Kimm 2010; OECD 2014). A review of childhood obesity concluded that its prevalence had increased over the last two to three decades in most industrialised countries and several lower income countries, particularly in urban areas, from 8% in 1980 to 13% in 2013 (Wang and Lobstein 2006; Ng *et al.* 2014). At a country-specific level, between the 1970s and late 1990s the prevalence of childhood obesity doubled or tripled in many countries, including Australia, Brazil, Canada, France, Germany, the UK, and the USA (Wang and Lobstein 2006; Ng *et al.* 2014). In 2014 an estimated 41 million children under the age of 5 years were overweight or obese globally (World Health Organisation 2017). In Africa, the number of children who are overweight or obese has nearly doubled from 5.4 million in 1990 to 10.6 million in 2014, and nearly half of the children under 5 who were overweight or obese in 2014 lived in Asia (WHO 2016).

Prevalence of Child Obesity in England

Estimates of obesity prevalence are usually derived from surveys or population studies. In England, child overweight/obesity levels are estimated using data from

the National Child Measurement Programme (NCMP). The NCMP measures the height and weight of children in reception (4-5 years old) and year 6 (10-11 years old) in England (NHS Digital n.d.). The programme began in the academic year 2006/2007. On the whole, the prevalence of childhood overweight/obesity has remained relatively stable in reception aged children since the programme began (22.9% in 2006/07 and 22.6% in 2016/17); although it is still considered to be high. For children in year 6 the prevalence increased from 31.6% to 34.2% over the same period (NHS Digital 2017b).

After primary school, children are not measured at a national level; however, van Jaarsveld and Gulliford (2015) analysed primary care records for 370,544 children and reported the prevalence of overweight and obesity in 11-15 year olds increased from 28.7% in 1994 to 37.2% in 2013. Taken in conjunction with the NCMP data, this would suggest children continue to gain weight into adolescence. UK child obesity levels remain amongst the highest in the world, with little evidence of the prevalence reducing in the near future (Ogden and Flegal 2010; Dinsdale, Ridler and Ells 2011; WHO 2014a; Public Health England 2015b; van Jaarsveld and Gulliford 2015; NHS Digital 2016).

Over recent years the prevalence of childhood obesity overall appears to have levelled off. However, analysis based on the 2016-2017 NCMP data has suggested a disparity in prevalence based on ethnicity and socioeconomic status. NCMP data has reported a higher prevalence of obesity in South Asian boys and girls, aged 4-5 and 10-11 years, and in Black girls aged 10-11 years, compared with White boys and girls, and a higher prevalence in those from more deprived areas (Hudda *et al.* 2017). Furthermore, comparing the most recent NCMP data with 2006/07 data indicated that the difference between obesity prevalence between the most and least deprived areas has increased from 4.5 to 6.8 percentage points in reception children and 8.5 to 15.0 percentage points in year 6 children; indicating widening inequalities between sub-populations (Hudda *et al.* 2017; Public Health England 2018a; Public Health England 2018b).

In addition to differences in prevalence between sub-populations, there has also been steady rise in the prevalence of severe obesity (Skinner and Skelton 2014; NHS 2015b). In UK children aged 10-11 years, between 2006/2007 and 2012/2013 the prevalence of severe obesity (BMI≥99.6[th] percentile) increased from 3.6% to 3.9% in boys and 2.5% to 2.9% in girls (Ells *et al.* 2015). In 2016/2017 the NCMP reported the prevalence of severe obesity in children for the first time. Data

indicated 2.4% of reception aged children and 4.0% of year 6 children were classed as severely obese, which received widespread media attention (BBC News 2018; Public Health England 2018a; Public Health England 2018b). Furthermore, Ells et al. (2015) reported that the prevalence of severe obesity varies geographically, with a higher prevalence in children from deprived areas, and in those from Black ethnic groups; in line with data from the NCMP regarding health inequalities (Hudda *et al.* 2017).

2.2.2 Definition of Overweight and Obesity in Adults versus Children

Obesity is characterised by an excess of body fat or adiposity; however, its measurement at a population level is predominantly assessed using BMI (kg/m^2) which, while correlated with adiposity, is not a direct measurement of it (Güngör 2014). In children, population-based, gender- and age-specific BMI percentile curves are often used to define overweight and obesity. This is due to differences in body composition in children resulting from puberty. In the UK, the 1990 reference charts are used in clinical practice and population monitoring, in which overweight is defined as BMI≥85th and <95th percentile, and obesity as BMI≥95th percentile (Table 1)(National Obesity Observatory (NOO 2011)).

Table 1: Classification of body weight in adults and children adapted from Hruby and Hu (2015)

	Age	Indicator	Healthy Weight	Overweight	Obese	Severe Obesity
Adults	≥18 years	BMI (kg/m^2)	18.5 - 24.9	25 - 29.9	≥30	≥40
Children	2-18 years	BMI Percentile	≥5th - <85th	≥85th - <95th	≥95th	≥99.6th

2.2.3 Aetiology and Risk Factors for Overweight/Obesity

At a simplistic level, overweight/obesity in children can be regarded as resulting from a prolonged period of energy imbalance, where energy intake is surplus to energy expenditure requirements (Wiskin *et al.* 2011; Güngör 2014). However this view does not account for the range of factors that determine the levels of energy intake and expenditure. These factors are multifactorial and can be regarded as a combination of individual, familial, and societal factors (Davison and Birch 2001;

Campbell 2016). For example, improvements in the amount, availability, and variety of food, socioeconomic status, and technological advancements have been associated with an increase in the prevalence of obesity (Chopra, Galbraith and Darnton-Hill 2002; Monteiro, Conde and Popkin 2004; Hruby and Hu 2015; World Health Organisation n.d.). The influence of these factors varies from person to person, which makes the aetiology of childhood obesity a complex issue (Davison and Birch 2001; Vandenbroeck, Goossens and Clemens 2007). An in-depth assessment of all possible risk factors is beyond the scope of this PhD, and only a brief overview of individual/genetic, family, social factors is provided.

At the individual level, research has considered genetic explanations for obesity. Twin studies estimated the heritability of body weight as being 30% to 70% (Borjeson 1976; Stunkard, Foch and Hrubec 1986; D'Amore 2006). Furthermore, genome-wide association studies have identified a number of genes, such as the FTO and MC4R genes, that have a positive association with increased weight (Mergen *et al.* 2001; Farooqi *et al.* 2003; Lubrano-Berthelier *et al.* 2003; Frayling *et al.* 2007; Tanofsky-Kraff *et al.* 2009; Fawcett and Barroso 2010; Karra *et al.* 2013). However on their own each gene is shown to have only a small effect on weight, accounting for less than 5% of cases of childhood obesity (Farooqi and O'Rahilly 2000; Anderson and Butcher 2006; Hindorff *et al.* 2009; Speliotes *et al.* 2010). Suggesting other factors are involved.

When considering families, research has indicated familial correlations with regards to obesity, a key part of which is parenting (Berge 2010). Parents have a key role in determining a child's behaviour and preferences related to food, physical activity and sedentary behaviour from an early age (Owen *et al.* 2000; Patrick and Nicklas 2005; Gustafson and Rhodes 2006; Budd and Hayman 2008; Bauman *et al.* 2012; Tzou and Chu 2012; Erkelenz *et al.* 2014; Sahoo *et al.* 2015; Vilchis-Gil *et al.* 2015; Wilkie *et al.* 2016; Garriguet, Colley and Bushnik 2017; Jago *et al.* 2017). Food preferences tend to develop at an early age and remain relatively stable (Skinner *et al.* 2002a; Skinner *et al.* 2002b; Story, Neumark-Sztainer and French 2002; De Cosmi, Scaglioni and Agostoni 2017). Children from overweight/obese families tend to have a higher preference for high fat foods, and a lower preference for fruits and vegetables than healthy weight children, and the children are more likely to become obese themselves (Wardle *et al.* 2001; Kral and Faith 2009). Furthermore being obese as a child increases the risk of adult obesity (Whitaker, Wright and Pepe 1997; The *et al.* 2010). Evidence has shown that parents can also have a positive influence on increasing levels of physical activity and reducing sedentary behaviour

in their children through role modelling (being active themselves), providing support for activities (financial, logistic, co-participation), setting clear rules and limits on sedentary activities, such watching television, and limiting access to electronic devices (e.g. computers, tablets, etc.) (Gustafson and Rhodes 2006; Moreno 2011; Bauman *et al.* 2012; Schmidt *et al.* 2012; Erkelenz *et al.* 2014). Evidence from systematic reviews has demonstrated that weight loss interventions are more successful when parents are actively involved (Ewald *et al.* 2014; Loveman *et al.* 2015).

Environmental influences on obesity are also well-documented (Kumanyika 2008; Ochoa and Berge 2017). As children age, the factors influencing behaviour expand to include friends, the media, and environmental factors, such as fewer parks and a decrease in active transport (Owen *et al.* 2000; Carver *et al.* 2011; Epstein *et al.* 2012; Hingle and Kunkel 2012; Salvy *et al.* 2012; Salvy and Bowker 2014; Mwaikambo *et al.* 2015). Additionally over the years there has been an increase in the consumption of both convenience foods and calorie-dense foods, in increasingly larger portion sizes (Harris and Shiptsova 2007; Wales 2009; Zheng *et al.* 2017). This, in combination with changes to the built-environment has led to a reduction in physical activity and an increase in sedentary behaviour (Owen *et al.* 2010; Janssen *et al.* 2016).

The Foresight Obesity Systems Map was developed to conceptualise the complex structure of factors associated with obesity, and identified 108 variables with over 300 interconnections (Vandenbroeck, Goossens and Clemens 2007). The impact of each variable on an individual's weight differs depending on the strength of connections between other variables. For instance, the variable palatability of food offerings will be influenced by variables such as food variety, food convenience, and individual dietary habits (Finegood, Merth and Rutter 2010). The wide-ranging factors associated with obesity and the varying associations between variables and individuals makes obesity a complex problem to address (Vandenbroeck, Goossens and Clemens 2007; Finegood, Merth and Rutter 2010; Frood *et al.* 2013).

Obesity as a disease

In May 2017, the World Obesity Federation released a position statement recognising obesity as a "chronic, relapsing, progressive disease" (Bray, Kim and Wilding 2017). This view is in accordance with the World Health Organization and the American Medical Association (WHO 2000; Pollack 2013). Furthermore in 2016, Members of the European Parliament called for the European Commission and

Council to work towards a Europe-wide recognition of obesity as a chronic disease; however this has not yet occurred (European Association for the Study of Obesity 2016).

Although it is not for the PhD to state whether or not obesity is a disease, there were potential ramifications for how the PhD was structured and the language used. The PhD is considering co-morbidities associated with obesity, therefore framing obesity as a disease would infer causation between the presence of obesity and the co-morbidity; however the thesis is not providing evidence of causality, but of association. Furthermore, obesity is not currently classed as a disease in the UK and the studies conducted as part of the PhD were completed with the view that obesity is not a disease; therefore it was decided to continue using language and definitions consistent with the time the PhD was commenced. Therefore for the PhD, obesity is seen as a contributor to disease and not a disease itself.

2.2.4 Implications of Obesity

Having an increasingly obese population has many implications for the individual and society. The following section first considers individual implications before going on to discuss societal implications.

Implications for the Individual

Research in adults indicates that being overweight/obese is associated with an increased risk of developing physical and psychological co-morbidities, such as type 2 diabetes, cardiovascular disease, depression, and some cancers (Must *et al.* 1992; van Dam *et al.* 2006; Narayan *et al.* 2007; Bjørge *et al.* 2008; Abdullah *et al.* 2010; Basen-Engquist and Chang 2011; Reilly and Kelly 2011; Must, Phillips and Naumova 2012; Schienkiewitz, Mensink and Scheidt-Nave 2012; Inge *et al.* 2013; Segula 2014; Pantalone *et al.* 2017; Pereira-Miranda *et al.* 2017). Co-morbidities are "any distinct clinical entity that has co-existed or that may occur" in an individual who is overweight/obese (Feinstein 1970). Comorbidities can be concurrent or successive depending on whether there is an overlap of conditions or one condition precedes the other (Angold, Costello and Erkanli 1999).

A meta-analyses of 89 studies identified statistically significant associations between BMI, a measure of adiposity, and the presence of 18 co-morbidities, including cardiovascular diseases, asthma, gallbladder disease, osteoarthritis, and chronic back pain (Guh *et al.* 2009). Previously, obesity-associated co-morbidities were only thought to affect adults; however evidence suggests that children are also

susceptible to a range of obesity-associated co-morbidities (Reilly *et al.* 2003; Hannon, Rao and Arslanian 2005; Al-Agha, Ocheltree and Shata 2012; Gunnarsdottir *et al.* 2012; Pulgarón 2013).

Child obesity can adversely affect multiple organ systems which can lead to the development of multiple co-morbidities, including hypertension, dyslipidaemia, insulin resistance, and type 2 diabetes (Daniels 2009; Han, Lawlor and Kimm 2010; Güngör 2014; Yoon 2014; Parker *et al.* 2016; Reuter *et al.* 2016; Brady 2017). Developing these co-morbidities at a younger age prolongs the duration of the co-morbidity and potentially influences the rate of its progression and the severity of any associated future complications (Gungor *et al.* 2005; Dean and Sellers 2007; Dabelea *et al.* 2017). The impact of prolonged exposure in children is particularly concerning as children can remain asymptomatic for long periods, which delays identification and early treatment of co-morbidities and associated complications (Daniels 2009; Wake *et al.* 2010; Mortensen *et al.* 2011; Öztürk 2017).

In addition to physical co-morbidities, evidence indicates an impact on one's quality of life by impacting social and psychological functioning, and their motor skills, affecting educational attainment and interpersonal relationships, which can persist into adulthood (Stunkard and Burt 1967; Gortmaker *et al.* 1993; Strauss 2000; Reeves, Postolache and Snitker 2008; Luppino *et al.* 2010; Esposito *et al.* 2014; Cheng *et al.* 2016a; Rankin *et al.* 2016).

Evidence has suggested that obesity can impact on multiple organs and systems and is associated with an increased risk for a wide spectrum of co-morbidities and associated complications (Dean and Sellers 2007; Daniels 2009; Han, Lawlor and Kimm 2010; Yoon 2014; Parker *et al.* 2016; Reuter *et al.* 2016; Brady 2017; Dabelea *et al.* 2017). These co-morbidities can severely impact one's health and wellbeing in childhood (Must *et al.* 1992; van Dam *et al.* 2006; Bjørge *et al.* 2008; Welsh, Karpen and Vos 2013). As a result, children are likely to require additional medical resources/treatment at an earlier age and will have a significantly decreased life expectancy (Hsia, Fallon and Brandt 2012).

Societal and economic burden of obesity

Having a high proportion of overweight and obese citizens also has socioeconomic implications for health services and the wider society.

Treatment and management of obesity and associated co-morbidities places a huge financial burden on the economy (NOO 2010; Rudisill *et al.* 2016). The global

economic impact from obesity has been estimated as approximately 2.8% of GDP, roughly $2.0 trillion (Dobbs *et al.* 2014). This is roughly equivalent to the global impact from smoking ($2.1 trillion) and armed violence, war, and terrorism ($2.1 trillion) (Dobbs *et al.* 2014). Evidence from the US indicates that men who are obese are thought to incur an additional $1,152 per year in medical spending compared to their non-obese counterparts, whereas women incur an average of $3,513 per year (Cawley and Meyerhoefer 2012). This is largely due to increased hospitalisations and medication costs (Cawley and Meyerhoefer 2012). Based on these values the authors estimated that 21% of annual US healthcare spending (approximately $190 billion) is due to treating obesity and related conditions (Arterburn, Maciejewski and Tsevat 2005; Cawley and Meyerhoefer 2012). These estimated costs are attributed to the treatment of obesity, such as bariatric surgery and associated conditions, such as type 2 diabetes (NOO 2010; Dobbs *et al.* 2014; Rudisill *et al.* 2016; NHS Digital 2017c; Diabetes UK n.d.).

In the UK, an economic analysis undertaken by the McKinsey Global Institute predicted that being overweight/obese costs society at least £27 billion each year (Dobbs *et al.* 2014). This figure attempted to calculate the wider societal costs of obesity, beyond NHS costs, such as lost working days due to sickness (Tunceli, Li and L.K. 2006). In 2007 obesity was estimated to cost the UK economy approximately £15.8 billion, with direct NHS costs estimated to be £10-12 billion per year by 2030 (NOO 2010; Health & Social Care Information Centre 2011; Wang *et al.* 2011; Dobbs *et al.* 2014).   In 2015/2016 the NHS carried out 6,438 bariatric surgery procedures, at a total cost of approximately £38.6 million (NICE 2014a; NHS Digital 2017c). Evidence suggests the cost of surgery may be recouped in the following years as a result of healthcare savings elsewhere. For instance if the healthcare savings from the cost of diabetes drugs alone is considered, then the cost of surgery is likely recouped within 2 to 3 years (NICE 2014a). However caution needs to be applied regarding the projected estimated costs of obesity. Subsequent evidence has suggested there was a mathematical error in the calculation due to a misinterpretation of the data upon which prevalence of obesity was estimated (Vadon, 2007; Smith, 2011; Jebb, 2017).

Being overweight/obese has implications for the individual (Enzi 1994). Children with obesity, particularly girls, are socially stigmatised, which adversely affects their education, and in adulthood affects their socioeconomic and marital status (Puhl and Heuer 2010; Kark and Karnehed 2012; Cohen *et al.* 2013). A systematic review of

educational attainment and obesity reported an inverse association, this was particularly evident in girls compared to boys (Cohen *et al.* 2013). However the authors reported that majority of studies in the systematic review did not adjust for confounders such as race/ethnicity, which limited generalisability of the results (Enzi 1994; Puhl and Brownell 2012; Cohen *et al.* 2013). Being overweight or obese as a child can also profoundly affect one's social and emotional well-being, and self-esteem (Strauss 2000; Reeves, Postolache and Snitker 2008; Sahoo *et al.* 2015). This in turn is associated with poor academic performance and a lower quality of life (Strauss and Pollack 2003; Sahoo *et al.* 2015).

These issues can perpetuate into adult life. Obesity also represents a major risk factor for premature resignation and has been associated with lower salaries compared with healthy weight counterparts, even when controlling for intelligence and social class, particularly for women (Sargent 1994; Reilly *et al.* 2003; Baum and Ford 2004; Jusot *et al.* 2008; Han, Norton and Stearns 2009). Moreover, the number of co-morbidities associated with obesity (see Section 2.2.4) means that obese employees are more likely to be sick, more often and for longer, impacting on their productivity, and subsequently increasing an employers' costs (Burton *et al.* 1998; Bhattacherjee *et al.* 2003; Chau *et al.* 2004; Degli Esposti *et al.* 2006; Tunceli, Li and L.K. 2006; Gates *et al.* 2008; Jusot *et al.* 2008). In the US, Gates *et al.* (2008) attempted to calculate the impact on employee productivity and sickness. They considered 341 manufacturing employees and reported that those with a BMI between 30-35kg/m^2 had the highest absenteeism, averaging 91 hours/year, and those with a BMI ≥35kg/m^2 had the greatest health-related work limitations, specifically regarding the time needed to complete certain tasks and ability to perform the physical demands of the job. This equated to a 4.2% loss in productivity, which annually resulted in an additional loss of $506 per employee.

2.3 NICE Recommendation: Screening for Obesity-associated Co-morbidities in Children attending Community Weight Management Services

In 2013, NICE released public health guidance 47 – "Obesity in children and young people: prevention and lifestyle weight management programmes" (NICE 2013). The guidance provided advice for community weight management services on the delivery of effective weight management programmes that "support children and young people to change their lifestyle and manage their weight" (NICE 2013).

Community weight management services were developed to provide lifestyle and behavioural strategies to manage weight, which include education regarding healthy diets, physical activity, and the risks of being overweight/obese, and setting realistic goals and appropriate rewards (NICE 2014c). However, the majority of child weight management services are not part of the NHS, therefore the staff running the services are not trained health professionals.

NICE health guidance 47 included a number of recommendations for child community weight management services, one of which was:

> "Assess each child or young person for obesity-associated diseases or conditions (co-morbidities). Use a locally approved co-morbidities assessment tool, where available. Assessment is particularly important if the child or young person and their family have self-referred to the programme, or have not been assessed by a health professional. Refer them to their GP if any concerns are identified." (NICE 2013)

Currently in the UK, child weight management services do not screen for obesity-related co-morbidities, although basic anthropometric measurements (height and weight) are taken to monitor progress. Some services, such as WatchIt, do assess self-esteem; however this is to assess the effectiveness of the programme and the results are not shared with families or the GP.

Screening children for obesity associated benefits may proffer some benefits. Firstly, children with co-morbidities would be identified sooner. Given that referral to community weight management services does not require a GP referral, the majority of children may not have had a clinical assessment. Thus this might be the child's first assessment. Currently weight management services conduct anthropometric measurements (i.e. BMI, waist circumference, waist-to-height ratio). However, evidence suggests that anthropometric measurements alone are not accurate markers for metabolic co-morbidities, indicating more comprehensive methods are required (Seibert, Allen and Carrel 2014; Morandi *et al.* 2014). Some services refer children, typically those who are extremely obese, to tertiary services for an in-depth clinical assessment (Sharma 2014). However, anecdotally the majority of service users do not attend these appointments, indicating that potential opportunities for early identification of co-morbidities may be missed (Sharma 2014).

Awareness of co-morbidities might encourage children and families to take weight-loss more seriously. Although there is no evidence of this in children, Ramirez *et al.*

(2017) reported that in adults awareness of modifiable risk factors for myocardial infarction increased the perceived need to change particular lifestyle behaviours. However, the association between risk factors and perceived need for change varied; for instance smoking, obesity, and low physical activity were most strongly associated with a perceived need to improve physical health, whereas hypertension and diabetes mellitus showed no association (Ramirez *et al.* 2017). In contrast, a systematic review and meta-analysis of 21 trials indicated that early intervention has been shown to prevent or delay the progression of impaired glucose tolerance to type 2 diabetes (Gillies *et al.* 2007). Although generalisability of the results to children is limited as the studies were based on adults; however the results do suggest that awareness may lead to positive lifestyle changes.

2.4 Introduction of Population Screening

Screening from a healthcare perspective, is "actively seeking to identify a disease or a pre-disease state in people who are asymptomatic but potentially have the disease" (Wilson and Jungner 1968). Screening grew in popularity in the twentieth century and in the UK there are currently 11 screening programmes which cover specific diseases at particular stages of the life cycle (Reiser 1978; Charap 1981; Ruf and Morgan 2008; Harris *et al.* 2011; Gov. UK n.d.). The benefits of screening programmes include early detection of conditions, sometimes prior to any symptoms (NHS Choices 2018). Early detection also means treatment is likely to be more effective (PubMed Health 2016). Additionally, awareness of a health problem may help people make informed decisions about their health, in terms of lifestyle choices. Finally, screening can save lives (NHS Choices 2018).

2.4.1 Definition of Screening

One of the first definitions of screening was put forward by the US Commission on Chronic Illness (1957):

> *"Screening is the presumptive identification of unrecognised disease or defect by the application of tests, examinations or other procedures, which can be applied rapidly. Screening tests sort out apparently well persons who apparently have a disease from those who probably do not."*

In 2013 the definition was simplified to "screening is the process of identifying healthy people who may be at increased risk of disease or condition" (UK National Screening Committee 2013). Although there is variation in terminology the core components remain consistent; the idea of actively testing an otherwise healthy, but potentially at risk individual, to identify a disease prior to the presentation of symptoms (McKeown 1968; Wilson and Jungner 1968; UK National Screening Committee 1998; UK National Screening Committee 2000; UK National Screening Committee 2013).

2.4.2 History of Screening Programmes

The first screening programme which demonstrated benefits was the use of MMR (mass miniature radiography) for identifying individuals with tuberculosis, (Semple 1953, 1960, 1966); Hawthorne 1964; Holland and Stewart 2005). The resulting global reduction in tuberculosis encouraged the use of screening for other chronic diseases (Breslow and Roberts 1955). The view was that regular screening, or preventative medical examination of adults, for a variety of conditions could reduce the costs and utilisation of medical services (Cutting 1985).

Despite the success of tuberculosis screening in the UK, there was a decline in screening programmes until Sir George Godber, Chief Medical Officer between 1960 and 1970, recognised the importance of screening as a method of delivering preventative healthcare (Holland and Stewart 2005). In the late 1970s, the UK's first steps towards implementing screening programmes began with cervical cancer screening (Godber 1975; NHS 2015a).

Over the years, screening programmes made their way to the forefront of the health agenda due to their appeal of early detection and therefore early treatment/prevention. However, programmes had varying degrees of success (Holland and Stewart 2005). An evaluation of 10 screening programmes found that in six programmes there was insufficient evidence with regard to one or more of:

    i)   The natural history of the disease;
    ii) The methods of diagnosis and treatment;
    iii) Operational problems;
    iv) Assessment of benefits and costs.

Secondly; the research and administrative framework for screening required developments in three areas:

i) Greater definition of the screening's requirements and evidence of existing programmes;

ii) More larger-scale, long-term research looking at the economics of screening;

iii) Meticulous attention to the introduction and development of programmes to ensure they were co-ordinated with the health service as a whole.

(McKeown 1968; Wilson and Jungner 1968; Holland and Stewart 2005).

Thus, the evaluation raised questions about the evidence base for screening programmes and criteria to assess the effectiveness of existing and future screening programmes (Harris *et al.* 2011).

At the same time as the evaluation, a Joint Standing Sub-committee on Screening in Medical Care was established. Their remit was to review the evidence for any screening programme and make recommendations on what was needed prior to introduction of the programme into the NHS (Holland and Stewart 2005). However the sub-committee only continued until 1980 as its future was being considered by the Standing Medical Advisory Committee. It was not until 1996 that the National Screening Committee was established. Along with this came an effective mechanism to influence the implementation of effective programmes and identify areas for further research (Holland and Stewart 2005).

2.4.3 Appraising the appropriateness of a Proposed Screening Programme

A key aim of screening is to bring disease detection and treatment to those who are undiagnosed whilst limiting harm to those who do not require treatment. However harm is a broad concept having physical and psychological components, and there is subjectivity in terms of what constitutes harm from person to person. As such, it is important to make specific considerations when developing and evaluating a screening programme that applies in the broadest sense. In 1968, Wilson and Jungner proposed 10 principles for evaluating screening programmes; the overall aim being to assess the benefit versus the cost of screening. In the UK, the National Screening Committee are tasked with assessing a suitability of implementing a screening programme (Kitchener *et al.* 2014). The National Screening Committee developed their own criteria, based on Wilson and Jungner's original criteria, but taking into account both the more rigorous standards of evidence required to improve effectiveness and the greater concern about the adverse effects of screening (Kitchener *et al.* 2014; Public Health England

2015a)(Box 1). The criteria are grouped into four categories, each of these are discussed below.

Box 1: Summary of the criteria for appraising the viability, effectiveness and appropriateness of a screening programme (Department of Health 2013)

The Condition:

- Important health problem
- Epidemiology, prevalence and natural history should be understood
- Implementation of cost-effective primary prevention interventions

The Test

- Simple, safe, precise and validated
- Suitable cut-off level defined and agreed
- Acceptable to the target population
- Agreed policy on further diagnostic investigation

The Treatment

- Effective intervention
- Pre-symptomatic intervention leads to better outcomes
- Agreed policies about what interventions are offered and to whom

The Screening Programme

- Effective in reducing mortality or morbidity
- Information provided about the test is readily understood
- Screening programme is acceptable to health professionals and public
- Benefits from the screening programme outweighs any harms
- Opportunity cost of the screening programme should be balanced in relation to expenditure on medical care as a whole

The Condition: Obesity-associated co-morbidities in children

The condition under consideration for screening should be an important health problem from the perspective of the individual, the wider community, and the health service in terms of length, course, and consequences of the condition (Public Health England 2015a). Furthermore, the epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood and there should be a detectable risk factor, or disease marker (Public Health England 2015a).

Adults who are obese are at risk of developing a number of obesity-associated co-morbidities, such as type 2 diabetes, hypertension and depression (Hannon, Rao and Arslanian 2005; Wardle and Cooke 2005; Loth *et al.* 2011; Maggio *et al.* 2014)(see section 2.2.4 (page 9)). Evidence suggests these co-morbidities are also present in children (Daniels 2009; Han, Lawlor and Kimm 2010; Yoon 2014; Parker *et al.* 2016; Reuter *et al.* 2016; Brady 2017), and if left untreated, can have serious health complications (de Gonzalez *et al.* 2010; Flegal *et al.* 2013). Obesity-

associated co-morbidities may meet the criteria for screening given the aforementioned evidence, in terms of being a significant problem, and the impact to the health service and wider society.

The Test: Obesity-associated co-morbidities in children

The National Screening Committee criteria state there should be a simple, safe, precise and validated screening test available, which is acceptable to the target population (Public Health England 2015a). Additionally, the distribution of test values in the target population should be known and a suitable cut-off level should be defined and agreed (Public Health England 2015a).

Given that the proposed screening programme is likely to screen for a number of different psychological and physical co-morbidities, for example, hyperglycaemia, hypertension and depression, and that a single test for all the potential co-morbidities is not available, the screening programme is likely to consist of a battery of tests. The number of screening tests will be contingent on the number of co-morbidities deemed suitable for inclusion in the proposed screening programme.

Additionally, each co-morbidity may potentially have multiple screening tests, for instance hyperglycaemia can be screened via fasted, 2-hour post-prandial, and random glucose measures, insulin resistance, and a risk factor screening questionnaire (Diabetes UK 2018b). Each of these screening tests will vary in terms of acceptability to the target population, i.e. children attending a community weight management service, and in terms of its cut-off and accuracy in screening for hyperglycaemia. Acceptability and accuracy of a screening test are considered in the following sections.

Screening Test Acceptability

Acceptability of screening tests amongst service users will likely vary by age, gender, and the type of screening test, e.g. blood test versus questionnaire (Gemmill *et al.* 2006; Eyles *et al.* 2013; Balán *et al.* 2016; Kirkøen *et al.* 2017). Research has indicated that acceptability of a screening test can be increased by engaging individuals in the screening process and informing them about the importance of screening and the potential benefits (Gemmill *et al.* 2006; Eyles *et al.* 2013; Balán *et al.* 2016; Kirkøen *et al.* 2017).

As mentioned above (Section 2.4.3) the proposed screening tool is likely to consist of multiple screening tests, one for each co-morbidity. Having multiple tests

increases the burden placed on service users, and each test is likely to have its own level of acceptability amongst the target population, for instance blood tests that require an overnight fast have lower acceptability than those that do not (Eborall *et al.* 2012; Diabetes UK 2016).

Another factor which affects acceptability of the test is the impact to the service user from the screening. The potential benefits of a screening test must outweigh the risks, both in terms of the potential for physical harm from the test, and psychological impacts from a true positive result, as well as from false positive and negative results (WHO 2003; Public Health England 2015a). A systematic review and meta-analysis considered the short (<4 weeks) and longer term (>4 weeks) emotional impact of screening (Collins, Lopez and Marteau 2011). Results indicated no significant impact of screening on longer term anxiety, depression, or quality of life (Collins, Lopez and Marteau 2011). However, only 12 studies were included in the systematic review, and even fewer were eligible for the meta-analysis. Furthermore, there were not sufficient studies assessing short term outcomes for meta-analysis. This outcome has been supported by other studies, which did not find psychological effects from screening or from false-positive results (Adriaanse and Snoek 2006; Eborall *et al.* 2007; Asif *et al.* 2014). Further research indicates that psychological impact can be reduced by providing appropriate information pre-screening, details of the screening test, the implications and risk, and information about what happens post-screening, in terms of impact of the results, and the next steps (Connelly *et al.* 1998; Adriaanse *et al.* 2002). The results indicate the importance of pre- and post-screening information given to the service users and their parents, in particular ensuring it is age-appropriate, as well as the communication skills of the staff member to go over the results with the family.

Screening Test Accuracy

The accuracy of a screening test is its ability to identify those with and without the co-morbidity of interest, and is usually quantified using sensitivity and specificity (Baratloo *et al.* 2015). Sensitivity of a test is the number of people correctly identified as having the co-morbidity, divided by the total number of people with the co-morbidity. Using a 2x2 table sensitivity can be expressed as $a \div (a + c)$ (Table 2). Specificity of a test is number of people correctly identified as not having the co-morbidity, divided by the total number of people without the co-morbidity; this can be expressed as $d \div (b + d)$. A perfect screening test would have 100% sensitivity and specificity.

Table 2: 2x2 table depicting the accuracy of a screening test

Best Proxy for Truth

		Positive	Negative	Total
Test Result	Positive	a	b	a+b
	Negative	c	d	c+d
	Total	a+c	b+d	a+b+c+d

However no screening test is likely to be perfect (100% sensitivity and specificity), therefore false positives and false negatives will occur. False positives occur when the screening test indicates presence of the co-morbidity even though the person does not have the co-morbidity (cell B in Table 2). False negatives occur when the screening test indicates absence of the co-morbidity, when in fact the person does have the co-morbidity (cell C in Table 2). As a screening test is not usually used to obtain a definitive diagnosis and is sometimes an indirect measure of the condition, it may have a higher degree of error than would be considered appropriate for diagnosis (Maxim, Niebo and Utell 2014). When selecting a screening test, consideration needs to be given to the practicalities of the test and the sensitivity and specificity of the test to minimise harm to the individual and potential impact to the screening programme as a whole (Raffle 2017a).

The Treatment: Obesity-associated co-morbidities in children

Chapter 1: Once co-morbidity has been identified, the NSC criteria states there should be an effective treatment or intervention available. The intervention should be supported by evidence indicating that early intervention leads to better outcomes than late intervention, with suitable provision of the treatment and clinical management of patient outcomes (Public Health England 2015a). For obesity, the first line of treatment for associated co-morbidities is typically weight-loss (NICE 2014c). For certain co-morbidities, however, such as the psychological ones or severe physical co-morbidities, additional treatment is recommended in addition to weight-loss. However treatment pathways and availability of services nationally for the treatment of co-morbidities specifically in children are less developed. For instance, for obstructive sleep apnoea there are currently only 32 sleep centres across the UK which are catered towards adults with sleep apnoea rather than children. Additionally the lack of centres nationally means they are not easily accessible to all. Furthermore for child mental health services there are typically long waiting lists, which means a child, once referred, may have to wait a long time

before being assessed and commencing treatment due to service specific constraints.

With regards to weight-loss, treatment strategies are designed to assist individuals in attaining and maintaining a healthy weight. Treatment interventions can be categorised as lifestyle and behavioural, pharmacological, and surgical (NICE 2014c).

Lifestyle and Behavioural Interventions

Lifestyle and behavioural interventions are intended to support individuals and families in adopting healthy lifestyle habits, increasing physical activity levels and improving diet to support weight-loss (Department of Health 2013). The lifestyle component typically includes education regarding the risks of being overweight/obese, information on the distinction between losing weight and maintaining weight loss, and the importance of developing skills for both, and realistic targets for outcomes other than weight loss, such as increased physical activity and healthier eating (NICE 2014c). Lifestyle strategies can also be supplemented with behavioural strategies, which provide training on self-monitoring of behaviours, goal setting, and deciding on appropriate rewards for reaching goals (NICE 2014c). To assist with lifestyle and behavioural interventions, community level services were established for children and adults (Department of Health 2013). A summary of the effectiveness of lifestyle interventions in children are considered below.

One systematic review and one meta-analysis were identified which considered the effectiveness of multicomponent lifestyle and behavioural interventions for weight loss in children (Al-Khudairy *et al.* 2017; Elvsaas *et al.* 2017). Although there was some variation in eligibility criteria, minimum duration of interventions, length of follow-up, and databases searched, results suggested multi-component lifestyle interventions, which consisted of a combination of diet, physical activity and behavioural components, achieved the greatest weight-loss when compared with controls (Al-Khudairy *et al.* 2017; Elvsaas *et al.* 2017). A systematic review of 28 studies in 2774 adolescents indicated the intervention group had on average a 1.18 kg/m^2 (95% confidence interval (CI) -1.67 to -0.69) reduction in BMI compared with control groups, suggesting a benefit from multi-component interventions (Al-Khudairy et al. 2017). Furthermore the BMI reduction was maintained at 18 to 24 months of follow-up, at which point BMI was on average 1.49 kg/m^2 (95% CI -2.56 to

-0.41) lower in the intervention groups compared with the control groups (Al-Khudairy et al. 2017).

A meta-analyses of 39 studies in children reported a significant difference in body mass index (BMI) after 6 months, 12 months, and 24 months in favour of multicomponent lifestyle interventions compared to standard, minimal, and no treatment (Elvsaas et al. 2017). A subgroup analysis indicated a greater effect in the short-term in studies conducted in specialist health care settings, which included a group treatment component; however, long-term benefits could not be confirmed. Although there was increased weight-loss, significant improvements in cardio-metabolic risk factors or psychological measures were not found (Al-Khudairy *et al.* 2017). Conclusions were that multicomponent lifestyle interventions resulted in a moderate improvement in BMI at 6, 12, and 24 months compared with standard and no treatment. In general the recommendations were that further research is required to assess the impact on cardio-metabolic risk factors and psychological measures.

Pharmacological Interventions

Only after lifestyle and behavioural interventions have been started and evaluated should pharmacological treatment be considered as an adjunct to an overall weight management plan (NICE 2014c; NICE 2014b). This is typically considered for people who have not reached their target weight loss or have reached a plateau using lifestyle and behaviour change techniques (NICE 2014c). Pharmacological treatment is not generally considered for children younger than 12 unless exceptional circumstances exist, such as the presence of severe co-morbidities which have been assessed by a multidisciplinary team (NICE 2014c).

A Cochrane Review of drug interventions for the treatment of childhood obesity identified 21 RCTs comparing investigational medical products against placebos, each group also had a behaviour change component such as diet, exercise, or both (Mead *et al.* 2016). The drugs considered by the studies included metformin (10 studies), sibutramine (six studies), orlistat (four studies) and one study group evaluated the combination of metformin and fluoxetine (Mead *et al.* 2016). Results indicated that the control groups had an average reduction in BMI of 0.45 kg/m^2, whereas the intervention groups had an average reduction of 1.3 kg/m^2. However the authors noted that many participants dropped out of the studies due to serious side effects, which limited the generalisability of the results, which were already based on relatively small samples. Mead *et al.* (2016) concluded that on average

the intervention groups lost 3.9kg more that the control groups, however many studies were of low quality, with short or no post-intervention follow-up period, and high dropout rates. They recommended future research focus have longer follow-ups to understand the long-term effects of any pharmacological intervention, and ensure sufficient power once drop-outs are considered (Mead *et al.* 2016). Of the drugs considered in the systematic review, only orlistat is currently available on the NHS (Wise 2016). The European Medicines Agency has approved the use of two additional medications in Europe, liraglutide and naltrexone-buproprion, but these are not available on the NHS (NICE 2017). This is due to a lack of data on the long-term effectiveness of the medications.

Surgical Interventions

Surgical treatment is typically for individuals who have a BMI ≥40 kg/m^2, or between 35 and 40 kg/m^2 and have a co-morbidity, such as type 2 diabetes or high blood pressure (Cheng *et al.* 2016). Bariatric surgery is considered once non-surgical interventions have been tried and the person has not achieved or maintained "adequate, clinically beneficial" weight loss (NICE 2014c). There are five main types of bariatric surgery (gastric band, gastric bypass, sleeve gastrectomy, intra-gastric balloon, and biliopancreatic diversion) (NHS 2017b). The mechanisms of effect vary between procedures, and include restriction of intake, diversion of food from the small intestine, and malabsorption of macronutrients (O'Brien 2016). At 10 years the weight-lost varies by type of surgery, ranging from 45-55% for gastric banding to 70% for biliopancreatic diversion (O'Brien 2016).

A meta-analysis of 637 children (age ranged 5-23) from 23 studies comparing surgical interventions (adjustable gastric band, sleeve gastrectomy, Roux-en-Y gastric bypass or biliopancreatic diversions) indicated significant reductions in BMI at one year, with a mean difference of 13.4kg/m^2 (Black *et al.* 2013). Results suggested that bariatric surgery led to significant short-term weight loss in obese children, with Roux-en-Y gastric bypass being associated with the largest reduction in BMI, followed by sleeve gastrectomy and then adjustable gastric band. With regards to co-morbidities, Black *et al.* (2013) reported that data on co-morbidity resolution were of very poor quality; studies examined different co-morbidities and the majority did not provide definitions. Only one of the included studies was an RCT (O'Brien et al. 2010). Results from the RCT indicated a mean weight loss of 34.6kg in the intervention group compared with 3kg in the control group. Furthermore 84% of those in the intervention group lost more than 50% of excess

weight compared with 3% in the control group. Gastric banding also had positive impacts on cardio-metabolic risk factors at 24 months and improved quality of life (O'Brien et al. 2010).

Despite the promising weight-loss resulting from surgical intervention, there is a lack of long-term data regarding the long term implications and risks from surgery for a still-developing adolescent (Hsia, Fallon and Brandt 2012). Furthermore, surgery is often not considered as treatment for children who are obese. A survey of 381 US paediatricians and family physicians indicated that 48% would not refer an obese adolescent for bariatric surgery (Woolford *et al.* 2010). The remaining would consider the severity of obesity, the child's age, and the duration of attendance at a weight-loss programme before referring an adolescent for surgery (Woolford *et al.* 2010).

The Screening Programme

The final NSC criteria relates to the screening programme itself, and requires evidence from high quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity (Public Health England 2015a). Related to this, the NSC requires evidence that the entire screening programme, the test, diagnostic procedures, and treatment) is "clinically, socially and ethically acceptable" (Public Health England 2015a). Part of this is ensuring that the benefits from the screening outweigh any harm, such as from false negatives or positives, and treatment issues, e.g. lack of suitable treatment. A lack of suitable treatment with a positive screening result can cause anxiety and worry, and not having a suitable treatment available makes the screening programme unethical (Raffle 2017b). The criterion is focused primarily on the effectiveness of the screening programme; although individual elements of the screening (test, diagnosis, and treatment) can be assessed individually, until a screening programme is implemented, the effectiveness of the programme, i.e. better outcomes for the target population, cannot be assessed.

Another element considered by the NSC is the opportunity cost of the screening programme. The screening programme (including testing, diagnosis, treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care for the condition as a whole (i.e. value for money). For screening programmes, there is a small cost incurred for each person screened, which is dependent on the cost of developing and administering of the programme, e.g. cost of the test, and the time and training of staff. The cost of the

conducting the screening test(s) can be reasonably well estimated, however the subsequent costs for diagnosis and treatment will depend on the accuracy of the test in the target population. Therefore the total cost of the screening programme overall may be high if there are a lot of false negatives (i.e. low prevalence and/or poor sensitivity of test) and/or a lot of false positives who then go on to have further unnecessary tests. Additionally there will also be additional costs associated with treatment for the true positives. The cost of the screening programme should be offset by reductions in costs associated with successfully treating/preventing a co-morbidity, but these savings will only be realised if morbidities can be prevented or delayed. This relies on the treatment being effective for a reasonable number of cases. If the co-morbidities may be identified outside of the screening programme with reasonable frequency then the programme is also not likely to save money.

Pharoah *et al.* (2013) assessed the overall effectiveness of the NHS breast screening programme, comparing 364 500 women aged 50 years, who would be eligible for screening, and a similar cohort who had regular mammographic screening. The results suggested that screening was cost effective due to an increase in quality adjusted life years (QALYs), however in 588 scenarios screening actually resulted in a reduction in QALYs (Pharoah *et al.* 2013). The authors concluded that the breast screening programme is only moderately likely to be cost effective and further primary research is required to provide definitive cost effectiveness data. This was supported by Morton, Sayma and Sura (2017), who focused on the cost-effectiveness of the breast cancer screening programme over a 20 year period. The results suggested that overall the breast cancer screening is cost-effective, however in the future this may change as more evidence becomes available over the risks of screening and chemotherapy drugs become cheaper. In contrast the cost-effectiveness of prostate cancer was considered to be less clear (Sanghera *et al.* 2018). The results from a systematic review of decision analytical models indicated there was no consensus on the optimal model type or approach to model prostate cancer progression and there was a general lack of data to enable analysis of cost-effectiveness.

With regards to co-morbidity screening, as the programme is likely to screen for multiple co-morbidities, each with its own treatment pathway involving different services to different extents, the cost effectiveness of the programme as a whole may be difficult to assess. Instead, the cost-effectiveness of screening for co-morbidities may need to be assessed individually.

Based on the criteria by the NSC in relation to obesity-related co-morbidities, there may be a potential benefit of screening for co-morbidities in children attending weight management services. However there are some unknowns, such as which co-morbidities and screening tests are suitable for inclusion in a screening programme and how many comorbidities may be prevented or ameliorated as a result of screening. Before assessing the suitability of screening for obesity-associated co-morbidities it is important to identify which comorbidities are important, whether there are suitable screening tests available, and whether or not such a programme is feasible.

2.5 Overview of Programme of Work

2.5.1 Aim and objectives of the PhD

To assess the feasibility of implementing an obesity-associated co-morbidities screening programme for children attending community weight management services in the UK. This aim will be meet via the following objectives, which are based on the criteria developed by the National Screening Committee (Public Health England 2015a):

- Co-morbidities: Identification of obesity-associated co-morbidities in children, that are suitable for inclusion in the proposed screening programme.
- Screening Tests: Identification of screening tests for the co-morbidities that are suitable for community weight management service settings.
- Screening Programme: Identification of the factors which should be considered when considering the feasibility of implementing a co-morbidity screening programme

The objectives and the chosen methods are summarised in Figure 2.
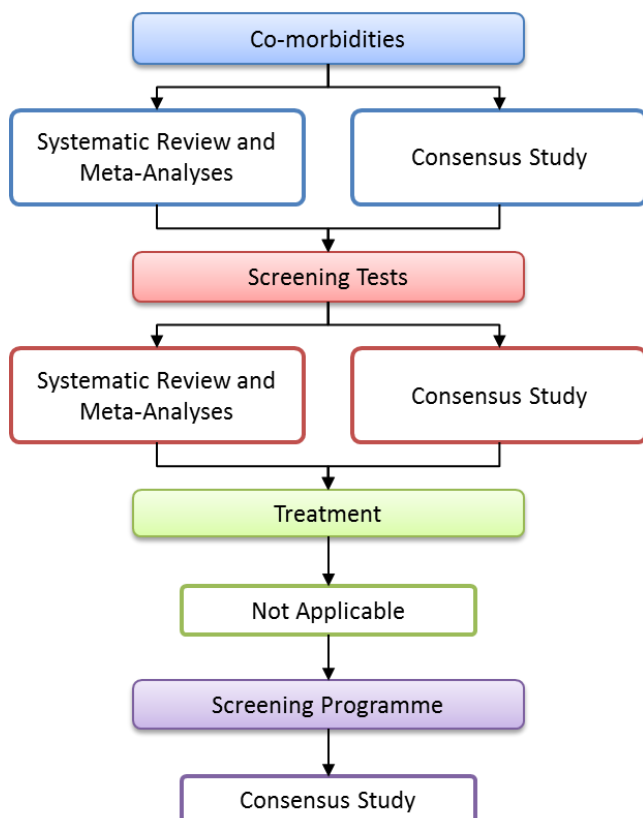
Figure 2: Stages of the PhD to address NSC criteria for developing a screening programme

2.6 Summary and Subsequent Thesis Chapters

Over recent decades there has been a steady rise in the global prevalence of obesity, to the point that it is now considered a global health epidemic (Racette, Deusinger and Desusinger 2003; Flegal *et al.* 2013). This rise has also been observed in children, and although the prevalence varies between specific sub-groups defined by, for example, ethnicity and socioeconomic status, the prevalence is alarmingly high throughout (Gatineau and Mathrani 2011; Devaux and Sassi 2013).

The rise in obesity has been followed by a rise in related co-morbidities in both adults and children (Abdullah *et al.* 2010; Parker *et al.* 2016). Each of these co-morbidities results in additional implications for the health and well-being of the population, which in turn has repercussions for society as a whole (de Gonzalez *et al.* 2010; Health & Social Care Information Centre 2011; Flegal *et al.* 2013).

The rise in obesity and its associated co-morbidities supported the NICE recommendation to screen children attending community weight management

services, for obesity-related co-morbidities (NICE 2013). Screening for the early identification of obesity-related co-morbidities may seem theoretically justified; such screening programmes should meet specific criteria to ensure their success in achieving the desired goals from a medical and financial standpoint (Sheehy, Coursin and Gabbay 2009; Kitchener *et al.* 2014). Furthermore, such criteria should be used as part of an ongoing evaluation of screening programmes to justify the continued investment of limited resources.

Subsequent chapters of the Thesis report work undertaken to assess the feasibility of developing an obesity-associated co-morbidity screening programme, aimed at children attending community weight management services in the UK:

- Chapters 3 and 4: discusses the methodology and presents the findings of a systematic review and meta-analyses to obtain a comprehensive list of obesity-associated co-morbidities and their estimated relative prevalence, for consideration for inclusion in a screening tool.
- Chapter 5: takes the results of Chapter 3, to obtain consensus on the co-morbidities and screening measures that are suitable for the intended population and settings (community weight management services in the UK).
- Chapter 6: presents the results of a thematic analysis to better understand the factors considered by the consensus panel in arriving at their decision in the consensus study (Chapter 5). The results are compared with other consensus studies and with the outcomes of current and previous screening programmes in the UK.
- Chapter 7: summaries the findings of the overall programme of work. Implications of the results and areas for future research are discussed.

Chapter 3: Systematic Review and Meta-Analyses of Obesity-associated Co-morbidities

3.1 Introduction to Systematic Review and Meta-Analyses

This chapter reports on a systematic review and meta-analyses, which was undertaken to identify a comprehensive list of co-morbidities associated with childhood obesity, their prevalence, and population prevalence ratio for children who were overweight and obese relative to those of a healthy weight. In addition, the review was used to identify co-morbidity screening methods. This was the first stage in identifying co-morbidities and screening methods that may be appropriate for inclusion in the proposed screening programme. Section 3.2 provides an overview of and justification for systematic reviews and meta-analyses. The chapter goes on to provide the aim (Section 3.3), details of methodological considerations (Section 3.4), methods (Section 3.5), and an overview of the results (Section 3.6). Results for the individual co-morbidities and discussion of the findings are provided in Chapter 4.

3.2 Systematic Review and Meta-Analyses

3.2.1 Background to Systematic Reviews

Systematic reviews, through a comprehensive process, aim to identify all studies pertaining to a focussed question (Egger, Smith and Altman 2001). The studies are appraised, their results summarised, and the key findings presented (Oxman, Cook and Guyatt 1994; Cook, Mulrow and Haynes 1997; Garg, Hackam and Tonelli 2008). Systematic reviews may also identify gaps in the literature, that future research should address in order to develop an understanding within the field (Egger, Smith and Altman 2001). The key strengths of systematic reviews are that they:

1. Identify, appraise and synthesise all available research relevant to a predefined research question. This reduces biases in the review and increases replicability of the results.
2. Collate all published data on a topic, providing a reliable foundation for decision makers by resolving conflicting information.
3. Provide details of a clearly documented review process. This enables replication and integration of new evidence as it becomes available.

(Oxman, Cook and Guyatt 1994; Cook, Mulrow and Haynes 1997; Egger, Smith and Altman 2001; Garg, Hackam and Tonelli 2008).

Justification for a Systematic Review Childhood Obesity-Related Co-morbidities

A systematic review was conducted to identify a comprehensive list of co-morbidities associated with childhood obesity, and to estimate their prevalence as part of a programme of work considering inclusion in a co-morbidity screening programme. Previous systematic reviews on childhood obesity-related co-morbidities have been conducted; however these had limitations (Guh *et al.* 2009; Pulgarón 2013; Sanders *et al.* 2015). Guh *et al.* (2009) considered co-morbidities over the entire lifespan, with participants aged 14 to 98. The exclusion of children aged 5-13 and inclusion of individuals over 18 meant a comprehensive list of obesity-related co-morbidities specific to children could not be constructed. Although the review was well-conducted and structured, the authors did not conduct a quality appraisal of the studies and only included prospective cohort studies, limiting the accuracy of the estimated population prevalence (Fletcher and Fletcher 2005). Pulgarón (2013) focused on paediatric obesity, with participants aged up to 22 years. However, Pulgarón's review lacked a clearly focused question, as well as details regarding the search strategy and study eligibility criteria, limiting the replicability of the review. Additionally, the review started with a predefined list of co-morbidities. Thus it does not provide an exhaustive list of comorbidities. A more recent systematic review was conducted by Sanders *et al.* (2015). Overall, this was well conducted with a clear, focused question and replicable search strategy; however, the review was restricted to children in Australia, limiting the generalisability of the results to the UK population and the identification of all obesity-related co-morbidities. Conducting a global review increases the probability of identifying all childhood obesity-related co-morbidities. However, data from non-UK countries are less likely to be representative of the UK population (where the proposed screening programme would be implemented), due to variation in culture, diet, lifestyle and ethnic composition. Thus caution is required when analysing results from non-UK countries.

3.2.2 Background to Meta-analyses

Building upon the structured nature of systematic reviews, meta-analyses are defined as the "statistical synthesis of results from a series of studies" (Borenstein *et*

al. 2009). Unlike narrative reviews where reviewers implicitly assign a degree of importance to each study, in meta-analyses each study is assigned a weighting (importance) based on the estimated precision of the study estimates (Egger, Smith and Phillips 1997; Haidich 2010). This weighting, in combination with statistical analyses, provides a transparent, objective, and replicable analysis of the literature (Egger, Smith and Phillips 1997). The outcome of a meta-analyses is a pooled (population) estimate; the population prevalence ratio of the co-morbidities in children in the overweight and obese groups relative to the healthy weight group (Klassen, Jadad and Moher 1998; Garg, Hackam and Tonelli 2008). This estimated population prevalence ratio would allow for conclusions to be drawn regarding the association between weight status and co-morbidity prevalence (Stroup *et al.* 2000).

Justification for Meta-analyses of comorbidities

To date, there has been no meta-analysis of the prevalence of obesity-related co-morbidities in children. One existing meta-analysis was identified, however it considered participants aged 14-98 years and therefore the results would not apply directly to children (Guh *et al.* 2009). Therefore meta-analyses were conducted to understand the extent of increased prevalence of co-morbidities in children who are overweight/obese relative to those of a healthy weight. This was to ensure the co-morbidity was sufficiently more prevalent in overweight/obese children to warrant inclusion in the proposed screening programme.

3.3 Aim

The aim of the systematic review and meta-analyses was to identify a comprehensive list of co-morbidities associated with childhood obesity, to estimate their prevalence and population prevalence ratio (primary aim), and identify screening methods (secondary aim). This was the first stage in identifying co-morbidities and screening methods that may be appropriate for inclusion in the proposed screening programme for community weight management services in the UK. This aim was met through the following objectives:

1. Identification of articles reporting co-morbidity prevalence in children, who were overweight and/or obese, compared to those of a healthy weight.
2. Extraction of study data, such as study design, participant characteristics, and prevalence by weight status.
3. Estimation of prevalence ratio of co-morbidities in overweight and obese weight groups (separately) relative to healthy weight.

4. Appraisal of study quality.

5. Development of a comprehensive list of obesity-related co-morbidities in children.

6. Identification of candidate screening/assessment techniques for the co-morbidities and cut-off points for referral.

3.4 Methodological Considerations

3.4.1 Study Design

Four observational study designs were considered for inclusion in this review:

1. **Cross-sectional studies** examine the prevalence of a disease in a specified population (MacMahon and Trichopoulos 1996). They may use existing patient data collected as part of routine clinical care, or may recruit and survey new participants. The former often being economical as the data have already been captured, but is more likely to suffer from missing data, retrieval bias and poorly defined outcomes (MacMahon and Trichopoulos 1996; Mann 2003).

2. **Cohort studies** identify a sample of participants from a population according to potential risk factors (e.g. weight status) and assess subsequent outcomes (e.g. the presence of disease) (MacMahon and Trichopoulos 1996; Fletcher and Fletcher 2005). Cohort studies can use prospective data collection or existing data (Fletcher and Fletcher 2005; Sedgwick 2014).

3. **Case-control studies** select groups of individuals based on whether they do (cases) or do not (controls) have the disease of interest (MacMahon and Trichopoulos 1996). The two groups form the basis of an evaluation of the relationship between the presence of a disease and existing characteristics amongst the participants, e.g. increased weight (Fletcher and Fletcher 2005).

4. **Case series** provide a description of the course of a series of patients with a similar diagnoses (Carey and Boden 2003; Chan and Bhandari 2011).

3.4.2 Weight Measurement

Measurement of a child's weight can be self-reported or measured by the clinical/research team. Self-report data is less reliable and less accurate than measured weight (Wang, Patterson and Hills 2002; Sherry, Jefferds and Grummer-Strawn 2007; Enes *et al.* 2009; Gokler *et al.* 2018). Weight measurement is used, along with height, to categorise children as healthy weight, overweight, or obese using national BMI percentile charts (NHS Digital 2017c).

3.4.3 Co-morbidity Measurement

Self-reporting of co-morbidities is common, especially for psychological co-morbidities such as depression and anxiety, usually via a questionnaire. Conditions screened for using questionnaires, rather than objective clinical tests, may be more susceptible to self-report/response bias, for example as a result of social desirability effects (Grimm 2010; Rosenman, Tennekoon and Hill 2011). Conversely, excluding these studies could exclude potentially relevant co-morbidities that can only be screened for through self-report.

3.4.4 Prevalence

Prevalence is defined as the proportion of a population who have a specified characteristic/disease of interest in a given timeframe, represented as n/N or as a percentage ((n/N)*100) (Egger, Smith and Altman 2001). The definition of the timeframe influences the type of prevalence that is calculated:

1. Point prevalence: proportion of a defined population affected by the disease in question at a specified point in time.
2. Period prevalence: proportion of a defined population affected by the disease in question at any point during a given time period, e.g. past 12 months.
3. Lifetime prevalence: the proportion of a population that, at some point in their lives, up to the time of assessment, have ever had the disease.

(Adapted from Egger, Smith and Altman (2001) and National Institute of Mental Health (2017)).

3.4.5 Prevalence Ratio

Given the prevalence for two or more comparison groups, the prevalence ratio (PR) can be estimated. This is the ratio of the proportion of people who have the disease in the exposed group, and the proportion with the disease in the unexposed group (Figure 3). A ratio equal to 1 indicates the prevalence is the same in the exposed (e.g. obese) and unexposed (e.g. healthy weight) groups. A ratio greater than 1 indicates greater prevalence in the exposed group, and a ratio less than 1 indicates a lower prevalence in the exposed group.

	Co-morbidity	
	Yes	No
Exposure Yes	a	b
(e.g. No	c	d
obese)		

Prevalence Ratio = $\left(\dfrac{a}{a+b}\right) \div \left(\dfrac{c}{c+d}\right)$

Figure 3: 2x2 prevalence table and prevalence ratio calculation

One thing to consider is that PR is influenced by the prevalence in the healthy weight group. For example, if the prevalence in the healthy weight group is 0.1 for one study, and 0.2 in another, and the prevalence in the obese group for both studies is 10, the PR would be 100 and 50, respectively. This would suggest a considerable difference, when in fact it is not; therefore the prevalence in the healthy weight group needs to be considered when interpreting the PR.

3.4.6 Quality Appraisal of Studies

A key part of a systematic review is the quality/critical appraisal of included studies, as the main source of bias is predominantly due to limitations or design issues in the original studies, which reduces confidence in and generalisability of results (Sanderson, Tatt and Higgins 2007; Bown and Sutton 2010; Dhillon and Gill 2014). The proposed systematic review considered observational study designs to assess the prevalence of obesity-related co-morbidities in children, thus a quality/critical appraisal tool suitable for observational studies was required.

A literature search identified two potential tools; the Newcastle-Ottawa Scale and the Joanna Briggs Critical Appraisal Checklist for Studies Reporting Prevalence Data (Hartling *et al.* 2013; Lo, Mertz and Loeb 2014; Munn *et al.* 2014; The Joanna Briggs Institute 2016). Both tools were assessed against four criteria that were developed from a systematic review of observational study appraisal tools (Sanderson, Tatt and Higgins 2007). The four criteria are, i) number of domains, ii) specific as possible, iii) checklist rather than a scale, and iv) evidence of careful development and their validity and reliability. Each criterion is discussed below.

Criterion 1: Number of Domains

Sanderson, Tatt and Higgins (2007) identified six key domains that should be addressed by a critical appraisal tool (Table 3). The domains were developed using the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) Guidelines.

The Newcastle-Ottawa Scale addressed four of the domains using eight questions; in contrast the Critical Appraisal Checklist addressed five with 10 questions (Table 3). Neither tool assessed whether there was a conflict of interest in the study, yet many articles publish this information and it is a requirement of many journals that this is disclosed. Additionally the Newcastle-Ottawa Scale does not ask about whether appropriate statistical methods were used for the primary analysis of effect.

Table 3: Sanderson et al.'s domains addressed by the Newcastle-Ottawa Scale and Critical Appraisal Checklist

Domains Identified by Sanderson, Tatt and Higgins (2007)	**Newcastle-Ottawa Scale (Wells et al. 2014)**	**Critical Appraisal Checklist (Munn et al. 2014)**
Methods for selecting study participants	✓ - 4 items	✓ - 4 items
Methods for measuring exposure and outcome variables	✓ - 2 items	✓ - 2 items
Design-specific sources of bias (excluding confounding)	✓ - 1 item	✓ - 1 item
Methods to control confounding	✓ - 1 item	✓ - 2 items
Statistical methods (excluding control of confounding)	✕	✓ - 1 item
Conflict of Interest	✕	✕

Criterion 2: Specific as Possible

Sanderson et al. (2007) did not provide detailed information regarding how this criterion should be assessed, however they did state that due consideration of the study's design and topic area was required, whilst still being applicable to all forms of observational studies. This links to some of the key domains reported in criterion one, specifically participant selection, appropriate measurement, and control of confounders.

The Newcastle-Ottawa Scale has separate versions for use in cohort, case-control and cross sectional study designs (Anglin *et al.* 2013; Hermont *et al.* 2014; Wells *et al.* 2014). Each version gives consideration to the representativeness and selection of participants, comparability between groups, and ascertainment of disease (e.g.

self-report versus clinical records). The response options available vary by question, for instance question one (Is the case definition adequate?) has three options (a - yes, with independent validation; b - yes, e.g. record linkage or based on self-reports; c - no description), whereas question four only has two response options. The number of potential responses ranges from two to four depending on the question. The Critical Appraisal Checklist also gives consideration to the representativeness and selection of participants, however neglects to consider how the presence of the disease was assessed. In contrast, the Critical Appraisal Checklist offers standard options of Yes, No, Unclear, and Not Applicable for all questions.

To assist with completion, both tools provide guidance to minimise subjectivity. The existing guidance for the Newcastle-Ottawa Scale, however, has been described as "vague" and "unhelpful", which increases subjectivity in responses (Devereaux *et al.* 2004; Soares *et al.* 2004; Stang 2010; Hootman *et al.* 2011; Oremus *et al.* 2012; Hartling *et al.* 2013). In contrast, the guidance of the Critical Appraisal Checklist appears to be more detailed, with information on which factors the reviewer needs to consider.

Criterion 3: Checklist rather than a Scale

Sanderson et al.'s (2007) stated the quality appraisal tool should be a checklist as opposed to a scale. A checklist identifies which elements are addressed by study, whereas a scale assigns a score to each item which is used to calculate a summary score. Although the summary score can be used to rank studies as high, medium, or low quality, summary scores do not reflect the numerous considerations that go into assessing a study's quality (Jüni *et al.* 1999; Dreyer *et al.* 2014). Additionally scales typically treat each item as being of equal importance, whereas two studies scoring six might differ greatly in terms of quality when the individual items are reviewed (Greenland and O'Rourke 2001; Stang 2010; Higgins *et al.* 2011; Higgins and Green 2011).

Despite the Newcastle-Ottawa Scale begin called a "scale", it can be used as a simple checklist without modification. The Critical Appraisal Checklist was designed as a checklist; however it has been amended for use as a scale (Stang 2010; González-Serrano *et al.* 2016).

Criterion 4: Evidence of Careful Development and their Validity and Reliability

This final criterion evaluates the tool's development methods and its validity and reliability. There is limited information available on the development of the Newcastle-Ottawa Scale, its first reported publication being an abstract at the Third Symposium for Systematic Review in Oxford, UK (Hartling *et al.* 2013). The lack of publication in a peer-reviewed journal detailing the development process makes it difficult to assess development methods and the tool itself. Data regarding reliability of the Newcastle-Ottawa Scale are also poor (Devereaux *et al.* 2004; Soares *et al.* 2004; Stang 2010; Hootman *et al.* 2011; Oremus *et al.* 2012; Lo, Mertz and Loeb 2014). Hartling *et al.* (2013) had reviewers independently apply the Newcastle-Ottawa Scale to 131 cohort studies included in 8 meta-analyses, the overall Kappa was 0.29 (95CI 0.10 to 0.47), indicating only fair agreement. In contrast, for the Critical Appraisal Checklist there is information on its development in a peer-reviewed journal (Munn *et al.* 2014). A working group identified and reviewed tools which had previously been used to assess quality in prevalence studies. This information was collated to develop the first version of the Critical Appraisal Checklist, which was piloted with 16 workshop participants at the Joanna Briggs Institute Convention (Munn *et al.* 2014). Based on feedback the tool was adapted and refined. However, despite use in multiple reviews, there is limited evidence on the validity and reliability of the Critical Appraisal Checklist.

Summary of Quality Tool Appraisal

Assessment against the criteria put forward by Sanderson, Tatt and Higgins (2007) indicated the Critical Appraisal Checklist covered a greater number of the key domains in comparison to the Newcastle-Ottawa Scale (Table 3, page 35), and was developed specifically for prevalence systematic reviews. Both are relatively similar in terms of being specific and both can be used as a scale or checklist. There is more information available on the development of the Critical Appraisal Checklist, but less evidence on its validity and reliability.

Based on the above analysis the studies included in the meta-analyses were assessed for quality using the Critical Appraisal Checklist.

3.4.7 Meta-Analyses

Summary Effect Models

Within meta-analyses there are two commonly-used statistical approaches for estimating the summary effect, the Common Fixed Effect Model and the Random Effects Model.

Common Fixed Effect Model

The common fixed-effect model starts with the assumption that there is one effect size (prevalence ratio) which underlies all the studies in the analysis, and any variation between studies is due to sampling error (Hedges and Vevea 1998; Borenstein *et al.* 2010). However, the assumption that the prevalence ratio is the same across all the studies is often not plausible. There may be sufficient commonality between the studies to enable a meta-analyses, but variation in the study design (e.g. participant ages, country of study, lifestyle and diet, and/or ethnicities) may result in variation in the underlying prevalence ratio (Borenstein *et al.* 2009).

Random Effects Model

The random-effects model assumes that the prevalence ratio may vary from study to study, and that the prevalence ratio for these studies follows a statistical distribution, often assumed to be normal on the natural, log-odds, or log scale (Hedges and Vevea 1998; Higgins, Thompson and Spiegelhalter 2009; Borenstein *et al.* 2010). The prevalence ratios reported by the studies are assumed to represent a random sample from this distribution of prevalence ratios. The average prevalence ratio from a random-effects meta-analyses has a greater standard error than the common fixed-effect model, since both within and between-study variation is accommodated (Hedges and Vevea 1998; Borenstein *et al.* 2010).

3.5 Methods

The systematic review search was conducted on the 6th of March 2015. A protocol was developed and approved prior to running searches. MOOSE guidelines were adopted to structure the search, review, and reporting of the systematic review (Stroup *et al.* 2000). The systematic review was registered with Prospero (registration number: CRD42015029997).

Scoping of the identified studies highlighted that there were two distinct populations sampled, i) population samples and ii) clinic samples; the latter being more likely to have a higher reported prevalence and be less generalizable to the wider population. This limited comparability between studies. With regards to data by weight status, some studies provided prevalence for the overweight and obese groups combined, whereas others did not report prevalence for the overweight and/or healthy weight groups. This increased heterogeneity between the studies and limited the understanding of the association between weight status and co-morbidity. Therefore it was decided to apply an additional level of eligibility criteria to increase homogeneity between the studies. As such the methods and results sections are separated in to stages 1 and 2.

3.5.1 Stage 1 Methods

Stage 1 Study Eligibility

Stage 1 Inclusion Criteria

Population: The sample included children aged 5-18 years (inclusive) to cover the varying ages seen by weight management services.

Study Design: Cross-sectional, cohort, case-control and case series study designs were eligible for the systematic review.

Data: Weight status had to be measured by the research/clinical team or obtained from clinical records to categorise the child as healthy weight (BMI <85th percentile), overweight (BMI ≥85th and <95th percentile), or obese (BMI ≥95th percentile).

Point prevalence provided as n/N, or sufficient information provided to calculate n/N.

Language/Country Restriction: No language or country restrictions were applied to the initial search. For non-English language articles, attempts were made to source English translations. If English translations were not available, the articles were not included in the systematic review, but were retained in a separate EndNote folder.

Stage 1 Exclusion Criteria

Population: Studies with the following populations were not included in the systematic review:

1. Studies related exclusively to the non-obese/non-overweight population, e.g. BMI <85th percentile.
2. Studies related exclusively to adults (aged over 18).
3. Studies related exclusively to infants (aged under 5).
4. Studies where obesity is a symptom of an underlying illness, e.g. Prader-Willi syndrome, Cushing Syndrome, or Hypothyroidism, or a side effect from medication, e.g. anti-depressants, anti-psychotics, anti-hypertensives, and steroids.

Study Design: RCTs, case reports, qualitative studies, editorials, commentaries, letters to editors, author replies, study protocols, and non-human studies were excluded as they would not provide prevalence data or the studies are conducted in a highly selected population. Systematic reviews were excluded. Conference abstracts and presentations were also excluded as the study quality could not be assessed and there can be discordance between results presented at a conference and subsequent published results (Martin *et al.* 2005; Tam and Hotte 2008).

Data: Studies that utilised self-reported weight were excluded, since these may result in biased prevalence estimates in different weight categories (Wang, Patterson and Hills 2002; Enes *et al.* 2009).

Search Strategy and Data Sources

Scoping searches were initially conducted to balance the search's sensitivity and specificity in identifying relevant articles. Search results were assessed for suitability with supervisors, and discussion led to refinements in search keywords and mesh terms.

Relevant articles were identified through a systematic search of MEDLINE, EMBASE, PsycINFO, and Web of Science since the date of inception. The search strategy for MEDLINE is available in Appendix 1. For each database search MESH and free-text terms were grouped into four categories (Figure 4). The Boolean operator "OR" joined the terms and subject headings within each category, and the operator "AND" combined the four categories. Search terms and keywords were altered as per each database's requirements. The search strategy included examples of specific co-morbidities associated with obesity, such as depression, hyperglycaemia, and sleep disorders (e.g. sleep apnoea). This was performed to increase the specificity of the results, but the search was not limited to these co-morbidities.
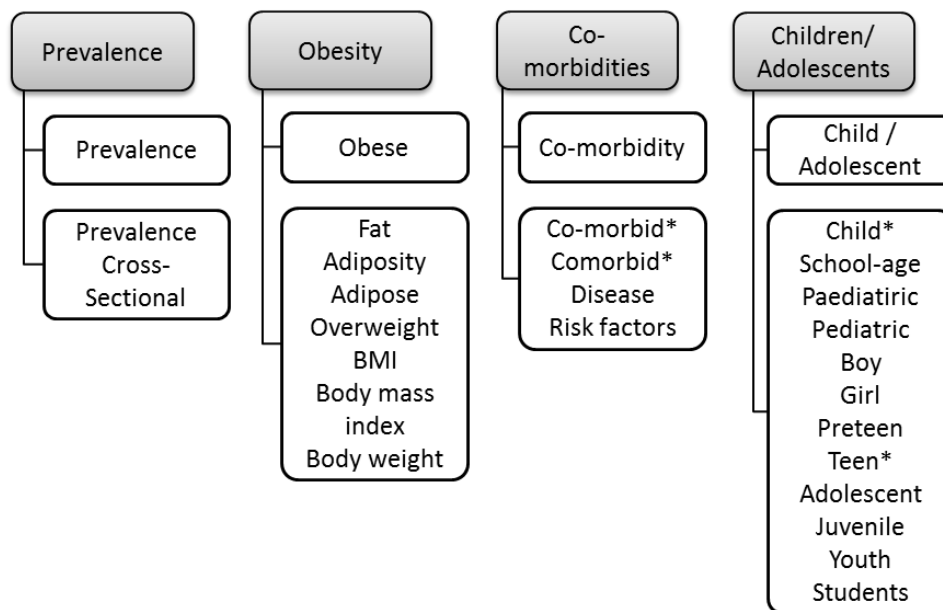
Figure 4: Subject groups for the search strategy

Title Abstract Review and Full-Text Review

After conducting the searches, amalgamating the results, and removing duplicates, title and abstracts were screened for eligibility. One of the supervisors (MB) acted as an independent second reviewer on 603 studies to assess agreement. Agreement was assessed over three rounds using Kappa values (Landis and Koch 1977; Viera and Garrett 2005). After each round, disagreements were discussed with all supervisors to ensure that all potentially eligible articles were taken forward. Review of the first round of articles indicated that a high proportion of articles were coded as "no", which may have artificially inflated the level of agreement. For the subsequent two rounds an equal number of articles coded as Yes/Maybe/No underwent a second review. Once all articles had been screened and the full-text obtained, agreement for full-text review was assessed over one round with one supervisor (MB) acting as an independent reviewer.

Data Extraction

Upon completion of full-text reviews, pertinent data from the articles were extracted. A modified version of The Joanna Briggs Institute (2014) form was utilised to capture details of the co-morbidities and prevalence by weight status (healthy weight, overweight, obese). The data extraction form was pilot tested, and additionally assessed by two supervisors (MB and SC) on six randomly selected articles to ensure the form was easy to use and appropriately designed to capture all

the relevant data. Key components of the data extraction form were study information, sample demographics, weight status categorisation, and prevalence of the co-morbidity by weight status. The full data extraction form is available in Appendix 2.

Quality Appraisal

Quality appraisal of the included studies was assessed using the Critical Appraisal Checklist (Munn *et al.* 2014) (Box 1). The checklist was selected as it can be utilised across multiple study designs and addressed more of the criteria reported by Sanderson, Tatt and Higgins (2007) (see section 3.4.6).

Box 2: The Joanna Briggs Critical Appraisal Checklist

Items in the Critical Appraisal Checklist
1. Was the sample representative of the target population?
2. Were study participants recruited in an appropriate way?
3. Was the sample size adequate?
4. Were the study subjects and the setting described in detail?
5. Was the data analysis conducted with sufficient coverage of the identified sample?
6. Were objective, standard criteria used for the measurement of the condition?
7. Was the condition measured reliably?
8. Was there appropriate statistical analysis?
9. Are all important confounding factors/subgroups/differences identified and accounted for?
10. Were subpopulations identified using objective criteria?

Stage 1 Data Analysis

Upon completion of the data extraction and quality appraisal, data from eligible studies were summarised in a flow diagram to report the number of studies identified, reviewed, and included in the systematic review. For the included articles, the year of publication, number of participants, and age range of participants were reported. Co-morbidities were separated into physical and psychological co-morbidities, and physical co-morbidities were further separated by the bodily system they impacted.

3.5.2 Stage 2 Methods

Due to heterogeneity between the identified studies in terms of the population sampled and the availability of prevalence data for all three weight statuses, additional eligibility criteria were applied, in addition those in Section 3.5.1 (page

39). The additional criteria (see below) were applied to increase homogeneity between the studies and enable more justifiable comparisons between studies and population prevalence ratio estimates.

Stage 2 Study Eligibility

Stage 2 Inclusion Criteria

Population: The sample included children aged 5-18 (inclusive) recruited from a general population.

Data: Prevalence was provided for all three weight categories (healthy weight, overweight, and obese) to enable estimation of prevalence ratio of the overweight group and obese group relative to the healthy weight group.

Stage 2 Exclusion Criteria

Population: Participants recruited from narrowly defined samples, e.g. clinical samples.

Data: Studies which did not report data for all three weight statuses, or combined the prevalence for the overweight and obese groups.

Stage 2 Data Analysis

Descriptive Analysis

Articles were grouped by co-morbidity (e.g. hyperglycaemia) and then by co-morbidity indicator/definition (e.g. fasting plasma glucose) and by cut-off (e.g. 100mg/dL). The extracted data were grouped by weight status (healthy weight: <85th percentile; overweight: ≥85th and <95th percentile; obese: ≥95th percentile) to enable comparison between subgroups (WHO 2014a).

Funnel Plots

Funnel plots were provided to indicate the prevalence of the co-morbidity by weight status for each study (with weight statuses differentiated by shape and colour) (Figure 5). The x-axis represented the sample size and the y-axis the prevalence of the co-morbidity per 1000 of the population (Spiegelhalter 2005; Fisher *et al.* 2012). In Figure 5, point A shows the prevalence of a co-morbidity in a fictitious healthy weight group of 301 out of 1000 (30.1%) for one study. Adjusting the prevalence to per 1000 enabled comparison between studies to identify variation between study

prevalence estimates, and visually assess whether the overweight and obese groups had substantially different prevalence relative to the healthy weight group; as well as variation in prevalence within weight status groups.
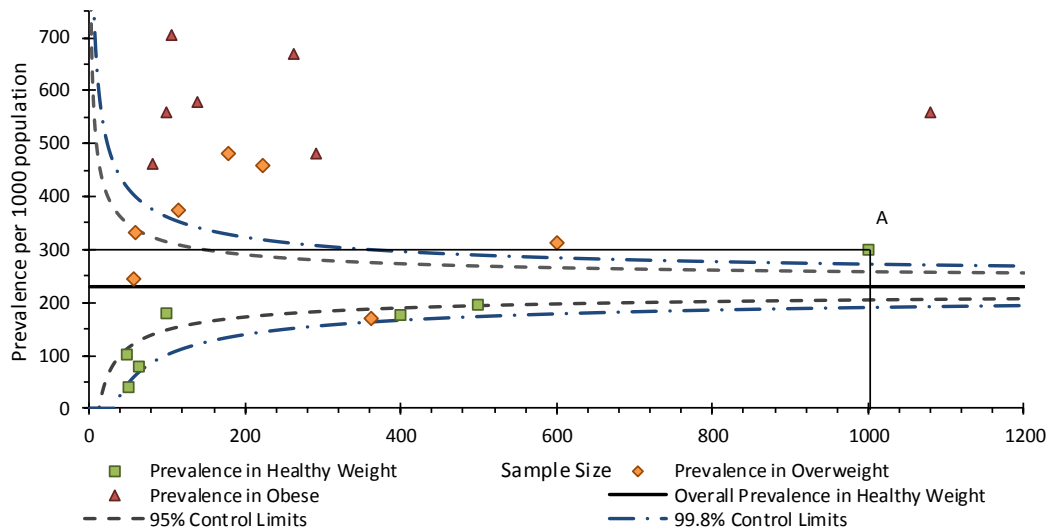
Figure 5: Example funnel plot using fictitious data

To assess between-group differences, the overall prevalence for the healthy weight data points was calculated (solid horizontal line at 231). Secondly the upper and lower 95% and 99.8% control limits were calculated for the healthy weight data points (the dotted curved lines above and below the overall prevalence (Figure 5)). Control limits are funnel shaped due to the smaller sampling variability in prevalence estimates for larger samples (Dover and Schopflocher 2011). These control limits had two functions:

1. For the healthy weight group, if the study participants are from a similar population, and vary only due to sampling error, then 95% of prevalence estimates should fall within the 95% control limits (Dover and Schopflocher 2011). If fewer points were contained within these limits, it would indicate between-study heterogeneity. In Figure 5 only one of the seven data points for the healthy weight group lies within the 95% control limits suggesting that the observed variation within the healthy weight group is likely due to study specific differences.

2. If the prevalence of the comorbidity differs by weight status (i.e. increased prevalence for the higher weight status groups), the estimates for the obese group should fall outside the 99.8% control limits, and data points for the overweight group should lie between the obese and healthy weight data group. In Figure 5, prevalence estimates for the obese groups are higher

than the healthy weight groups. In this example, children with obesity have higher co-morbidity prevalence than those of a healthy weight.

Funnel plots for the prevalence ratios have not been provided, since forest plots of these estimates are presented for the co-morbidities.

Meta-Analyses

A meta-analysis was performed to estimate the population prevalence ratio for the overweight and obese groups relative to the healthy weight group. Since the studies included in the meta-analyses were from diverse populations and study designs, a random-effects model on a logarithmic scale was adopted, and data were analysed using MedCalc version 17.9 (MedCalc Software 2017).

Prevalence Ratio

For each co-morbidity eligible for meta-analysis, the total number of participants and the number with a positive test result in each weight status were entered into MedCalc (MedCalc Software 2017). The average prevalence ratio was then calculated using a random effects model based on the DerSimonian and Laird method (Deeks and Higgins 2010). The prevalence ratio was calculated for the overweight and obese groups relative to the healthy weight group. For the prevalence ratio the 95% confidence intervals were calculated on a logarithmic scale $(\hat{\theta} \pm SE\{\hat{\theta}\}\Phi(1 - \frac{\alpha}{2}))$, where $\hat{\theta}$ is the estimated log prevalence ratio and Φ is the standard normal deviate (Deeks and Higgins 2010). For the co-morbidity the following are provided:

- Table summarising the prevalence ratio
- Forest plot providing the individual study prevalence ratio (95% confidence intervals) and summary effect
- The Q statistic and I^2 statistic

Forest Plot

Forest plots were used to present the prevalence ratio and 95% confidence interval for each study, along with the average prevalence ratio (Fletcher and Fletcher 2005). This enabled comparison between studies and assessment of the overall effect.

Q and I² Statistic

The Q statistic summarises the variation between study results and can be compared with a Chi-squared statistic to test the null hypothesis that all studies are evaluating the same effect, i.e. absence of between-study heterogeneity (Huedo-Medina *et al.* 2006; Borenstein *et al.* 2009). However the statistic is known to have low power for detecting between-study heterogeneity, particularly if the number of studies included in a meta-analyses is small, and is over-sensitive when a large number of studies is included (Higgins *et al.* 2003; Huedo-Medina *et al.* 2006).

To overcome the shortcomings of the Q Statistic the I^2 statistic was developed; I^2 measures the proportion of the total variation in observed effects that is due to between-study variation rather than due to sampling error (Borenstein et al. 2009; Borenstein et al. 2017). Therefore it provides the proportion of total variation that would remain if sampling error could be accounted for/removed (Borenstein et al. 2009; Borenstein et al. 2017). I^2 ranges from 0-100% with higher values indicating that most of the observed variance is due to between-study heterogeneity, and investigation of the reasons for heterogeneity may be warranted (Borenstein et al. 2009). Higgins et al. (2003) tentatively assigned adjectives of low, moderate, and high to I^2 values of 25%, 50%, and 75%, respectively.

Potential Reasons for Heterogeneity

The third stage of the analysis considered the potential reasons for heterogeneity that might explain the observed differences in prevalence between studies. Reasons may include: participant or sampling differences, country of study, sample size, gender, setting and/or study design. The focus was directed by information reported in the articles, e.g. if prevalence was provided by gender or age, or if a number of the studies were conducted in the same country.

3.6 Results

3.6.1 Stage 1 Results

Study Characteristics

The database searches identified 10,391 articles, after removal of duplicates 8,173 were eligible for title and abstract review. Kappa agreement at the title and abstract review stage indicated substantial agreement for all three rounds, 0.84 (95% CI, 0.75-0.93, p<0.001), 0.92 (95% CI, 0.87-0.98, p<0.001) and 0.85 (95% CI, 0.78-

0.93, p<0.001). Of the 8,173 articles, 769 were obtained for full-text review (Figure 6). Of these, 162 fulfilled eligibility criteria for data extraction. These 162 studies were published between 1996 and 2014, and included a total of 1,801,388 participants (median 484; range 33 - 825,964), including children ranging in age from 1 to 22 years.

N = 10391	Studies identified through database searches	→	N = 2218 Duplicates
N = 8173	Title and Abstract Screening	→	N = 7404 Not eligible
N = 769	Full-text Review	→	N = 607 Not eligible 186 – No Prevalence Data 99 – Non-English Articles 89 – Abstract/Posters 75 – No Access/Cannot find 16 – Duplicates 11 – Intervention Study 13 – Reviews 8 – Self-reported Weight 53 – No prevalence by weight status 57 - Other
N = 162	Data Extraction and Quality Assessment		

Figure 6: Stage 1: systematic review flowchart

The systematic review identified 22 physical and four psychological co-morbidities; physical co-morbidities were further grouped by the bodily system they impacted (Table 4). For some of the co-morbidities multiple indicators/definitions were identified; for example six indicators of hyperglycaemia were reported (fasting plasma glucose, oral glucose tolerance test, Hb1Ac, HOMA-IR, insulin resistance, and the presence of acanthosis nigricans).

Table 4: List of co-morbidities identified via the systematic review

Physical Co-morbidities	
Endocrine	Hyperglycaemia (fasting plasma glucose, oral glucose tolerance test, Hb1Ac, HOMA-IR, insulin resistance, and acanthosis nigricans)
Gastrointestinal	Non-alcoholic fatty liver disease Non-alcoholic steatohepatitis Gallstones
Pulmonary	Obstructive Sleep Apnoea Asthma

Physical Co-morbidities	
	Exercise induced wheeze/cough
Cardiovascular	Carotid-Intima media thickness
	Metabolic syndrome
	High blood pressure
	High C-reactive protein
	Dyslipidaemia (total cholesterol; LDL cholesterol; HDL cholesterol; triglycerides)
Musculoskeletal	Bone Fractures
	Joint Pain
	Flat foot
Other	Vitamin D deficiency
	Enuresis
	Iron deficiency
	Elevated uric acid
	Traumatic dental injuries
Psychological Co-morbidities	
Attention Deficit Hyperactive Disorder	
Anxiety	
Depression	
Low Self-Esteem	

3.6.2 Stage 2 Results

The following section provides an overview of the results, beginning with study characteristics and study quality for the studies eligible for Stage 2. In Stage 2 additional eligibility criteria were applied (see section 3.5.2) to focus on studies conducted on general populations with data for all three weight status groups (healthy weight, overweight, and obese). The section goes on to summarise the co-morbidities excluded from analysis, based on the additional eligibility criteria applied in Stage 2. After which descriptive analysis of the included co-morbidities is provided. Results of the meta-analysis are provided in Chapter 4.

Study Characteristics

Of the 162 studies identified in Stage 1, 90 referred to a general population, of which 45 provided data for all three weight statuses (Figure 7). Whilst data were extracted

from all 162 papers meeting the stage 1 eligibility criteria, the results presented in this thesis focus on the 45 papers providing data for all three weight statuses in children recruited from a general population. Summary of the 45 studies is provided in Table 5. This enabled a more focused account in papers that are most likely to be relevant to the aim of the systematic review and thesis overall.
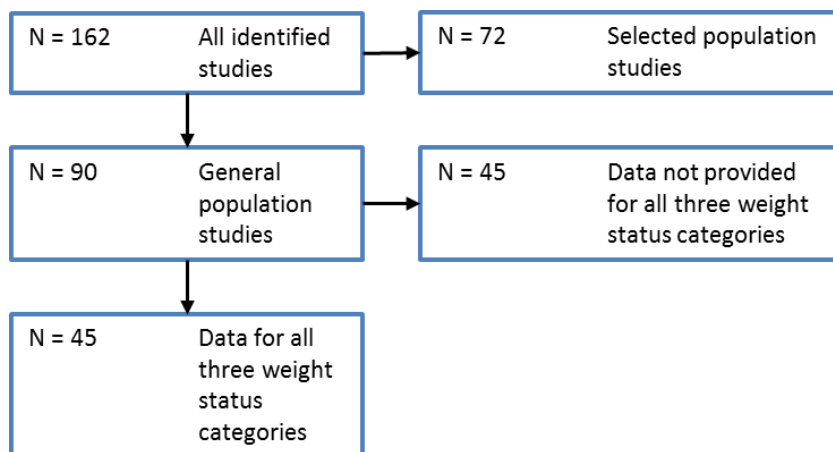
Figure 7: Eligible studies based on the additional eligibility criteria

Table 5: Summary of studies included in meta-analyses

Number of studies	45
Number of participants	1,597,733
Studies by continent*:	
Asia	22
Australia	3
Europe	12
North America	5
South America	7
Year of data collection (range)	1998-2014
Study Design:	
Cross Sectional	43
Cohort	2

*some studies collected data across multiple continents hence the total is 49

Study Quality

Quality appraisal results for the 45 studies included in stage 2, using the Critical Appraisal Checklist, are shown in Table 6. Two studies fulfilled all the criteria and were classed as high quality studies, whereas seven studies met only four of the 10 criteria. The criterion studies were least likely to address was "Was the condition measured reliably?", with 29 studies rated as "Unclear". Another area where studies did not provide sufficient information related to the adequacy of the sample size.

Although details of the final sample size and dropouts were provided, 26 studies did not provide details of an a priori sample size calculation. A quality categorisation was defined using Gonzalez-Serrano et al.'s (2016) criteria, <6 – low quality; ≥6 – high quality. The categorisation was used to assess the association between quality score and reported prevalence.

Table 6: Critical Appraisal Checklist responses for the 45 studies

Author (Year)	1. Was the sample representative of the target population?	2. Were study participants recruited in an appropriate way?	3. Was the sample size adequate?	4. Were the study subjects and the setting described in detail?	5. Was the data analysis conducted with sufficient coverage of the identified sample?	6. Were objective, standard criteria used for the measurement of the condition?	7. Was the condition measured reliably?	8. Was there appropriate statistical analysis?	9. Are all important confounding factors/subgroups/differences identified and accounted for?	10. Were subpopulations identified using objective criteria?	Quality Score
Adibi et al. (2009)	Yes	Yes	Unclear	No	Yes	Yes	Yes	Yes	No	N/A	6
Alavian et al. (2009)	Yes	Unclear	Unclear	No	Yes	Yes	Unclear	Yes	No	N/A	4
Al-Bajjali and Rajab (2014)	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	9
Ayonrinde et al. (2011)	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	N/A	7
Bar Dayan et al. (2005)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	9
Booth et al. (2008)	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	No	Yes	8
Caserta et al. (2010)	Unclear	Unclear	Unclear	No	No	Yes	Unclear	Yes	Yes	Yes	4
Chen, Chung and Wang (2009)	No	Unclear	No	No	Yes	Yes	Yes	Yes	No	N/A	4
Chen et al. (2012)	Yes	Yes	Yes	No	Yes	Yes	Unclear	Yes	No	N/A	6
Chu and Pan (2007)	Yes	Unclear	Yes	No	Yes	Yes	Unclear	Yes	No	Yes	6
Davis et al. (2005)	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Del-Rio-Navarro et al. (2008)	Yes	Unclear	Unclear	No	Unclear	Yes	Unclear	Yes	No	Yes	4

Author (Year)	1. Was the sample representative of the target population?	2. Were study participants recruited in an appropriate way?	3. Was the sample size adequate?	4. Were the study subjects and the setting described in detail?	5. Was the data analysis conducted with sufficient coverage of the identified sample?	6. Were objective, standard criteria used for the measurement of the condition?	7. Was the condition measured reliably?	8. Was there appropriate statistical analysis?	9. Are all important confounding factors/subgroups/differences identified and accounted for?	10. Were subpopulations identified using objective criteria?	Quality Score
Duzova et al. (2013)	Yes	Unclear	Unclear	No	Yes	Yes	Unclear	Yes	No	Yes	5
Dyson et al. (2014)	Yes	Unclear	Yes	No	Yes	Yes	Unclear	Yes	No	Yes	6
Elizondo-Montemayor et al. (2014)	Yes	Unclear	No	No	No	Yes	Unclear	Yes	Yes	Yes	5
Franklin et al. (2006)	Yes	Yes	Unclear	No	Yes	No	No	Yes	No	Unclear	4
Genovesi et al. (2010)	Yes	Unclear	Yes	No	Yes	Yes	Yes	Yes	No	Yes	7
Gong et al. (2013)	Yes	Unclear	Unclear	No	No	Yes	Yes	Yes	No	Yes	5
Guo et al. (2012)	Yes	Unclear	Yes	No	Yes	Yes	Unclear	Yes	No	Yes	6
Kajbaf, Asar and Alipoor (2011)	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	N/A	6
Khader et al. (2011)	Yes	Yes	No	Yes	Unclear	Yes	Unclear	Yes	Yes	Yes	7
Koebnick et al. (2012a)	Yes	N/A	Yes	No	Yes	Unclear	Unclear	Yes	No	Yes	5
Kwon et al. (2006)	Yes	Yes	Unclear	No	Yes	Yes	Unclear	Yes	Yes	Yes	7

Author (Year)	1. Was the sample representative of the target population?	2. Were study participants recruited in an appropriate way?	3. Was the sample size adequate?	4. Were the study subjects and the setting described in detail?	5. Was the data analysis conducted with sufficient coverage of the identified sample?	6. Were objective, standard criteria used for the measurement of the condition?	7. Was the condition measured reliably?	8. Was there appropriate statistical analysis?	9. Are all important confounding factors/subgroups/differences identified and accounted for?	10. Were subpopulations identified using objective criteria?	Quality Score
Lawlor et al. (2014)	Yes	Unclear	Unclear	Yes	No	Yes	Yes	Yes	No	Yes	6
Li et al. (2008)	Yes	Yes	Yes	Yes	No	Yes	Unclear	Yes	No	Yes	7
Manios et al. (2008)	Unclear	Yes	Unclear	Unclear	Yes	Yes	Unclear	Yes	No	Yes	5
Mehairi et al. (2013)	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	No	Yes	8
Messiah et al. (2008)	Yes	Unclear	Yes	No	Yes	Yes	Unclear	Yes	Yes	Yes	7
Moura et al. (2004)	Yes	Yes	Yes	No	Yes	Yes	Unclear	Yes	No	N/A	6
Noonan et al. (2010)	Yes	Yes	Unclear	No	Yes	Yes	No	Yes	Yes	Yes	7
Nur et al. (2008)	Yes	Unclear	Unclear	No	Yes	Yes	Unclear	Yes	No	Yes	5
Papoutsakis et al. (2012)	Unclear	Unclear	Unclear	No	Yes	Yes	Unclear	Yes	No	Yes	5
Pecin et al. (2013)	Yes	Unclear	Unclear	No	Yes	Yes	Unclear	Yes	No	N/A	4
Pereira et al. (2009)	Yes	Yes	Unclear	No	Yes	Yes	Unclear	Yes	Yes	Yes	7
Rafraf, Pourghassem Gargari	Yes	Yes	Unclear	No	Yes	Yes	Unclear	Yes	No	Yes	6

Author (Year)	1. Was the sample representative of the target population?	2. Were study participants recruited in an appropriate way?	3. Was the sample size adequate?	4. Were the study subjects and the setting described in detail?	5. Was the data analysis conducted with sufficient coverage of the identified sample?	6. Were objective, standard criteria used for the measurement of the condition?	7. Was the condition measured reliably?	8. Was there appropriate statistical analysis?	9. Are all important confounding factors/subgroups/differences identified and accounted for?	10. Were subpopulations identified using objective criteria?	Quality Score
and Safaiyan (2010)											
Ramos and Barros (2005)	Yes	Yes	Unclear	No	Yes	Yes	Unclear	Yes	No	Yes	6
Ribeiro-Silva et al. (2013)	Unclear	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes	6
Seki, Matsuo and Faria Carrilho (2008)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10
Tandon et al. (2013)	Yes	Yes	Unclear	Yes	Yes	Yes	Unclear	Yes	No	Yes	7
Tenenbaum et al. (2013)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10
Wang et al. (2013)	Yes	Yes	Unclear	Yes	Yes	Yes	Unclear	Yes	No	Yes	7
Wiegand et al. (2010)	Yes	Yes	Yes	No	Yes	Yes	Unclear	Yes	No	N/A	6
Xu and Ji (2008)	Yes	Yes	Unclear	Yes	Yes	Yes	Unclear	Yes	No	Yes	7
Xu et al. (2012)	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	9
Zakeri et al. (2012)	Yes	Unclear	Yes	Yes	Yes	Unclear	No	Yes	No	Yes	6

Co-morbidities excluded from Stage 2

Restricting analysis to studies conducted in general populations and with data for all three weight statuses meant nine co-morbidities/indicators were excluded from further analysis. For hyperglycaemia and metabolic syndrome, 4 indicators and 1 definition, respectively, were also excluded (Table 7). This meant that the assessment of the prevalence of co-morbidities potentially associated with increased weight may not be comprehensive. Additional research in general populations, stratified by weight category is required to obtain an understanding of the impact of increased weight on the prevalence of these co-morbidities.

Table 7: Co-morbidities/indicators excluded from the systematic review

Comorbidity	Indicator/Criteria	Reason for Exclusion
Hyperglycaemia	Oral Glucose Tolerance Test (≥140mg/dL and 140-200mg/dL)	No data for some weight categories
	Fasting Insulin	No data for some weight statuses
	HbA1c	No data for some weight categories
	Acanthosis Nigricans	No general population studies
Metabolic Syndrome	WHO, 1998	No data for some weight statuses
Non-alcoholic steatohepatitis	Ultrasound	No general population studies
Obstructive Sleep Apnoea	Questionnaire	No data for some weight statuses
Bone Fractures	Case note review	No general population studies
Joint pain	Questionnaire	No data for some weight statuses
Vitamin D Deficiency	Blood test	No data for some weight statuses
Dental Caries	Physical examination	No data for some weight statuses
Enuresis	Questionnaire	No general population studies
Iron Deficiency	Blood test	No general population studies
ADHD	Questionnaire	No general population studies

Co-morbidities included in Stage 2

Descriptive Analysis

Nineteen co-morbidities/indicators remained in the analysis for Stage 2. The reported results from the studies were used to estimate the prevalence per 1000 individuals in each group (Table 8). This enabled comparison within and between co-morbidities by weight category. For some of the co-morbidities it was observed that there was a low prevalence in the healthy weight group, and therefore the increase observed in the overweight and obese groups, which may be statistically significant, may not be sufficiently important clinically to justify screening of children in weight management clinics. For example, when considering fasting plasma glucose, 66/1000 children of a healthy weight would exceed the threshold of 100mg/dL (based on the overall reported prevalence in the identified studies), whereas in the overweight and obese groups 93/1000 would reach the cut-off. Thus, although the prevalence ratio is statistically significant, from a clinical perspective identifying an additional 27 children per 1000 may not be sufficient to justify the implementation of a screening programme, given the potential cost and time implications of developing and running a screening programme (Watmough and Kumar 1994). On the other hand, an increase in prevalence from 31 to 124 per 1000 in high blood pressure (>90th percentile) as weight category increases, together with a simple screening test, may be sufficient to warrant screening, provided that an effective intervention strategy is available, and the test is deemed suitable and appropriate for a weight management setting.

Table 8: Estimated prevalence per 1000 children in each weight status

Co-morbidity	Indicator / Measure	Prevalence in Healthy Weight	Prevalence in Overweight	Prevalence in Obese
Hyperglycae-mia	Fasting Plasma Glucose	66 (62 to 71) out of 1000	93 (80 to 106) out of 1000	93 (80 to 113) out of 1000
	2 Hour Plasma Glucose	0 (0 to 0) out of 1000	0 (0 to 1) out of 1000	3 (1 to 8) out of 1000
	HOMA-IR	22 (6 to 68) out of 1000	16 (1 to 128) out of 1000	108 (25 to 462) out of 1000
Dyslipidaem-ia	Total Cholesterol (≥200mg/dL)	148 (134 to 163) out of 1000	178 (118 to 267) out of 1000	267 (148 to 459) out of 1000
	Low HDL Cholesterol (<40mg/dL)	81 (71 to 84) out of 1000	163 (122 to 212) out of 1000	236 (179 to 334) out of 1000
	High LDL Cholesterol (≥130mg/dL)	39 (25 to 58) out of 1000	77 (35 to 166) out of 1000	127 (62 to 262) out of 1000
	High Triglycerides (≥150mg/dL)	46 (40 to 52) out of 1000	115 (87 to 151) out of 1000	193 (147 to 257) out of 1000
High Blood Pressure	≥90th percentile	31 (30 to 32) out of 1000	65 (53 to 80) out of 1000	124 (86 to 176) out of 1000
Metabolic Syndrome	Cook et al. 2003	10 (6 to 15) out of 1000	24 (8 to 73) out of 1000	202 (11 to 984) out of 1000
	IDF 2007	7 (5 to 9) out of 1000	90 (13 to 612) out of 1000	371 (67 to 561) out of 1000
	NCEP ATP III 2001	5 (4 to 8) out of 1000	85 (40 to 182) out of 1000	312 (149 to 653) out of 1000
	de Ferranti et al. 2004	23 (18 to 30) out of 1000	234 (150 to 365) out of 1000	352 (227 to 546) out of 1000
NAFLD	Ultrasound	26 (21 to 32) out of 1000	157 (85 to 288) out of 1000	672 (242 to 938) out of 1000
	Elevated ALT	0 (0 to 100) out of 1000	7 (0 to 124) out of 1000	20 (1 to 326) out of 1000
	Elevated AST	43 (25 to 70) out of 1000	109 (51 to 232) out of 1000	297 (103 to 855) out of 1000
Asthma	Questionnaire	109 (96 to 123) out of 1000	174 (130 to 207) out	185 (152 to 228) out of 1000

Co-morbidity	Indicator / Measure	Prevalence in Healthy Weight	Prevalence in Overweight	Prevalence in Obese
			of1000	
Exercise Induced Wheeze/ Cough	Questionnaire	5 (2 to 14) out of 1000	369 (134 to 601) out of 1000	687 (255 to 943) out of 1000
Flat foot	Physical examination	135 (134 to 136) out of 1000	175 (175 to 189) out of 1000	243 (189 to 310) out of 1000
Anxiety	Questionnaire	131 (123 to 140) out of 1000	124 (102 to 149) out of 1000	97 (74 to 126) out of 1000
Depression	Questionnaire	301 (290 to 313) out of 1000	302 (271 to 337) out of 1000	309 (270 to 352) out of 1000
Self-Esteem	Questionnaire	45 (37 to 56) out of 1000	82 (55 to 121) out of 1000	146 (100 to 213) out of 1000
Carotid-Intima Media Thickness	Ultrasound	242 (196 to 294) out of 1000	369 (280 to 485) out of 1000	427 (311 to 587) out of 1000
Elevated Uric Acid	Blood Test	145 (129 to 162) out of 1000	283 (231 to 345) out of 1000	436 (367 to 519) out of 1000
Gallstones	Ultrasound	1 (1 to 1) out of 1000	2 (1 to 2) out of 1000	3 (3 to 4) out of 1000
High C-Reactive Protein	Blood Test	26 (12 to 53) out of 1000	76 (71 to 337) out of 1000	323 (149 to 699) out of 1000
Traumatic Dental Injuries	Physical Examination	172 (146 to 202) out of 1000	185 (125 to 272) out of 1000	126 (72 to 219) out of 1000

Of 19 co-morbidities/indicators only 10 were eligible for the meta-analyses; nine were not eligible as only one general population study was identified. Twelve of the studies reported the prevalence of hypertension (≥95th percentile) and nine reported hyperglycaemia (fasting plasma glucose ≥100mg/dL). Eight studies were conducted in China, five in America and only one in the UK. Review of the studies indicated

that overall children who were obese did have a higher prevalence for many of the co-morbidities compared with those of a healthy weight. For children who were overweight, the increase in prevalence compared with those of a healthy weight was not as substantial. There were considerable differences between the studies which may influence the prevalence of some co-morbidities, such as differences in sample sizes, ranging from 211 to 825,964 (Davis *et al.* 2005; Tenenbaum *et al.* 2013). Such differences impact the precision and accuracy of population estimates of prevalence ratios. For some of the co-morbidities/co-morbidity indicators, however, only a small number of studies were eligible for the meta-analyses, thus caution must be applied when interpreting the results.

3.7 Chapter Summary

This chapter provided a detailed account of the methods undertaken to conduct the systematic review and meta-analyses, and a brief overview of the systematic review results. The results suggest children with obesity have a higher prevalence of all co-morbidities/indicators included in the meta-analyses, though there is considerable variation between co-morbidities. The results for those who are overweight also suggested a marked increase in prevalence relative to the healthy weight group. Furthermore, considerable between-study heterogeneity was observed, related to factors such as country of study, sample size, and measurement method. The following chapter provides details of the 19 co-morbidities/indicators included in the meta-analyses, and discusses the factors that may explain some of the between-study variation in reported prevalence. This is followed by a discussion of the results and implications for the programme of work.

Chapter 4: Results of the Systematic Review and Meta-Analyses

4.1 Introduction

Chapter 3 presented the methods and an overview of the results of a systematic review conducted to ascertain the prevalence of obesity-associated co-morbidities in children who are overweight/obese relative to those of a healthy weight. This chapter focuses on the 19 co-morbidities from the 45 studies eligible for Stage 2 of the systematic review and meta-analysis. Prevalence ratios for the co-morbidities are presented and factors which may have influenced the prevalence ratios are discussed (Section 4.2). This is followed by a discussion of the findings, their implications and the strengths and limitations of the systematic review and meta-analyses (Section 4.3) and conclusion (Section 4.4).

4.2 Results

The following section provides the results of the systematic review and meta-analysis for each individual co-morbidity/indicator.

4.2.1 Hyperglycaemia

Hyperglycaemia is defined as an elevation of blood glucose, which is a biomarker for diabetes. Over prolonged periods high glucose can cause damage to internal organs (Diabetes UK 2018a). Eleven studies were included in the systematic review using three different screening methods (Figure 8).
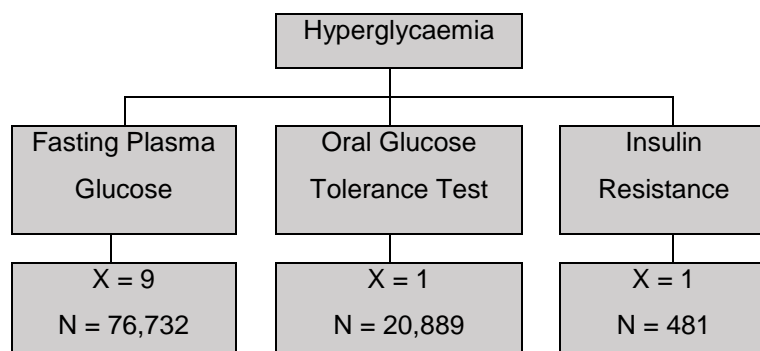
Figure 8: Overview of Hyperglycaemia and screening methods (X: number of studies; N: number of participants)

Fasting Plasma Glucose

Nine studies that measured the prevalence of fasting plasma glucose ≥100mg/dL (5.5mmol/l) were identified (Table 9). The funnel plot displayed considerable heterogeneity between the studies, with no clear clustering of studies by weight status (Figure 9). Point estimates for prevalence ranged from 0.5% to 26% in the healthy weight group, suggesting that study methodology varied widely. Seven studies reported fairly low prevalence of fasting plasma glucose, whereas Chu and Pan reported prevalence that was substantially higher for all weight statuses. There was also considerable overlap between weight statuses, suggesting that the relationship between weight status and fasting plasma glucose is complex. Study quality scores ranged from four to 10, with five studies being classified as high quality (Davis *et al.* 2005; Chu and Pan 2007; Seki, Matsuo and Faria Carrilho 2008; Khader *et al.* 2011; Wang *et al.* 2013). Study quality was not related to reported prevalence as the two highest scoring studies reported markedly different prevalence for the weight groups (Davis *et al.* 2005; Seki, Matsuo and Faria Carrilho 2008).

Overall the studies suggested that prevalence of fasting plasma glucose increased with weight status (Table 9); however the relationship between weight status and prevalence was not consistent. Five of the nine studies reported that prevalence increased in line with weight status, whereas three reported a lower prevalence in the obese group compared to the overweight group, e.g. Chu and Pan (2007) reported a prevalence of 36% in the overweight group and 29% in the obese group. However in all cases the prevalence in the obese group was higher than the healthy weight group.

Table 9: Summary of studies reporting prevalence data for Hyperglycaemia (Fasting Plasma Glucose ≥100mg/dL)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean Age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Caserta et al. (2010)	Italy	575	11-13 (NR)	Education	0.6 (1/164)	1.7 (3/179)	3.7 (3/82)	4
Chu and Pan (2007)	Taiwan	2405	6-12 (NR)	Education	26.0 (455/1753)	35.5 (128/361)	29.2 (85/291)	6
Davis et al. (2005)	America	211	7-18 (NR)	Education	12.0 (10/83)	15.2 (5/33)	18.2 (8/44)	8
Del-Rio-Navarro et al. (2008)	Mexico	1819	6-13 (9.8)	Education	1.3 (12/923)	4.4 (18/411)	3.4 (15/438)	4
Gong et al. (2013)	China	538	9-15 (12.0)	Education	3.5 (10/283)	4.3 (5/115)	7.9 (11/140)	5
Khader et al. (2011)	Jordan	1034	7-18 (NR)	Community	4.5 (38/837)	3.6 (4/111)	7.0 (6/86)	7
Seki, Matsuo and Faria Carrilho (2008)	Brazil	2170	6-16 (11.3)	Education	0.5 (9/1755)	1.0 (3/299)	0.9 (1/116)	10
Wang et al. (2013)	China	3373	6-18 (12.0)	Education	9.8 (151/1541)	13.5 (86/637)	13.7 (164/1195)	7
Xu et al. (2012)	China	8764	7-11 (8.6)	Education	2.5 (120/4813)	3.3 (22/675)	4.0 (25/629)	9
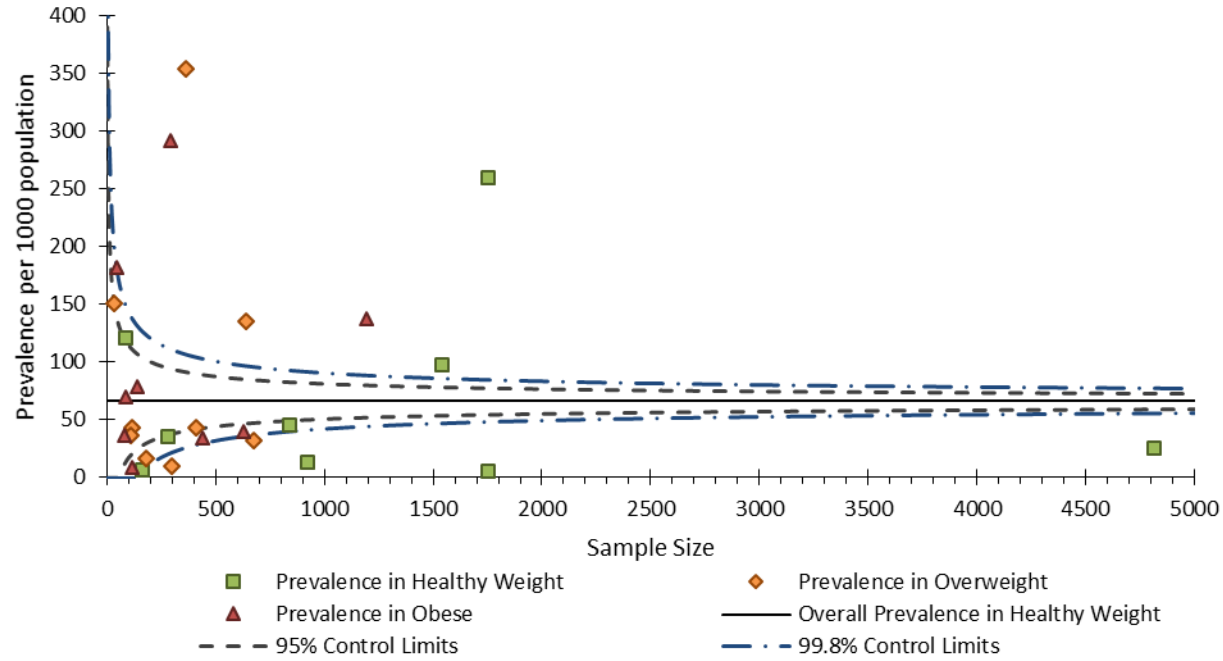
Figure 9: Funnel plot showing prevalence of High Fasting Plasma Glucose per 1000 population

All nine studies were included in the meta-analyses, the average prevalence ratio (PR) indicated that for every 1 child of a healthy weight with high fasting plasma glucose there would be 1.4 (CI 1.2 to 1.6, p<0.001) who were overweight (Figure 10) and 1.4 (CI 1.2 to 1.7, p<0.001) with obesity (Figure 11) who have high fasting plasma glucose (assuming equal population sizes in the three groups). The I^2 statistic in Figure 10 is 0% indicating little heterogeneity between studies, whilst the I^2 Statistic in Figure 11 was higher at 23.4%, indicating low heterogeneity. Overall the results show a moderate but significant increase in prevalence of high fasting plasma glucose in the higher weight categories, with similar increases in prevalence in the overweight and obese groups relative to the healthy weight.

Study	Overweight Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Caserta et al. (2010)	3	179	1	164	0.3	2.7 (0.3 to 26.2)
Chu and Pan (2007)	128	361	455	1753	59.25	1.4 (1.2 to 1.6)
Davis et al. (2005)	5	33	10	83	1.53	1.3 (0.5 to 3.4)
Del-Rio-Navarro et al. (2008)	18	411	12	923	2.92	3.4 (1.6 to 6.9)
Gong et al. (2013)	5	115	10	283	1.37	1.2 (0.4 to 3.5)
Khader et al. (2011)	4	111	38	837	1.49	0.8 (0.3 to 2.2)
Seki et al. (2008)	3	299	9	1755	0.9	2.0 (0.5 to 7.2)
Wang et al. (2013)	86	637	151	1541	24.65	1.4 (1.1 to 1.8)
Xu et al. (2012)	22	675	120	4813	7.59	1.3 (0.8 to 2.1)
Total (random effects)	274	2821	806	12152	100.00	1.4 (1.2 to 1.6)

Heterogeneity: Q = 7.8; df = 8 (P = 0.4528); I² = 0%
Test for overall effect: Z = 5.28 (P <0.001)

Figure 10: Forest plot of prevalence ratio for overweight relative to healthy weight, children for High Fasting Plasma Glucose

Study	Obese Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Caserta et al. (2010)	3	82	1	164	0.65	6.0 (0.6 to 56.8)
Chu and Pan (2007)	85	291	455	1753	33.91	1.1 (0.9 to 1.4)
Davis et al. (2005)	8	44	10	83	4.21	1.5 (0.6 to 3.5)
Del-Rio-Navarro et al. (2008)	15	438	12	923	5.34	2.6 (1.2 to 5.6)
Gong et al. (2013)	11	140	10	283	4.43	2.2 (1.0 to 5.1)
Khader et al. (2011)	6	86	38	837	4.43	1.5 (0.7 to 3.5)
Seki et al. (2008)	1	116	9	1755	0.78	1.7 (0.2 to 13.2)
Wang et al. (2013)	164	1195	151	1541	32.3	1.4 (1.1 to 1.7)
Xu et al. (2012)	25	629	120	4813	13.95	1.6 (1.0 to 2.4)
Total (random effects)	318	3021	806	12152	100	1.4 (1.2 to 1.7)

Heterogeneity: Q = 10.45; df = 8 (P = 0.2350); I² = 23.43%
Test for overall effect: Z = 3.79 (P <0.001)

Figure 11: Forest plot of prevalence ratio for obese relative to healthy weight, children for High Fasting Plasma Glucose

Although heterogeneity between prevalence ratios was not statistically significant, prevalence estimates varied widely, which may relate to key differences in population characteristics. Variation was seen between and within countries,

perhaps due to regional differences. Three of the studies examining fasting plasma glucose were conducted in China, with comparable age ranges (Xu *et al.* 2012; Gong *et al.* 2013; Wang *et al.* 2013). Wang et al. reported a prevalence of 14% for the obese group compared with 8% and 4% reported by Gong et al. and Xu et al., respectively.

Studies with smaller samples would be expected to have lower precision and report a more extreme prevalence estimate for high fasting plasma glucose. However, examination of studies did not indicate a pattern between sample size and prevalence, with similar prevalence reported by Caserta et al. and Xu et al., despite having markedly different sample sizes.

The other main difference between the studies was participant eligibility criteria. Seki, Matsuo and Faria Carrilho (2008) and Xu *et al.* (2012) excluded participants with pre-existing conditions that may affect metabolism, or influence the results of the tests, which might explain the lower reported prevalence. The other studies may have included participants with pre-existing diagnoses/conditions that can affect results but this was not recorded in the eligibility criteria.

Overall the results of the meta-analyses suggest that the prevalence of high fasting plasma glucose may be related to weight status, although the increase in prevalence was small. Variation between studies may be due to factors such as sampling methods, country and region of study, sample size, and eligibility criteria, although the prevalence ratio estimates were consistent. There were no studies based in the UK, so generalisability of results to the UK population is limited due to population, cultural, and dietary differences.

Oral Glucose Tolerance Test

One study considered the prevalence of elevated glucose (≥200mg/dl) using an oral glucose tolerance test (OGTT) and had a quality score of nine (Bar Dayan *et al.* 2005). The prevalence estimate in the recruited sample ranged from 0.0% (healthy weight and overweight groups) to 0.3% (obese group) (Table 10, page 67). Bar Dayan et al. recruited 76,732 Israeli 17 year olds who were screened by the Israeli Defence Force for military service, which is mandatory for all Israelis. As only one study was eligible, definitive conclusions regarding the association between weight status and elevated fasting plasma glucose using an OGTT could not be drawn and a meta-analysis was not possible.

Insulin Resistance

One study of the prevalence of insulin resistance (>3.16) was included in the systematic review (Manios *et al.* 2008). The study showed that the prevalence increased from 2% in the healthy weight group to 11% in the obese group (Table 11), with a lower prevalence in the overweight group compared to the healthy weight group (1.4%). The high prevalence reported for the obese group may be a factor of lower precision due to the small sample, 37 participants, compared with the 137 participants in the healthy weight group. Manios et al. recruited 481 children aged 10-12 from primary schools in Crete, although insulin levels were only measured in 248, which was not based on an a priori sample size calculation. Therefore it is unclear if the study was sufficiently powered to be representative. This, along with a lack of detail on the sample, lowered the quality score and limits generalisation of the results beyond the study population.

Table 10: Summary of studies reporting prevalence data for Hyperglycaemia (Oral Glucose Tolerance Test ≥200mg/dl)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Bar Dayan *et al.* (2005)	Israel	76732	17-17 (17)	Community	0.0 (7/53684)	0.0 (4/9202)	0.3 (8/2897)	9

Table 11: Summary of studies reporting prevalence data for Hyperglycaemia (Insulin Resistance >3.16)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Manios *et al.* (2008)	Greece	248	10-12 (11.3)	Education	2.2 (3/137)	1.4 (1/74)	10.8 (4/37)	5

4.2.2 Dyslipidaemia

Dyslipidaemia is defined as an abnormal amount of lipids in the blood (National Heart Lung and Blood Institute 2005). Lipids are composed of total cholesterol (consisting of Low Density Lipoprotein (LDL) and High Density Lipoprotein (HDL) cholesterol) and triglycerides. Higher values for LDL cholesterol and triglycerides, and low levels of HDL cholesterol have been linked to adverse outcomes (National Heart Lung and Blood Institute 2005; British Heart Foundation 2012). Eight studies examined total cholesterol and/or the individual lipid profiles (Figure 12).
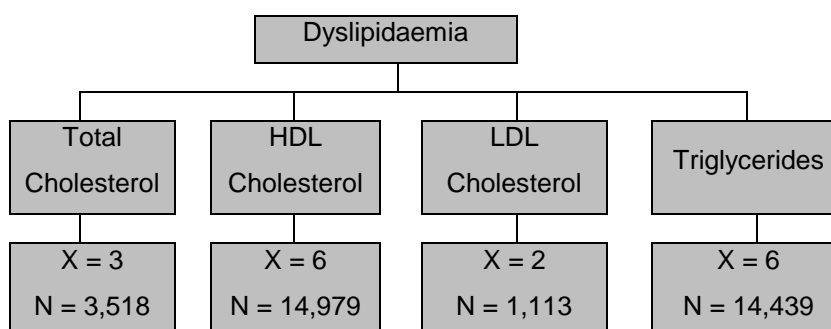
Figure 12: Overview of studies of Dyslipidaemia (X: number of studies; N: number of participants) (number of studies is higher as some reported data for multiple lipid profiles)

High Total Cholesterol:

Three articles were identified using the cut-off ≥200mg/dL (5.2mmol/L) to define cases (the normal range should be <200mg/dL (5.2mmol/L) (NHS 2013)). All studies reported a higher prevalence in the obese group compared with the healthy weight group (Table 12). All three studies neglected to provide sufficient details regarding the recruitment of participants and for Chu and Pan (2007) and Gong *et al.* (2013) there was insufficient information regarding the control of potential confounding factors, such as gender and age.

The funnel plot did not indicate a clear association between weight status and prevalence of high total cholesterol (Figure 13). The two data points above the 99.8% control limits were from Chu and Pan (2007). This suggested the prevalence was not due to random variation between studies and that the data from Chu and Pan (2007) may arise from a different population or a different assay than the other two studies. The data for the other two studies did suggest some clustering by weight groups and the prevalence in the healthy weight group for these two studies was close to the expected 5% from the normal range.

Table 12: Summary of studies reporting prevalence of High Total Cholesterol (≥200mg/dL)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Quality Score
Caserta *et al.* (2010)	Italy	575	11-13 (NR)	Education	5.1 (16/314)	7.3 (13/179)	8.5 (7/82)	4
Chu and Pan (2007)	Taiwan	2405	6-12 (NR)	Education	18.2 (319/1753)	16.9 (61/361)	24.1 (70/291)	6
Gong *et al.* (2013)	China	538	9-15 (12)	Education	4.6 (13/283)	8.7 (10/115)	14.3 (20/140)	5
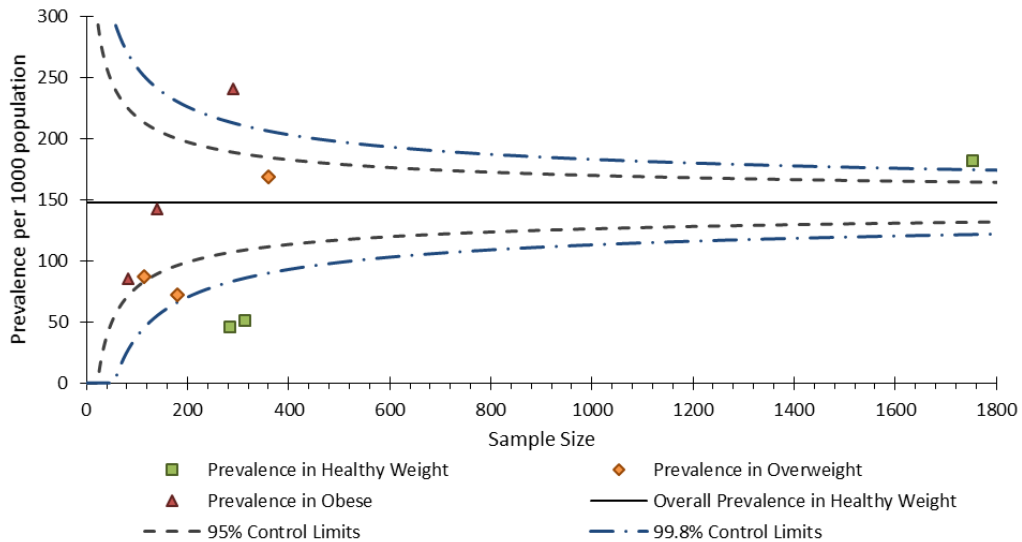
Figure 13: Funnel plot showing prevalence of High Total Cholesterol per 1000 population

All three studies were included in the meta-analyses, the average PR indicated that (assuming equal sample sizes in the three groups) for every 1 child of a healthy weight who has high total cholesterol there would be 1.2 (CI 0.8 to 1.8, p = 0.4) who are overweight (Figure 14) and 1.8 (CI 1.0 to 3.1, p=0.04) who are obese (Figure 15). The I^2 Statistics were 46.4% (Figure 14) and 65.5% (Figure 15), suggesting moderate heterogeneity between the studies. Excluding Chu and Pan (2007) from the meta-analyses increased the PR to 1.6 (CI 0.9 to 2.7, p = 0.08) and lowered the I^2 Statistic to 0% for the overweight group relative to the healthy weight group. For the obese group relative to the healthy weight, the PR increased to 2.4 (CI 1.3 to 4.4, p = 0.004) and the I^2 Statistic decreased to 20.1%. This supports the view that Chu and Pan used different methodology from the other studies, although the small number of studies meant definitive conclusions could not be made.
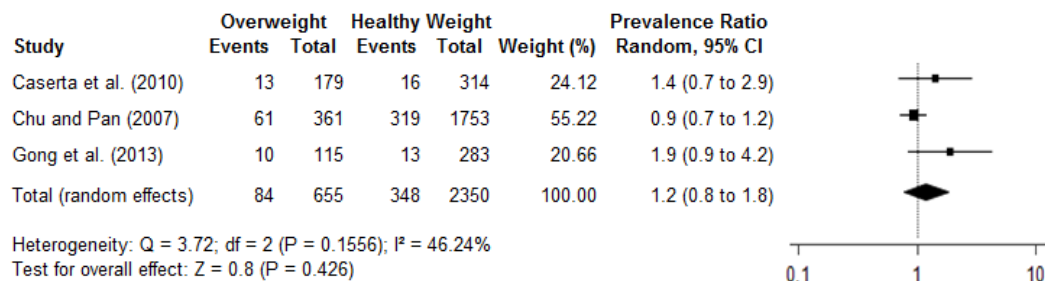
Figure 14: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for High Total Cholesterol

Study	Obese Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Caserta et al. (2010)	7	82	16	314	23.17	1.7 (0.7 to 3.9)
Chu and Pan (2007)	70	291	319	1753	47.38	1.3 (1.1 to 1.7)
Gong et al. (2013)	20	140	13	283	29.45	3.1 (1.6 to 6.1)
Total (random effects)	97	513	348	2350	100.00	1.8 (1.0 to 3.1)

Heterogeneity: Q = 5.8; df = 2 (P = 0.0551); I² = 65.49%
Test for overall effect: Z = 2.07 (P = 0.038)

Figure 15: Forest plot of prevalence ratio for obese, relative to healthy weight, children for High Total Cholesterol

Review of the studies indicated variation between the studies in terms of sample size. Caserta et al. and Gong et al. had similar sample sizes and comparable prevalence across the weight statuses. Chu and Pan had a much larger sample size, which increased precision, but the reported prevalence was two to three times greater than that reported in the other studies, suggesting the population studied was at greater risk or the test used was more sensitive.

Chu and Pan adopted a two-stage stratified sampling method targeting 6 to 13 year olds enrolled in 104 public or private schools registered with the Ministry of Education. From each school 24 children were randomly selected (Tu *et al.* 2007). In contrast Gong *et al.* (2013) recruited 538 children aged 9 to 14 from 14 schools in two of Shanghai's districts. Caserta *et al.* (2010) recruited 575 randomly selected children aged 11 to 13 in Reggio Calabria, Italy, from the school census list. The samples were of a comparable age, and were appropriately selected to obtain a representative sample of the target population. With regards to eligibility criteria Caserta *et al.* (2010) excluded one participant from analysis due to the presence of type 1 diabetes. As neither Caserta *et al.* (2010) nor Chu and Pan (2007) explicitly stated their eligibility criteria it is difficult to determine if their samples only included healthy children or if those with diagnosed conditions were included, which may have resulted in a biased estimate of prevalence. However it is unlikely that this difference alone would account for the higher prevalence reported by Chu and Pan.

Prevalence data by gender was only provided Chu and Pan and Caserta et al. Chu and Pan reported that females had a higher prevalence of high cholesterol than males across the weight statuses, this was only significant in the obese group (Fisher's exact test, p=0.015), whereas Caserta *et al.* (2010) found that females had a higher prevalence in the healthy weight group. However as weight increased, the relationship reversed and males had a higher prevalence in the overweight and obese groups. None of these comparisons were significant. The variation between

studies may be indicative of population differences or may be partially attributable to the smaller sample sizes in the higher weight status groups reported by Caserta et al.

Overall the results show a higher prevalence of total cholesterol in obese children relative to the healthy weight group. However the variable prevalence between studies suggested other factors are involved, such as study methodology. There was insufficient data to determine an association with gender and there were no UK-based studies, so generalisability to the UK population is limited.

Low High Density Lipoprotein (HDL) Cholesterol

Six articles measured the prevalence of low HDL Cholesterol, three studies utilised the threshold of <40mg/dL (1.03mmol/l), two studies used ≤40mg/dL (1.03mmol/l) and one study provided data for both (Wang *et al.* 2013), resulting in seven data points (Table 13). Review of Wang et al.'s data indicated they reported a higher prevalence for the <40mg/dL cut-off than for the ≤40mg/dL cut-off across the weight statuses. Review of the data showed the estimated prevalence for girls over 10 years was notably higher for the <40mg/dL cut-off than the ≤40mg/dL cut-off, casting doubt over the accuracy of the data. Therefore data for the <40mg/dL was removed from the analysis, leaving six prevalence estimates, which were analysed together.

The quality scores ranged from four to nine, and there did not appear to be a clear association between prevalence and quality score. Tandon et al., Wang et al. and Khader et al. all had a score of 7, yet the reported prevalence varied between the studies, with Wang et al. consistently reporting a lower prevalence across the weight statuses, suggesting other study specific factors may be associated with the variation in prevalence. Five of the studies had not conducted an a priori sample size calculation, with additional uncertainty regarding the control of potential confounding factors for three of the studies (Caserta *et al.* 2010; Khader *et al.* 2011; Gong *et al.* 2013; Tandon *et al.* 2013; Wang *et al.* 2013).

The studies suggested that the prevalence of low HDL cholesterol increased with weight category (Table 13). However the prevalence varied considerably within the weight statuses. In the healthy weight group prevalence ranged from 1.1% (Gong et al.) to 23.1% (Khader et al.); similar variation was noted for overweight and obese categories. Tandon et al. and Khader et al. reported a consistently higher prevalence regardless of weight status compared with the other studies. The overall

variation between studies suggested that factors other than weight may influence HDL cholesterol levels.
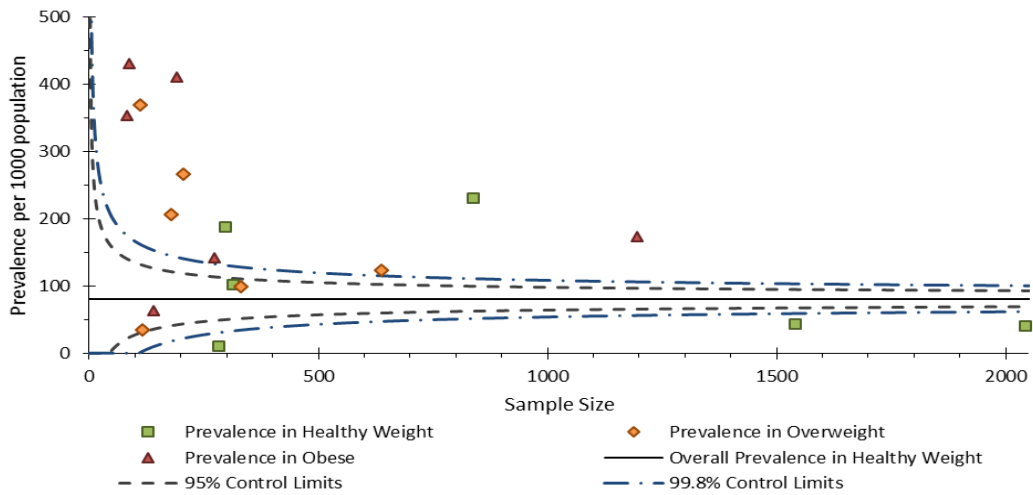
Figure 16: Funnel plot showing prevalence of Low HDL Cholesterol (≤40mg/dL) per 1000 population

The funnel plot suggested considerable heterogeneity between the studies, with no clear clustering of studies by weight status (Figure 16). For the healthy weight group only one of the six data points was within the 95% control limit, with the others distributed outside the 99.8% control limits (Caserta *et al.* 2010). Similar variation was observed in the overweight and obese data points. The between study variation in prevalence suggested the relationship between weight status and low HDL cholesterol was not straightforward.

All six data points were included in the meta-analysis. The combined average PR indicated that for every 1 child of a healthy weight who had low HDL cholesterol there would be 2.0 (1.5 to 2.6, p<0.001) who were overweight (Figure 17) and 2.9 (2.1 to 4.1, p<0.001) who were obese (assuming equally sized populations) (Figure 18). The plots indicated significant heterogeneity between studies with I^2 statistics of 63% (Figure 17) and 79% (Figure 18), indicating moderate to high heterogeneity.

Table 13: Summary of studies reporting prevalence data for Low HDL Cholesterol
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
<40mg/dL								
Gong *et al.* (2013)	China	538	9-15 (12)	Education	1.1 (3/283)	3.5 (4/115)	6.4 (9/140)	5
Tandon *et al.* (2013)	India	695	10-18 (13.4)	Education	18.9 (56/297)	26.7 (55/206)	41.1 (79/192)	7
Wang *et al.* (2013)a	China	3373	6-18 (12)	Education	6.2 (95/1541)	15.4 (98/637)	17.8 (213/1195)	7
Xu *et al.* (2012)	China	8764	7-11 (8.6)	Education	4.0 (82/2044)	10.0 (33/330)	14.3 (39/273)	9
≤40mg/dL								
Caserta *et al.* (2010)	Italy	575	11-13 (NR)	Education	10.2 (32/314)	20.7 (37/179)	35.4 (29/82)	4
Khader *et al.* (2011)	Jordan	1034	7-18 (NR)	Community	23.1 (193/837)	36.9 (41/111)	43.0 (37/86)	7
Wang *et al.* (2013)b	China	3373	6-18 (12)	Education	4.3 (67/1541)	12.4 (79/637)	17.3 (207/1195)	7

Study	Overweight Events	Total	Healthy Weight Events	Total	Weight	Prevalence Ratio Random, 95% CI
Caserta et al. (2010)	37	179	32	314	12.98	2.0 (1.31 to 3.14)
Gong et al. (2013)	4	115	3	283	2.13	3.3 (0.75 to 14.43)
Khader et al. (2011)	41	111	193	837	18.38	1.6 (1.22 to 2.1)
Tandon et al. (2013)	55	206	56	297	16.46	1.4 (1.02 to 1.96)
Wang et al. (2013)	79	637	67	1541	16.98	2.9 (2.09 to 3.9)
Xu et al. (2012)	33	330	82	2044	14.46	2.5 (1.69 to 3.67)
Total (random effects)	347	2215	528	6857	100.00	2.0 (1.5 to 2.6)

Heterogeneity: Q = 13.53; df = 5 (P = 0.0189); I² = 63.06%

Test for overall effect: Z = 5.23 (P <0.001)

Figure 17: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for Low HDL Cholesterol (<40mg/dL)

Study	Obese Events	Total	Healthy Weight Events	Total	Weight	Prevalence Ratio Random, 95% CI
Caserta et al. (2010)	29	82	32	314	16.41	3.5 (2.2 to 5.4)
Gong et al. (2013)	9	140	3	283	4.94	6.1 (1.67 to 22.05)
Khader et al. (2011)	37	86	193	837	20.20	1.9 (1.42 to 2.45)
Tandon et al. (2013)	79	192	56	297	19.82	2.2 (1.63 to 2.92)
Wang et al. (2013)	207	1195	67	1541	20.37	4.0 (3.06 to 5.19)
Xu et al. (2012)	39	273	82	2044	18.25	3.6 (2.49 to 5.1)
Total (random effects)	400	1968	433	5316	100.00	2.9 (2.1 to 4.1)

Heterogeneity: Q = 23.32; df = 5 (P = 0.0003); I² = 78.6%

Test for overall effect: Z = 6.6 (P <0.001)

Figure 18: Forest plot of prevalence ratio for obese, relative to healthy weight, children for Low HDL Cholesterol (<40mg/dL)

Heterogeneity may be due to factors such as country of study and regional variation. Three of the studies were conducted in China (Xu *et al.* 2012; Gong *et al.* 2013; Wang *et al.* 2013), with fairly comparable age ranges. Wang et al. and Xu et al. reported similar prevalence for all three weight statuses, with Gong et al. reporting substantially lower prevalence. Gong et al. was conducted in Shanghai (a coastal city), Wang et al in Beijing (China's capital), and Xu et al. in six provincial capital cities, including Beijing and Shanghai. Regional differences in these populations may explain some of the variation in reported prevalence.

Overall the meta-analysis indicated that children who are overweight or obese have a higher prevalence of low HDL cholesterol than those of a healthy weight.

Heterogeneity between studies and lack of UK-based studies again limits generalisability.
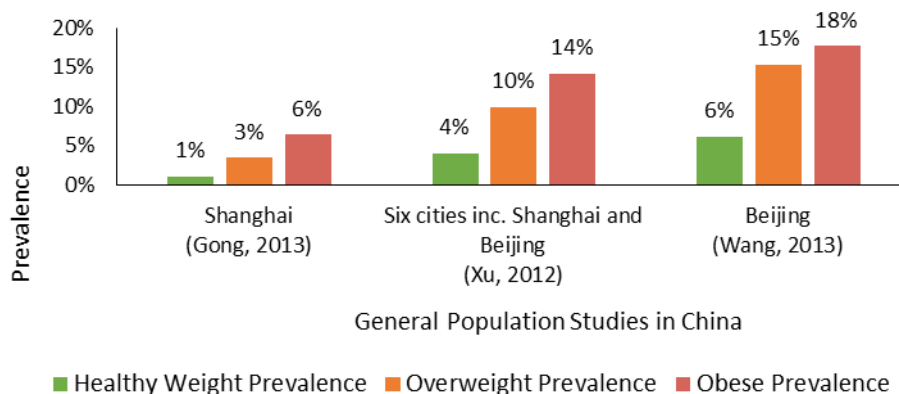
Figure 19: Prevalence of low HDL cholesterol, restricted to studies conducted in China

High Low Density Lipoprotein (LDL) Cholesterol

Two articles measured the prevalence of high LDL cholesterol using the cut-off of ≥130mg/dL (3.4mmol/L) and both reported that the prevalence of high LDL cholesterol increased with weight status (Table 14), with comparable prevalence for the overweight and obese groups, and minor difference in prevalence for the healthy weight groups. This suggested that the prevalence of LDL cholesterol is associated with weight; however the small number of studies and the failure to address confounding limited conclusions. The quality scores for the studies were four (Caserta *et al.* 2010) and five (Gong *et al.* 2013), however the lack of studies limited definitive conclusions being drawn. Neither study conducted an a priori sample size calculation, nor did they provide sufficient detail regarding the sample, which lowered their quality rating.

Table 14: Summary of studies reporting prevalence data for High LDL Cholesterol (≥130mg/dL))
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Caserta *et al.* (2010)	Italy	575	11-13 (NR)	Education	5.1 (16/314)	7.3 (13/179)	12.2 (10/82)	4
Gong *et al.* (2013)	China	538	9-15 (12)	Education	2.5 (7/283)	7.8 (9/115)	12.1 (17/140)	5
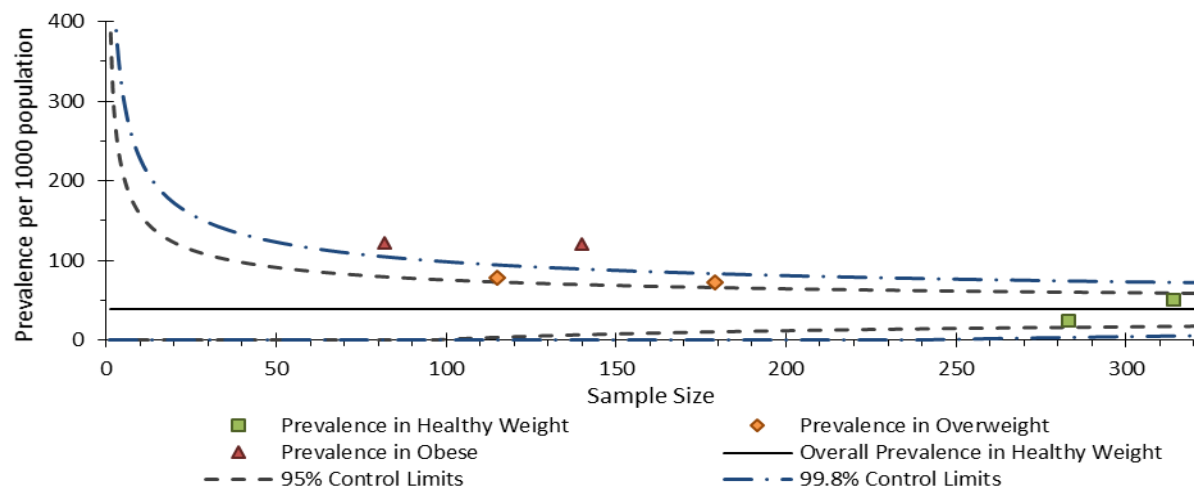
Figure 20: Funnel plot showing prevalence of High LDL Cholesterol per 1000 population

The funnel plot showed that the prevalence rate in the overweight and obese weight categories were very comparable (Figure 20). The data points for the obese group were almost identical and both fell outside the 99.8% control limits, with the healthy weight data points within the 95% control limits. This would be expected if there is an association between weight status and prevalence of high LDL cholesterol.

A meta-analysis of the two studies was conducted. For equally large populations, the average PR indicated that for every 1 child of a healthy weight there would be 2.0 (CI 0.9 to 4.3, p=0.082) who were overweight (Figure 21) and 3.3 (CI 1.6 to 6.8, p=0.001) with obesity (Figure 22) who have high LDL cholesterol. The I^2 Statistic ranged from 35.6% (Figure 22) to 41.5% (Figure 21) indicating moderate heterogeneity between the studies. The results suggested a moderate increase in prevalence of high LDL cholesterol in the higher weight categories, with a higher increase observed in the obese group than the overweight group. However the non-significance of the PR for the overweight group relative to the healthy group may be the result of the small number of studies and the low prevalence of high LDL. Further studies would be required to assess whether the significance is achieved, or if the healthy weight and overweight groups have similar prevalence of high LDL cholesterol.

Study	Overweight Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Caserta et al. (2010)	13	179	16	314	58.73	1.4 (0.7 to 2.9)
Gong et al. (2013)	9	115	7	283	41.27	3.2 (1.2 to 8.3)
Total (random effects)	22	294	23	597	100.00	2.0 (0.9 to 4.3)

Heterogeneity: Q = 1.71; df = 1 (P = 0.1912); I² = 41.46%
Test for overall effect: Z = 1.74 (P = 0.082)
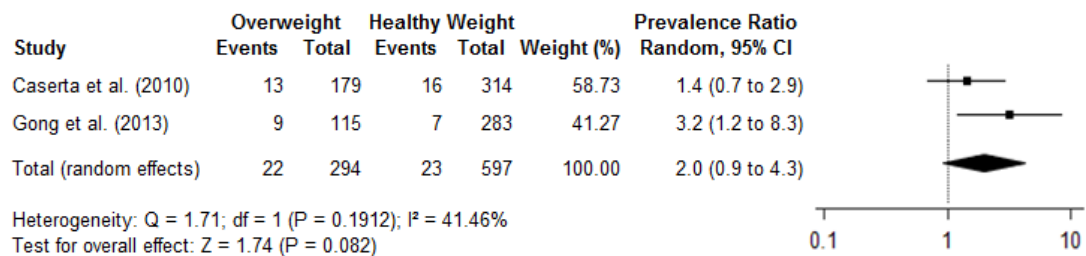
Figure 21: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for High LDL Cholesterol (≥130mg/dL)

Study	Obese Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Caserta et al. (2010)	10	82	16	314	54.18	2.4 (1.1 to 5.1)
Gong et al. (2013)	17	140	7	283	45.82	4.9 (2.1 to 11.6)
Total (random effects)	27	222	23	597	100.00	3.3 (1.6 to 6.8)

Heterogeneity: Q = 1.55; df = 1 (P = 0.2127); I² = 35.61%
Test for overall effect: Z = 3.33 (P = 0.001)

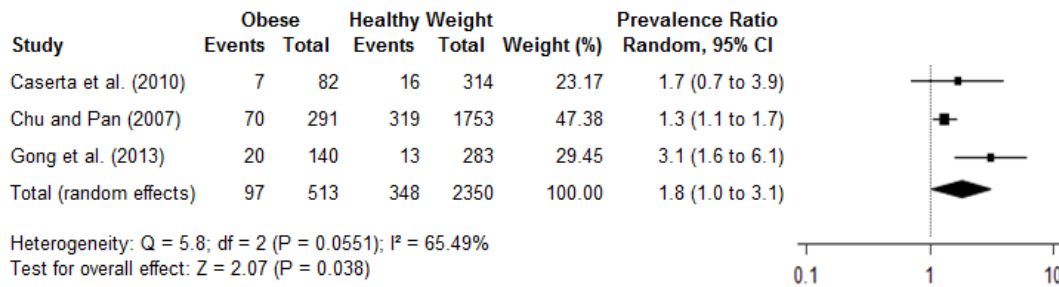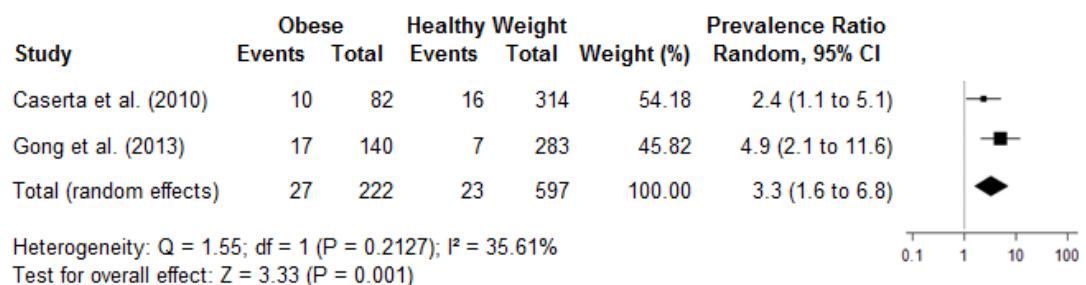**Figure 22: Forest plot of prevalence ratio for obese, relative to healthy weight, children for High LDL Cholesterol (≥130mg/dL)**

The studies were conducted in different continents, albeit with comparable age groups and sample sizes. Gong et al. recruited 538 children aged 9-14 from 14 elementary and junior high schools in two of Shanghai's (China) districts, without any explicitly stated eligibility criteria. In contrast, Caserta et al. sampled 575 11-13 year olds in Reggio Calabrai, Italy who were randomly selected from the school census list. Caserta et al. did not state any specific inclusion criteria, however one participant with type 1 diabetes was excluded. Despite these differences, the reported prevalence estimates are very similar, particularly for the overweight and obese groups (Table 14).

Overall the results suggest that the prevalence of high LDL cholesterol was higher in the obese groups than in the healthy weight groups. However the lack of studies, in particular with a UK based population, limits generalisability of the results.

High Triglycerides

Six studies measured the prevalence of high triglycerides using the cut-off ≥150mg/dL (1.7mmol/L) (Table 15). Of the six studies, Tandon et al. reported a substantially higher prevalence for all weight statuses than the other studies; e.g. 16% in the healthy weight group versus the second highest of 5% (Wang *et al.* 2013)(Table 15). The reported prevalence in the other five studies was more comparable. Study quality scores ranged from four to nine. The main areas where studies did not provide sufficient information related to adequacy of the sample size and reliability of the screening method. There did not appear to be an association between study quality and reported prevalence, as the three studies which scored seven reported considerably different prevalence, e.g. 4.7%, 5.2%, and 16.8% prevalence in the healthy weight group (Pereira *et al.* 2009; Tandon *et al.* 2013; Wang *et al.* 2013). Similar variation was observed for the overweight and obese groups.

The funnel plot suggested considerable heterogeneity between the studies, with no clear clustering of data points by weight status (Figure 23). The three highest data points for each weight category were from Tandon et al., suggesting the study may be an outlier. Only two of the healthy weight data points fell within the 95% control limits, which also included some overweight and obese data points, suggesting the prevalence may not be due to weight status alone. However the low number of studies meant definitive conclusions could not be drawn.

Table 15: Summary of studies reporting prevalence data for High Triglycerides (≥150mg/dL)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Caserta *et al.* (2010)	Italy	575	11-13 (NR)	Education	1.3 (4/314)	6.7 (12/179)	8.5 (7/82)	4
Gong *et al.* (2013)	China	538	9-15 (12)	Education	1.1 (3/283)	8.7 (10/115)	10.7 (15/140)	5
Pereira *et al.* (2009)	Brazil	494	2-19 (9.9)	Education	4.7 (18/383)	14.6 (7/48)	27.0 (17/63)	7
Tandon *et al.* (2013)	India	695	10-18 (13.4)	Education	16.8 (50/297)	43.7 (90/206)	61.5 (118/192)	7
Wang *et al.* (2013)	China	3373	6-18 (12)	Education	5.2 (80/1541)	10.8 (69/637)	15.1 (181/1195)	7
Xu *et al.* (2012)	China	8764	7-11 (8.6)	Education	3.3 (68/2044)	6.1 (20/330)	16.5 (45/273)	9
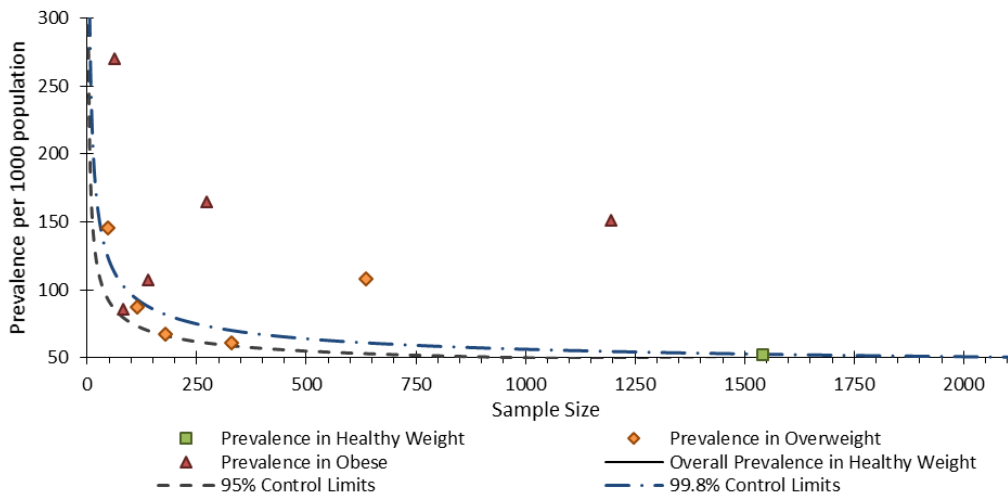
Figure 23: Funnel plot showing prevalence of High Triglycerides per 1000 population, excluding Tandon *et al.* (2013)

Although the funnel plot suggested Tandon et al. might be an outlier, the forest plots suggest the PR is comparable with other studies; therefore all six studies were included in the meta-analyses. Assuming equal population sizes in the three groups, the average PR indicated that for every 1 child of a healthy weight with high triglycerides there would be 2.5 (CI 1.9 to 3.3, p≤0.001) who were overweight (Figure 24) and 4.2 (CI 3.2 to 5.6, p≤0.001) who were obese (Figure 25). The I^2 Statistic in Figure 24 was 38.9% indicating low-moderate heterogeneity between studies, whilst in Figure 25 the I^2 indicated high heterogeneity (57.3%). The results suggested a moderate and significant increase in prevalence of high triglycerides in children who were overweight or obese, with a larger increase in the obese than the overweight group.

Study	Overweight Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Caserta et al. (2010)	12	179	4	314	5.47	5.3 (1.7 to 16.1)
Gong et al. (2013)	10	115	3	283	4.33	8.2 (2.3 to 29.3)
Pereira et al. (2009)	7	48	18	383	9.25	3.1 (1.4 to 7.0)
Tandon et al. (2013)	90	206	50	297	31.17	2.6 (1.9 to 3.5)
Wang et al. (2013)	69	637	80	1541	30.28	2.1 (1.5 to 2.8)
Xu et al. (2012)	20	330	68	2044	19.52	1.8 (1.1 to 3.0)
Total (random effects)	208	1515	223	4862	100.00	2.5 (1.9 to 3.3)

Heterogeneity: Q = 8.18; df = 5 (P = 0.1466); I² = 38.88%
Test for overall effect: Z = 6.53 (P <0.001)

Figure 24: Forest plot of prevalence ratios for overweight, relative to healthy weight, children for High Triglycerides (≥150mg/dL)

Study	Obese Events	Obese Total	Healthy Weight Events	Healthy Weight Total	Weight (%)	Prevalence Ratio Random, 95% CI
Caserta et al. (2010)	7	82	4	314	4.85	6.7 (2.01 to 22.3)
Gong et al. (2013)	15	140	3	283	4.73	10.1 (3.0 to 34.3)
Pereira et al. (2009)	17	63	18	383	13.59	5.7 (3.1 to 10.5)
Tandon et al. (2013)	118	192	50	297	26.44	3.7 (2.8 to 4.8)
Wang et al. (2013)	181	1195	80	1541	27.59	2.9 (2.3 to 3.8)
Xu et al. (2012)	45	273	68	2044	22.79	5 (3.5 to 7.1)
Total (random effects)	383	1945	223	4862	100.00	4.2 (3.2 to 5.6)

Heterogeneity: Q = 11.7; df = 5 (P = 0.0391); I² = 57.27%
Test for overall effect: Z = 9.89 (P <0.001)

Figure 25: Forest plot of prevalence ratios for obese, relative to healthy weight, children for High Triglycerides (≥150mg/dL)

Three studies were conducted in China (Xu *et al.* 2012; Gong *et al.* 2013; Wang *et al.* 2013); Wang et al. and Xu et al. reported comparable prevalence in the obese group, 15% and 17%, whereas Gong et al. reported a prevalence of 11%. Gong et al. recruited a smaller sample than the other two studies (Table 17), thus prevalence estimates may be more subject to sampling error. There may also be geographical variation; Gong et al. recruited from two districts in Shanghai (China) and Wang et al. recruited children from Beijing. Xu et al. recruited from Shanghai and Beijing, however city level prevalence was not provided. Other methodological differences were ages of the participants, dietary differences and sampling methods. Wang et al. recruited the broadest age range, six to 18 years in comparison to Xu et al. (seven to nine years) and Gong et al. (nine to 15 years). Although Caserta et al. and Wang et al. also provided prevalence data by gender, no clear association between prevalence and weight status was observed. This is likely due to the small number of studies.

Tandon et al. reported the highest prevalence across the weight statuses, 17% in healthy weights and 62% in obese, despite recruiting 695 out of 900 randomly selected children from schools located in four different geographical areas of Delhi, India. The higher prevalence may be an indicator of higher risk in South Asians, or differences in lifestyle and diet. The prevalence is considerably higher than Caserta et al., who reported a prevalence of 9% in the obese group.

Overall, a higher prevalence of high triglycerides was observed in children who were obese than in those of a healthy weight; however the reported prevalence is likely confounded by many factors such as study methodology, country of study, and diet and lifestyle factors. The lack of UK based studies meant that generalisations were limited.

4.2.3 High Blood Pressure

High blood pressure is the pressure exerted on blood vessel walls; consistent high pressure can cause extra strain on the blood vessels, heart and other organs. Three definitions for high blood pressure were identified (Figure 26).
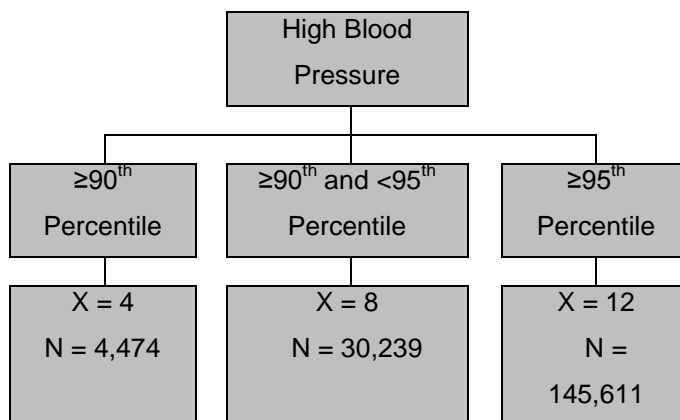
Figure 26: Overview of High Blood Pressure and the three cut-offs used in the studies (X: number of studies; N: number of participants) (number of studies is higher as some studies reported the prevalence across multiple cut-offs)

Nineteen studies were identified across the three definitions, with three studies providing prevalence data for two or more definitions (Ramos and Barros 2005; Genovesi *et al.* 2010; Pecin *et al.* 2013). Another study provided prevalence data for three countries; each country was included as a separate data point (Dyson *et al.* 2014). In total 24 data points across the three definitions were available for the meta-analysis. Analysis indicated that the prevalence ratios of overweight/obese relative to healthy weight were similar for each definition (Table 16). Therefore the data for the three definitions were combined for analysis.

Table 16: Summary of prevalence ratios by definition of High Blood Pressure

Definition of high blood pressure	Prevalence Ratio: overweight relative to healthy weight	Prevalence Ratio: obese relative to healthy weight
≥90th Percentile	2.5 (1.4 to 4.5)	4.4 (2.2 to 8.8)
≥90 and <95th Percentile	2.2 (1.4 to 3.4)	3.9 (1.9 to 8.0)
≥95th Percentile	2.0 (1.5 to 2.6)	4.0 (2.5 to 6.5)

All 24 data points suggested prevalence increased with weight status (Table 17); however there was considerable variation within weight statuses. For instance, in the healthy weight group prevalence ranged from 0.1% (Nur *et al.* 2008) to 17.5% (Ramos and Barros 2005), similar variation was observed for the overweight and

obese groups (Table 17), suggesting that estimated prevalence may be the result of other factors in addition to increased weight. There was considerable variation in quality scores, with a modest negative association between quality score and reported prevalence. The main quality criteria that studies did not address were a lack of detail regarding the subjects and setting, uncertainty over the reliability of measurement, and control for confounding factors. Eleven studies did not state if the people taking measurements were suitably trained, the remaining eight studies stated that measurements were conducted by trained physicians or nurses (Moura *et al.* 2004; Bar Dayan *et al.* 2005; Nur *et al.* 2008; Seki, Matsuo and Faria Carrilho 2008; Genovesi *et al.* 2010; Rafraf, Pourghassem Gargari and Safaiyan 2010; Guo *et al.* 2012; Xu *et al.* 2012).

The funnel plot of all studies indicated moderate clustering of prevalence by weight status, suggesting a positive association between weight status and prevalence, particularly when comparing the healthy weight and obese groups (Figure 27). As Bar Dayan had a sample size of 53,000, a funnel plot with a truncated X axis is also presented to show dispersion of prevalence at smaller sample sizes (Figure 28, page 88). Only four of the healthy weight data points were inside the 95% control limits and there appeared to be greater variation in prevalence for the overweight groups, which were distributed predominantly to the outer edges of the 95% and 99.8% control limits, whereas the obese data points were outside the 99.8% control limits, indicating considerable difference between the groups in prevalence.
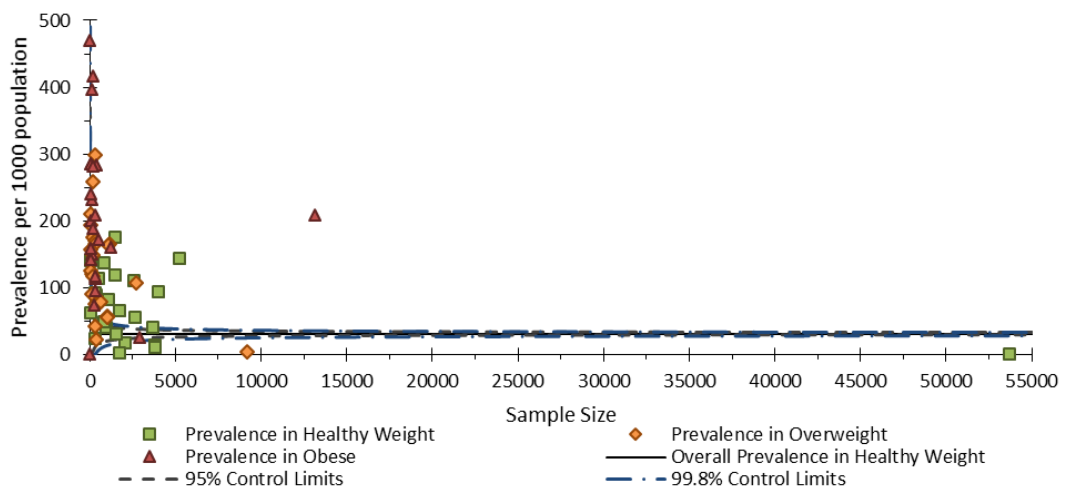
Figure 27: Funnel plot showing prevalence of High Blood Pressure per 1000 population

Table 17: Summary of studies reporting prevalence data for High Blood Pressure
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
≥90th Percentile								
Caserta *et al.* (2010)	Italy	575	11-13 (NR)	Education	9.2 (29/314)	11.7 (21/179)	15.9 (13/82)	4
Khader *et al.* (2011)	Jordan	1034	7-18 (NR)	Community	4.9 (41/837)	9.0 (10/111)	15.1 (13/86)	7
Seki, Matsuo and Faria Carrilho (2008)	Brazil	2170	6-16 (11.3)	Education	6.6 (116/1755)	17.1 (51/299)	39.7 (46/116)	10
Tandon *et al.* (2013)	India	695	10-18 (13.4)	Education	2.4 (7/297)	17.5 (36/206)	28.1 (54/192)	7
≥90th and <95th Percentile								
Chu and Pan (2007)	Taiwan	2405	6-12 (NR)	Education	0.3 (5/1753)	2.2 (8/361)	9.6 (28/291)	6
Genovesi *et al.* (2010)	Italy	5131	5-11 (NR)	Education	1.1 (41/3780)	5.6 (57/1025)	11.7 (38/326)	7
Guo *et al.* (2012)	China	6802	5-18	Education	14.4	16.4	17.1	6

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
			(10.9)		(752/5230)	(178/1088)	(83/484)	
Pecin *et al.* (2013)	Croatia	756	15-18 (15.9)	Education	6.3 (4/64)	15.8 (3/19)	24.0 (6/25)	4
Rafraf, Pourghassem Gargari and Safaiyan (2010)*	Iran	985	14-17 (15.7)	Education	13.7 (109/795)	14.8 (24/162)	14.3 (4/28)	6
Ramos and Barros (2005)	Portugal	2023	13-13 (NR)	Education	11.9 (175/1476)	16.6 (55/331)	18.9 (39/206)	6
Wang *et al.* (2013)	China	3373	6-18 (12)	Education	3.0 (47/1541)	7.8 (50/637)	16.1 (192/1195)	7
Xu *et al.* (2012)	China	8764	7-11 (8.6)	Education	1.7 (35/2044)	4.2 (14/330)	7.3 (20/273)	9
≥95th Percentile								
Bar Dayan *et al.* (2005)	Israel	76732	17-17 (NR)	Community	0.1 (70/53684)	0.5 (44/9202)	2.6 (75/2897)	8
Duzova *et al.* (2013)	Turkey	3622	5-18 (11.9)	Community	5.5 (146/2656)	7.6 (25/331)	11.4 (36/317)	5

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Dyson *et al.* (2014)a (China)	China	4444	12-18 (NR)	Education	4.0 (148/3706)	7.7 (46/596)	23.2 (33/142)	6
Dyson *et al.* (2014)b (India)	India	4197	12-18 (NR)	Education	9.4 (378/4025)	25.8 (40/155)	47.1 (8/17)	6
Dyson *et al.* (2014)c (Mexico)	Mexico	4089	12-18 (NR)	Education	11.1 (286/2572)	16.6 (195/1174)	28.3 (97/343)	6
Genovesi *et al.* (2010)	Italy	5131	5-11 (NR)	Education	1.3 (50/3780)	5.7 (58/1025)	20.9 (68/326)	7
Moura *et al.* (2004)	Brazil	1253	7-17 (12.4)	Education	8.1 (88/1081)	12.1 (14/116)	28.6 (16/56)	6
Nur *et al.* (2008)	Turkey	1020	14-18 (15.9)	Education	4.1 (39/962)	19.4 (7/36)	0.0 (0/2)	5
Pecin *et al.* (2013)	Croatia	756	15-18 (15.9)	Education	14.1 (9/64)	21.1 (4/19)	20.0 (5/25)	4
Pereira *et al.* (2009)	Brazil	494	2-19 (9.9)	Education	8.9 (34/383)	12.5 (6/48)	28.6 (18/63)	7
Ramos and Barros (2005)	Portugal	2023	13-13 (NR)	Education	17.5 (259/1476)	29.9 (99/331)	41.7 (86/206)	6
Wiegand *et al.*	Germany	16390	1-20	Clinical	11.4	10.8	20.9	6

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
(2010)	Austria Switzerland		(12.5)		(60/528)	(291/2697)	(2748/13165)	
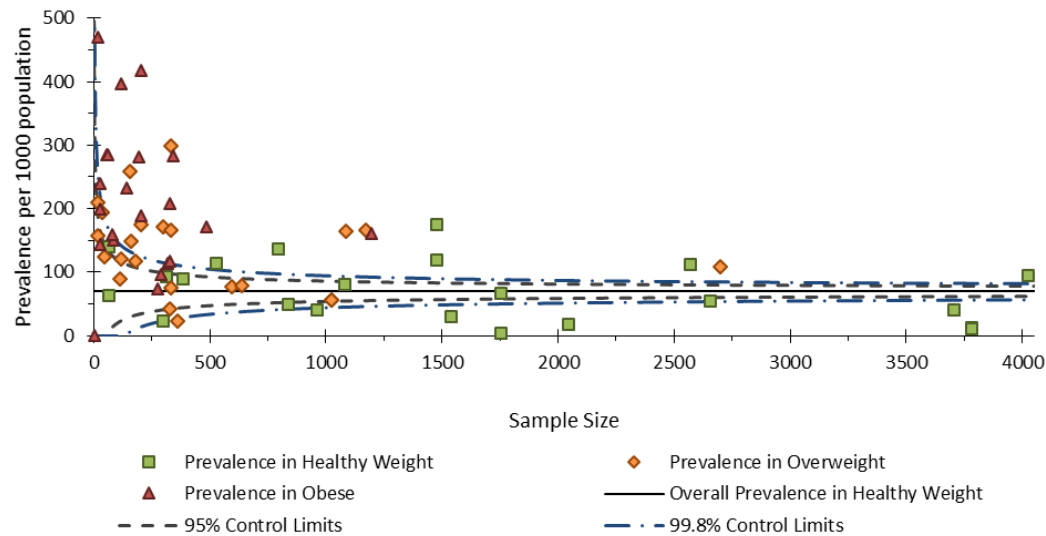
Figure 28: Funnel plot showing prevalence of High Blood Pressure per 1000 population, truncated x-axis

All 24 data points were included in the meta-analyses, the average PR indicated that (for equally sized groups) for every 1 child of healthy weight there would be 2.1 (CI 1.7to 2.6, p<0.001) who are overweight (Figure 29) and 4.0 (CI 2.8 to 5.7, p<0.001) with obesity (Figure 30) who have high blood pressure. The I^2 Statistics indicated high heterogeneity (Higgins *et al.* 2003).

Study	Overweight Events	Overweight Total	Healthy Weight Events	Healthy Weight Total	Weight	Prevalence Ratio Random, 95% CI
Bar-Dayan et al. (2005)	44	9202	70	53684	4.71	3.7 (2.52 to 5.34)
Caserta et al. (2010)	21	179	29	314	4.09	1.3 (0.75 to 2.16)
Chu and Pan (2007)	8	361	5	1753	2.16	7.8 (2.56 to 23.61)
Duzova et al. (2013)	25	331	146	2656	4.58	1.4 (0.91 to 2.07)
Dyson et al, (2014) (India)	40	155	378	4025	5.04	2.7 (2.07 to 3.65)
Dyson et al. (2014) (China)	46	596	148	3706	4.92	1.9 (1.4 to 2.66)
Dyson et al. (2014) (Mexico)	195	1174	286	2572	5.37	1.5 (1.26 to 1.77)
Genovesi et al. (2010)	57	1025	41	3780	4.63	5.1 (3.45 to 7.61)
Genovesi et al. (2010)	58	1025	50	3780	4.73	4.3 (2.95 to 6.21)
Guo et al. (2012)	178	1088	752	5230	5.41	1.1 (0.98 to 1.32)
Khader et al. (2011)	10	111	41	837	3.56	1.8 (0.95 to 3.57)
Moura et al. (2004)	14	116	88	1081	4.09	1.5 (0.87 to 2.52)
Nur et al. (2008)	7	36	39	962	3.3	4.8 (2.31 to 9.98)
Pecin et al. (2013)	3	19	4	64	1.58	2.5 (0.62 to 10.31)
Pecin et al. (2013)	4	19	9	64	2.28	1.5 (0.52 to 4.32)
Pereira et al. (2009)	6	48	34	383	3.01	1.4 (0.62 to 3.18)
Rafraf et al. (2010)	24	162	109	795	4.58	1.1 (0.72 to 1.63)
Ramos and Barros (2005)	55	331	175	1476	5.06	1.4 (1.06 to 1.85)
Ramos and Barros (2005)	99	331	259	1476	5.29	1.7 (1.4 to 2.08)
Seki et al. (2008)	51	299	116	1755	4.97	2.6 (1.9 to 3.5)
Tandon et al. (2013)	36	206	7	297	3.09	7.4 (3.37 to 16.34)
Wang et al. (2013)	50	637	47	1541	4.67	2.6 (1.75 to 3.79)
Wiegand et al. (2010)	291	2697	60	528	5.11	0.9 (0.73 to 1.23)
Xu et al. (2012)	14	330	35	2044	3.77	2.5 (1.35 to 4.55)
Total (random effects)	1336	20478	2928	94803	100	2.1 (1.71 to 2.59)

Heterogeneity: Q = 181.23; df = 23 (P < 0.0001); I^2 = 87.31%
Test for overall effect: Z = 6.99 (P = <0.001)

Figure 29: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for High Blood Pressure
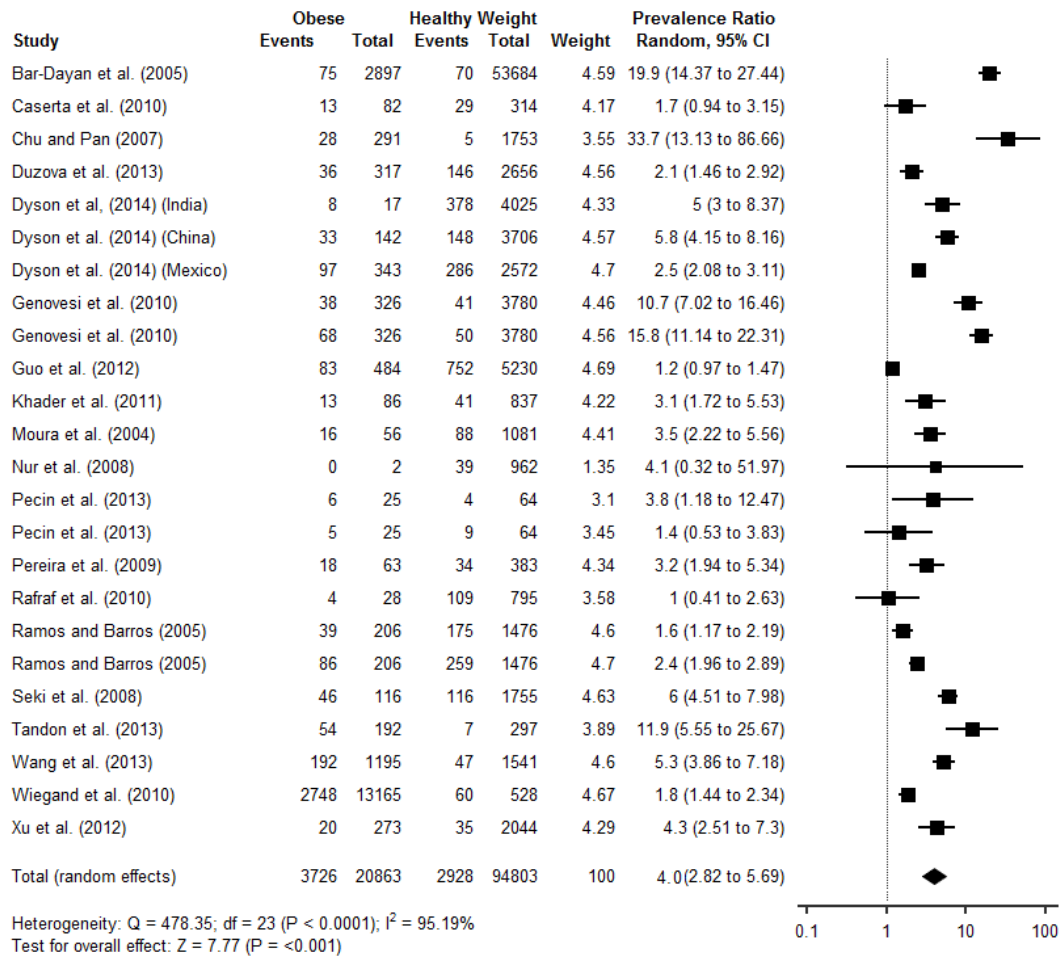
Figure 30: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for High Blood Pressure

Some of the variation in reported prevalence may be due to measurement error. Blood pressure is difficult to measure accurately due to the range of factors which might lead to a false positive, although some studies provided details regarding the methods, e.g. the average of multiple measurements (Karnath 2002; Bauldry, Bollen and Adair 2015). As well as potential misclassification due to staff, there may be variation in the accuracy of the measurement device. Most of the studies reported using a manual calibrated mercury sphygmomanometer, whereas Dyson reported using an automated device. Previous studies have reported that automatic devices are less accurate (lower sensitivity and specificity) than manual devices, so that comparisons between automatic and manual devices may not be appropriate (Mansoor *et al.* 2016; Shahbabu *et al.* 2016). However the reported prevalence by Dyson et al. was comparable with other studies, suggesting other factors are involved in determining the prevalence of high blood pressure.

Differences in eligibility criteria may explain some of the observed variation in prevalence. Caserta *et al.* (2010) recruited a random sample of 576 children aged 11-13 from the school census list in Reggio Calabrai, Italy, without eligibility criteria. Similarly Khader et al. did not provide eligibility criteria for the 1034 children aged 7-18 recruited in Jordan; neither did Tandon et al. for the 695 participants aged 10-18 years recruited from multiple schools in Delhi, India. In contrast, Seki et al. reported excluding those with severe illness/disease that may affect results, those who were pregnant, as well as irregularities in data/sample collection, though this is not elaborated upon. By excluding particular individuals Seki et al. had a more narrowly defined sample, thus differences in prevalence are more likely due to increases in weight status. Some of the observed variation may be due to country-specific factors. For instance Dyson et al. recruited from three countries (China, India, Mexico) using the same methodology, and reported marked differences in prevalence, particularly for India (Figure 31). Thus a generic screening tool, which does not account for potential confounders, may not be appropriate.
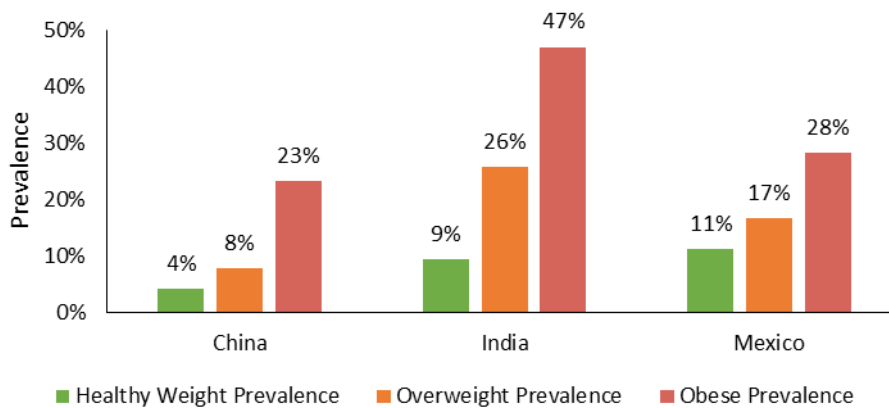
Figure 31: Prevalence of High Blood Pressure (≥95th Percentile), reported by Dyson *et al.* (2014)

Overall the results suggested higher prevalence of high blood pressure in those with obesity compared to the healthy weight and overweight groups. The meta-analyses suggested a significant prevalence ratio for both comparisons, yet the high I^2 statistics suggested that there was considerable heterogeneity between the studies which may be due to factors such as study methodology, measurement error, subjectivity of the measurer, and/or country of study. Additionally the lack of studies from the UK restricted generalisability of the results.

4.2.4 Metabolic Syndrome

Metabolic Syndrome is not a condition itself, but is a composite of high blood pressure, high blood sugar, excess body fat, and dyslipidaemia. Four definitions of metabolic syndrome were identified during the review, with no current consensus on which definition to use, and different definitions resulting in somewhat different estimates of prevalence (Kassi *et al.* 2011)(Figure 32). Body size or weight is included in all commonly-used definitions, but for three definitions this is not a mandatory criterion (Table 18). Furthermore, there is variation between the definitions in how body size is measured; Cook et al and de Ferranti assess obesity using waist circumference, but using different cut-offs, whereas the International Diabetes Federation (IDF) and the National Cholesterol Education Program (NCEP) definitions are based on BMI percentiles, again using different cut-offs, limiting comparability of the results. Therefore the definitions were investigated separately.

Table 18: Definitions of Metabolic Syndrome

	Cook et al.	de Ferranti et al.	IDF	NCEP
Criteria	Three of five	Three of five	Obesity, plus two others	Three of five
1. Obesity	Waist circumference: ≥90th percentile	Waist circumference: ≥75th percentile	BMI > 90th percentile	BMI >95th Percentile
2. Hyperglycaemia	Fasting glucose: ≥100mg/dL	Fasting glucose: ≥110mg/dL	Fasting glucose: ≥100mg/dL	Fasting glucose: ≥100mg/dL
3. Dyslipidaemia	Triglycerides: ≥110mg/dL	Triglycerides: ≥100mg/dL	Triglycerides: ≥150mg/dL	Triglycerides: ≥110mg/dL
4. Dyslipidaemia (second criteria)	HDL Cholesterol: ≤40mg/dL	HDL Cholesterol: <45mg/dL	HDL Cholesterol: <40mg/dL	HDL Cholesterol: <40mg/dL
5. Hypertension	Blood Pressure: ≥90th Percentile	Blood Pressure: ≥90th Percentile	Blood Pressure: ≥130/85mmhg	Blood Pressure: >95th Percentile
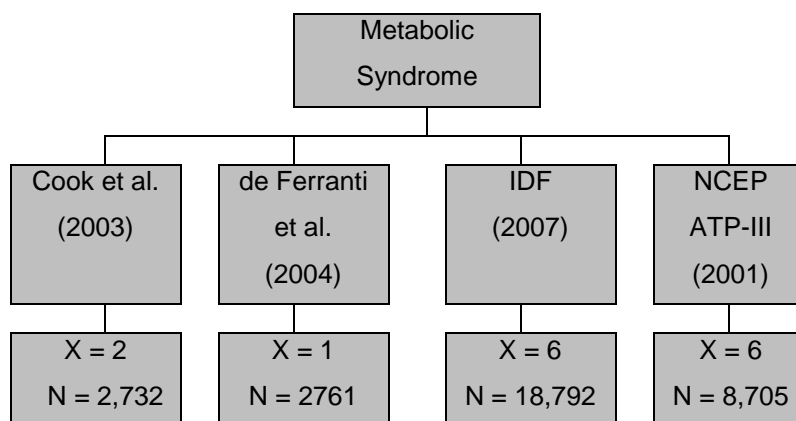
Figure 32: Overview of Metabolic Syndrome and the four definitions (X: number of studies; N: number of participants) (number of studies is higher as some studies reported the prevalence across multiple screening methods)

Metabolic syndrome according to Cook et al.

Two studies measured the prevalence of metabolic syndrome using the cut-off developed by Cook *et al.* (2003), both had a quality score of seven (Table 19). Khader et al. had not conducted an a priori sample size calculation, whereas Messiah et al. selected all Mexican American, non-Hispanic white, and non-Hispanic black boys and girls aged 8 to 14 years from the combined 1999 to 2000 and 2001 to 2002 NHANES (National Health and Nutrition Examination Survey) data, potentially providing a more representative sample. Both studies indicated that the prevalence of metabolic syndrome increased with weight status (Table 19). The prevalence in the obese group ranged from 11.6% (Khader *et al.* 2011) to 20.8% (Messiah *et al.* 2008).

The funnel plot showed that the studies were fairly similar in terms of reported prevalence by weight status of the healthy weight and overweight groups (Figure 33). The healthy weight estimate for Messiah *et al.* (2008) fell below the 95% control limit whereas for Khader et al. it was just inside the control limit. The data points for the overweight group were also within the 95% control limits; in contrast the data points for the obese were considerably above the 99.8% control limits. This suggested the prevalence of metabolic syndrome increased dramatically as weight reached higher levels; however without continuous data it is not possible to identify if there is a tipping point, and the small number of studies means the results offer limited generalisability.

Table 19: Summary of studies reporting prevalence data for Metabolic Syndrome (Cook et al.)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (Mean Age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Khader *et al.* (2011)	Jordan	1034	7-18 (NR)	Community	1.8 (15/837)	2.7 (3/111)	11.6 (10/86)	7
Messiah *et al.* (2008)	America	1698	8-14 (NR)	Community	0.3 (3/1049)	1.4 (4/289)	20.8 (75/360)	7
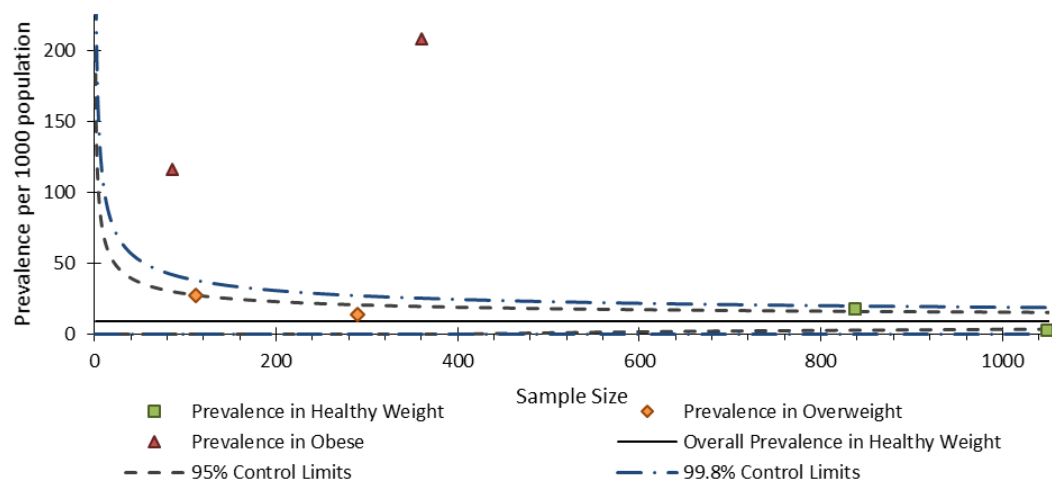
Figure 33: Funnel plot showing prevalence of Metabolic Syndrome (Cook et al.) per 1000 population

The meta-analyses of both studies indicated that (assuming equal population size) for every 1 child of a healthy weight there would be 2.5 (CI 0.8 to 7.7, p=0.1) who were overweight (Figure 34) and 21.2 (CI 1.2 to 381.6, p=0.04) with obesity (Figure 35) who have metabolic syndrome based on Cook et al.'s definition. The I^2 Statistic was 28.8% in Figure 34 and 94.3% in Figure 35 indicating low to high heterogeneity.

Study	Overweight Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Khader et al. (2011)	3	111	15	837	56.96	1.5 (0.4 to 5.1)
Messiah et al. (2008)	4	289	3	1049	43.04	4.8 (1.1 to 21.5)
Total (random effects)	7	400	18	1886	100.00	2.5 (0.8 to 7.7)

Heterogeneity: Q = 1.4; df = 1 (P = 0.2361); I² = 28.75%
Test for overall effect: Z = 1.58 (P = 0.114)
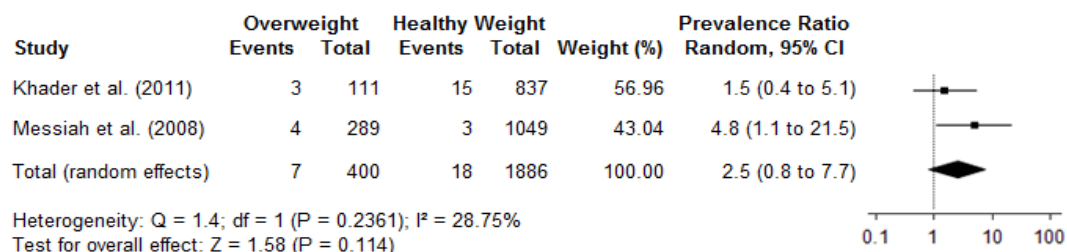
Figure 34: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for Metabolic Syndrome (Cook et al. 2003)

Study	Obese Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Khader et al. (2011)	10	86	15	837	51.09	6.5 (3.01 to 14.0)
Messiah et al. (2008)	75	360	3	1049	48.91	72.8 (23.1 to 229.6)
Total (random effects)	85	446	18	1886	100.00	21.2 (1.2 to 381.6)

Heterogeneity: Q = 17.53; df = 1 (P < 0.0001); I² = 94.30%
Test for overall effect: Z = 2.07 (P = 0.038)
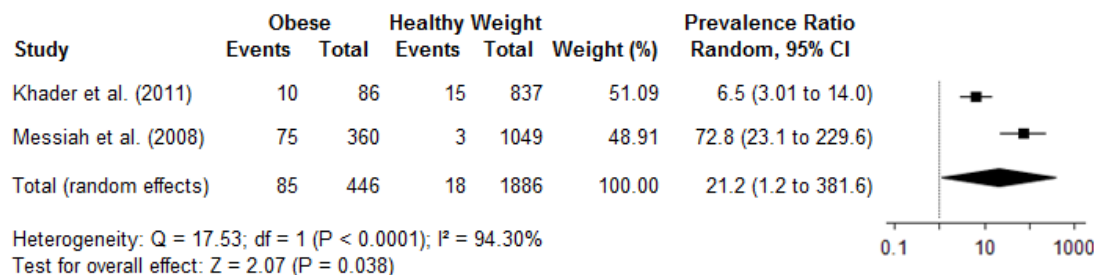
Figure 35: Forest plot of prevalence ratio for obese, relative to healthy weight, children for Metabolic Syndrome (Cook et al. 2003)

Khader et al. conducted a population survey of households in Jordan, recruiting 1034 children aged 7-18, whereas Messiah et al. utilised NHANES data from 1999-2002 of 1698 children aged 8-14 years. Messiah et al. included Mexican American, non-Hispanic white, and non-Hispanic black boys and girls in their study, whereas Khader et al. participants were selected from the 12 governorates of Jordan. In the latter study, no details of ethnicity were provided, so its impact on prevalence could not be determined.

Both Khader et al. and Messiah et al. provided prevalence data by age group, Khader et al. for 7-12 and 13-18 year olds and Messiah et al. for 8-11 and 12-14 year olds. Given that obesity tends to increase with age it would be expected that the prevalence of metabolic syndrome would increase with age if it is associated with weight (NHS Digital 2016). This was reflected in the data (Figure 36). However the data from Messiah et al. suggested the prevalence increased from 6% to 44% between the age groups in the obese, much steeper than that reported by Khader et

al. This may partly be due to the low number of children in the obese group (86) affecting the precision.
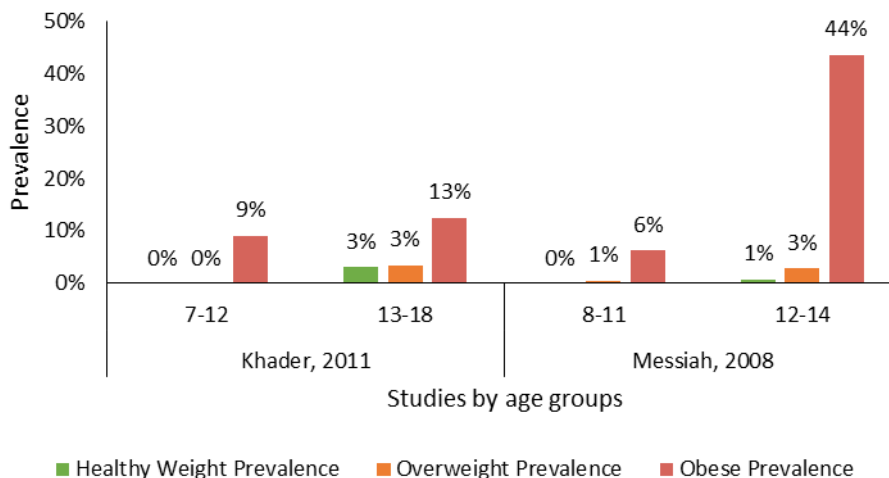
Figure 36: Prevalence of Metabolic Syndrome (Cook et al.) by age sub-groups

Overall the results suggested a higher prevalence of metabolic syndrome in children with obesity, although there was heterogeneity between studies and these analyses are based on only two studies. Additionally there were no UK based studies, further limiting application of the results to a UK population.

Metabolic syndrome according to de Ferranti

One study reported the prevalence of metabolic syndrome using de Ferranti's definition (de Ferranti *et al.* 2004). Li et al. reviewed data of 2761 Chinese adolescents from the 2002 China National Nutrition and Health Survey, which is a nationally representative cross-sectional survey covering thirty-one provinces. The estimated prevalence in this study increased with an increase in weight, from 2.3% in the healthy weight group to 35% in the obese group (Table 20). Although weight is not a mandatory criterion for de Ferranti's definition, it is included as one of the components. Therefore those who are overweight or obese already meet one of the three requirements for metabolic syndrome. The lack of data from other studies meant it was not possible to assess if this would be consistent across studies, nor was it possible to conduct a meta-analyses.

Table 20: Summary of studies reporting prevalence data for Metabolic Syndrome (de Ferranti, et al., 2004)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Li *et al.* (2008)	China	2761	15-19 (17.1)	Community	2.3 (60/2613)	23.4 (22/94)	35.2 (19/54)	7

Metabolic syndrome according to the International Diabetes Federation (IDF)

Seven studies reported the prevalence of metabolic syndrome using the International Diabetes Federation's cut-off (Zimmet *et al.* 2007) (Table 21, page 99). Two articles were based on the same population, thus the data from Garg *et al.* (2014) was removed (Tandon *et al.* 2013; Garg *et al.* 2014). For the remaining six studies, quality scores ranged from four to nine, with four classed as high quality; there was no association between quality score and prevalence.

The studies suggested that the prevalence of metabolic syndrome increased with weight status (Table 21). Five of the six studies reported a prevalence in the healthy weight group of less than 1%, whilst Mehairi et al. estimated it at 5% (Chen *et al.* 2012; Papoutsakis *et al.* 2012; Xu *et al.* 2012; Tandon *et al.* 2013; Wang *et al.* 2013). For the overweight and obese groups the data from Mehairi et al. was more similar to the other studies, whereas the prevalence reported by Papoutsaki et al. and Xu et al. were consistently lower.

The funnel plot suggested considerable variation between the studies, with no clear clustering by weight status (Figure 37). Only two of the healthy weight data points fell within the 95% control limits, one was within the 99.8% control limits, and the rest were outside the 99.8% control limits, suggesting considerable heterogeneity (Papoutsakis *et al.* 2012; Xu *et al.* 2012; Tandon *et al.* 2013). Similar between study variation was observed for the overweight and obese groups, two of the overweight data points fell within the 95% control limits for the healthy weight group (Papoutsakis *et al.* 2012; Xu *et al.* 2012).

Table 21: Summary of studies reporting prevalence data for Metabolic Syndrome (International Diabetes Federation)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Chen *et al.* (2012)	China	3814	10-18 (NR)	Community	0.2 (4/1745)	10.0 (116/1155)	27.6 (252/914)	6
Mehairi *et al.* (2013)	United Arab Emirates	1018	12-18 (15.4)	Education	5.1 (33/645)	13.1 (23/175)	40.3 (77/191)	8
Papoutsakis *et al.* (2012)	Greece	1128	9-13 (11.2)	Education	0.0 (0/698)	0.0 (0/327)	7.8 (8/103)	4
Tandon *et al.* (2013)	India	695	10-18 (13.4)	Education	0.3 (1/297)	13.6 (28/206)	46.4 (89/192)	7
Wang *et al.* (2013)	China	3373	6-18 (12)	Education	0.0 (0/1541)	3.1 (20/637)	11.9 (142/1195)	7
Xu *et al.* (2012)	China	8764	7-11 (8.6)	Education	0.5 (10/2044)	0.9 (3/330)	6.6 (18/273)	9
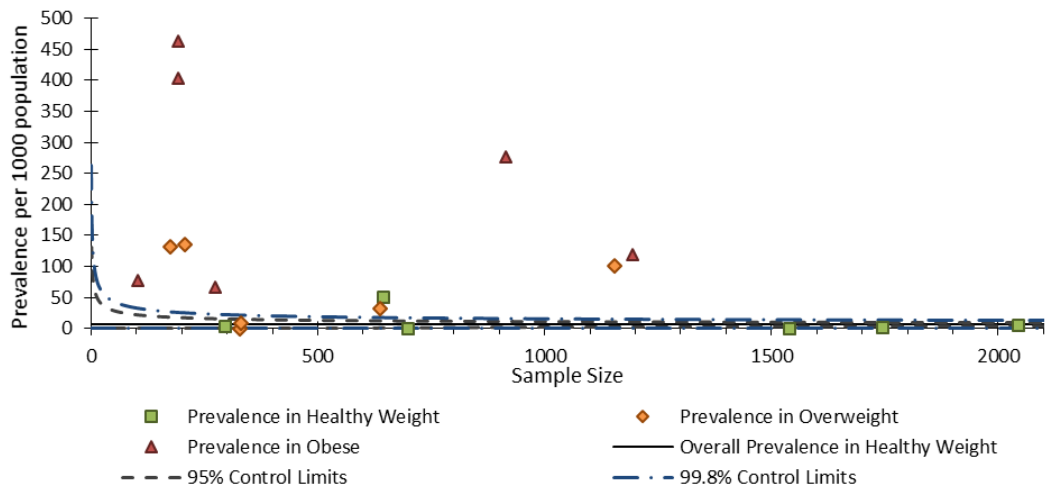
Figure 37: Funnel plot showing prevalence of Metabolic Syndrome (IDF) per 1000 population

All six studies were included in the meta-analyses, the average PR indicated that (assuming equal population sizes) for every 1 child of a healthy weight there were 13.1 (CI 1.9 to 88.9, p=0.008) who were overweight (Figure 38) and 53.9 (CI 9.7 to 298.0, p<0.001) with obesity (Figure 39) who had metabolic syndrome based on the IDF definition. The I^2 statistic was high in Figure 38 (92.3%) and Figure 39 (93.7%), indicating high between-study heterogeneity.

Study	Overweight Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Chen et al. (2012)	116	1155	4	1745	21.91	43.8 (16.2 to 118.4)
Mehairi et al. (2013)	23	175	33	645	22.91	2.6 (1.6 to 4.3)
Papoutsakis et al. (2012)	0	327	0	698	0	Not Available
Tandon et al. (2013)	28	206	1	297	18.6	40.4 (5.5 to 294.4)
Wang et al. (2013)	20	637	0	1541	15.51	99.1 (6.0 to 1636.0)
Xu et al. (2012)	3	330	10	44	21.07	1.9 (0.51 to 6.7)
Total (random effects)	190	2830	48	6970	100.00	13.1 (1.94 to 88.86)

Heterogeneity: Q = 52.21; df = 4 (P < 0.0001); I² = 92.34%
Test for overall effect: Z = 2.64 (P = 0.008)

Figure 38: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for Metabolic Syndrome (IDF 2007)

Study	Obese		Healthy Weight			Prevalence Ratio
	Events	Total	Events	Total	Weight (%)	Random, 95% CI
Chen et al. (2012)	252	914	4	1745	18.84	120.3 (44.9 to 321.9)
Mehairi et al. (2013)	77	191	33	645	19.9	7.9 (5.4 to 11.5)
Papoutsakis et al. (2012)	8	103	0	698	12.92	114.3 (6.6 to 1965.0)
Tandon et al. (2013)	89	192	1	297	15.89	137.7 (19.3 to 979.9)
Wang et al. (2013)	142	1195	0	1541	13.14	367.5 (22.9 to 5897.5)
Xu et al. (2012)	18	273	10	44	19.32	13.5 (6.3 to 28.9)
Total (random effects)	586	2868	48	6970	100.00	53.8 (9.7 to 298.0)

Heterogeneity: Q = 79.39; df = 5 (P < 0.0001); I² = 93.70%
Test for overall effect: Z = 4.57 (P <0.001)

Figure 39: Forest plot of prevalence ratio for obese, relative to healthy weight, children for Metabolic Syndrome (IDF 2007)

Three of the six studies were conducted in China, with fairly large sample sizes (Chen et al., Wang et al, and Xu et al.). Although the prevalence in the healthy weight group was consistently low, there was considerable variation in the overweight and obese groups (Figure 40). Part of this variation may be due to the population sampled. Both Chen et al. and Wang et al. recruited a sub-sample of the Beijing Child and Adolescent Metabolic Syndrome (BCAMS) cohort. The complete cohort consisted of 19,593 children, thus it was difficult to ascertain if any of the participants were included in both studies. In contrast, Xu et al. recruited from six provincial capital cities (including Beijing). However they did not provide prevalence by city limiting comparison, therefore the between study variation may be due to regional variations, perhaps in diet and lifestyle, as well as methodological differences.
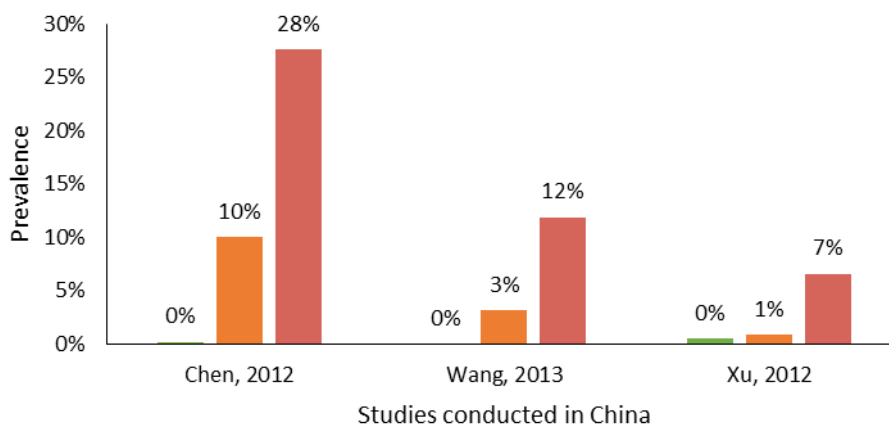
Figure 40: Prevalence of Metabolic Syndrome (IDF), restricted to studies conducted in China

Mehairi et al. and Tandon et al. reported similar prevalence for the overweight and obese groups, and both recruited a random sample of schoolchildren. Mehairi et al. recruited 1018 schoolchildren from eight selected schools in Al Ain Abu Dhabi, United Arab Emirates (UAE), and Tandon et al. recruited 695 schoolchildren across Delhi, India. A large proportion of UAE citizens are from India, thus there may be similarities between the two study populations related to diet, ethnicity, and lifestyle (Global Media Insights 2018). However, the studies did not report ethnicity data, and without additional data definitive conclusions cannot be made.

Overall, the results suggest that a higher prevalence of Metabolic Syndrome using the IDF criteria was observed in children with obesity than those of a healthy weight (Zimmet *et al.* 2007). The association is likely influenced by multiple factors, such as study methodology, country and city of study. Again, there were no studies conducted on a UK sample, limiting generalisability of the results.

Metabolic syndrome according to NCEP

Six studies that measured the prevalence of metabolic syndrome based on the NCEP definition were identified (Expert Panel on Detection 2001). Quality scores ranged from five to 10, with no clear association between quality score and reported prevalence (Table 22). Five studies did not conduct an a priori sample size calculation, thus it was unclear if the sample was sufficient to provide a representative assessment of prevalence. Additionally there was uncertainty around the reliability of the measure, however this related predominantly to measurements of blood pressure.

Data suggested that the prevalence of metabolic syndrome increased with weight status (Table 22). There was within weight group variation, in the healthy weight group prevalence ranged from 0.0% to 1.3%, 1.7% to 18% in the overweight group, and 28% to 49% in the obese group. This indicated that weight status group alone may not explain the observed variation.

The funnel plot showed considerable variation in prevalence between the weight categories (Figure 41). All bar one of the data points for the healthy weight group fell within the 95% control limits, suggesting fairly homogenous studies (Xu and Ji 2008). Additionally for the overweight and obese groups the data points were clustered together, supporting an association between prevalence and weight status.

Table 22: Summary of studies reporting prevalence data for Metabolic Syndrome (NCEP)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Davis *et al.* (2005)	America	211	7-18 (NR)	Education	1.3 (1/77)	15.6 (5/32)	41.9 (18/43)	8
Elizondo-Montemayor *et al.* (2014)	Mexico	236	6-12 (9)	Education	0.0 (0/44)	1.7 (1/58)	40.3 (54/134)	5
Seki, Matsuo and Faria Carrilho (2008)	Brazil	2170	6-16 (11.3)	Education	0.3 (5/1755)	10.7 (32/299)	34.5 (40/116)	10
Tandon *et al.* (2013)	India	695	10-18 (13.4)	Education	1.0 (3/297)	18.4 (38/206)	49.0 (94/192)	7
Wang *et al.* (2013)	China	3373	6-18 (12)	Education	1.0 (16/1541)	7.7 (49/637)	29.4 (351/1195)	7
Xu and Ji (2008)	China	2020	14-16 (14.6)	Education	0.2 (3/1490)	6.0 (23/385)	28.3 (41/145)	7
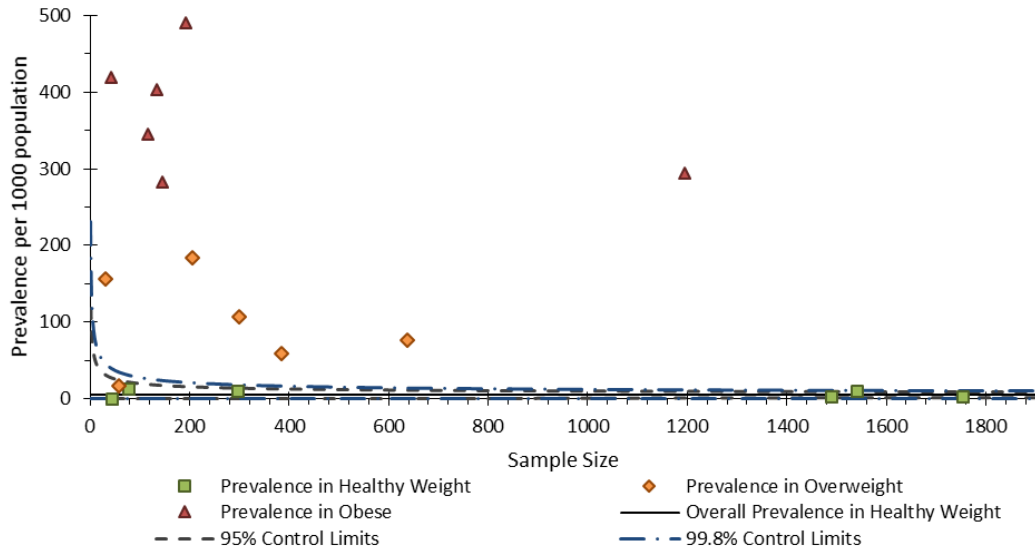
Figure 41: Funnel plot showing prevalence of Metabolic Syndrome (NCEP) per 1000 population

All six studies were included in the meta-analyses, the average PR indicated that (assuming equal population sizes) for every 1 child of a healthy weight there were 15.8 (CI 7.4 to 33.8, p<0.001) who were overweight (Figure 42) and 57.9 (CI 27.7 to 121.3, p<0.001) with obesity (Figure 43) who have metabolic syndrome based on the NCEP ATP III definition. The I^2 statistic indicated moderate heterogeneity between studies in both Figure 42 (59%) and Figure 43 (61%).

Study	Overweight Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Davis et al. (2005)	5	32	1	77	9.3	12.0 (1.46 to 99.0)
Elizondo-Montemayor et al. (2014)	1	58	0	44	4.87	2.3 (0.1 to 54.9)
Seki et al. (2008)	32	299	5	1755	21.77	37.6 (14.8 to 95.6)
Tandon et al. (2013)	38	206	3	297	18.46	18.3 (5.7 to 58.4)
Wang et al. (2013)	49	637	16	1541	27.62	7.4 (4.3 to 12.9)
Xu et al. (2012)	23	385	3	1490	17.97	29.7 (9.0 to 98.3)
Total (random effects)	148	1617	28	5204	100.00	15.8 (7.4 to 33.8)

Heterogeneity: Q = 12.22; df = 5 (P = 0.0319); I² = 59.08%
Test for overall effect: Z = 7.11 (P <0.001)

Figure 42: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for Metabolic Syndrome (NCEP ATP III 2003)

Study	Obese Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Davis et al. (2005)	18	43	1	77	9.65	32.2 (4.5 to 233.2)
Elizondo-Montemayor et al. (2014)	54	134	0	44	5.82	36.3 (2.3 to 576.4)
Seki et al. (2008)	40	116	5	1755	21.26	121.0 (48.7 to 300.9)
Tandon et al. (2013)	94	192	3	297	18.04	48.5 (15.6 to 150.8)
Wang et al. (2013)	351	1195	16	1541	27.53	28.3 (17.2 to 46.4)
Xu et al. (2012)	41	145	3	1490	17.7	140.4 (44.0 to 447.9)
Total (random effects)	598	1825	28	5204	100.00	57.9 (27.7 to 121.3)

Heterogeneity: Q = 12.82; df = 5 (P = 0.0251); I^2 = 61%
Test for overall effect: Z = 10.77 (P <0.001)

Figure 43: Forest plot of prevalence ratio for obese, relative to healthy weight, children for Metabolic Syndrome (NCEP ATP III 2003)

A primary reason for the variation may be the low reported prevalence in the healthy weight group, which would increase the prevalence ratio in comparison to studies with a slightly higher prevalence. For instance, for Davis et al. the prevalence in the healthy weight group is 0.01 (1/77) in comparison the prevalence for Seki et al. is 0.002 (5/1755). Thus, dividing the prevalence in the overweight/obese groups by a smaller healthy weight prevalence would alter the prevalence ratio for each study, and thus impact the estimated population prevalence.

Another reason for the variation in prevalence may include participant eligibility criteria. Tandon et al. recruited 695 children aged 10-18 years in Delhi, India, without any explicitly stated eligibility criteria. Wang et al. recruited a much larger cross-section of 3373 children aged 6-18 year olds from the BCAMS cohort, and excluded those with physical disability, congenital diseases, and any diagnosed medical condition that might influence metabolism and therefore test results. With this narrowly defined sample a lower prevalence would be expected in comparison to Tandon et al., which is observed. Two studies recruited from a Chinese population, with similar age distributions and large sample sizes (3370 versus 2020), and reported fairly similar prevalence across the weight statuses (Wang et al. and Xu et al.). This would suggest the variation in prevalence may be due, in part, to differences in the population sampled, as well as study specific differences. Davis et al. considered differences between ethnic groups, reporting a lower prevalence among Black children than White children, 10% versus 17%, respectively. However, these comparisons are crude and no adjustment is made for potential confounders, such as age, sex and sampling characteristics.

Both Tandon *et al.* (2013) and Wang *et al.* (2013) reported prevalence using both the IDF and NCEP definitions (Table 23). In Wang et al. the prevalence among obese children increased from 12% using the IDF definition to 29% for NCEP.

However, for Tandon et al. a similar increase is not observed, suggesting interplay between many factors.

Table 23: Comparing prevalence based on two definitions of Metabolic Syndrome

Study	Prevalence (%) in Obese Group (IDF)	Prevalence (%) in Obese Group (NCEP)
Tandon et al. (2013)	46.4	49.0
Wang et al. (2013)	11.9	29.4

Overall, the meta-analysis estimated a considerably higher prevalence of metabolic syndrome in children who were overweight and obese than in those of a healthy weight. Potential factors for this include country of study, ethnicity (though this was not often reported) and the definition used. Again, there were no UK studies.

4.2.5 Non-Alcoholic Fatty Liver Disease

Non-alcoholic fatty liver disease (NAFLD) is the accumulation of fat in the liver not due to excessive alcohol consumption. Seven studies assessed for NAFLD using three methods (Figure 44).
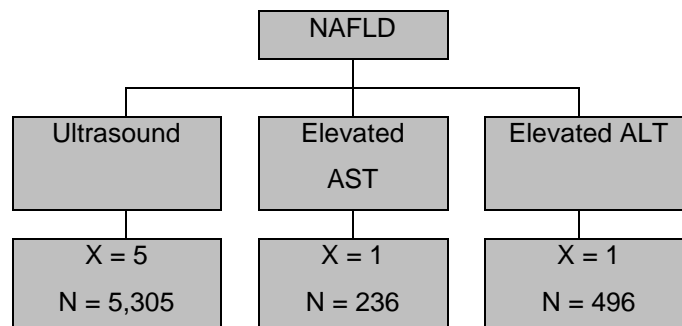
Figure 44: Overview of NAFLD and the 3 screening methods (X: number of studies; N: number of participants)

NAFLD screened by ultrasound

Five studies diagnosed NAFLD via ultrasound and all five suggested that the prevalence of NAFLD increased with increasing weight group (Table 24), yet there was considerable variation in prevalence between the studies. The prevalence for the healthy weight group ranged from 0.4% to 6.8%, with similar variation in the overweight and obese groups, suggesting between-study heterogeneity (Ayonrinde *et al.* 2011; Lawlor *et al.* 2014).

Table 24: Summary of studies reporting prevalence data for NAFLD (ultrasound)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Adibi *et al.* (2009)	Iran	952	6-18 (12.6)	Education	1.0 (4/408)	10.5 (33/314)	53.9 (124/230)	6
Alavian *et al.* (2009)	Iran	966	7-18 (NR)	Education	2.6 (18/700)	11.0 (21/191)	32.0 (24/75)	4
Ayonrinde *et al.* (2011)	Australia	1138	16-17 (17)	Other	6.8 (60/878)	21.3 (34/160)	62.0 (62/100)	7
Gong *et al.* (2013)	China	538	9-15 (12)	Education	1.1 (3/283)	13.0 (15/115)	52.1 (73/140)	5
Lawlor *et al.* (2014)	UK	1711	NR-NR (NR)	Community	0.4 (5/1226)	4.3 (12/279)	22.2 (26/117)	6

Quality scores ranged from four to seven, with little evidence of an association between prevalence and quality score. The key areas where studies lost points related to a lack of information regarding the adequacy of the sample size, a lack of detail regarding the participants and setting, and not accounting for important confounding factors, such as ethnicity.

The funnel plot indicated some clustering by weight category (Figure 45). Two of the five healthy weight prevalence estimates fell within the 95% control limits (Alavian *et al.* 2009; Gong *et al.* 2013), with Adibi *et al.* (2009) falling just outside the 95% control limit, and Ayonrinde *et al.* (2011) and Lawlor *et al.* (2014) falling above and below the 99.8% control limits, respectively. This suggested a fair degree of homogeneity between the studies, though this was difficult to assess with five studies.
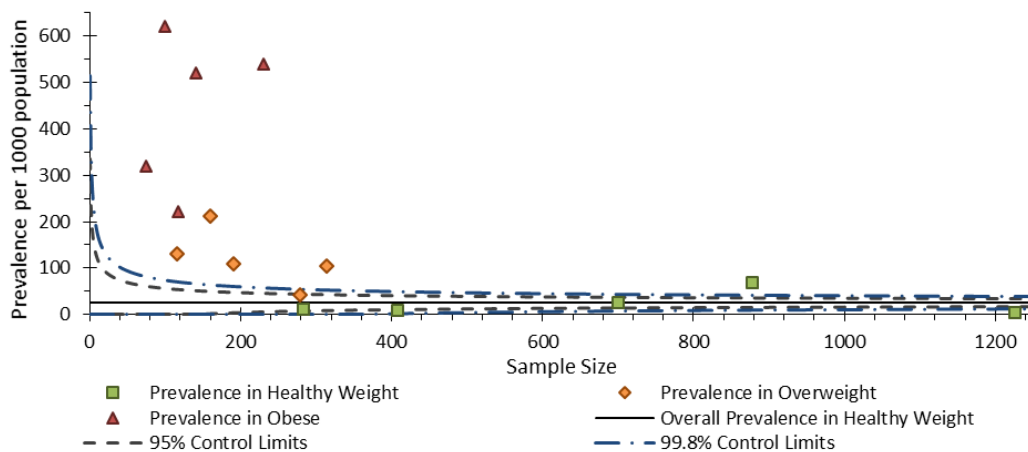
Figure 45: Funnel plot showing prevalence of NAFLD (Ultrasound) per 1000 population

All five studies were included in the meta-analyses, the average PR indicated that (assuming equal population sizes) for every 1 child of a healthy weight there would be 6.1 (CI 3.3 to 11.2, p<0.001) who were overweight (Figure 46) and 26.1 (CI 9.4 to 72.3, p<0.001) with obesity (Figure 47) who have NAFLD. The I^2 Statistic indicated high heterogeneity between studies, 67.9% (Figure 46) and 91.4% (Figure 47). Reasons for this high heterogeneity are unclear, but may in part be due to differences in study methodology, such as eligibility criteria or measurement of NAFLD.

Study	Overweight Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Adibi et al. (2009)	33	314	4	408	16.72	10.7 (3.8 to 29.9)
Alavian et al. (2009)	21	191	18	700	24.28	4.3 (2.3 to 7.9)
Ayonrinde et al (2011)	34	160	60	878	28.45	3.1 (2.1 to 4.6)
Gong et al. (2013)	15	115	3	283	13.96	12.3 (3.6 to 41.7)
Lawlor et al. (2014)	12	279	5	1226	16.59	10.5 (3.8 to 29.7)
Total (random effects)	115	1059	90	3495	100.00	6.1 (3.3 to 11.2)

Heterogeneity: Q = 12.47; df = 4 (P = 0.0142); I² = 67.91%
Test for overall effect: Z = 5.86 (P <0.001)

Figure 46: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for NAFLD (Ultrasound)

Study	Obese Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Adibi et al. (2009)	124	230	4	408	18.92	55.0 (20.6 to 146.9)
Alavian et al. (2009)	24	75	18	700	21.46	12.4 (7.1 to 21.9)
Ayonrinde et al (2011)	62	100	60	878	22.54	9.1 (6.8 to 12.1)
Gong et al. (2013)	73	140	3	283	17.86	49.2 (15.8 to 153.3)
Lawlor et al. (2014)	26	117	5	1226	19.22	54.5 (21.3 to 139.2)
Total (random effects)	309	662	90	3495	100.00	26.1 (9.4 to 72.3)

Heterogeneity: Q = 46.74; df = 4 (P < 0.0001); I² = 91.44%
Test for overall effect: Z = 6.27 (P <0.001)

Figure 47: Forest plot of prevalence ratio for obese, relative to healthy weight, children for NAFLD (Ultrasound)

Two of the studies were conducted in Iran with similar sample sizes (952 and 966, respectively) and similarly aged participants (6-18 and 7-18 years, respectively) (Adibi *et al.* 2009; Alavian *et al.* 2009). They reported comparable prevalence for the healthy weight and overweight groups, yet considerably different prevalence for the obese group (32% versus 54%). The studies were conducted in different parts of Iran, Tehran (the capital) and Isfahan, and the number of children with obesity differed. There were also differences between the studies in eligibility criteria; Adibi et al. excluded participants with a history of diabetes or metabolic diseases, taking medications or symptoms of liver disorders, whereas Alavian et al. only excluded those with a history of alcohol intake.

Lawlor et al., the only UK study, recruited a cross-section of 1711 children from the Avon Longitudinal Study of Parents and Children, which is a prospective, UK population-based cohort study of 14,541 pregnancies (http://www.alspac.bris.ac.uk). Ayonrinde et al. recruited 1771 participants from the Western Australian Pregnancy Cohort (Raine) Study (the Raine Cohort), a prospective cohort of pregnancy, childhood, and adolescents (http://www.rainestudy.org.au). There were substantial differences in results, with the estimated prevalence for the obese group in

Ayonrinde et al. almost three times that reported by Lawlor et al. The results are consistent with the idea that country specific factors, such as diet and lifestyle, may be related to the prevalence of NAFLD, in addition to increased weight.

Overall, the results suggest that obese children were at greater risk of developing NAFLD than their overweight and healthy weight counterparts in these studies. This may be partly due to other factors such as geographical locations and age distribution. One study was conducted in a UK population, and reported the lowest prevalence across the weight categories, although it found a strong association between weight group and prevalence of NAFLD, one study is not sufficient to draw conclusions for the whole of the UK (Lawlor et al.).

NAFLD screened by AST and ALT

As there was only one study screening for NAFLD using AST (>33U/L for males, >26U/L for females) and one for ALT (>40U/L), both will be considered together (Booth *et al.* 2008; Elizondo-Montemayor *et al.* 2014). AST (Aspartate transaminase) is an enzyme produced by the liver; high levels of AST indicate liver damage. Whereas ALT (alanine aminotransferase) is an enzyme used to break down food into energy; high levels of ALT indicate liver damage.

Both studies reported similar results, with increasing prevalence of elevated AST and ALT with weight status (Table 25). Booth *et al.* (2008) recruited 496 children aged 14-17 from 48 randomly selected secondary schools in New South Wales, Australia. However special schools, schools with less 180 students and remote schools were excluded, which limits the generalisability of the results. Furthermore, the obese group was considerably smaller than the healthy weight group, 28 versus 376, and more likely to be biased due to small numbers of positive results, and also impacting the precision of prevalence estimates. Within the obese group, results suggested a (non-statistically significant) gender difference, none of the 10 girls in the obese group had NAFLD according to AST levels, compared with 26% of 19 boys in the obese group (Fisher's exact test, p=0.1336). Elizondo-Montemayor *et al.* (2014) recruited 236 children aged six to 12 years from schools in Monterrey, Mexico. Children with known diabetes mellitus, blood pressure, glucose or lipid altering medications, alcohol intake, and medications that can alter liver enzymes were excluded. Elizondo-Montemayor et al. also reported a slightly higher prevalence in boys than girls (20% versus 16% in the obese group) but as with Booth et al., this is based on very small numbers and was not significant (Fisher's exact test, p=0.5729).

Compared with data from studies using an ultrasound as an indicator of NAFLD (Table 24), the reported prevalence for Booth et al. and Elizondo-Montemayor et al. is much lower. The overall prevalence in the obese group using an ultrasound was 47%, compared with 17% and 22% for Booth et al. and Elizondo-Montemayor et al., respectively. This highlights the potential impact that the chosen tool has on reported prevalence, though comparison of the tests within the same population would be required to see if the difference remains.

Overall, both Booth et al. and Elizondo-Montemayor et al. suggest an association between weight category and prevalence of elevated AST or ALT, respectively. However further studies are required in order to draw definitive conclusions regarding the association with weight status.

4.2.6 Asthma

Asthma occurs when the airways narrow and swell, producing extra mucus. This can make breathing difficult and trigger coughing, wheezing, and shortness of breath. Three articles considered the prevalence of asthma, measured by self-report questionnaire. The studies suggested that the prevalence of asthma increased with weight status (Table 26). In two studies the prevalence in the overweight and obese groups was fairly similar, with a much lower reported prevalence in the healthy weight group, whereas Riberio-Silva et al. reported similar prevalence for all weight statuses (Kwon et al.; Noonan et al.). Furthermore the prevalence estimates varied considerably within weight statuses; for example, for the healthy weight group prevalence ranged from 7% to 17%. In terms of quality, all three studies rated fairly well. The main concern was the reliability of the measure, which was a child or parent completed questionnaire and is subject to responder bias. Additionally, a lack of detail regarding the target population and an a priori sample size calculation also lowered the quality rating.

The funnel plot showed little difference in the prevalence of asthma between weight statuses (Figure 48). The three data points above the 99.8% control limits were all from Kwon et al., implying the study may have important differences from the other two studies.

Table 25: Summary of studies reporting prevalence data for NAFLD (AST and ALT)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
AST (>33U/L for males, >26U/L for females)								
Booth, 2008	Australia	496	14-17 (15.4)	Education	4.3 (16/376)	10.9 (10/92)	17.2 (5/29)	9
ALT (>40U/L)								
Elizondo-Montemayor, 2014	Mexico	236	6-12 (9)	Education	0.0 (0/44)	6.9 (4/58)	22.4 (30/134)	9

Table 26: Summary of studies reporting prevalence data for Asthma (self-report)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Kwon *et al.* (2006)	America	853	2-11 (7.5)	Community	17.7 (85/479)	28.8 (40/139)	30.9 (58/188)	7
Noonan *et al.* (2010)	America	1852	9-22 (13.2)	Education	7.1 (50/703)	12.1 (41/339)	11.7 (62/528)	7
Ribeiro-Silva *et al.* (2013)	Brazil	1307	6-12 (NR)	Education	10.2 (97/952)	12.0 (15/125)	14.9 (13/87)	6
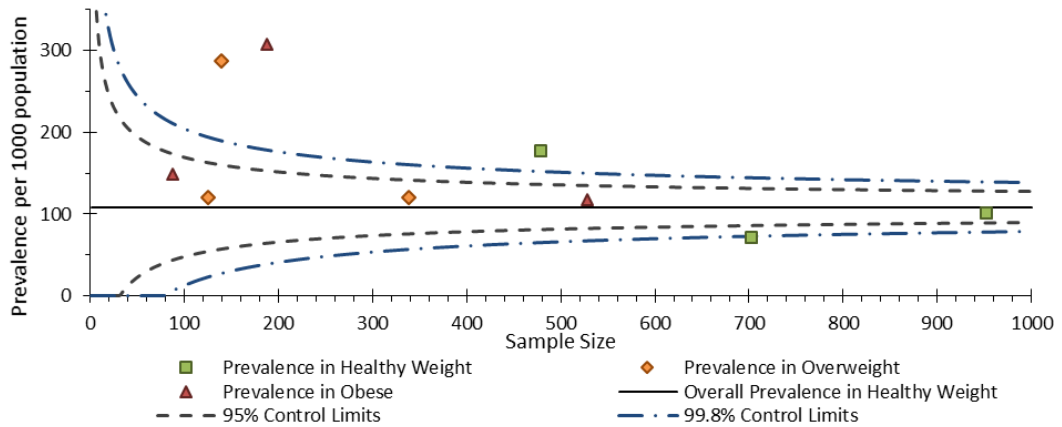
Figure 48: Funnel plot showing prevalence of Asthma per 1000 population

All three studies were included in the meta-analyses. The estimated PR indicated that (assuming equal population sizes) for every 1 child of a healthy weight there would be 1.5 (CI 1.2 to 1.9, p<0.001) children who were overweight (Figure 49) and 1.7 (CI 1.4 to 2.0, p<0.001) children with obesity (Figure 50) who have asthma. Both forest plots (Figure 49 and Figure 50) reported an I^2 Statistic of 0%, indicating little heterogeneity between the studies. For both forest plots the PR for Ribeiro-Silva et al. crossed the line of no effect suggesting no difference between the healthy weight and the overweight and obese groups.

Study	Overweight Events	Overweight Total	Healthy Weight Events	Healthy Weight Total	Weight (%)	Prevalence Ratio Random, 95% CI
Kwon et al. (2006)	40	139	85	479	47.82	1.6 (1.2 to 2.2)
Noonan et al. (2010)	41	339	50	703	32.84	1.7 (1.1 to 2.5)
Ribeiro-Silva et al. (2013)	15	125	97	952	19.35	1.2 (0.7 to 2)
Total (random effects)	96	603	232	2134	100.00	1.5 (1.2 to 1.9)

Heterogeneity: Q = 1.4; df = 2 (P = 0.4964); I^2 = 0%
Test for overall effect: Z = 3.81 (P <0.001)

Figure 49: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for Asthma

Study	Obese Events	Obese Total	Healthy Weight Events	Healthy Weight Total	Weight (%)	Prevalence Ratio Random, 95% CI
Kwon et al. (2006)	58	188	85	479	51.35	1.7 (1.3 to 2.3)
Noonan et al. (2010)	62	528	50	703	33.8	1.7 (1.2 to 2.4)
Ribeiro-Silva et al. (2013)	13	87	97	952	14.85	1.5 (0.9 to 2.5)
Total (random effects)	133	803	232	2134	100.00	1.7 (1.4 to 2.0)

Heterogeneity: Q = 0.3; df = 2 (P = 0.8586); I^2 = 0%
Test for overall effect: Z = 4.85 (P <0.001)
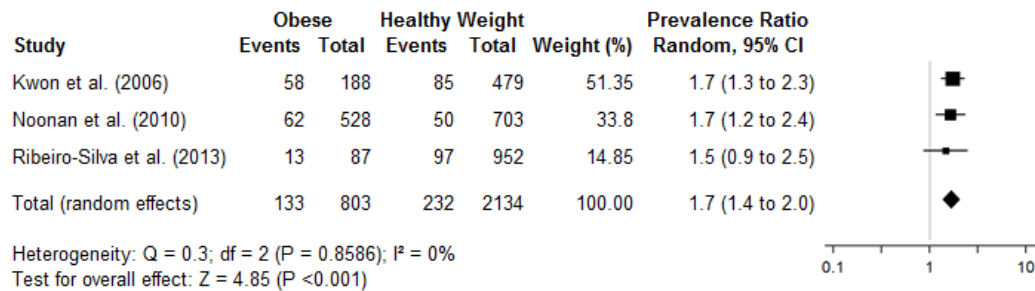
Figure 50: Forest plot of prevalence ratio for obese, relative to healthy weight, children for Asthma

None of the studies explicitly stated any exclusion criteria for the participants. However Kwon et al. did state inclusion criteria; only children who were part of the Harlem Children's Zone Asthma Initiative (http://www.hcz.org) could take part. The Harlem Children's Zone was established to address the high asthma-related morbidity observed among children (0–12 years) within a 60-block area of Central Harlem, New York City (Centers for Disease Control and Prevention 2005). Thus the population was considered to be at a higher risk of developing asthma due to factors unrelated to weight, which may have confounded the results. Kwon et al.'s sample consisted of Hispanic children (19.7%) and Black children (80.3%), with comparable asthma being reported in both groups (25% in Hispanics versus 22.2% in Blacks). Kwon et al. was conducted in New York city, a location with high levels of pollution (World Health Organisation 2014), which may explain the very high levels of asthma across all weight categories in comparison to Noonan et al. and Ribeiro-Silva et al. (Table 26). Removing Kwon et al. from the meta-analysis did not considerably impact the PR for the overweight group relative to the healthy weight group (1.5 (1.0 to 2.1), p=0.032) or the obese group relative to the healthy weight group (1.5 (1.2 to 2.1), p=0.002). However these results are based on only two studies, thus caution is required when interpreting them.

Some of the variation in prevalence between studies may, in part, be explained by the use of self-report questionnaires, resulting in either intentional or unintentional bias. Unintentional bias may be due either to a lack of clinical knowledge or to misclassification of a general lack of fitness as asthma (Cortés-Télles *et al.* 2015; Yang *et al.* 2017).

The results suggest a higher prevalence of asthma in children with obesity compared to the healthy weight group; however the evidence base is small and confounded by imperfect methods of screening for asthma and biased samples.

4.2.7 Less Established Co-morbidities

The following section provides an overview of co-morbidities with predominantly one study. Each co-morbidity will be discussed, however the lack of studies means in-depth analysis was not feasible.

Exercise Induced Wheeze/Cough

Exercise induced wheeze/cough is the narrowing of the airways causing difficulty moving air out of the lungs during physical activity. One study considered prevalence of exercise induced wheeze/cough using a self-report questionnaire

(Table 27). The study suggested a considerable increase in the prevalence of exercise induced wheeze/cough with increasing weight category. The substantial increase in prevalence with increasing weight status is likely the result of using self-report data from parents; therefore there is a risk of self-report bias. As such further research is required before any conclusions can be drawn.

Flat foot

Flat foot is a condition whereby feet have low or no arches and press almost completely flat against the floor. There are two types, rigid flat feet, where the arch is not present at all, and flexible flat feet where the arch normalises when the foot is not bearing weight, e.g. lying down. Two studies measured the prevalence of flat foot assessed by physical examination by a trained professional (Table 28). Quality scores ranged from 4 to 10, with a lower prevalence reported in the higher quality study (Chen, Chung and Wang 2009; Tenenbaum *et al.* 2013). Both studies reported that prevalence increased with increasing weight category (Table 28). However the high prevalence in the healthy weight group would suggest flat foot is due to factors other than obesity, particularly in terms of the data reported by Chen, Chung and Wang (2009), where around one-quarter of children were diagnosed with flat foot. Due to Tenenbaum et al.'s large sample size the funnel plot is incredibly narrow (Figure 51), with none of the data points for any of the weight statuses lying within the 95% or the 99.8% control limits. The lack of studies and the substantial differences in sample sizes make interpretation of the funnel plot difficult.
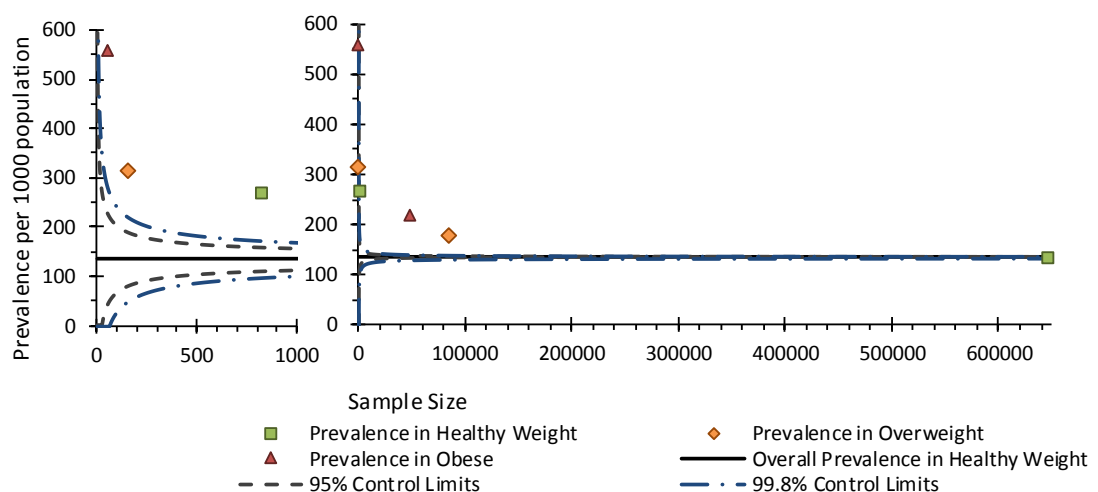
Figure 51: Funnel plot showing prevalence of Flat foot per 1000 population. Left-hand side displays prevalence for Chen et al. with an adjusted x-axis. Right-hand side displays prevalence for Chen et al. and Tenenbaum et al.

Table 27: Summary of studies reporting prevalence data for Exercise Induced Wheeze/Cough (self-report)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Kajbaf, Asar and Alipoor (2011)	Iran	903	7-11 (9.0)	Education	0.5 (4/755)	36.9 (31/84)	68.8 (44/64)	7

Table 28: Summary of studies reporting prevalence data for Flat foot (physical examination)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Chen, Chung and Wang (2009)	Taiwan	1024	5-13 (NR)	Education	26.7 (219/819)	31.4 (48/153)	55.8 (29/52)	4
Tenenbaum et al. (2013)	Israel	825964	16-19 (16.8)	Community	13.5 (86836/645357)	17.8 (15275/85652)	21.7 (10365/47710)	10

The two studies were combined in a meta-analysis. The average PR indicated that (assuming equal population sizes) for every 1 child of a healthy weight there would be 1.3 (CI 1.3 to 1.4, p<0.001) who were overweight (Figure 52) and 1.8 (CI 1.4 to 2.3, p<0.001) with obesity (Figure 53) who have flat foot. The I^2 statistic was 0% in Figure 52, indicating no heterogeneity between studies, and 71.5% in Figure 53, indicating high heterogeneity.

Study	Overweight Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Chen et al. (2009)	48	153	219	819	0.36	1.2 (0.9 to 1.5)
Tenenbaum et al. (2013)	15275	85652	86836	645357	99.64	1.3 (1.3 to 1.4)
Total (random effects)	15323	85805	87055	646176	100.00	1.3 (1.3 to 1.4)

Heterogeneity: Q = 0.84; df = 1 (P = 0.3595); I² = 0%
Test for overall effect: Z = 35.29 (P <0.001)
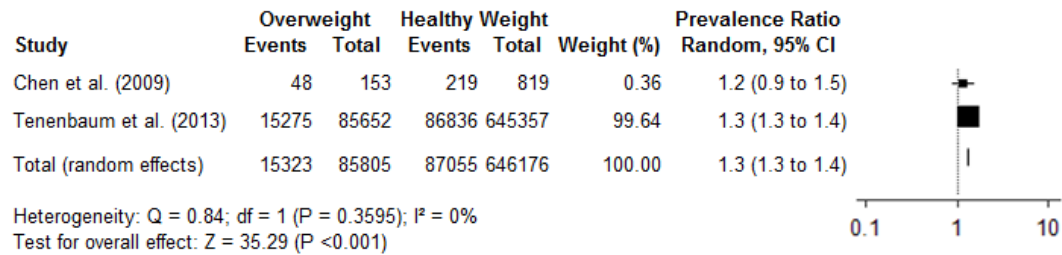
Figure 52: Forest plot of prevalence ratio for overweight, relative to healthy weight, children for Flat foot

Study	Obese Events	Total	Healthy Weight Events	Total	Weight (%)	Prevalence Ratio Random, 95% CI
Chen et al. (2009)	29	52	219	819	35.89	2.1 (1.6 to 2.7)
Tenenbaum et al. (2013)	10365	47710	86836	645357	64.11	1.6 (1.6 to 1.6)
Total (random effects)	10394	47762	87055	646176	100.00	1.8 (1.4 to 2.3)

Heterogeneity: Q = 3.51; df = 1 (P = 0.0610); I² = 71.52%
Test for overall effect: Z = 4.65 (P <0.001)
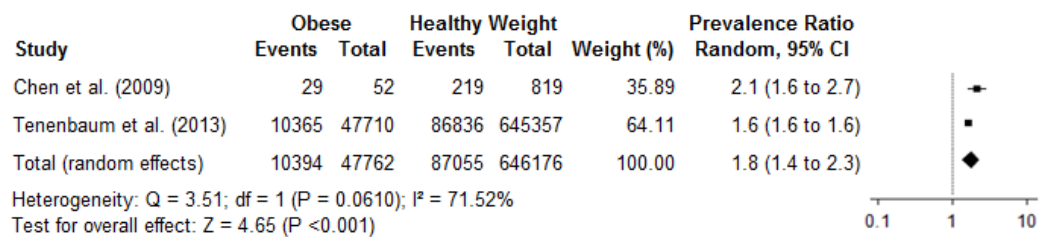
Figure 53: Forest plot of prevalence ratio for obese, relative to healthy weight, children for Flat foot

One of the main distinctions between the studies is the sample size. Chen et al. recruited 1024 children aged 5-13 years from a preschool and primary school in Taiwan, whereas Tenenbaum et al. recruited 825,964 adolescents who were being evaluated for the Israeli Defence Forces. It would be expected that estimates from studies with smaller samples would have lower precision and be more likely to be biased due to small sample variation. This is supported by the data, with Chen et al. reporting a much higher prevalence across the weight statuses than Tenenbaum et al.

The variation in prevalence may also be related to the different populations recruited and the eligibility criteria. Chen et al. excluded those with either a history of lower extremity injury, or of foot or ankle surgery, as these might confound results, whereas Tennenbaum et al. excluded those with rigid pes planus (where the alignment of the arch does not normalise when the feet are not bearing body weight). A further distinction between the studies is the age of the participants.

Chen et al.'s participants were aged 5-13 years and Tenenbaum et al.'s 16-19 years. Given that the prevalence of flat foot decreases with age, a lower prevalence would be expected in the older participants, as is observed (NHS 2017a). Furthermore Chen et al. reported that prevalence decreased in both boys and girls up to the age of 12. However, this data were not provided by age and weight status, thus prevalence may be a factor of both.

Overall, there was a higher prevalence of flat foot in children with obesity. However the association between flat foot and weight status may be confounded by other factors such as participant eligibility criteria, age, and gender. Further UK population-based studies are required to confirm these results.

Carotid Intima-Media Thickness

One study considered the prevalence of thickening of the inner two layers of the carotid artery, the intima and media, which is used to diagnose the extent of carotid atherosclerotic vascular disease (Touboul and Crouse 1997). Caserta *et al.* (2010) suggested that the prevalence increased with weight status, from a considerable prevalence of 24% in the healthy weight group to 43% in the obese group (Table 29). The high prevalence in the healthy weight group would suggest the data is from a high risk population. Caserta et al. randomly recruited 575 children aged 11 to 13, without any specific eligibility criteria, and assessment was conducted by a trained operator using ultrasound, further suggesting the population may be at high risk. The fact that the population may be at higher risk casts doubt over the accuracy of the prevalence of other co-morbidities. Caserta et al. did not provide sufficient information that might explain the higher than expected prevalence. As such definitive conclusions cannot be made and further research is required, ideally on UK populations.

Elevated Uric Acid

One study considered the prevalence of elevated uric acid in 2405 primary school children in Taiwan. Uric acid is produced by the body when breaking down purines, which are found in certain foods and are known to cause gout and are linked to kidney and heart disease (Marker 2016). Chu and Pan (2007) suggested that prevalence was associated with weight status, increasing from 15% in the healthy weight group to 44% in the obese group (Table 30). The prevalence may be a factor of diet rather than weight status, however without additional information; research conclusions cannot be made (Zgaga *et al.* 2012; UK Gout Society 2014;

Marker 2016). Moreover the lack of data from other countries, including the UK, meant that a clear understanding of the association with weight status could not be determined.

Gallstones

One study considered the prevalence of gallstones.   Gallstones are small, hard crystalline masses that form in the gall bladder or bile ducts (NHS Choices 2015). Typically no treatment is required, however in a small number of cases complications may occur. Koebnick et al. (2012b) reported that the prevalence of gallstones is below 1% in all weight categories and there was very weak evidence of an increase in prevalence with increasing weight status (Table 31).

Table 29: Summary of studies reporting prevalence data for Carotid-Intima Media Thickness (Carotid ultrasonography)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Caserta et al. (2010)	Italy	575	11-13 (NR)	Education	24.2 (76/314)	36.9 (66/179)	42.7 (35/82)	7

Table 30: Summary of studies reporting prevalence data for Elevated Uric Acid (≥7mg/dL)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Chu and Pan (2007)	Taiwan	2405	6-12 (NR)	Education	14.5 (254/1753)	28.3 (102/361)	43.6 (127/291)	7

Table 31: Summary of studies reporting prevalence data for Gallstones (patient records)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Koebnick et al. (2012b)	America	510,816	10-19 (NR)	National Survey	0.1 (215/301549)	0.2 (179/99987)	0.3 (372/109280)	9

High C-Reactive Protein

One study reported the prevalence of high C-reactive protein, a biomarker of inflammation (Sadovsky 1998); the underlying condition that causes increases in C-reactive protein can range from an infection to cancer (NHS Choices 2016). Caserta *et al.* (2010) suggested that the prevalence of high C-reactive protein in their study was associated with weight status (Table 32), with considerably higher prevalence in the overweight and obese groups than the healthy weight group. Caserta et al. did not provide sufficient information regarding the selected sample, and it was unclear how representative the sample was of the wider population.

Traumatic Dental Injuries

One study considered the association between weight status and traumatic dental injuries, which refers to potentially irreversible damage to the tooth enamel/dentine due to accident or injury. Al-Bajjali and Rajab (2014) reported that prevalence decreased with increasing weight status (Table 33), 17% in healthy weight group and 11% in the obese group. However, these prevalence rates are implausibly high for a population study and it is difficult to draw firm conclusions regarding the association from this single study. Given the proposed screening programme is intended for co-morbidities of higher prevalence in the obese population, the results would suggest traumatic dental injuries should not be included. However further UK based research is required, before definitive conclusions can be made.

Table 32: Summary of studies reporting prevalence data for High C-Reactive Protein (≥3mg/L)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Caserta _et al._ (2010)	Italy	575	11-13 (NR)	Education	2.6 (8/306)	15.5 (24/155)	32.3 (20/62)	4

Table 33: Summary of studies reporting prevalence data for Traumatic Dental Injuries (physical examination)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Al-Bajjali and Rajab (2014)	Jordan	1015	12-12 (NR)	Education	17.2 (127/738)	15.8 (26/165)	10.7 (12/112)	8

Anxiety

One study considered the prevalence of anxiety using the Global School-based Health Survey (GSHS) (World Health Organisation 2018). Anxiety is feelings of worry, nervousness, and/or unease about certain events/situations above what would be expected. Zakeri *et al.* (2012) suggested a decline in prevalence as weight category increased (Table 34). The prevalence in the 10-18 year old Iranian cohort was high compared with the adult (>16 years) population prevalence of 5.9% (McManus *et al.* 2016). Although there is evidence that prevalence of anxiety decreases with age, suggesting the results may be appropriate for the age group considered, the GSHS is not a screening tool for anxiety (Martin 2003; Mental Health Foundation 2016; Remes *et al.* 2016). The non-specific nature of the questions affects its accuracy and suitability in comparison to other more established tools such as the Revised Children's Anxiety and Depression Scale.

Depression

Zakeri et al. also considered the prevalence of depression using two questions from the GSHS questionnaire (World Health Organisation 2018). Depression is a mood disorder resulting in persistent feelings of sadness and loss of interest in previously enjoyed activities. Results suggested prevalence of depression was very high (30%) in all three weight categories (Table 34). The reported prevalence of depression was implausible, given that the reported population prevalence in individuals aged 16 in the UK is 3.3% and reflects the lack of specificity of the two questions used to define depression (Mental Health Foundation 2016). The GSHS is not a screening tool for depression; as such its accuracy for screening is unlikely to be comparable to more established screening tools such as the Beck's Depression Inventory for Children (CDC 2016; World Health Organisation 2018).

Low Self-Esteem

One low quality study measured low self-esteem via self-report (Franklin *et al.* 2006). Low self-esteem is associated with feelings of incompetency, incapability, and unworthiness in one or more area of one's life. Results suggested that prevalence increased with increasing weight category (Table 34). Self-esteem was assessed based on perceived competence in five domains (scholastic competence, social acceptance, athletic competence, physical appearance, and behavioural conduct) using the Self-perception Profile for Children.

Table 34: Summary of studies reporting prevalence data for Anxiety (self-report), Depression (self-report), and Low Self-Esteem (self-report)
(NR – Not Reported)

Author (Year)	Country	Sample Size	Age Range (yrs.) (mean age)	Setting	Prevalence (%) in Healthy Weight (n/N)	Prevalence (%) in Overweight (n/N)	Prevalence (%) in Obese (n/N)	Study Quality Score
Anxiety								
Zakeri *et al.* (2012)	Iran	8460	10-18 (NR)	Education	13.1 (806/6154)	12.4 (106/858)	9.7 (52/538)	6
Depression								
Zakeri *et al.* (2012)	Iran	8640	10-18 (NR)	Education	30.1 (1853/6154)	30.2 (259/858)	30.9 (166/538)	6
Low Self-Esteem								
Franklin *et al.* (2006)	Australia	2491	9.2-13.7 (11.3)	Education	4.5 (86/1905)	8.9 (32/359)	14.5 (33/227)	4

4.2.8 Overview of Prevalence Ratios

A summary of the estimated average prevalence ratios and 95% confidence intervals for children who are overweight relative to healthy-weight or obese relative to healthy weight for each of the 10 co-morbidities/indicators included in the meta-analyses is shown in Table 35 with a summary plot provided in Figure 54. The average prevalence ratio for the overweight group, relative to healthy weight, ranged from 1.2 (total cholesterol) to 15.8 (metabolic syndrome (NCEP)). Greater variation was observed when considering the prevalence ratio for the obese group relative to the healthy weight group, the prevalence ratio ranged from 1.4 (fasting plasma glucose) to 58.0 (metabolic syndrome (NCEP)).

Table 35: Summary of the combined estimates of the prevalence ratios (95% confidence intervals) for co-morbidities in overweight groups and obese groups, relative to healthy weight

Co-morbidity	Indicator/Measure	Prevalence Ratio: overweight relative to healthy weight		Prevalence Ratio: obese relative to healthy weight	
		(95% CI)	**p**	**(95% CI)**	**p**
Hyperglycaemia	Fasting Plasma Glucose (≥100mg/dL)	1.4 (1.2 to 1.6)	<0.001	1.4 (1.2 to 1.7)	<0.001
Dyslipidaemia	Total Cholesterol (≥200mg/dL)	1.2 (0.8 to 1.8)	0.4	1.8 (1.0 to 3.1)	0.04
	Low HDL Cholesterol (<40mg/dL)	2.0 (1.5 to 2.6)	<0.001	2.9 (2.1 to 4.1)	<0.001
	High LDL Cholesterol (≥130mg/dL)	2.0 (0.9 to 4.3)	0.082	3.3 (1.6 to 6.8)	0.001
	High Triglycerides (≥150mg/dL)	2.5 (1.9 to 3.3)	<0.001	4.2 (3.2 to 5.6)	<0.001
High blood pressure	≥90th percentile	2.1 (1.7 to 2.6)	<0.001	4.0 (2.8 to 5.7)	<0.001
Metabolic Syndrome	Cook et al. 2003	2.5 (0.8 to 7.7)	0.1	21.2 (1.2 to 381.6)	0.04
	IDF, 2007	13.1 (1.9 to 88.9)	0.008	53.9 (9.7 to 297.9)	<0.001
	NCEP ATP III 2001	15.8 (7.4 to 33.8)	<0.001	58.0 (27.7 to 121.3)	<0.001
Non-Alcoholic Fatty Liver Disease	Ultrasound	6.1 (3.3 to 11.2)	<0.001	26.1 (9.4 to 72.2)	<0.001

Co-morbidity	Indicator/Measure	Prevalence Ratio: overweight relative to healthy weight		Prevalence Ratio: obese relative to healthy weight	
		(95% CI)	p	(95% CI)	p
Asthma	Questionnaire	1.6 (1.2 to 1.9)	<0.001	1.7 (1.4 to 2.1)	<0.001
Flat foot	Physical examination	1.3 (1.3 to 1.4)	<0.001	1.8 (1.4 to 2.3)	<0.001
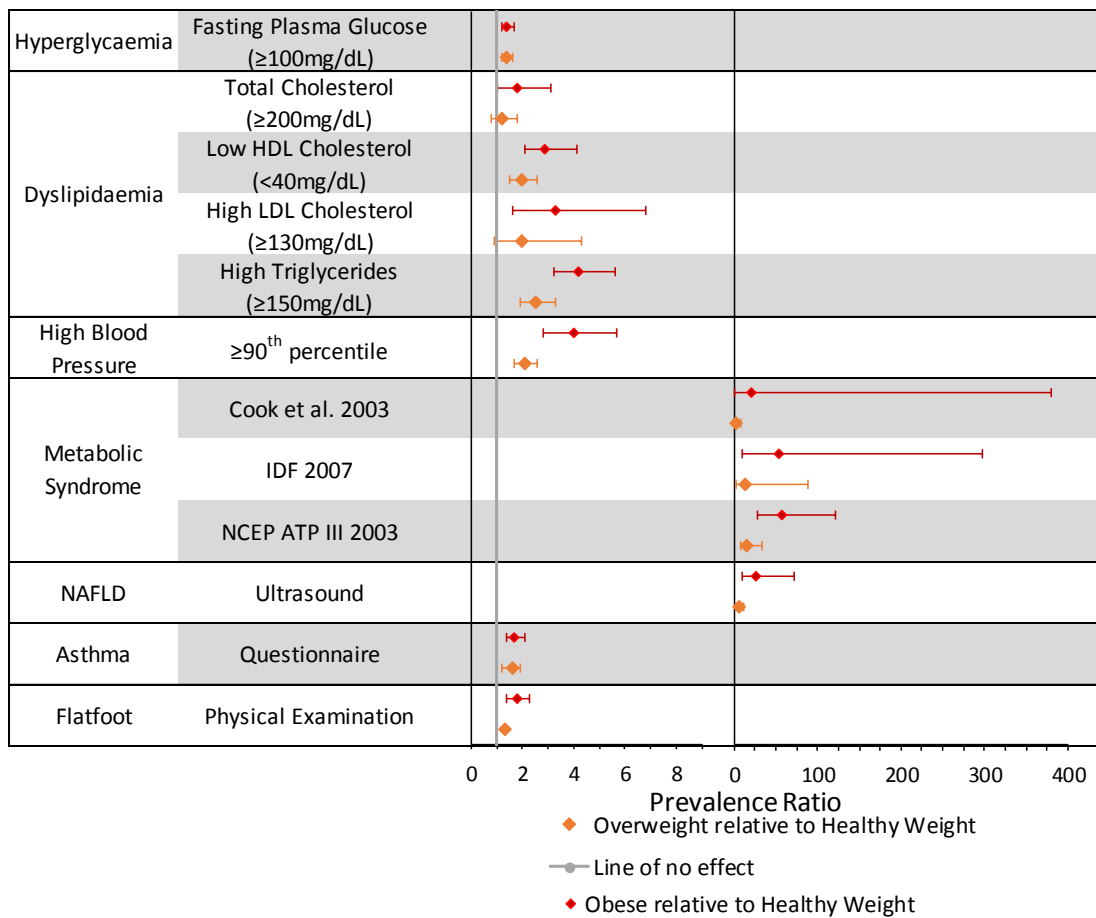
Figure 54: Average random effects estimates of prevalence ratios and 95% confidence intervals for the co-morbidities/co-morbidity indicators. Left-hand side excludes metabolic syndrome and NAFLD due to differences in scale

4.3 Discussion

4.3.1 Summary of Findings

The systematic review initially identified 26 co-morbidities (22 physical and four psychological) from 162 studies. The review indicated variation in the populations sampled and studies not reporting data for all three weight status groups. This was not congruent with the aim of the systematic review, as it would not be possible to estimate the population prevalence ratio in children who were overweight and obese, relative to those of a healthy weight. Therefore studies with highly selected populations and those without data for all three weight statuses were excluded from the meta-analysis to decrease between-study heterogeneity; this left 45 articles and 19 co-morbidities.

Overall, the results suggested that children with obesity had a higher prevalence of the majority of the identified co-morbidities included in the meta-analysis, relative to those of a healthy weight. The review indicated that there were small to large increases in prevalence ratio between co-morbidities. For instance, for hyperglycaemia, asthma, flat foot, total cholesterol, and low HDL cholesterol there was a small increase, with a PR between 1.4 and 2.9. For high LDL cholesterol, triglycerides, and high blood pressure, there was a moderate increase with a PR between 3.3 and 4.4. For NAFLD and metabolic syndrome the increase was extremely large, with the PR ranging from 26 to 58.

4.3.2 Previous Systematic Reviews

Results from previous systematic reviews supported the current findings, that children who were overweight/obese were more likely to have a number of co-morbidities when compared to those of a healthy weight (Guh *et al.* 2009; Pulgarón 2013; Sanders *et al.* 2015). The previous reviews identified between 15 and 20 co-morbidities, fewer than the current review's 26. Co-morbidities not identified by some or all of the previous reviews include enuresis, gallstones, vitamin D deficiency, uric acid, anxiety, flat foot, NAFLD, and carotid-intima media thickness; however the results for some of these were not significant. Additionally the other reviews identified co-morbidities which were not identified in the current review; e.g. Guh et al. identified a number of cancers associated with obesity (e.g. colorectal, kidney, prostate, breast, and ovarian), although, Guh et al. predominantly considered participants aged 40 and above in whom cancers are more prevalent. Furthermore, cancers tend to have long latent periods and the ones identified are

rare in childhood (Nadler and Zurbenko 2014). Other reasons for the variation in co-morbidities identified could be the types of studies included; Guh et al. only included prospective cohort studies, and Sanders et al. only considered observational studies conducted in Australia.

4.3.3 Between-study Heterogeneity

Despite the current results suggesting a higher prevalence of co-morbidities in those with childhood obesity, the association between weight status and prevalence of co-morbidities was often difficult to ascertain due to between-study heterogeneity. Some areas of heterogeneity are now considered.

Differences in Terminology and Definition

In the current systematic review there was a lack of consistency between studies in definitions and cut-offs for co-morbidities. For high blood pressure three cut-offs were identified ($\geq 90^{th}$ percentile, $\geq 90^{th}$ and $<95^{th}$ percentile, and $\geq 95^{th}$ percentile) and for low HDL cholesterol both <40mg/dL and ≤40mg/dL were used in studies. For each of these two co-morbidities studies with different cut-offs were combined into one group, since the PR for each definition was comparable. For Metabolic syndrome, however, it was not possible to combine definitions for analysis due to differences in definitions and PRs between definitions. Currently there is no consensus on which definition of Metabolic Syndrome should be adopted, and different definitions resulted in different estimates of population prevalence, limiting the ability to obtain a single estimate (Kassi *et al.* 2011).

Differences in Sample Size and Weight Status Group Sizes

There was also considerable variation in sample sizes, which ranged from 211 to 825,964 (Davis *et al.* 2005; Tenenbaum *et al.* 2013). This variation was more evident when comparing group sizes for weight statuses, which ranged from 44 to 645,357 for healthy weight, 19 to 99,987 for overweight, and 2 to 109,280 for obese (Nur *et al.* 2008; Koebnick *et al.* 2012b; Pecin *et al.* 2013; Tenenbaum *et al.* 2013; Elizondo-Montemayor *et al.* 2014). Variations in group size can influence the reported prevalence, as smaller groups would only require small numbers of positive results to inflate the reported prevalence, which would subsequently impact the prevalence ratio, and therefore the accuracy of the population estimate. However this is controlled for to some extent by weighting studies as part of the meta-

analysis, where smaller studies are given less weighting than larger studies (Borenstein *et al.* 2009).

Geographical Variation

The systematic review identified studies across the globe. Conducting a global search means that a more comprehensive list of co-morbidities can be identified, but the results from other countries are less likely to be generalizable to the UK population, the target population for the proposed screening programme. In total, the systematic review identified studies from 49 different countries, including eight from China, five from the USA, and one from the UK. The lack of studies from the UK does limit relevance to this country, particularly for studies from populations that are very different in culture, lifestyle and diet, such as Iran and Brazil, all of which may affect the prevalence of some co-morbidities. Data from Western Europe, North America and Australia are more likely to be generalizable to the UK population; however characteristics of the samples and differences in diet and lifestyle would need consideration prior to generalising results. Only one identified study was conducted in the UK, which was for NAFLD (Lawlor *et al.* 2014). The reported prevalence in other studies was markedly different, ranging from 32% (Iran) to 62% (Australia). Suggesting country-specific factors are important. For instance, in relation to uric acid levels, evidence indicates a positive association between uric acid levels and consumption of sugar-sweetened beverages. A 2015 global review indicated marked differences in sugar-sweetened beverage consumption between countries (Singh *et al.* 2015). The systematic review identified one study considering elevated uric acid conducted in Taiwan, which according to Singh *et al.* (2015) consume approximately 0.05 8oz. servings/day compared with the UK's 0.6 8oz. servings/day (Zgaga *et al.* 2012; UK Gout Society 2014; Singh *et al.* 2015; Marker 2016). Given this differences, additional research with a UK sample is required to assess the prevalence of obesity-related co-morbidities in children.

Differences in Ethnicity

Although ethnicity has been associated with obesity in adult studies, results were rarely reported by different ethnic subgroups in the studies (Louthan *et al.* 2005; Franklin *et al.* 2006; Fox *et al.* 2010; Lang *et al.* 2012). Therefore, confounding due to ethnicity could not be investigated. The importance of this is particularly evident where studies indicated that certain ethnicities had a higher prevalence. For instance Tandon *et al.* (2013) consistently reported a higher prevalence of co-morbidities in South Asians than other ethnic groups, and Davis *et al.* (2005)

reported White children had a higher prevalence of metabolic syndrome than Black children. Therefore an awareness of the impact of ethnicity on prevalence is important, particularly in a multi-ethnic society such as the UK.

Gender Differences

Previous research has indicated that in adults, females and males differ in prevalence for some conditions, e.g. high total cholesterol and cardiovascular diseases (Regitz-Zagrosek 2012; Centers for Disease Control and Prevention 2015; Gupta *et al.* 2016). Within the systematic review there were some differences between male and female prevalence estimates, for example Chu and Pan (2007) reported that females had a higher prevalence of high cholesterol than males, but there was little consistency between studies. Similar inconsistencies between genders were identified for NAFLD (elevated ALT and AST), elevated triglycerides, flat foot, high blood pressure, and self-esteem, suggesting any screening programme should be suitable for both males and females.

Socioeconomic Status

There is considerable evidence of higher prevalence of obesity in individuals of lower socioeconomic status, particularly in women (Kuntz and Lampert 2010; Stamatakis, Wardle and Cole 2013; Newton, Braithwaite and Akinyemiju 2017). Thus if the co-morbidities are associated with obesity, a higher prevalence would be expected in those from lower socioeconomic groups. However, the studies identified did not report co-morbidity prevalence by socioeconomic status, which may be important given the socioeconomic disparities observed in the UK.

Quality Appraisal

For each study a quality appraisal was undertaken using the critical appraisal checklist, using simplistic scoring (0-5 – poor quality; 6-10 – high quality) (Munn *et al.* 2014; González-Serrano *et al.* 2016). The overall scores ranged from four to 10, with 32 of the 45 studies being classed as high quality. Although quality appraisal is an important part of a systematic review/meta-analyses there are limitations in the process. Appraisal tools are reliant on the information provided in the article, which may not be sufficient to answer the items on the checklist. For example in the systematic review 26 studies did not provide details of an a priori sample size calculation, so an assessment of the power of the study could not be determined. Additionally, 29 studies did not explicitly state that the measurement of the co-morbidity was conducted by trained staff. Previous research has shown that authors

and independent reviewers differed in their assessment of a study's quality, with authors typically giving lower quality scores (Lo, Mertz and Loeb 2014). This would suggest information bias between reviewers and authors which limits the assessment of study quality and the potential to make fully-informed conclusions. This asymmetry may explain results by Katikireddi, Egan and Petticrew (2015) who reported that, of 59 systematic reviews from 14 high-ranked medical journals, 20 did not use quality appraisal to inform results and conclusions.

Control of potential confounders

Few studies completed an in-depth analysis of the likely causal nature of any associations between weight and co-morbidities. Well known correlates of obesity in adults, such as socioeconomic status, ethnicity, and gender were rarely controlled for in analysis of prevalence ratios, so that the associations may have been inflated or completely induced by known or unknown confounders. There were issues pertaining to the control of other confounding factors. For instance, when considering carotid-intima media thickness, research has suggested that elevated blood pressure plays a major role in the thickening of the carotid artery (Rumińska *et al.* 2017); however Caserta et al. did not control for this in their results. Thus the prevalence of carotid-intima media thickness reported by Caserta et al. may be biased given their reported prevalence of high blood pressure.

4.3.4 Implications for the proposed screening programme

The Conditions (Co-morbidities)

The current systematic review and meta-analyses found an increase in prevalence of many co-morbidities in children who are overweight/obese relative to those of a healthy weight. However, increased prevalence alone is not sufficient to warrant screening for co-morbidities. The National Screening Committee's criteria for appraising the viability of a screening programme considers the co-morbidities' prevalence and severity (Public Health England 2015a).

Previous research has also reported an increased prevalence of obesity-associated co-morbidities in children. For instance, Abdullah *et al.* (2010) reported that being overweight raised the risk of developing type 2 diabetes by a factor of three, and being obese by a factor of seven, compared to being a healthy weight (Abdullah *et al.* 2010). Other research indicated that obesity in children has been associated with twofold or higher risk of adult hypertension, coronary heart disease, and stroke

(Reilly and Kelly 2011). This is supported by analysis of data from four prospective cohort studies (Juonala *et al.* 2011). Participants who were overweight/obese as children, and remained overweight/obese as adults, had considerably higher risk of high-risk dyslipidaemia, high blood pressure, and higher carotid-intima media thickness (Singh *et al.* 2011). Additionally, the presence of type 2 diabetes has been associated with other serious conditions such as hypertension, depression and nephropathy, supporting the screening and management of diabetes sooner rather than later (Springer *et al.* 2013).

Previous research has also suggested a positive association between increased weight and the prevalence of psychological co-morbidities (Bell *et al.* 2007; Gibson *et al.* 2008; Bell *et al.* 2011). This is contrary to the current meta-analyses' findings and the findings from other studies where the prevalence of depression was only modestly or weakly associated with increasing weight status (Anton *et al.* 2006; Goldstein *et al.* 2008; Rankin *et al.* 2016).

Overall, there is evidence of a positive association between childhood obesity and co-morbidity prevalence; moreover, some of the co-morbidities can have severe implications if left unmanaged (Sabin, Crowne and Shield 2002; Narayan *et al.* 2007; Abdullah *et al.* 2010; Reilly and Kelly 2011; Staimez *et al.* 2013; Kodama *et al.* 2014). Given the reported prevalence and the potential long-term implications to one's health and well-being there is some evidence to support the screening for obesity-related co-morbidities in children. Screening for all the identified co-morbidities would not be practical, thus further research is required to gain consensus on which co-morbidities are considered appropriate for the proposed screening programme in terms of prevalence and the impact to the child's health, in the short- and long-term. The value of screening will also depend on the availability of cost-effective and acceptable screening test and an appropriate management strategy.

The Screening Tests

The NSC criteria also consider the accuracy, suitability and acceptability, amongst clinicians and patients, of screening tests. The current systematic review identified potential screening tests for the co-morbidities. For some, such as hyperglycaemia, multiple tests were identified, and the accuracy is likely to vary between screening tests. Accuracy consists of sensitivity (the probability of a positive test in children with the co-morbidity) and specificity (the probability of a negative test in children without the co-morbidity) (Lalkhen and McCluskey 2008). Thus the prevalence

estimate for any co-morbidity will depend on the accuracy of the screening test. If specificity of a test drops from, say 80% to 70%, the number of false positives will increase from 20% to 30%. For example, previous studies have reported that automatic blood pressure devices are less accurate than manual devices (lower sensitivity and specificity) (Mansoor *et al.* 2016; Shahbabu *et al.* 2016). Thus studies adopting an automated device are likely to have a higher false positive rate than those utilising a manual device with appropriately trained staff.

The value of a test will also depend on the prevalence of the comorbidity in the screened population, which is particularly important for children in which serious co-morbidities are rare. For example, for an excellent test with 95% sensitivity and specificity, the probability that a child with a positive test has the comorbidity will be 68% if the prevalence is 1-in-10, 16% if the prevalence is 1-in-100 and 2% if the prevalence is 1-in-1000. Screening tests that are likely to be used in a weight management service are likely to be much less specific than this, so that any programme must estimate the expected number of false positive diagnoses and the potential impact that this would have on a screening programme.

For the proposed screening programme, the increase in false positives has a downstream impact on services that have to sort people into those who do and do not require treatment. According to the NICE recommendation referral would be made to the child's GP. Therefore accurate screening methods would be required to avoid an influx of false positive results at GP surgeries, and to ensure the screening programme causes minimal harm to children from incorrect screening results. For instance, with regards to impaired fasting glucose the World Health Organisation state a cut-off of ≥110mg/dL, whereas in the literature the cut-off ≥100mg/dL is adopted, which is in line with recommendations from the American Diabetes Association (The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus 2003). This change was intended to improve the sensitivity to predict impaired fasting glucose, but resulted in a 2- to 4-fold increase in the prevalence of impaired fasting glucose and created the term "pre-diabetes" (Valdés *et al.* 2008). This resulted in increased anxiety amongst screen-positives (Valdés *et al.* 2008). Furthermore, historically the NSC has rejected potential screening programmes due to impractical and unsuitable screening tests. For instance, in their recent report the NSC recommended not screening for alcohol misuse as the screening would be a self-report questionnaire, which they feared would lead to a higher number of false positives and potentially overwhelm services (UK National Screening Committee 2017). Therefore, once agreement on co-morbidities has

been made, additional work would be required to identify suitable screening tests and cut-offs to limit harms and increase the benefits of implementing a screening programme.

4.3.5 Strengths and Limitations

Strengths of the systematic review and meta-analyses include the development of an a priori protocol, which explicitly detailed the inclusion and exclusion criteria for the studies and the steps undertaken in each stage. This allows for replicability of the review in light of future research, and for more accurate estimates of population prevalence. The search criteria included multiple study designs which enabled greater sensitivity in detecting relevant literature in order to develop an exhaustive list of obesity-related co-morbidities. This was a limitation of previous systematic reviews in the area (Guh *et al.* 2009; Pulgarón 2013; Sanders *et al.* 2015). This is also the first systematic review and meta-analyses considering the prevalence of co-morbidities in children who are overweight or obese relative to those of a healthy weight.

Some limitations are worthy for consideration. Firstly, only a small number of articles pertaining to psychological co-morbidities and a number of physical co-morbidities were identified. It is possible that other articles may have been identified through a co-morbidity specific search; however it may also be an indication of a lack of prevalence studies in this population. The lack of studies for some co-morbidities prevented a detailed understanding of the association between weight status and prevalence. Related to this, there was a lack of population based studies that provided data for all three weight statuses. This further reduced the number of co-morbidities eligible for the meta-analyses. However the exclusion of highly selected population studies meant the results are more generalizable and relevant to the aims of the thesis. Secondly, studies related to all psychological co-morbidities and some physical co-morbidities, utilised self-report questionnaires, some of which were not validated. This hindered the accuracy of the data, due to self-report bias and non-specific tests, and reduced the generalisability of the results. Although it should be noted that alternative screening methods are not available for many of these co-morbidities. Thirdly, non-English language articles were not included in the review, potentially excluding relevant articles. Previous reviews have offered conflicting information about the impact of language restrictions on results (Grégoire, Derderian and Le Lorier 1995; Juni *et al.* 2002; Bown and Sutton 2010; Morrison *et al.* 2012). Dhillon and Gill (2014) proposed that whenever feasible all relevant

studies should be included but all non-English articles should be identified with "language" as the reason for exclusion if translation was not feasible (Centre for Reviews and Dissemination 2009). All non-English articles were retained in a separate Endnote folder for possible future research. Finally, a global review was conducted, although this increased the probability of identifying all childhood obesity-related co-morbidities, it also reduced generalisability of data from non-UK countries to the UK population (where the proposed screening programme would be implemented). This is due to variation in culture, diet, lifestyle and ethnic composition (as discussed in Section 4.3.3). Thus caution is required when interpreting results from non-UK countries.

There are some recommendations for future research. Firstly, additional research in general populations, stratified by weight category is required to obtain an understanding of the impact of increased weight on the prevalence of these co-morbidities in UK children. This would provide additional guidance regarding the suitability of implementing a screening programme. Secondly, additional population-based research is required to enable generalisability of the results beyond the study sample. Finally, consensus is required on the definitions and cut-offs used for co-morbidities and screening tests. This would enhance between-study homogeneity and enable a more accurate population prevalence ratio estimate to be calculated.

4.4 Conclusions

This systematic review and meta-analyses provided a comprehensive list of co-morbidities associated with child obesity. The results from the meta-analyses indicated that children with obesity are at risk of a number of physical and psychological co-morbidities; however this is only a small part of developing a screening programme. Due to the number of co-morbidities identified, it would be impractical to include them all in a screening programme. Thus the number would need to be reduced for implementation within a screening programme. Additionally, further work is required to identify suitable screening tests for use within community weight management settings. The next chapter builds on the results of the systematic review and meta-analyses to obtain consensus via a structured and transparent method on which co-morbidities and screening methods are suitable for inclusion in the proposed co-morbidity screening programme.

Chapter 5: Consensus on Co-morbidities and Screening Methods for the Proposed Screening Programme

5.1 Introduction

Chapters 3 and 4 identified the prevalence of obesity-related co-morbidities and prevalence ratio in children who are overweight or obese relative to those of a healthy weight. From the 162 observational studies identified by the systematic review, 22 co-morbidities were identified. Of these, 45 studies and 19 co-morbidities were eligible for the meta-analysis, with results suggesting an increased prevalence in children who are overweight or obese relative to those of a healthy weight, for many of the co-morbidities. However screening for all the identified co-morbidities would not be practical and due to a lack of clear evidence, a consensus study was undertaken with the aim to gain agreement on a) co-morbidities where there may be patient benefit through screening and early identification, and b) optimal screening methods for use within weight management services (Section 5.2). Section 5.3 discusses methodological considerations for the study, including defining informal and formal screening methods, and reviews four commonly used formal consensus methods (Sections 5.3.2). Following this, the methods (Section 5.4), data collection (Section 5.4.3), data analysis (Section 5.5) and results (Section 5.6) are provided. Finally, a discussion of the results and implications of the results to the PhD are presented in Section 1.1.

5.1.1 Justification for Consensus Methods

The systematic review and meta-analyses (Chapters 3 and 4) allowed for the identification of co-morbidities associated with obesity. Although the results suggested an increased prevalence in those who were overweight or obese, relative to those of a healthy weight for many of the co-morbidities, there were gaps in the literature regarding some potentially important confounds, such as ethnicity, socioeconomic status and gender. In particular only one study was conducted on a UK population, this limits the applicability of the results to the UK population, which is the target population for the proposed screening programme. Furthermore a number of studies were not eligible for inclusion in the meta-analysis, thus could not be included when calculating the prevalence ratio. This does not meet the National Screening Committee's (NSC) criteria, which states there should be robust evidence

about the association between the risk, e.g. obesity, and the condition, e.g. obesity-associated co-morbidities, which should be considered an important health problem (Public Health England 2015a). Increased prevalence is only one aspect of assessing a co-morbidity's importance, the consequences of the co-morbidity to the individual's psychological and physical well-being also needs consideration. Furthermore, the NSC's criteria stated that acceptability amongst health professionals and service users was an important part in the development of a screening programme (Public Health England 2015a). Therefore to address the NSC's criteria, a consensus study was conducted to understand the views and opinions of health professionals and researchers regarding suitable co-morbidities and screening tests for inclusion in the proposed screening programme.

5.2 Aim

The aim of the consensus study was to achieve consensus on co-morbidities where there may be service user benefit from earlier identification and consensus on optimal screening methods for inclusion in a screening programme within UK community weight management services. This aim was met through the following objectives:

1. Collation of obesity-associated co-morbidities identified via the systematic review and meta-analysis, their prevalence and health consequences.
2. Collation of screening tests, for the co-morbidities identified, their sensitivity and specificity, and cut-off points for referral.
3. Presentation of the co-morbidities and screening tests to health professionals, researchers, and service users to obtain consensus.

5.3 Methodological Considerations

5.3.1 Informal versus Formal Consensus Methods

Consensus approaches are the means by which group opinions are synthesized to obtain consensus or make a decision. There are two broad approaches to consensus development, informal and formal.

Informal approaches tend to reach agreement through open discussion by a committee, sometimes in a single meeting, and are typically based on poorly defined or even undefined criteria and without the incorporation of scientific evidence (Rycroft-Malone 2001). They do not document research methodologies and do not perform analytical assessment of the decision to assess the level of agreement

between participants (Eccles *et al.* 1996; Woolf 1992). This lack of transparency and structure means it is difficult to assess the validity and reliability of the decisions made; thus the quality of the decisions may be poor due to reliance on subjective judgement and the lack of systematic procedures to assess the decision (Woolf 1992).

On the other hand, formal approaches tend to have a clear, controlled, identifiable, and structured processes which are supplemented by scientific evidence from published literature (Trickey *et al.* 1998; Black *et al.* 1999b). This results in transparency regarding how the consensus process is conducted, the synthesis and analysis of judgements, reproducibility, and the ability to critique the methodology, meeting the requirements of scientific methods (Woolf 1992; Eccles *et al.* 1996; Black *et al.* 1999b; Nair, Aggarwal and Khanna 2011). Formal methods are particularly useful where a state of uncertainty exists, i.e. evidence is lacking, limited, or contradictory, or the topic area is regarded as complex (Field and Lohr 1992; Black *et al.* 1999b; Nair, Aggarwal and Khanna 2011; Humphery-Murto *et al.* 2017). Other advantages of formal methods are that the structure is designed to reduce the influence of dominant personalities and provides participants the opportunity to alter their opinion in light of feedback and/or group discussion (Jones and Hunter 1995; Murphy *et al.* 1998). Although the approaches are intended to capture collective knowledge, they are subjective judgements using the best available evidence, thus the reliability of decisions may be limited outside the decision panel (Black *et al.* 1999b).

Given the complexity of childhood obesity and the potential number of co-morbidities associated with childhood obesity, formal consensus methods were adopted (Guh *et al.* 2009; Finegood, Merth and Rutter 2010; Pataky, Bobbioni-Harsch and Golay 2010; Frood *et al.* 2013). This enabled the incorporation of scientific evidence, along with a robust and transparent process for decision making, which would subsequently support the content validity of the results (Field and Lohr 1992; Murphy *et al.* 1998; SAC 2002; USDHHS *et al.* 2009).

5.3.2 Overview of Formal Consensus Methods

Within public health there are four main formal consensus methods, which are discussed below (Nair, Aggarwal and Khanna 2011).

Delphi Method

Developed in 1951 by the Research ANd Development (RAND) Corporation in conjunction with the United States Air Force, the Delphi Method was originally intended to objectively "obtain the most reliable consensus of opinion" from a group of multidisciplinary experts using a series of intensive questionnaires. The questionnaires are completed in private over multiple rounds with controlled feedback being provided after each round (Dalkey and Helmer 1962).

Over the years the Delphi method has been used in many areas, including healthcare (Broomfield and Humphris 2001; Beattie and Mackway-Jones 2004; Iqbal and Pipon-Young 2009; Eubank *et al.* 2016; Berman *et al.* 2017; Sudore *et al.* 2017; Yousuf 2007). Within healthcare screening the Delphi has been used to gain consensus on screening tests for glaucoma; referral criteria for school dental screening; consensus on PSA (prostate specific antigen) screening in asymptomatic men; and identifying components of a proposed blood-borne virus population screening programme, including acceptability of a reduced consent procedure for screening programmes (Kearney-Mitchell *et al.* 2006; Campbell *et al.* 2012; Crane, Henderson and Chadwick 2016; Prostate Cancer UK 2016).

Over multiple rounds participants are able to re-rate statements based on new information and/or the overall views of other panel members (Hsu and Sandford 2007)(Figure 55). The process continues until the pre-determined number of rounds has been completed.
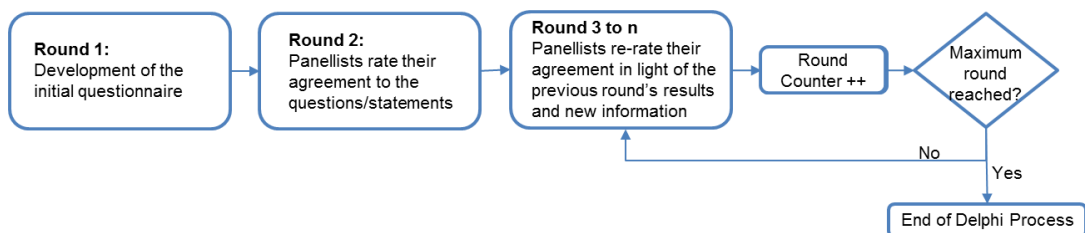
Figure 55: Steps in the Delphi Method

There have been modifications to specific elements of Delphi. The original approach asked for answers to specific questions, a subsequent modification was to have the panel rate their agreement to a statement on a Likert scale (Diamond *et al.* 2014). In their original study, Dalkey and Helmer (1962) noted that convergence of opinion was more likely for numerical answers (e.g. the number of bombs required) than non-numerical ones, (e.g. which targets to hit). Furthermore a numerical answer was suitable for statistical analysis, e.g. calculating the median response, to

determine the group's decision (Stitt-Gohdes and Crews 2004; Iqbal and Pipon-Young 2009). When rating agreement to a statement the Likert scale typically ranges from three to nine depending on the requirements of the researchers (Cassar Flores, Marshall and Cordina 2014; van Rijssen *et al.* 2017). The number of rounds in Delphi also varies according to the complexity of the topic and to minimise participant fatigue. Generally researchers propose three to four rounds due to diminishing returns, however up to 10 rounds have been reported (Thangaratinam and Redman 2005; Kearney-Mitchell *et al.* 2006; Iqbal and Pipon-Young 2009; Campbell *et al.* 2012; Sudore *et al.* 2017). Dalkey and Helmer (1962) stated that additional rounds may have improved convergence of opinion in their study, whereas others have stated that multiple rounds leads to participant fatigue and greater risk of participants dropping out (Linstone and Turoff 1975; Yousuf 2007; Shariff 2015).

The benefits of the Delphi technique include allowing participants to complete the questionnaires privately, enabling honest opinions to be captured without the influence of strong personalities, thereby reducing conformity (Hsu and Sandford 2007; Jones and Twiss 1978; Birko, Dove and Özdemir 2015). Participants are also offered the opportunity to re-rate their initial judgements based on the results and feedback from previous rounds (Yousuf 2007). In addition, as questionnaires are sent out via email/post, recruitment of participants is not limited by number or geography, keeping costs low (Murphy *et al.* 1998; Raine, Sanderson and Black 2005).

The lack of face-to-face meetings, however, is the primary limitation of the method, as areas of disagreement/uncertainty cannot be discussed, preventing a shared understanding from being reached. This places additional emphasis on the skills of the research team to present structured feedback to encourage a shared understanding without the researcher's own views and prejudices influencing feedback and, therefore, potentially the outcome of the study (Linstone and Turoff 1975; Murry Jr. and Hammons 1995). Furthermore, there are logistical issues of arranging and facilitating a face-to-face meeting, particularly with geographically dispersed participants. There is also evidence that statistical feedback between rounds induces conformity to the median value, without changing the median (Dalkey 1969; Salancik 1973; Ford 1975). Conformity also occurs when the median value is false, limiting the validity of the results, though this can be reduced by keeping the number of rounds and consensus statements to a minimum (Scheele 1975; Riggs 1983). Finally the lack of face-to-face contact also distances

participants from accountability in their ratings, which again influences the validity and reliability of the results (Sackman 1974; Murphy *et al.* 1998).

Delphi Technique provides a structured approach to gaining consensus over multiple rounds, with flexibility over the exact number of rounds, and the number and location of participants. Concerns surround the impact that the lack of face-to-face interaction might have on the consensus decision, and the increased emphasis on the researcher's ability to collate, understand, and present the feedback from previous rounds in a structured manner, without introducing bias which may affect the outcome of the study.

Nominal Group Technique (NGT)

Developed in the 1960s in the United States, the Nominal Group Technique (NGT) was intended to facilitate effective group decision making in social psychological research (Van De Ven and Delbecq 1971). A key element of the NGT is that individuals work in the presence of others, but do not immediately interact (Delbecq and Van De Ven 1971). Instead the method relies on the private generation of potential ideas/solutions to the proposed problem, after which the ideas are shared and discussed (Delbecq and Van De Ven 1971).

The NGT has been utilised in many areas, such as social services, education, government, industry and healthcare (Scott and Black 1991; Ziemba *et al.* 1991; Hunter *et al.* 1994; Spencer 2010; Porter 2012; Parthasarathy and Sharma 2014; Evans *et al.* 2017). In healthcare screening, NGT has been used to identify a suitable vision screening test for service members with traumatic brain injury, and identifying patient preferences for genetic testing for colorectal cancer (Radomski *et al.* 2014; Veldwijk *et al.* 2016).

Early approaches of NGT consisted of four rounds, typically conducted over a day, incorporating facilitated group discussions (Figure 56) (Van De Ven and Delbecq 1971; Nair, Aggarwal and Khanna 2011). In the first round participants privately generate ideas related to the proposed question, e.g. "which obesity-related co-morbidities should be included in a screening programme for children attending a UK weight management service?". In the second round the ideas are shared with the group in round-robin manner and are recorded for everyone to see. The ideas are then discussed and clarified to ensure group understanding in Round 3, during which additional ideas may be generated. The final round involves participants

rating/ranking the ideas privately, with the results used to determine consensus (Fink *et al.* 1984; Jones and Hunter 1995).

| Round 1: Silent generation of ideas | → | Round 2: Round robin sharing of generated ideas | → | Round 3 Clarification and discussion of ideas | → | Round 4 Voting/Ranking of idea |
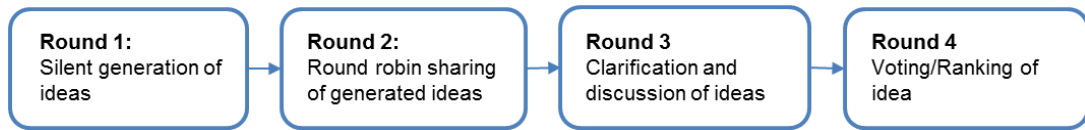
Figure 56: Steps in the Nominal Group Technique

The number of rounds has remained consistent over the years; however there have been modifications in other aspects such as the number of participants. Typically an NGT ranges from two to 14 participants, with seven regarded as the ideal (McMillan *et al.* 2014; McMillan, King and Tully 2016). However to overcome concerns that results from consensus studies are dependent on the participants included, Redman *et al.* (1997) recruited between 18 and 33 participants for each of 13 workshops held across Australia when gaining consensus on priorities for the National Breast Cancer Centre. Similarly Dobbie *et al.* (2004) utilised groups of up to 40 participants to elicit feedback when gathering student feedback regarding a medical education programme. Dobbie et al. also simplified the classic NGT procedure by providing specific questions for participants to answer rather than have them generate ideas, allowing for specific information to be captured, although this also restricted new idea generation. Other studies have generated ideas using a literature review, incorporating evidence, or exploratory surveys (Delbecq and Van De Ven 1971; Allen, Dyas and Jones 2004; Dobbie *et al.* 2004; Hiligsmann *et al.* 2013). Other modifications include the re-ranking of ideas over subsequent rounds, with the original participants either in the initial meeting, or a post meeting questionnaire (Gallagher *et al.* 1993; Allen, Dyas and Jones 2004). Alternatively, re-ranking has been completed with another group of participants to assess the results' reliability (Vella *et al.* 2000; Hiligsmann *et al.* 2013). The reasons for re-ranking ranged from assessing the representativeness of the results, to allowing changes based on the group discussion (Gallagher *et al.* 1993; Vella *et al.* 2000; Allen, Dyas and Jones 2004). The modifications highlight the flexibility of the approach, whilst still maintaining structure and transparency (Dobbie *et al.* 2004; Raine *et al.* 2004; Hiligsmann *et al.* 2013).

The advantages of the NGT include the opportunity for equal participation of group members, everyone's voice is heard as part of the face-to-face discussion and the voting process is done privately, limiting the influence of dominate personalities (Sample 1984; Murphy *et al.* 1998; Jones 2004). Additionally, the idea generation

round allows for, potentially, a broader and richer generation of ideas, which is further enhanced through the group discussion (Jones 2004). This in turn may lead to the development of other ideas through a democratic and open discussion (Raine *et al.* 2004).

The approach also has some disadvantages, it is reliant on the ability of the facilitator to ensure specific individuals do not dominate discussions and coerce consensus, and to keep the group engaged in the process for a prolonged period (Nair, Aggarwal and Khanna 2011). Due to the short timescales of the approach the original design does not allow participants time to consider the ideas, therefore it has been suggested that the process can feel rushed to those participants who require time to reflect, which may lead to a false consensus decision (van Teijlingen *et al.* 2006; Humphrey-Murto *et al.* 2017). The lack of anonymity from face-to-face meetings may result in participants being unwilling to share their views and ideas openly, due to dominant personalities or perceived superiority. Related to this, stronger personalities may influence the conversation and potentially influence the rankings, even when these are done privately (Nair, Aggarwal and Khanna 2011; Kosloff *et al.* 2017). However these issues can be overcome with experienced facilitation to encourage participation and management of dominant personalities (Kosloff *et al.* 2017).

The method classically relies on ideas generated by the panel, which may be suitable in areas with limited to no evidence, however may not be appropriate when some, possibly conflicting, evidence exists (Raine, Sanderson and Black 2005). This lack of evidence integration may hinder the process, as participants may not be aware of all the available evidence to make an informed decision. Finally there are practical issues relating to the time and financial resources required to undertake face-to-face meetings with a potentially large group of experts (Raine, Sanderson and Black 2005). Moreover, the requirement of face-to-face meetings limits attendance from geographically dispersed participants, affecting validity and reliability of the results.

Overall, the NGT does provide some benefits of gaining consensus, with the added flexibility of modifications to overcome some of the method's weaknesses. The main concerns are the lack the evidence integration in the standard method, and the reliance on participants' knowledge, which may not always be sufficient.

RAND/UCLA Appropriateness Method (RAM)

The RAND/UCLA Appropriateness Method (RAM) was developed in the 1980s as part of the RAND Corporation/University of California Los Angeles (UCLA) Heath Services Utilisation Study, primarily as a means to assess the over- and underuse of medical and surgical procedures (Fitch *et al.* 2001; Bourree, Michel and Salmi 2008). To support the process the best available scientific evidence is combined with the collective judgement of experts to provide a statement regarding the appropriateness of an intervention/procedure based on the characteristics of a specific patient population (Fitch *et al.* 2001; Nair, Aggarwal and Khanna 2011; Lawson *et al.* 2012). In the context of the approach, "appropriateness" referred to the relative weight of the benefits and harms of an intervention for a specific population (Park *et al.* 1986; Fitch *et al.* 2001).

As well as being used to identify the over- and underuse of surgical procedures, the RAM has also been used in other aspects of healthcare (Basger, Chen and Moles 2012; Bell *et al.* 2014; De Schreye *et al.* 2017). In healthcare screening, the approach has been used to identify an appropriate screening tool for depression for use in primary care, screening for colorectal cancer using colonoscopy, and assessing the appropriateness of imaging and treatment procedures when screening for ovarian cancer (Arditi *et al.* 2009; Nabbe *et al.* 2016; Expert Panel on Women's Imaging *et al.* 2017).

The RAM consists of three rounds (Figure 57). In the first round participants privately rate statements on a nine-point Likert scale, e.g. "based on the available evidence is hyperglycaemia appropriate for inclusion in the co-morbidities screening programme?". Typically the group median value is taken to assess agreement with the statement, with a median of 1-3 indicating 'inappropriate', 4-6 'uncertain', and 7-9 'appropriate'. Additional analysis is conducted to assess the level of agreement between participants. In round 2, participants discuss the results of round 1 in a face-to-face meeting, focusing on statements where there is uncertainty and/or disagreement. This allows for clarification of any misunderstanding, sharing of knowledge between participants and the discussion of the presented evidence. The final round allows participants to re-rate the statements in private, based on the discussions, the results of which determine the final consensus decision (Park *et al.* 1986; Park *et al.* 1989). Finally each statement is classified as 'appropriate', 'uncertain' or 'inappropriate' based pre-determined consensus criteria (Fitch *et al.* 2001).
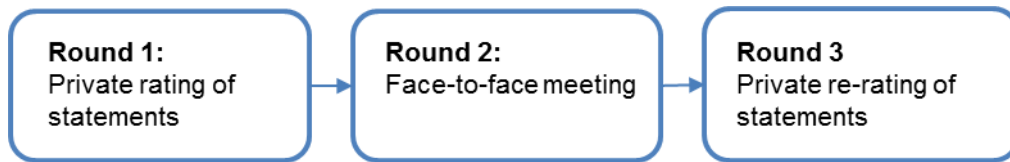
Figure 57: Steps in the RAM

Since its initial use, the procedure has undergone some modifications; however the underlying stages remain inherently the same. Stegbauer et al. (2017) incorporated two face-to-face meetings and three stages to rate the indicators. This extended process enabled Stegbauer et al. to assess the relevance (strength of the existing evidence and the benefit to the patient) and feasibility (suitability of the assessment instrument and implementation barriers) of the indicators. The number of experts has also varied with seven to 119 being reported (Khodyakov *et al.* 2011; Basger, Chen and Moles 2012; Brar *et al.* 2014; Stegbauer *et al.* 2017); however retention of experts varied, for example of the 119 experts recruited by Khodyakov *et al.* (2011) only 66% (78 panellists) engaged in all three rounds.

The RAM proffers many benefits. Firstly, the approach encourages the synthesis of published data to ensure participants are fully informed of the available evidence, and therefore able to make informed decisions. The RAM allows for a confidential rating and re-rating of statements and a face-to-face discussion to encourage a shared understanding.

In terms of disadvantages, there are time and cost implications of a face-to-face meeting; furthermore discussions can be dominated by strong personalities (Boulkedid *et al.* 2011; Nair, Aggarwal and Khanna 2011). Although this can be mitigated through appropriate facilitation. There is also reliance on the accuracy of the evidence presented to the panel, which is particularly of concern in areas where the evidence is contradictory or goes against one's held beliefs. Related to this, is how the data is presented to ensure any quantitative data can be clearly interpreted by a panel with varying levels of expertise and knowledge (Nair, Aggarwal and Khanna 2011).

Although the approach has acceptable validity and reliability, there are some concerns over the generalisability of the results which are typically based on the views of a small number of experts (Lawson *et al.* 2012). This can be overcome with larger group sizes, or by assessing the reliability of the results with a separate group of experts (Vella *et al.* 2000; Khodyakov *et al.* 2011). Additionally there are concerns over participant fatigue if a large number of statements are presented, for

example, Shekelle *et al.* (1998) requested ratings for more than 1000 indications related to hysterectomy. Additionally, a large number of statements will influence the length of the face-to-face meeting, thus increasing the potential for participant drop-out.

The RAM can be adopted for any project that needs a formal method for combining scientific evidence with expert consensus via a face-to-face meeting, and can be used in multiple situations. Some of the disadvantages can be overcome with appropriate facilitation, and due consideration needs to be given to the number of statements to minimise the possibility of participant fatigue and drop-out.

National Institutes of Health (NIH) Consensus Development Conference (CDC)

The Consensus Development Conference (CDC) was developed by the United States National Institute of Health in the 1980s (Fink *et al.* 1984). The method brings together stakeholders to reach consensus regarding a particular issue over the course of a two to three day meeting. Stakeholders are typically experts in the area; however, members of the public are sometimes also invited (Fink *et al.* 1984; Sorrell *et al.* 2009; Nair, Aggarwal and Khanna 2011; Buick *et al.* 2014).

As the approach was developed by the National Institutes of Health it has predominantly been used within healthcare, such as in an assessment of the associations between obesity, diabetes, and dyslipidaemia and cardiovascular disease (Consensus Development Conference on Antipsychotic Drugs and Obesity and Diabetes 2004). In healthcare screening, the approach has been used to assess the effectiveness of mammography screening for women aged 40-49, obtaining consensus on neonatal hearing screening, and guidance on the screening for, and management of, phenylketonuria (Sickles 1997; Grandori and Lutman 1998; NIH Consensus Statement 2000).

Although the method has undergone modifications over the years the general stages involve a group of experts developing questions or scenarios for which consensus is required (McGlynn, Kosecoff and Brook 1990) (Figure 58). A separate group of experts then presents information and evidence regarding the topic in question to a third group of experts, known as the decision-making group (Murphy *et al.* 1998). The decision-making group consists of experts from various fields, not always related to the topic in question, and occasionally includes members of the public (McGlynn, Kosecoff and Brook 1990). The evidence and presented

information is discussed by attendees, before the decision-making group convene to privately discuss the evidence (McGlynn, Kosecoff and Brook 1990). The decision-making group draft a consensus statement which is reviewed by conference attendees (Murphy *et al.* 1998). Following the review and further evidence or discussion, the decision-making group may modify their decision(s) before a final report is disseminated (Murphy *et al.* 1998). Both the open conference and the private group discussion are facilitated by an experienced chairperson (Black *et al.* 1999a; WHO 2014b).

Figure 58: Steps in the Conference Development Conference

Modifications to the approach typically relate to the number and background of the attendees. For example, two conferences considering prophylaxis and the treatment of osteoporosis had 700 and 2,000 attendees, and the decision making panel of 13 and 14 people, with representatives from the medical profession, the pharmaceutical industry, the press and ministries of health (Consensus development conference: prophylaxis and treatment of osteoporosis 1987; Consensus development conference: prophylaxis and treatment of osteoporosis 1991).

The benefits of the CDC include the use of a multidisciplinary group of experts, such as practicing physicians, researchers, consumers, and others, who assemble and evaluate the existing evidence before making a decision. Additionally, the inclusion of members of the public means that decisions take into account the views of the public and patients. Furthermore the decision making panel consists of experts in areas other than the conference topic and thus, are less likely to be biased. A key part of a CDC is the circulation of the results, typically via a press conference or a press release.

The primary limitation of the method is that a formal feedback system is lacking and the interaction is not generally structured, although this can be modified as part of the process (Nair, Aggarwal and Khanna 2011; Solomon 2015). Additionally, no formal guidance is given as to how consensus is ultimately reached; however specific methods, such as voting or rating statements on a Likert scale, such as in the RAM, can be adopted (Nair, Aggarwal and Khanna 2011). Finally, the

conferences take a substantial amount of time and cost to set up and run, due to the high number of potential attendees, and co-ordination of a large number of experts for two or three days. For instance Doo, Hoofnagle and Rodgers (2009) stated that their conference took two years of active planning for the three day conference. This is not always feasible due to budget and time constraints.

Overall, CDC conferences are particularly useful for providing guidance when a controversy exists, there is a high degree of public interest, or there is an impact on health care cost. However the limitations associated with the time and cost taken to set-up and run the conferences does limit its suitability for time-sensitive projects.

5.3.3 Enhancing the validity and reliability of Consensus Methods

Validity

Validity refers to whether the consensus decisions made through a consensus study are correct, and there are many approaches for assessing validity, including comparisons with gold standards, predictive validity, and concurrent validity (Murphy *et al.* 1998; Drost 2011). However, it has been recognised that it may not be possible to assess the validity of the consensus decision, i.e. whether it is a good or bad decision, at the time it is made (Murphy *et al.* 1998; Black *et al.* 1999b). Consensus studies are undertaken when there is uncertainty, therefore is no gold standard with which to compare the results. Predictive validity could be assessed once new evidence is made available, but cannot be assessed at the time of the consensus study. Concurrent validity can be assessed by comparing group judgements with existing evidence, and deviation from evidence may indicate a lack of concurrent validity.

The difficulty in assessing validity of consensus study results means the focus shifts to the method, and the steps undertaken so that the chosen method produces more good decisions and fewer bad decisions than alternative methods (Black *et al.* 1999b; Raine, Sanderson and Black 2005). Therefore it is imperative that the process is rigorous and transparent, by adhering to good practice in the planning, delivery, and reporting (Dixon-Woods *et al.* 2004; Raine, Sanderson and Black 2005). Waggoner, Carline and Durning (2016), after a review of consensus methods, suggested steps to improve future consensus studies, this included having clear concise questions where consensus is required, recruiting a panel of experts in the field, ensuring sufficient panel size, and including predefined statistical analysis

to determine whether or not consensus is achieved. These steps were adopted for the study to improve validity of the results.

Reliability

Reliability refers to the method, using the same information and questions, achieving the same results with a different group of people (Drost 2011). Previous research has indicated low levels of reliability when using consensus methods (Shekelle *et al.* 1998). This may be due to the reliance on the views of a small sample of experts. One method to overcome this is to include a diverse and representative group of experts in the study, and including a larger number of people to increase reproducibility of the results (Raine, Sanderson and Black 2005; USDHHS *et al.* 2009). Additionally, having a transparent and structured approach to synthesising the judgements of the panel enables replication of the study to test reliability of the results.

5.3.4 Integration of scientific evidence

Consensus methods are typically used in areas of uncertainty, due to a lack of evidence or conflicting evidence (Waggoner, Carline and Durning 2016). Therefore, to facilitate participant decision making it is important to provide current scientific evidence in a comprehensible format. Without evidence provision it has been suggested that participants are more likely to rely solely on their own experience and knowledge, which may be limited and bias decisions (Fink *et al.* 1984; Raine *et al.* 2004). Furthermore, the provision of evidence has been demonstrated to encourage knowledge transfer amongst the group and increase the content validity of the results (SAC 2002; Hutchings and Raine 2006; USDHHS *et al.* 2009). In relation to the development of a screening programme, the National Screening Committee (NSC) advocates the integration of evidence in the process, as this provides scientific justification for the committee's decisions, and highlights areas where research is insufficient, or further research is required (UK National Screening Committee 2017). For instance, the NSC did not recommend screening kernicterus in newborns due to a lack of evidence regarding the test's predictive accuracy (UK National Screening Committee 2017).

5.3.5 Expert Group

A key part of a consensus study relates to the participants recruited. Raine *et al.* (2004) indicated greater agreement in single discipline groups (GPs only) than in

multi-disciplinary groups (GPs and mental health professionals. Part of this difference in agreement may be due to the experience and training of the participants. Akins, Tolson and Cole (2005) reported that participants with similar training and knowledge were more likely to agree with each other than those with different training backgrounds. Although a single discipline group may increase consensus, it goes against recommendations from the NSC, who advocate for a multi-disciplinary group to increase the validity and reliability the results Waggoner, Carline and Durning (2016); (Raffle 2017b). Furthermore, Fink *et al.* (1984) stated participants should be selected because "they are representative of their profession, have power to implement the findings, or because they are not likely to be challenged as experts in the field" (p. 981). Based on this definition, the group should consist of individuals with experience of childhood obesity, have an awareness of associated co-morbidities, and be involved with weight management services. This definition would support the inclusion of a multi-disciplinary panel due to the broad impacts obesity has on the health and well-being of children.

In addition to the composition of the group, care is also required in how participants are identified, to reduce the potential for selection bias, which will reduce the validity of the method and therefore the validity of the results (Jones and Hunter 1995; Brett *et al.* 2017). Therefore it is important that the participant identification and selection process is transparent to reduce bias and enable replication of the study. Moreover, selection bias can be further reduced by having a multi-disciplinary group and allows for the consideration of a wider range of opinions and potentially greater knowledge transfer amongst participants, which enhances the content validity and reliability of the results (Hutchings and Raine 2006; USDHHS *et al.* 2009).

The final consideration is the panel size. Each consensus method has its own recommendations for the ideal panel size, therefore the number of participants will vary according to the consensus method adopted (Fink *et al.* 1984; Waggoner, Carline and Durning 2016). For face-to-face panels, such as the RAM and NGT, smaller sizes are recommended to minimise co-ordination and facilitation issues (Murphy *et al.* 1998; Waggoner, Carline and Durning 2016). In practice, panel sizes typically range from five to 12 members (Hanekom *et al.* 2015; Locke *et al.* 2015). The exception to this is the CDC, although the panel size typically ranges from nine to 18 member, once conference attendees are considered the total can be over a thousand (McGlynn, Kosecoff and Brook 1990). For Delphi studies panels of up to 3000 participants have been reported (Thangaratinam and Redman 2005). Inviting more participants increases the variety of expertise, and thus improves the reliability

of the findings, however eventually leads to diminishing returns and can limit consensus being achieved due to divergence of opinion (McMillan, King and Tully 2016). Therefore, when deciding on group sizes a balance needs to be maintained between maximising reliability of the findings, co-ordination problems, and diminishing returns (Black *et al.* 1999b).

5.3.6 Service User Group

In their definition of "expert", Fink et al. recommended including potential "consumers" whenever appropriate. In the context of the PhD and the proposed screening programme, "consumers" refers to the children and their parents attending the weight management service. This is known as service user involvement, or "participatory research", which refers to the process in which current or historic service users become involved in the planning, development and delivery of that service (Gray *et al.* 2000; NHS England 2015). There has been a growing emphasis over the years in the importance of obtaining views from service users, via patient/public involvement, which has increased in all aspects of healthcare and health-related research, including the development of healthcare treatments and decisions about treatment (Gray *et al.* 2000; McLaughlin 2010; Wainwright, Boichat and McCracken 2013). This commitment to patient involvement is evident from documents such as 'Choosing Health and Our health, Our Care, Our Say' (Department of Health 2004; Department of Health 2005; Department of Health 2006; Department of Health 2010).

The benefit of involving service users/patients is that it allows them to offer an insight into the nature of their condition, what is important to them, and how it should be addressed, thereby increasing the validity and reliability of the consensus results (Minogue *et al.* 2005; McLaughlin 2010). Furthermore service users' views may differ to the views of healthcare professionals who may lack this understanding (van Wersch and Eccles 2001; Campbell *et al.* 2004; EPP Evaluation Team 2005; Cheng *et al.* 2010).

However, the complex nature of facilitating mixed groups of service users/patients and professionals may be a barrier to effective service user/patient involvement (Van De Ven and Delbecq 1972). It has been suggested that some service users/patients, particularly young people, may not speak openly if healthcare professionals are present (Graham *et al.* 2014). They may lack the confidence to challenge someone in a position of authority who uses technical/medical terminology, thereby making aspects of the discussion inaccessible to non-experts

(Van De Ven and Delbecq 1972; Worrall-Davies and Marino-Francis 2008). This also relates to the detail of information provided and how it is formatted for service users, e.g. providing plain English versions. With regards to the inclusion of children, there is evidence that children under 12 may not have the required expressive language and social interaction skills (Greenbaum 1988; Clark 1996; Vaughn, Schumm and Sinagub 1996). Additionally developmental differences within children mean that participants should be within a two-year age span and mixed-gender groups are not always recommended (Greenbaum 1988; Spethman 1992; Charlesworth and Rodwell 1997; Maccoby 1998; Kennedy, Kools and Krueger 2001).

5.3.7 Opportunity to privately re-rate initial ratings

All the methods advocate the re-rating of initial decisions in private in light of face-to-face discussion and/or new information (Nair, Aggarwal and Khanna 2011). Although multiple rounds may increase participant fatigue and drop-out; to maintain reliability and validity of the results, at least two rounds of rating would be recommended, the maximum number depends on the method adopted (Nair, Aggarwal and Khanna 2011; Waggoner, Carline and Durning 2016).

5.3.8 Pre-determined definition of consensus

When defining consensus consideration should be given to levels of agreement, which has been defined in two ways:

1. Agreement with a statement: The extent to which each member of the consensus panel agrees with a statement, e.g. should a co-morbidity be considered for inclusion in a screening tool?
2. Agreement with each other: The extent to which the panellists agree with each other regarding a co-morbidity's inclusion in a screening tool?

(Jones and Hunter 1995)

Within the literature there does not appear to be a universally accepted definition of consensus, in their systematic review Diamond *et al.* (2014) identified 11 definitions of consensus from 98 studies. The definitions ranged from percent agreement to a mean value above a predefined cut-off. Data from the 98 studies also indicated that consensus was never defined as full agreement between participants. In fact consensus study guidelines support this, saying full agreement amongst participants is not required to achieve consensus; however it is recommended that an a priori

definition of consensus is agreed (Linstone and Turoff 1975; Beattie and Mackway-Jones 2004; Campbell *et al.* 2004; Butler *et al.* 2009). Definitions of consensus typically refer to some measure of central tendency, e.g. the median, along with a measure of dispersion, e.g. range; the former providing the level of agreement with the statement and the latter the level of agreement/disagreement between participants. Of the four methods only the RAM provides explicit and transparent guidance on methods for statistically analysing the data to derive the group's consensus judgement and to highlight areas where there is uncertainty and/or disagreement amongst the panel members (Fitch *et al.* 2001). Methods do provide guidance, however there is variation between studies with no approach being consistently adopted (Diamond *et al.* 2014). Waggoner, Carline and Durning (2016) recommend a predefined method to determine consensus to increase the validity and reliability of the results.

5.3.9 Summary of Methodological Considerations and chosen Consensus Methods

Studies comparing different consensus methods vary in their results, some studies report that there is little difference between methods based on the results, e.g. Washington *et al.* (2003) tested whether a mail-only process could substitute the standard in-person process normally used in a RAM. They used both to assess the appropriateness of coronary revascularisation and hysterectomy, and concluded that both, mail-only and in-person, gave comparable results and suggested that future studies could be conducted by mail-only. Similar results were found by Kadam, Jordan and Croft (2006) when comparing NGT with Delphi, suggesting limitations associated with face-to-face meetings, i.e. geographically dispersed participants can be overcome using other methods. In contrast, when comparing single to multi-disciplinary groups, Raine *et al.* (2004) obtained consensus in treatment for irritable bowel syndrome, chronic fatigue syndrome, and chronic back pain, comparing a GP-only group with a mixed speciality group (GP and mental health professionals). Results demonstrated that agreement was more likely if the group was GP-only, a literature review was provided, or the evidence was in accordance with clinicians' existing beliefs. Highlighting the importance of integrating evidence, and the potential drawbacks of having a multi-disciplinary group and contradictory evidence. However, the broad impact of the identified co-morbidities and the potential range of stakeholders, e.g. GPs and weight

management services, in the proposed screening programme would require the views from a multidisciplinary panel.

To address the aim of the consensus study it was important that the chosen method met specific criteria. Table 36 presents an overview of the aforementioned consensus methods in light of the elements deemed important for the study, and additional strengths and weaknesses (Waggoner, Carline and Durning 2016). The key criteria for the method are the integration of scientific evidence, a multidisciplinary panel, including service users, the ability to privately rate and re-rate statements, and predefined methods to determine the group's decision and agreement with the decision (as detailed in Sections 5.3.3 to 5.3.8). Of the four methods reviewed, RAM met all but one of the criteria – the inclusion of service users; however previous studies have modified the method to include service user views in the decision making process (Van De Ven and Delbecq 1971; USDHHS *et al.* 2009; Coleman *et al.* 2014). Based on Table 36 the RAM seemed the most appropriate method for obtaining consensus on co-morbidities and screening methods to include in the proposed screening programme.

Table 36: Comparison of the consensus methods against key criteria

Method Criteria	Delphi	NGT	RAM	CDC
Integration of scientific evidence	�×	�×	✓	✓
Multidisciplinary Group	✓	✓	✓	✓
Face-to-face meeting	✗	✓	✓	✓
Private rating and re-rating	✓	✓	✓	✗
Service User Involvement	✗	✗	✗	✓
Pre-defined definition of consensus	✗	✓	✓	✗
Structured interaction	✗	✓	✓	✗
Formal Feedback methods	✓	✓	✓	✗
Other Strengths	• Cheap and	• Democratic prioritisation	• Encourages knowledge	• Panel are not topic

Method Criteria	Delphi	NGT	RAM	CDC
	convenient • Large number of participants	of ideas	transfer between panel members	experts, minimising bias
Other Weaknesses	• Emphasis on questionnaire design • Lack of accountability	• Group voting can occur • Dominant personalities may influence results	• Dominant personalities may influence results • Time and cost implications	• Time and cost implications • Co-ordination difficulties

5.4 Methods

5.4.1 Design

Based on the aforementioned requirements (detailed in sections 5.3.3 to 5.3.8 and summarised in Table 36), this PhD utilised the RAM with a multidisciplinary panel of experts (GPs, paediatricians, academics and researchers), and a modification to incorporate service user views (Van De Ven and Delbecq 1971; USDHHS *et al.* 2009; Coleman *et al.* 2014). The study consisted of two consensus cycles; the first cycle obtained consensus on which co-morbidities to include in the screening programme and the second cycle obtained consensus on screening methods for the agreed co-morbidities (Figure 59). The stages of the consensus study are described below.

NHS ethical approval was obtained prior to commencing the study (HRA (Health Research Authority) reference: 15/YH/457).
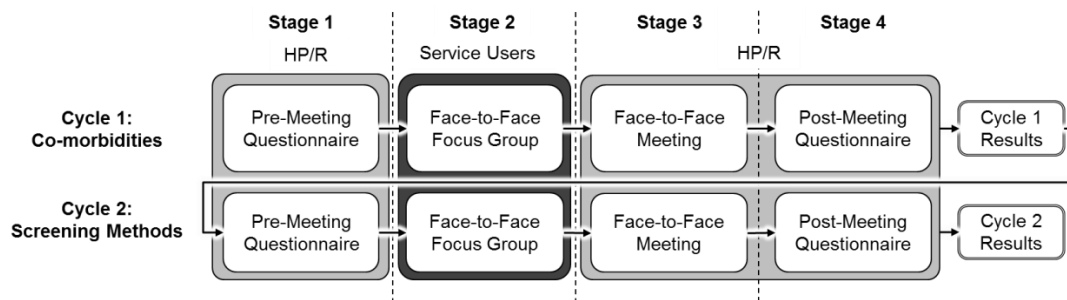
Figure 59: Flowchart for Stage 2 (HP/R – Health Professionals and Researchers)

Stage 1: Pre-meeting Questionnaire

Data from the systematic review and meta-analyses (Chapters 3 and 4) provided the basis for a questionnaire sent to the expert panel who rated their agreement/disagreement for the inclusion of the co-morbidities in the screening programme on a 9 point Likert scale. A summary of the systematic review and meta-analyses' results were made available within the questionnaire. After the pre-meeting questionnaire was returned, the median response for each statement was calculated along with the RAND Disagreement Index (Fitch *et al.* 2001). The median provided the panel's decision as to whether the co-morbidity should be included in the screening programme, and the disagreement index considered the dispersion of scores to identify disagreement between participants, i.e. where participants rated at both ends of the 9-point scale.

Stage 2: Service User Focus Group

A face-to-face meeting with a service user panel was facilitated to review the results of the pre-meeting questionnaire, and discuss the acceptability of the co-morbidities thought to be important by the expert group and those where there was uncertainty and disagreement (Figure 62, page 161). Participant comments were captured and summarised for presentation to the expert panel.

Stage 3: Face-to-Face Meeting

A face-to-face meeting with the expert panel was facilitated to review the results of the pre-meeting questionnaire, with a focus on those co-morbidities where there was disagreement or uncertainty about its inclusion, alongside comments from the service user panel and results from the systematic review and meta-analyses.

Stage 4: Post-Meeting Questionnaire

Following the meeting the expert members were asked to re-rate their initial decisions about the inclusion/exclusion of co-morbidities in the screening programme using an updated questionnaire from stage 1. The update included comments from the service user panel (Stage 2) and an overview of discussion from the expert panel's face-to-face meeting (stage 3). The results from the post-meeting questionnaire provided a list of potential co-morbidities to include in the screening programme based on predefined analysis methods (Section 5.5.1).

The above stages were then repeated for Cycle 2 to obtain consensus on which screening methods were deemed suitable for the agreed upon co-morbidities.

5.4.2 Participants

Expert Panel

The expert panel comprised of national clinicians and researchers with international standing, who were identified by their expertise in child obesity and its co-morbidities. An initial group were selected through discussion with supervisors and contacted to take part in the study, at which point they were asked to recommend others with expertise in the topic area. The group was purposively sampled to include the perspectives of nurses, doctors (paediatric endocrinologist, paediatrician, endocrine specialist nurse and general practitioner), and researchers/academics to consider the wide range of co-morbidities associated with child obesity (Fraser *et al.* 1994; Fitch *et al.* 2001; Waggoner, Carline and Durning 2016). The characteristics listed in Table 37 were considered relevant to the aim of the study and a multi-specialty group was selected due to the potentially broad implications resulting from the screening programme, and in order to incorporate a wide range of opinions and increase the validity of the results (Hutchings *et al.* 2006).

Thirty-five experts were invited to take part in the study. The optimum number in order to prevent co-ordination problems, whilst maximising reliability of the findings is 6 according to Brook (1995), whereas others state groups of between 7-15 (Fraser *et al.* 1994; Fitch *et al.* 2001). Given the broad range of co-morbidities and the number of services potentially impacted, inviting 35 expert participants was considered justified to enable a proper assessment and consideration of the evidence, as well as allowing for attrition over the two cycles.

Service User Panel

The service users were recruited from one weight management service, which has seven sites across Leeds. Weight management services are community based services that provide diet and lifestyle advice to help an individual lose weight. The service offers a community based weight management programme for children aged 5 to 18 years. This includes individual and family interventions related to adopting healthy lifestyles.

Due to potential problems that may arise from having experts and service users in the same group (Section 5.3.6, page 151), expert and service user groups were held separately (Campbell *et al.* 2004; Worrall-Davies and Marino-Francis 2008; Coleman *et al.* 2014; Graham *et al.* 2014). Although research recommends single

gender groups (Maccoby 1998), discussion with the staff at the weight management service and parents and children indicated that they would be happy participating in mixed gender groups. This was largely due to the exercise sessions being held in mixed-gender groups, thus the children felt more comfortable with one another. Children were also of similar ages, meeting recommendations that children should be within a two-year age range due to developmental differences (Greenbaum 1988; Charlesworth and Rodwell 1997; Maccoby 1998).

Prospective parents/carers and children were identified and initially approached by weight management staff based on the eligibility criteria (Table 37). Children under 12 were excluded due to research indicating they may lack language abilities (Clark 1996; Vaughn, Schumm and Sinagub 1996); however their parents were invited to take part to offer the perspective of younger children. After they agreed to speak to the researcher, they were approached by the researcher who went over the details of the study with a participant information sheet and provided contact details for the researcher should they have any question. All parents were consented on to the study and children provided assent.

Table 37: Participant eligibility criteria

Expert Panel	Service User Panel
Inclusion Criteria: • Experience of working in the field of childhood obesity, e.g. clinician, researcher, or weight management services. • Knowledge/awareness of co-morbidities associated with childhood (aged 5-18) obesity. • Able to take part in all parts of the study (pre-meeting questionnaire, face-to-face meeting, post-meeting questionnaire). • Able to speak English. Due to the nature of the study participants must understand English and be able to contribute to the face-to-face meeting.	Inclusion Criteria: • A child or parent/carer of a child who currently attends a weight management service. They are aware of current assessments, and can offer their views on any proposed assessments. • Service user is aged 12 - 18 years at consent. • The parent/carer age> 18. • Either the service users under 16 are able to give assent AND parent/carer is able to give informed consent OR service users 16 and over are able to give informed consent. • Able to speak English. Participants must understand English to contribute to the meeting.
Exclusion Criteria: • No exclusion criteria	Exclusion Criteria: • Aged under 12. • Aged over 18. • Unable to give informed assent (12<age<16) or consent (age>16).

5.4.3 Cycle 1: Consensus on Co-morbidities

Stage 1: Expert Panel Pre-meeting Questionnaire

Results of the systematic review and meta-analyses were presented within a questionnaire and sent out to the expert panel of health professionals and researchers. The questionnaire included details of the co-morbidity and associated health implications, and prevalence and prevalence ratio information from the systematic review and meta-analyses (Figure 60). The expert panel were asked to review the evidence and privately rate the appropriateness of a co-morbidity's

inclusion in the screening programme on a 9-point scale (1 – strongly disagree to 9 – strongly agree), as per existing RAM literature (Fitch *et al.* 2001) (Figure 61). The expert panel were encouraged to consider the evidence alongside their knowledge and experience to inform their decision. At the end of the questionnaire participants were provided the opportunity to suggest other co-morbidities that they felt had not been identified by the systematic review.

Co-morbidity 4: Flatfoot

Definition:	Feet have low or no arches and press almost completely flat against the ground
Impact(s):	Pain in the feet, ankles, lower legs, knees, hips or lower back. The feet roll inwards too much (over-pronate) – this can lead to injuries. Problem with the bones, muscles or connective tissues in and around the feet
Number of studies:	2, 2 eligible for analysis
Age range:	5-19
Data collection:	Not stated
Methods:	Physical examination and assessment
Cut-off(s):	Mild to severe depression of the arch with a score of >0.26 on the Arch Index
Design/Setting:	Cross sectional (1 prospective, 1 retrospective) studies conducted across community and educational settings
Prevalence Ratio:	For every 1 healthy weight child/adolescent there would be 1.3 (1.3 to 1.4) overweight children/adolescents with flatfoot For every 1 healthy weight child/adolescent there would be 1.8 (1.4 to 2.3) obese children/adolescents with flatfoot
Graph:	Prevalence and 95% CI of flat foot among healthy weight, overweight and obese children/adolescents

Author, Year	Country	Sample Size (M/F)*	Age Range	Healthy Weight* BMI <85th %tile	Overweight* BMI≥85th <95th %tile	Obese* BMI≥95th %tile
Tenenbaum, 2013	Israel	825964 (467412/358552)	16-19	6.7 (6.7-6.8)	8.9 (8.8-9.1)	10.9 (10.7-11.1)
Chen, 2009	Taiwan	1024 (549/475)	5-13	26.7 (23.2-30.3)	31.4 (22.5-40.2)	55.8 (35.5-76.1)
Summary	N/A	N/A	N/A	16.7 (-36.3)	13.6 (-17.7-44.9)	32.1 (-11.8-76.1)

*NR - Not Reported (the value was not provided in the article)
*Prevalence and 95% CI
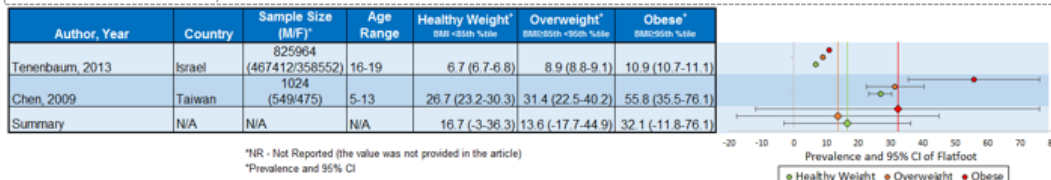
Figure 60: Example of how data was presented in the expert panel questionnaire

Based on the above evidence and your knowledge/experience data item(s)/clinical measure(s) relating to **Flatfoot** should be recorded in the co-morbidity screening tool

Strongly Disagree			Neutral: Neither agree nor disagree					Strongly Agree
○	○	○	○	○	○	○	○	○
1	2	3	4	5	6	7	8	9

Figure 61: Example question from the Cycle 1 questionnaire

Stage 2: Service User Panel Meeting

Once results of the expert panel pre-meeting questionnaire had been received and analysed according to existing literature to calculate the disagreement index (Section 5.5.1) (Fitch *et al.* 2001). The results were presented to a focus group of

service users. As well as the questionnaire results, the service users were provided with plain English versions of the systematic review and meta-analyses results (Figure 62). Service users were asked to discuss the importance of a co-morbidity's inclusion based on the evidence, the expert panel's initial views, and their own personal perspective.

1. FLATFOOT

Description: feet have low or no arches and press almost completely flat against the ground

Consequences: may cause pain in feet, ankles, knees, hips and lower back

Percentage:
17% of healthy weight
14% of overweight
32% of very overweight children/adolescents

Assessment Method: questionnaire

Should we screen for Flatfoot?

Care Professionals/Researchers' Initial View: Unsure

Your Comments:

Figure 62: Example of how data was presented to the Service User Panel

Stage 3: Expert Panel Meeting

The first expert panel meeting took place on the 21st of June 2016 and was facilitated by the researcher (VS) with support from a supervisor (SC). The aim of the meeting was to discuss all the co-morbidities, focusing on those where there was uncertainty and/or disagreement over the co-morbidities inclusion based on the results of the pre-meeting questionnaire and the comments from the service users. To assist with discussions the experts were provided with a results pack which contained information about the co-morbidities, as provided in the pre-meeting questionnaire, comments from the service user group, and results of the pre-meeting questionnaire. The pre-meeting questionnaire results consisted of the group median and disagreement index result, the absolute range of responses, and the member's individual response (this individual response was not shared with the others). The meeting was audio-recorded with the participants' consent for subsequent transcription and thematic analysis.

During the meeting, discussion considered the acceptability of certain screening tests, specifically blood tests for co-morbidities such as hyperglycaemia and dyslipidaemia. Acceptability related to both service users having blood taken and

weight management staff taking blood. After some discussion it was felt that additional information was required in line with the NSC's criteria (Public Health England 2015a). To assess the acceptability of blood tests amongst service users and staff a questionnaire was proposed, details of this are in section 5.4.4.

Stage 4: Expert Panel Post-meeting questionnaire

Following the face-to-face meeting, the panel members were asked to privately re-rate their initial responses regarding the appropriateness of including each co-morbidity in the screening programme. In the post-meeting questionnaire all co-morbidity statements were preceded by a summary of the systematic review and meta-analyses evidence, as per the pre-meeting questionnaire. In addition the panel were provided with comments from the service users and a summary of the discussions which took place during the expert panel's face-to-face meeting. As per the pre-meeting questionnaire, each statement was rated on a 9-point scale and analysed in the same way.

5.4.4 Acceptability of Blood Tests: a sub-Study

Discussions in the cycle 1 face-to-face meeting between the experts raised questions regarding the acceptability amongst staff, parents and children of blood tests. The panel were concerned that children would not want to have a blood test conducted, and that staff would not feel comfortable taking blood. As such a sub-study was proposed, with the aim of assessing acceptability of blood tests, finger prick and venous, amongst weight management staff and service users, parents and children. To assess this a questionnaire was developed with assistance from a member of the expert panel. Both questionnaires assessed whether the staff/parents and children were comfortable/happy taking/giving blood, and their preferred method, finger prick, venous, or other (Appendix 3). Furthermore space was provided for any comments. For staff to be eligible, they had to have face-to-face contact with service users, as these were likely to be the staff taking bloods under the proposed screening programme. For service users, the eligibility criteria related to attendance at an existing weight management programme. The questionnaires were distributed to staff and service users across three regional weight management services in Yorkshire (Shine, MEND and WatchIT). Staff questionnaires were distributed by senior staff, and service user questionnaires were distributed by front-line staff to service users as they attended their regular

appointments. Completion of the questionnaires was optional and consent was assessed based on return of a completed questionnaire.

5.4.5 Cycle 2: Consensus on Screening Methods

Cycle 2 focused on gaining consensus on suitable screening methods for use within a weight management service for the co-morbidities agreed upon in cycle 1. Potential screening methods were identified by reviewing articles from the systematic review and supplemented by a web search to identify new methods. A web search was also conducted to identify sensitivity and specificity information, studies where the tool has been used, and questions for self-report tools. Cycle 2 followed the same steps as Cycle 1 (detailed above) with a pre-meeting questionnaire, a service user panel, a face-to-face meeting with the expert panel, and a post-meeting questionnaire.

5.5 Data Analysis

5.5.1 Quantitative Analysis for Pre- and Post-Meeting Questionnaires

Questionnaire statements were summarised using the median group response as a measure of central tendency, in line with the RAMs (Fitch *et al.* 2001). The medians were grouped into tertiles; 1-3 – disagree; 4-6 – uncertain; 7-9 – agree. Additionally agreement and disagreement between the expert panel was calculated using the RAND Disagreement Index (Fitch *et al.* 2001), which considers the dispersion of individual scores and identifies statements where panellists rated at both ends of the 9-point Likert scale. For example, to calculate the group consensus and disagreement index for the data presented in Table 38 the following steps are taken:

Table 38: Example data to demonstrate calculating the disagreement index

Participant:	1	2	3	4	5	6	7	8	9
Rating:	7	8	7	4	9	8	8	9	6

- Calculate the median for the responses, which is 8 and gives the group consensus rating.
- Calculate the 70th and 30th percentile values, which are 8 and 7, respectively
- Calculate the inter-percentile range (IPR), which is the 70th percentile – 30th percentile, e.g. 8-7 = 1

- Calculate the Central Point IPR, which is the 70% percentile + 30th percentile / 2, e.g. (8 + 7)/2 = 7.5

- Calculate the Asymmetry Index which is the Central value on the 9-point scale – the Central Point IPR; e.g. 5 - 7.5 = 2.5. Note the absolute value is taken

- Calculate the IPRAS (Inter-percentile Range Adjusted for Symmetry) which is 2.35 + (1.5* Asymmetry Index); e.g. 2.35 + (1.5 * 2.5) = 6.1

- The IPR and the IPRAS are then compared. If IPR > IPRAS there is disagreement, if the IPR < IPRAS there is agreement.
  For the example the IPR (1) < IPRAS (6.1), therefore there is agreement over the group decision (median value of 8).

- Using the group median and disagreement index results, the following rules were applied to the post-meeting questionnaire in both cycles to indicate the consensus decision:
    a. Median 1-3 (disagree) without disagreement were considered to have face validity and would be excluded
    b. Median 7-9 (agree) without disagreement were considered to have face validity and would be included
    c. Median 4-6 or where there was disagreement would be excluded but are potential areas for further research

- As the median is between 7 and 9, and there is agreement the co-morbidity in the example would be included in the screening programme.

5.6 Results

5.6.1 Participants

Expert Panel

Thirty-five experts were invited, 21 (60%) agreed to take part. The expert panel comprised researchers with international standing, healthcare professionals (endocrine nurses, doctors (paediatric endocrinologist, paediatrician, and general practitioner)), and weight management staff from community weight management services.

Service User Panel

The service user panel comprised of children and their parents attending a weight management service in Leeds. Eight (four adults and four children) took part in Cycle 1 and six (three adults and three children) in Cycle 2.

The results relating to the co-morbidities (Cycle 1) and the screening methods (Cycle 2) are detailed below.

5.6.2 Cycle 1: Consensus on Co-morbidities

All of the 21 experts who agreed to take part completed the Cycle 1 pre-meeting questionnaire, 12 (57%) panel members attended the Cycle 1 face-to-face meeting, and 18 (86%) completed the post-meeting questionnaire.

Results of the pre-meeting questionnaire indicated there was agreement to include 10 co-morbidities, to not include two, and there was uncertainty over inclusion of eight co-morbidities. For the remaining co-morbidity (Joint Pain), there was disagreement over the group's decision of uncertainty about its inclusion (Table 39). The results of the pre-meeting questionnaire were then presented to the service user panel on the 6th June 2016. The service users discussed the co-morbidities one-by-one along with the results from the expert panel pre-meeting questionnaire. The service users agreed with the inclusion of the 10 co-morbidities identified by the expert panel, and with the initial decision to not include two co-morbidities (bone fractures and traumatic dental injuries). The service user group's reasoning was to not include co-morbidities that they or someone else, e.g. a teacher, could identify, and include those that may not have obvious symptoms. Using this approach they were able to offer their opinion on the nine co-morbidities where the expert panel were uncertain; choosing to include four (asthma, elevated uric acid, exercise induced wheeze/cough, and iron deficiency) and not to include the remaining five (ADHD, enuresis, flat foot, gallstones, and joint pain).

At the face-to-face meeting with the expert panel (21st of June 2016), co-morbidities were discussed in turn, taking into consideration the results of the pre-meeting questionnaire and the comments from the service user group. Initially dyslipidaemia was grouped under cardiovascular risk factors for the pre-meeting questionnaire and the service user focus group, however at the expert face-to-face meeting the discussion indicated it should be considered separately, hence there is no pre-meeting rating for dyslipidaemia. As well as considering the evidence presented and the pre-meeting questionnaire results, the panel also considered the potential

screening tests for the co-morbidities, to assess its suitability for administration in a weight management setting.

The final stage in cycle 1 was the post-meeting questionnaire; results indicated the expert panel did listen to some of the service users views. For instance, for the five co-morbidities the expert panel were initially uncertain about and the service users said to not include, the post-meeting results indicated that expert panel agreed to not include those co-morbidities. However, for other co-morbidities the expert panel went against the views of the service users, for example, the service users wanted to include elevated uric acid, however the face-to-face discussion prompted the expert panel to not include elevated uric acid. Review of the post-meeting questionnaire indicated there was still some uncertainty over specific co-morbidities, such for cardiovascular risk factors, which was previously ranked as include. Analysis of the post-meeting questionnaire indicated that the expert panel agreed for the inclusion of 7 co-morbidities in the proposed screening programme; hyperglycaemia, dyslipidaemia, hypertension, obstructive sleep apnoea, depression, anxiety and low self-esteem (Table 39).

Table 39: Summary of results for cycle 1 - co-morbidities
(There was agreement to include those highlighted in green; *Dyslipidaemia was originally grouped under cardiovascular risk factors, however discussion at the expert face-to-face meeting indicated it should be separated out, and hence there is no pre-meeting rating for dyslipidaemia.)

Co-morbidity	Pre-meeting (median (range))	Pre-Meeting Decision (Disagreement Index)	Service User Group's Opinion	Post-Meeting (median (range))	Post-Meeting Decision (Disagreement Index)
ADHD	5 (2-7)	Uncertain (Agreement)	Do Not Include	2 (1-7)	Do Not Include (Agreement)
Anxiety	8 (3-9)	Include (Agreement)	Include	7 (2-9)	Include (Agreement)
Asthma	6 (1-8)	Uncertain (Agreement)	Include	5 (1-8)	Uncertain (Disagreement)
Bone Fractures	3 (1-6)	Do Not Include (Agreement)	Do Not Include	1 (1-5)	Do Not Include (Agreement)
Cardiovascular Disease	7 (3-9)	Include (Agreement)	Include	5 (2-9)	Uncertain (Agreement)
Depression	7 (3-9)	Include (Agreement)	Include	8 (2-9)	Include (Agreement)
Dyslipidaemia*	N/A	N/A	N/A	7 (2-9)	Include (Agreement)
Elevated Uric Acid	5 (2-8)	Uncertain (Agreement)	Include	2 (1-7)	Do Not Include (Agreement)
Enuresis	4 (1-7)	Uncertain (Agreement)	Do Not Include	2 (1-4)	Do Not Include (Agreement)
Exercise Induced Wheeze/Cough	6 (1-8)	Uncertain (Agreement)	Include	4 (1-7)	Uncertain (Agreement)
Flat foot	5 (2-8)	Uncertain	Do Not	2 (1-7)	Do Not Include

Co-morbidity	Pre-meeting (median (range))	Pre-Meeting Decision (Disagreement Index)	Service User Group's Opinion	Post-Meeting (median (range))	Post-Meeting Decision (Disagreement Index)
		(Agreement)	Include		(Agreement)
Gallstones	4 (2-8)	Uncertain (Agreement)	Do Not Include	2 (1-5)	Do Not Include (Agreement)
Hepatic Disorders	7 (1-9)	Include (Agreement)	Include	5 (2-8)	Uncertain (Agreement)
Hyperglycaemia	8 (2-9)	Include (Agreement)	Include	8 (4-9)	Include (Agreement)
Hypertension	8 (1-9)	Include (Agreement)	Include	7 (3-9)	Include (Agreement)
Iron Deficiency	6 (2-9)	Uncertain (Agreement)	Include	3 (1-7)	Do Not Include (Agreement)
Joint Pain	5 (1-9)	Uncertain (Disagreement)	Do Not Include	3 (1-7)	Do Not Include (Agreement)
Low Self-Esteem	8 (3-9)	Include (Agreement)	Include	8 (1-9)	Include (Agreement)
Metabolic Syndrome	8 (1-9)	Include (Agreement)	Include	4 (1-8)	Uncertain (Agreement)
Obstructive Sleep Apnoea	7 (2-8)	Include (Agreement)	Include	7 (3-8)	Include (Agreement)
Traumatic Dental Injuries	2 (1-5)	Do Not Include (Agreement)	Do Not Include	1 (1-3)	Do Not Include (Agreement)
Vitamin D Deficiency	7 (1-9)	Include (Agreement)	Include	5 (1-8)	Uncertain (Disagreement)

5.6.3 Acceptability of Blood Tests: a sub-Study

Results of Staff Member Questionnaire

Thirteen staff members from the three participating services completed the questionnaire, only five gave their job title, four saying they were "facilitators" and one a senior nurse. Of the 13, 12 said they would be comfortable taking a blood sample from a child. Of these 12, eight had a preference for just taking a finger prick sample, whereas the remaining 4 were happy with either a venous or finger prick sample. The staff member who said no to taking blood samples was concerned that taking blood would affect recruitment and retention of service users, and furthermore was worried about the impact it might have on the on-going relationship with families. In contrast, those who were comfortable taking blood said:

> *"I feel the child will feel more comfortable with someone they know*
> *and trust and they won't have to wait as long to get a diagnosis"*

"Think this would be really cost effective and save long waits as GPs don't do blood screening, they refer to consultant at SCH so families have to wait."

"Think the children have more trust in the staff and many are anxious about having bloods taken so more likely to attend for bloods. Also better to receive results from us as we have developed relationships with families."

Two others stated they were already trained to take blood.

Results of Service User Questionnaire

Forty-six service users (23 parents and 23 children) took part in the survey. The average age of the children was 11.3 (range 9-14), and 57% were male (13/23). Nineteen of the 23 children were happy to have a blood sample taken, with a greater preference for either a finger prick test (7/19) or either finger prick or venous (7/19). The remaining five had a preference of a venous blood sample. Of the four kids who said they would not want a blood sample to be taken; only two gave a reason which was a fear of needles.

In terms of the parents, 91% (21/23) were happy for a blood sample to be taken from their child by weight management staff, either via finger prick or venous. Comments as to why they would be comfortable with this, included, "because we trust the staff", "less hassle", and "it can be time consuming and paying for parking to always attend hospitals, I think if staff are trained its ok to be done outside of clinical settings".

Overall both staff and service users were comfortable with blood samples being taken within a weight management service setting.

5.6.4 Cycle 2: Consensus on Screening Methods

During Cycle 2 the panel considered 29 screening methods for the 7 co-morbidities agreed upon during Cycle 1. Seventeen of the 21 (81%) expert members completed the Cycle 2 pre-meeting questionnaire, 11/21 (52%) expert members attended the Cycle 2 face-to-face meeting, and 16/21 (76%) completed the post-meeting questionnaire.

Two experts who were unable to attend the Cycle 1 meeting attended the Cycle 2 meeting, which changed the dynamic of the group. As such during the face-to-face

meeting two of the co-morbidities (dyslipidaemia and low self-esteem) were no longer deemed appropriate for inclusion in the screening programme despite previously being agreed for inclusion. Furthermore as part of the pre-meeting questionnaire two other measures were recommended for depression and anxiety by the expert panel, the Strengthens and Difficulties Questionnaire and the Short Mood and Feelings Questionnaire. These were discussed at the face-to-face meeting and included in the post-meeting questionnaire.

Results of the pre-meeting questionnaire indicated that there was agreement for the inclusion of two of the proposed screening tests (assessment tool for screening children (hyperglycaemia) and upper arm BP measurement (hypertension)). For the remaining screening tests, there was either uncertainty over its inclusion or the panel felt it should not be included in the proposed screening programme (Table 40, page 171). The results of pre-meeting questionnaire were presented to a panel of service users (17th October 2016). The service users discussed each screening test, for which they were presented with sensitivity and specificity information in plain English and for questionnaires they were provided with example questions. The service user panel agreed on the inclusion of one screening test for the co-morbidities. The service user group's decision making process considered how quickly the test could be administered and what they felt was a more scientific test, as such they opted for a blood test over a questionnaire for hyperglycaemia. The service user panel also considered the views of the children present; hence a wrist/forearm BP measurement was preferred over the upper arm measurement, which the children stated was painful.

At the face-to-face meeting with the expert panel (25th October 2016), each screening test was considered in turn, taking in to consideration the results of the pre-meeting questionnaire, information provided about the screening tests, such as questions and the test's reported sensitivity and specificity, and comments from the service user panel. As part of the meeting the results from the sub-study looking at the acceptability of blood tests were also presented to aid discussions. In addition the panel also re-discussed the co-morbidities and the practical considerations of implementing a screening programme within a weight management services.

The final stage in cycle 2 was the post-meeting questionnaire, results indicated the face-to-face panel discussion did not improve consensus, in fact in the post-meeting questionnaire the panel had changed their minds from "include" to "uncertain" for assessment tool for screening children (hyperglycaemia) and upper arm BP

measurement (hypertension). Overall 10 of the screening tests were considered to not be appropriate for inclusion in the proposed screening programme, and there was uncertainty of the appropriateness of the remaining 14 screening methods; thus consensus was not reached for the inclusion of any of the screening methods (Table 40, page 171).

5.7 Discussion

5.7.1 Summary of Findings

The consensus study built upon the results of the systematic review and meta-analyses (Chapters 3 and 4), which provided information on the difference in prevalence for the co-morbidities between those who are overweight or obese, relative to those who are a healthy weight. This is the starting point in developing a screening programme (Public Health England 2015a). According to the criteria put forth by the NSC, as well as prevalence, consideration is required on the impact of the co-morbidity to the individual, and whether the co-morbidity is considered an important health condition to warrant screening (Public Health England 2015a).

The consensus study adopted a modified version of the RAM to reach consensus on which co-morbidities and screening tests to include in the proposed screening programme for obesity-associated co-morbidities in children. A consensus study was conducted as results from the systematic review and meta-analyses indicated there were gaps in the literature regarding some potentially important confounds, such as ethnicity, socioeconomic status and gender. Furthermore, only one study was conducted on a UK population, this limited the applicability of the results to the UK population, which is the target population for the proposed screening programme. Although the systematic review and meta-analyses indicated an increased prevalence for many of the co-morbidities, increased prevalence is only one aspect of assessing a co-morbidities suitability for inclusion in a screening programme. The consequences of the co-morbidity to the individual's psychological and physical well-being also needs consideration (Public Health England 2015a). Furthermore, the NSC's criteria stated that acceptability of the screening test amongst health professionals and service users was an important part in the development of a screening programme (Public Health England 2015a).

Table 40: Summary of results for Cycle 2 – Screening Methods
*The Strengths and Difficulties Questionnaire and the Short Mood and Feelings questionnaire were suggested as part of the pre-meeting questionnaire by the expert panel; hence pre-meeting ratings for them are not available.

Co-morbidity	Screening Test	Pre-meeting (median (range))	Pre-Meeting Decision (Disagreement Index)	Service User Group's Opinion	Post-Meeting (median (range))	Post-Meeting Decision (Disagreement Index)
Hyperglycaemia	ADA Screening Criteria	6 (2-8)	Uncertain (Agreement)	Do Not Include	4 (1-7)	Uncertain (Disagreement)
	Assessment tool for Screening Children	7 (3-9)	Include (Agreement)	Do Not Include	5 (1-7)	Uncertain (Agreement)
	Finger Prick Blood Test	3 (1-7)	Do not Include (Agreement)	Include	3 (1-8)	Do not Include (Agreement)
	Urine Test	2 (1-6)	Do not Include (Agreement)	Do Not Include	2 (1-6)	Do not Include (Agreement)
Hypertension	Upper Arm BP Measurement	7 (3-8)	Include (Agreement)	Do Not Include	6 (2-9)	Uncertain (Agreement)
	Wrist/ Forearm BP Measurement	5 (1-8)	Uncertain (Agreement)	Include	4 (1-6)	Uncertain (Agreement)

Obstructive Sleep Apnoea	Neck Circumference	4 (2-7)	Uncertain (Agreement)	Do Not Include	3 (1-7)	Do Not Include (Agreement)
	Pediatric Daytime Sleepiness Scale	6 (2-8)	Uncertain (Agreement)	Do Not Include	6 (3-8)	Uncertain (Agreement)
	Epworth Sleepiness Scale – Revised for Children	4 (3-8)	Uncertain (Agreement)	Do Not Include	6 (3-8)	Uncertain Agreement
	Teen-Stop Bang	6 (3-8)	Uncertain (Agreement)	Include	3 (1-6)	Do Not Include (Agreement)
	Pediatric Sleep Questionnaire	6 (4-8)	Uncertain (Agreement)	Do Not Include	6 (3-9)	Uncertain (Agreement)
Depression	Children's Depression Inventory 2 (12 items)	5 (2-8)	Uncertain (Agreement)	Do Not Include	2 (1-7)	Do Not Include (Agreement)
	Children's Depression Inventory 2 (28	3 (1-6)	Do Not Include (Agreement)	Do Not Include	2 (1-5)	Do Not Include Agreement

	items)					
	Depression Scale for Children	5 (3-8)	Uncertain (Agreement)	Do Not Include	4 (1-7)	Uncertain (Agreement)
	Revised Child Anxiety and Depression Scale (47 items)	3 (1-6)	Do Not Include (Agreement)	Do Not Include	2 (1-6)	Do Not Include (Agreement)
	Revised Child Anxiety and Depression Scale (25 items)	4 (1-6)	Uncertain (Agreement)	Include	5 (1-8)	Uncertain (Agreement)
	Strengths and Difficulties Questionnaire	N/A	N/A	N/A	5 (1-6)	Uncertain (Disagreement)
	Short Mood and Feelings Questionnaire	N/A	N/A	N/A	6 (2-8)	Uncertain (Agreement)
Anxiety	Multi-dimensional Anxiety Scale for	4 (1-7)	Uncertain (Agreement)	Do Not Include	2 (1-6)	Do Not Include (Agreement)

	Children 2nd Eds.					
	Revised Child Anxiety and Depression Scale (25 items)	5 (1-7)	Uncertain (Agreement)	Include	5 (2-7)	Uncertain (Agreement)
	Revised Child Anxiety and Depression Scale (47 items)	2 (1-6)	Do Not Include (Agreement)	Do Not Include	3 (1-5)	Do Not Include (Agreement)
	Spence Children's Anxiety Scale	3 (1-7)	Do Not Include (Agreement)	Do Not Include	2 (1-5)	Do Not Include (Agreement)
	Strengths & Difficulties Questionnaire	N/A	N/A	N/A	4 (1-8)	Uncertain (Agreement)

Therefore to address the NSC's criteria, a consensus study was conducted to understand the views and opinions of health professionals, researchers and service users regarding suitable co-morbidities and screening tests for inclusion in the proposed screening programme.

Overall, the results indicated that consensus was reached for the inclusion of five co-morbidities, diabetes, hypertension, obstructive sleep apnoea, depression and anxiety. Yet agreement was not reached on screening methods. The results of each cycle are reviewed below.

5.7.2 Cycle 1: Co-Morbidities

Initially the results of the cycle 1 post-meeting questionnaire indicated consensus on the inclusion of seven co-morbidities, diabetes, hypertension, dyslipidaemia, obstructive sleep apnoea, depression, anxiety and low self-esteem. However, during the cycle 2 face-to-face meeting, there was a change in the group's decision regarding the inclusion of dyslipidaemia and low self-esteem. One factor which might explain this is the change in the composition of the cycle 2 expert panel for the face-to-face meeting. Specifically the cycle 2 meeting included two members unable to attend the cycle 1 meeting, and their contribution appeared to alter the conversation regarding the co-morbidities whilst discussing screening methods. This relates back to prior research that strong personalities, or in this case subject matter experts, may influence the outcomes of the results; impacting the validity of the results (Birko, Dove and Özdemir 2015). The impact of panel composition is a central criticism of all consensus methods, that the results may be contingent on the specific composition of the expert panel. Furthermore each expert's individual responses are based on their experience, leading to uncertainty in the validity of the results as it cannot be assumed that when rating statements every panellist is considering the evidence in the same manner (Phelps 1993; Hicks 1994). To assess the reproducibility of results with different panels Shekelle *et al.* (1998) simultaneously collated ratings from three equivalent groups when considering the over- and underuse of medical procedures. The results indicated considerable variation between the groups with Kappa ranging from 0.51 to 0.52. The results would suggest that as well as the panel composition, the statements being considered also impact replicability of results. For the current study, the change in inclusion of dyslipidaemia and low self-esteem supports Shekelle et al.'s results and highlights that further work may be required to assess the reliability of the results.

Additionally, comparison of the results from pre- and post-meeting questionnaire indicated a change in the panel's views for some of the co-morbidities. Typically areas of uncertainty change to agreement or disagreement once the panel have discussed the evidence and their views. For instance initially the panel were uncertain about the inclusion of ADHD and elevated uric acid in the pre-meeting questionnaire; however results of the post-meeting questionnaire indicated they agreed to not include them in the screening programme. However the results also indicated opposite, for instance the panel initially agreed on the inclusion of vitamin D deficiency, however after the face-to-face discussion they were uncertain about its inclusion (Table 39). The panel discussion indicated that vitamin D deficiency is a wider problem, beyond those who are overweight and obese. Thus offering screening to one population and ignoring the others would not be ethical. Related to this, in July 2016 (after the face-to-face meeting) the Scientific Advisory Committee on Nutrition recommended for "a reference nutrient intake of 10 micrograms of vitamin D per day, throughout the year, for everyone in the general population aged 4 years and older", which could be met through supplementation (2016). This is an example of where a public health intervention was required, rather than a targeted approach for a sub-population.

When reviewing the co-morbidities the expert panel found it difficult to separate the discussion of the co-morbidities from the potential screening methods, the subject of cycle 2, and the consequences of a positive screening result, such as the referral pathway for the child. As such there were times they pre-empted future discussions about the suitability of the screening method for use within a weight management service, and the consequences of a positive screening. Whereas the service user panel gave some consideration to the practical implications of conducting the screening, such as staff training and costs, they were less likely to consider the consequences post-screening and whether or not a suitable treatment pathway existed.

The expert panel were able to agree on the inclusion of five co-morbidities which is an important step in developing a co-morbidity screening programme for children attending a UK community weight management service. Results indicated that the five co-morbidities were also considered important by the service user panel, indicating some convergence of agreement between the two groups. Yet, conversely the expert panel also went against the views of the service users regarding other co-morbidities, including hepatic disorders and iron deficiency. Some of this may be explained by the differences in knowledge and perception.

The service users were less likely to have medical training/knowledge and were considering the co-morbidities from the perspective of their child, rather than the wider viewpoint likely adopted by the expert panel, as well considerations to the available referral pathways and treatment options. Previous research has indicated that the inclusion of service users does appear to influence the discussion and decisions of the expert panel (Coleman *et al.* 2014) although such research is limited even when considering other consensus methods (Jones *et al.* 2017). From the results of this study it is unclear if and how much impact the service user comments had on the expert panel's decisions.

The final results of cycle 1 indicated consensus for the inclusion of diabetes, hypertension, obstructive sleep apnoea, depression and anxiety in the screening programme.

5.7.3 Cycle 2: Screening Methods

In Cycle 2 the discussion focused on the practical implications of screening for the co-morbidities in terms of the suitability of the screening methods in a predominantly non-clinical setting and the subsequent impact to services, such as referrals to GPs and other possible services. One of the main factors for not including a screening method was the lack of a suitable care pathway within the UK if a child screened positively, this was influenced by known variations in service provision by city or county, such as for mental healthcare.

Although the health professionals and researchers panel were unable to achieve consensus on screening methods to include in a screening programme for use within a community weight management service; they were able to exclude particular methods on the basis the methods were unsuitable for the weight management setting. For instance, for blood tests there was concern over issues of taking bloods in a community setting, along with the required training for staff and acceptability and compliance from service users. With regards to staff training, guidance is available on the safe sampling of blood and training could be made available for service staff, particularly where the service is part of an NHS trust (World Health Organisation 2010), and 92% (12/13) of staff stated they would be happy to conduct a finger prick blood test, according to the sub-study. Though this was based on small samples, therefore it is unclear if this would have held up if the programme were implemented. With regards to the compliance from service users, 91% (21/23) of parents and 83% (19/23) of children would be happy with a blood

test. Although this was based on a small and selected sample, thus generalisability of the results is limited.

As well as the practicalities of blood tests, the expert panel were also not comfortable with the use of questionnaires, particularly the combined length if multiple questionnaires were selected. There were also concerns in relation to correct administration of self-report questionnaires to ensure the information was useful in obtaining an accurate assessment of the child's health. This was particularly important as research indicates that mode of administration can impact the quality of the data obtained, particularly when comparing self-administration with clinician/researcher guided administration (Bowling 2005). Therefore to ensure validity and reliability of the data appropriate staff training would be required for the administration of questionnaires to ensure inter-rater reliability. This relates to the particular concern that was raised multiple times, staff at weight management services are not clinical staff. Thus administering questionnaires and being able to separate the subtle nuances of what is normal versus abnormal behaviour may not be something they can feasibly do.

5.7.4 Implications of the Results

Co-morbidities

The results of the consensus study do have implications for policy and practice. The results recommended screening for co-morbidities deemed important health conditions based on scientific evidence and the views of a panel of multidisciplinary experts. The results indicated there were five co-morbidities for which there was consensus for inclusion in the proposed screening tool, hyperglycaemia, hypertension, obstructive sleep apnoea, depression and anxiety.

A previous consensus has not be conducted in a similar area, however a co-morbidity screening algorithm was identified, although details of its development were not available (Patel n.d.). The screening algorithm recommended screening for hypertension, cholesterol, diabetes, non-alcoholic steatohepatitis (NASH) and obstructive sleep apnoea in children. Cholesterol (dyslipidaemia) and NASH were initially deemed appropriate for inclusion in the screening programme; however in cycle 2's face-to-face meeting the panel changed their mind about screening for cholesterol, and for NASH the cycle 1 post-meeting questionnaire indicated the panel were uncertain about its inclusion.

Furthermore The Endocrine Society developed guidance on screening for co-morbidities in children (August *et al.* 2008). They recommend screening for hyperglycaemia, dyslipidaemia, hypertension and NAFLD (a mild form of liver disease compared with NASH) (Benson, Baer and Kaelber 2012). This guidance was developed using expert opinion, though the precise methods were not detailed. Interestingly, as per Patel's screening algorithm, the guidance does not recommend screening for psychological co-morbidities. However neither state a reason for excluding psychological co-morbidities, and the methods used to create the screening algorithm were not provided, thus critique and assessment of the result's validity is limited.

Screening Tests

The NSC criteria also considers the accuracy, suitability and acceptability of screening tests amongst clinicians and patients (Public Health England 2015a). The systematic review identified potential screening tests, which was supplemented by an additional search of the literature to identify other tests. The consensus study assessed acceptability of the tests amongst service users and experts. Results indicated that the service users were able to clearly decide on a single screening test for each co-morbidity and appeared to be more accepting of specific tests, such as blood tests, in contrast to the expert panels.

Results of the pre-meeting questionnaire indicated the expert panel reached consensus for a screening questionnaire for hyperglycaemia and upper arm BP measurement for hypertension. The post-meeting questionnaire, however, indicated uncertainty for these and 11 other screening tests and the decision to not include the remaining 10 screening tests.

In terms of implications, the results suggest that current screening tests are not suitable for use in weight management services, due to the practical implications and the training requirements of staff. The co-morbidity screening algorithm proposed by Patel (n.d.) and the guidance from The Endocrine society recommended tests and cut-offs for the tests. For the three co-morbidities identified by Patel and The Endocrine Society, and agreed upon by the expert panel, (hypertension, diabetes and obstructive sleep apnoea) the recommended screening tests were upper arm BP measurement (hypertension), fasting glucose (hyperglycaemia), and a positive history (obstructive sleep apnoea). Although details of the questions considered as part of the history were not stated, limiting comparability to the screening questionnaires identified for the consensus study.

In terms of the psychological co-morbidities, the panel for the most part agreed that specific screening tests should not be included. Screening tests for psychological conditions are predominantly questionnaires, which are not objective measures. In fact the NSC has previously not implemented a screening programme as the screening test was a questionnaire, which they regarded as being subjective and thus at risk of a high number of false positives that would potentially overwhelm services (UK National Screening Committee 2017). Therefore including psychological co-morbidities in the proposed screening programme may not be feasible until more objective measures are available.

5.7.5 Strengths and Limitations

Waggoner, Carline and Durning (2016) recommended steps to increase the validity and reliability of consensus studies, these steps were adopted as part of the study. Strengths of the study include the adoption of a robust and structured approach which has been used previously and has shown face validity and reliability (Lawson *et al.* 2012). Related to this the expert panel were purposively sampled to include a range of professions, with differing expertise and knowledge to enhance the overall process and increase the face validity of the results (Waggoner, Carline and Durning 2016). This diversity in sample enabled the expert panel to consider the wider scientific evidence as well as their own clinical and research experience, which encouraged knowledge transfer. The study was further strengthened through the inclusion of a service user panel who provided a different perspective for consideration by the expert panel (Fraser *et al.* 1994). Additionally, the results of the systematic review and meta-analyses evidence (Chapters 3 and 4) provided the foundation of the evidence base for the consensus study and through facilitation in the face-to-face meetings enabled the exploration of the data and knowledge sharing between the panel and a shared understanding of other panellists views (USDHHS *et al.* 2009). Finally the private rating of the questionnaires and anonymised results allowed expert members to make their ratings without peer pressure, thereby reducing the potential for bias in the final results and predefined statistical analysis was conducted to determine whether or not consensus was reached (Hsu and Sandford 2007; Birko, Dove and Özdemir 2015; Waggoner, Carline and Durning 2016).

The primary limitation of the study is the complexity of the topic, which made the study challenging in terms of having substantial amounts of conflicting information to discuss in fairly short timeframes (MacKinnon III and Swanson 2002). Part of this

complexity may be due to the study considering each co-morbidity in isolation, and ignoring the individuality of the patient, e.g. family history and the presence of other health conditions which may influence the actions taken in a clinical setting (Hopkins 1993). Moreover, the results of a consensus study are contingent on the available evidence at the time and the composition of the panel (Shekelle *et al.* 1998; Raine *et al.* 2015). Evidence of how changing the panel composition can influence the results was observed in cycle 2 when low self-esteem and dyslipidaemia were no longer regarded as suitable for inclusion. The small sample for the service user groups may have hindered the generalisability of the findings due to a degree of self-selection bias (Heckman 2010). In fact the expert panel raised concerns over the size of the service user groups, and this doubt may partially explain why the expert panel did not always take on board the comments from the service users regarding the inclusion/exclusion of specific co-morbidities and screening methods. Finally the separation of the co-morbidity from the screening methods whilst logical may have been somewhat artificial. As in both face-to-face meetings the expert panel discussed the entire care pathway to understand the advantages and disadvantages of including co-morbidities in the proposed screening programme. This may be an artefact of the training and knowledge of the panellists to view the process as a whole to understand the implications and potential consequences. However the decision to have two focused cycles was deemed appropriate to target discussions on co-morbidities and screening methods.

5.7.6 Future Research

One of the concerns raised by the expert panel was the feasibility and practical implications of implementing a screening programme within a weight management service, which future research could answer via a feasibility study. The feasibility study could also give an indication of the number of children referred to their GP due to a positive screening, to assess the impact of the screening programme on other services. Additionally, future research could also assess the reliability of the results with a different group of experts and service users, particularly in light of new evidence regarding the prevalence of the co-morbidities in children and the development of new screening tests (Shekelle *et al.* 1998). The consensus study highlighted that current screening tools for the selected co-morbidities were not deemed appropriate for use within a weight management service, either due to training implications and/or the tool's sensitivity and specificity. Therefore to implement a screening programme alternative screening methods would be required which are better suited to the proposed screening environment.

5.8 Conclusion

This is the first known study which aimed to develop a list of obesity-related co-morbidities appropriate for inclusion in a screening programme for children attending UK community weight management services. Using a modified RAM consensus was achieved on obesity-related co-morbidities deemed suitable for inclusion in a screening programme. Five co-morbidities (hyperglycaemia, hypertension, obstructive sleep apnoea, depression and anxiety) where selected by a panel of experts; however consensus on screening methods for the co-morbidities was not achieved. The discussion indicated there were a number of concerns regarding the implementation of a screening programme. The next chapter builds on the results of this consensus study by conducting a thematic analysis of the expert panel's face-to-face discussions in cycles 1 and 2 to enable an understanding of their concerns and provide guidance on how best to proceed with implementing the proposed screening programme.

Chapter 6: Expert Opinion on the Feasibility of the Proposed Screening Programme

6.1 Introduction

The previous chapter provided an overview of the work undertaken to achieve consensus on co-morbidities and screening methods for inclusion in the proposed co-morbidities screening programme. This chapter builds on the results of the consensus study by undertaking a thematic analysis of the transcripts from the expert panel face-to-face meetings to understand the factors considered when assessing the feasibility of the proposed screening programme. Firstly, a background and rationale for thematic analysis is provided (Section 6.2), which includes strengths and limitations of thematic analysis (Section 6.2.1). This is followed by the aim of the thematic analysis (Section 6.3) and analytical considerations for conducting a thematic analysis (Section 6.4). The chapter then describes the methods (Section 6.5), and analysis (Section 6.6), followed by a summary of the National Screening Committee's criteria (Section 6.7). The results of the thematic analysis are then presented (Section 6.8). Finally, the discussion compares the themes in relation to other qualitative studies and the National Screening Committee's criteria (Section 6.9). As part of this, the implications of the results are also discussed along with strengths and limitations of thematic analysis and conclusion.

The results of the consensus study (Chapter 5) indicated that the expert panel were able to achieve consensus on the co-morbidities to include in the proposed screening programme (hyperglycaemia, hypertension, obstructive sleep apnoea, depression, and anxiety), however consensus was not achieved on suitable screening tests for use within a community weight management service. Furthermore, throughout the consensus study the expert panel raised concerns regarding the feasibility of developing a screening programme. In order to understand the key themes relating to this further analysis of the expert group discussion was required. The analysis would highlight areas which required additional consideration prior to the implementation of a screening programme, which in turn would inform future work in the area. To enable identification of these factors thematic analysis was selected as it allows for the analysis and synthesis of data from multiple participants into key themes, which would allow for the

identification of any barriers and facilitators to implementing the programme, the results of which could inform future research (Boyatzis 1998).

6.2 Background to Thematic Analysis

Thematic analysis is a method for the identification, analysis and reporting of interesting and/or important patterns (themes) within qualitative data (Braun and Clarke 2006; Fereday and Muir-Cochrane 2006; Clarke and Braun 2013). Thematic analysis is one of the most commonly used methods of analysis for qualitative data; however the history and development of thematic analysis is not clear (Guest, MacQueen and Namey 2012). Thematic analysis, it appears, was first named as an approach in 1975 by Merton, who in turn had based it on the work of Holton (1973).

Thematic analysis is independent of any particular theory or epistemological position (Clarke and Braun 2013). This independence means thematic analysis can be applied as an analytic method across many forms of qualitative data, e.g. interviews, media reports, and focus groups, without any prior knowledge that is essential for other qualitative approaches, such as interpretative phenomenological analysis and discourse analysis (Boyatzis 1998; Braun and Clarke 2006; Fielden, Sillence and Little 2011; Alhojailan 2012; Clarke and Braun 2013; Strachan, Yellowlees and Quigley 2015; MacLean *et al.* 2015).

To date there has not been any qualitative examination of an expert panel's discussions related to screening for co-morbidities associated with childhood obesity. However thematic analysis has been used to assess the feasibility of other screening programmes.

Liles *et al.* (2015) utilised semi-structured interviews and focus groups with multi-level stakeholders to explore internal and external barriers for colorectal cancer (CRC) screening. Results of a thematic analysis indicated that the stakeholders reported a number of barriers to implementing CRC screening, such as a lack of consensus on a suitable screening test and the practicalities of training staff. In addition to barriers, the stakeholders also identified internal and external facilitators for screening, including previous success experienced in implementing a centralised mammography screening programme, and recent emphasis on increasing access to colonoscopy. Similar barriers and facilitators were identified by Schneider *et al.* (2016) via stakeholder interviews to assess facilitators and barriers for successful Lynch syndrome screening (a common forms of hereditary CRC).

6.2.1 Strengths and Limitations of Thematic Analysis

One of the primary strengths of thematic analysis, as mentioned above, is that it is not tied to a particular epistemological or theoretical perspective unlike other qualitative methodologies, such as grounded theory or discourse analysis (Braun and Clarke 2006). This freedom from epistemological perspectives makes thematic analysis a more accessible and flexible method which can be adapted to the needs of many studies, whilst still providing rich and detailed description of the whole data set (Boyatzis 1998; King 2004; Alhojailan 2012). This flexibility also means thematic analysis is particularly suitable for those early in their qualitative research career (Braun and Clarke 2006). Thematic analysis enables the researcher to highlight similarities and differences between participants, and generate insights which may not be identified through other analytical methods (Braun and Clarke 2006); making thematic analysis a useful approach when exploring new or under-researched areas, largely due to its systematic methodology which provides transparency for replication, and makes it a useful method to summarise the key features of a large data set (Braun and Clarke 2006; Alhojailan 2012; Chapman, Hadfield and Chapman 2015; Nowell *et al.* 2017).

However this flexibility may lead to inconsistency in how themes are developed, this is partly due to theme development being a product of the researcher's interpretation of the text which is bounded by his/her knowledge and experience (Patton 2002; Holloway and Todres 2003; Braun and Clarke 2006). Therefore thematic analysis can be open to bias, limiting its scientific rigour in relation to validity and reliability (Morse 1999; Healy and Perry 2000; Stenbacka 2001). Section 6.4 considers validity and reliability in context of qualitative research, and describes the step undertaken to limit bias in the analysis and results (Golafshani 2003).

6.3 Aim

To understand what factors experts perceive to be important when considering the feasibility of a screening programme for children with obesity attending a community weight management service.

6.4 Analytical Considerations

This section describes how concepts of validity and reliability can be applied to qualitative research. Historically qualitative research has been evaluated against

criteria defined for quantitative research, using terms such as validity and reliability (Lincoln and Guba 1985). But given that the nature and purpose of qualitative research is different to that of quantitative research, it has been argued that it is erroneous to apply quantitative criteria to qualitative research. For example, external validity is a key criteria of quantitative research and considers the generalisability of results from the sample to the population (Shadish, Cook and Campbell 1979). However in qualitative research, the purpose is to generate hypothesis for further investigation, rather than test them, so such criteria are not suitable (Sandelowski 1986). Instead criteria associated with trustworthiness are proposed, which if applied assist in demonstrating scientific rigour in qualitative research (Krefting 1991; Nowell *et al.* 2017).

Trustworthiness, as defined by Lincoln and Guba (1985), is demonstrating to others the results are worth paying attention to and taking account of. To demonstrate trustworthiness Lincoln and Guba (1985) discuss the concepts of credibility, transferability, dependability, and confirmability. Additionally, Elliott, Fischer and Rennie (1999) provided evolving guidelines for reviewing qualitative research which acts as a guide to the reporting of qualitative research. These guidelines were adopted to enhance the reporting of the thematic analysis (Madill and Shirley 2000).

The next sub-sections discuss the four concepts of trustworthiness and the concept of rigour, and their application to qualitative research. After which analytical considerations specific to thematic analysis are discussed.

6.4.1 Credibility

Credibility, comparable to internal validity, addresses the issue of "fit" between participants' views and how these are presented by the researcher (Schwandt 2001; Tobin and Begley 2004). Credibility can be demonstrated in a number of ways according to Lincoln and Guba (1985):

- Prolonged engagement: the investment of sufficient time to achieve the aims of the project to understand the participants' views regarding the screening programme by understanding participants' views within context and exploring potential areas for misunderstandings.
- Persistent observation: is the idea of identifying the most relevant factors to the issue being considered, and focusing on them in detail.
- Triangulation: refers to the act of taking data from multiple sources to verify the views of the participants. For instance, if a view of a particular

participant's comment could not be verified by other participants or documentation, the view may be regarded as less credible.

- Peer debriefing: is the process of speaking to a "disinterested peer" to explore hypotheses, particular viewpoints, and the process undertaken to analyse the data. This enables the researcher to clear their mind and see the data from a different perspective, enabling greater understanding of the data.

- Member checks: is where the participants have the opportunity to review interpretations and conclusions made by the researcher. This is regarded by Lincoln and Guba (1985) as the most important aspect of credibility, as it offers participants the opportunity to ensure the interpretations and conclusions are correct.

6.4.2 Transferability

Transferability, comparable with external validity, refers to the generalisability of results; however is restricted to the development of working hypotheses, provided with a description of the time and context within which they apply (Lincoln and Guba 1985; Tobin and Begley 2004). This is referred to as a "thick description". Lincoln and Guba (1985) do not provide a clear definition of a "thick description", although Geertz (1978) stated that thick description refers to there being sufficient detail in how a phenomenon is described that others can decide if the context is sufficiently close to their context to deem the findings as relevant. This goes beyond reporting what people said, and includes who these people are, in terms of their role, and the meaning they are trying to get across.

6.4.3 Dependability

Dependability, comparable with reliability, requires all aspects of the research are clearly documented (Schwandt 2001; Tobin and Begley 2004). This provides an audit trail, which others can examine, ensuring the steps undertaken are logical and traceable (Lincoln and Guba 1985).

6.4.4 Confirmability

Confirmability, comparable with objectivity or neutrality, is used to establish that the interpretations are clearly and accurately derived from the data, and whether the results can be corroborated in another way (Tobin and Begley 2004). Lincoln and Guba (1985) stated that confirmability is established once credibility, transferability,

and dependability are achieved, and is crucial in a rigorous qualitative study (Nowell *et al.* 2017). A common method to also demonstrate confirmability is by the use of reflexivity (Malterud 2001). Reflexivity is the self-awareness the researcher has about their own knowledge and biases and how these might interfere with data analysis and interpretation of the results (Malterud 2001).

6.4.5 Considerations Specific to Thematic Analysis

Inductive versus Deductive analysis

Within thematic analysis data is typically analysed in two ways: inductive (bottom up) or deductive (theoretical/top down) (Boyatzis 1998; Gabriel 2013). An inductive approach means that identification of the themes is driven by the data, and not influenced by the questions asked or the researcher's theoretical interest in the area or topic (Braun and Clarke 2006). Therefore inductive analysis is a method of coding the data without trying to fit the data into a predefined framework, theory, or researcher preconceptions. In contrast, a deductive approach is driven more by a predefined framework or theory, thus is analyst-driven (Patton 2002; Hsieh and Shannon 2005; Vaismoradi, Turunen and Bondas 2013). This approach generally tends to provide detailed analysis of some aspects of the data, and less about the data overall (Braun and Clarke 2006).

Although typically researchers adopt either inductive (Smedley and Coulson 2017; Parnell, Stanton and Plant 2018) or deductive analysis (Lambert and O'Halloran 2008; Ainsworth, Vargo and Petróczi 2018), there are instances of where both are combined. For example, Maine *et al.* (2017) considered the experiences of people with intellectual disabilities self-managing type 2 diabetes adopting an inductive approach to reflect the participants' experiences, after which, a deductive approach was adopted to organise the themes into the constructs of Bandura's Four Sources Model (which captures the four sources of self-efficacy enhancement; Bandura (1977)).

For this analysis I adopted an inductive approach. The rationale being that it was unclear what elements of the data would be important, thus an inductive (bottom up) approach would ensure the findings remained grounded in the data and that a participant driven account was provided

Semantic versus Latent themes

Identification of themes within the data can occur at two levels: the semantic (explicit) level or latent (interpretive) level (Braun and Clarke 2006). A semantic approach means the researcher is not looking for anything beyond what was said by the participants, essentially providing a description of the discussion which took place; although attempts should be made to theorise the significance of the patterns and their broader meanings and implications (Clarke and Braun 2013; Maguire and Delahunt 2017). Latent level themes, on the other hand, look beyond what was said and start to identify the "underlying ideas, assumptions, and conceptualisations" that give the semantic level themes their form and meaning (Braun and Clarke 2006). Therefore latent thematic analysis involves more interpretive work and goes beyond a description of the data.

For this analysis I decided to identify semantic level themes, the rationale for this was that a semantic approach allows the perspectives of the participants to be heard without the researcher's own biases and understanding influencing the outcome.

Essentialist versus Constructionist Approach

Thematic analysis can be conducted within two paradigms, essentialist or constructionist. A paradigm is defined as "commitments, beliefs, values, methods, outlooks and so forth shared across a discipline" (Schwandt 2001). An essentialist (realist) approach allows you to theorise motivations, experience, and meaning in a straightforward way to understand a person's "truth", the idea being that a person's language reflects and enables them to articulate their truth (Braun and Clarke 2006). However each individual's experience of truth will be influenced by their own subjective interpretation of that truth. In a constructionist approach meaning and language are socially produced and reproduced, rather than existing within individuals (Marshall and Rossman 1999), i.e. one person's truth is the same as another's, therefore it discounts the impact of subjective interpretation.

For this analysis I adopted an essentialist paradigm, as the aim of the study was to identify common, as well as conflicting truths from a number of health professionals and researchers, it was important to allow for divergence in opinion to be present and captured.

6.5 Method

The data corpus consisted of the two face-to-face meetings with the panel of experts conducted as part of the consensus study detailed in Chapter 5. The panel consisted of health professionals (paediatric endocrinologist, paediatrician, endocrine specialist nurse, and general practitioner), obesity researchers and academics, and weight management staff (Table 41). As mentioned in Chapter 2, community weight management services are not typically part of the NHS, therefore the staff are not trained medical professionals.

The first meeting was held on the 21st of June 2016 and considered the co-morbidities associated with obesity based on the results of the systematic review and meta-analyses (Chapters 3 and 4). The second meeting was held on the 17th October 2016 and considered screening methods for the agreed upon co-morbidities. Further details regarding the consensus study are available in Chapter 5.

Table 41: The expert panel members' roles and the code assigned to them

Expert Panel Role	Assigned Code
Researcher (Obesity)	EXP01
Medic (Endocrinologist)	EXP02
Researcher (Obesity)	EXP03
Medic (GP)	EXP04
Researcher (Obesity)	EXP05
Researcher (Obesity)	EXP06
Researcher (Public Health)	EXP07
Weight Management Staff	EXP08
Medic (Endocrinologist and Diabetologist)	EXP09
Medic (Endocrine Nurse)	EXP10
Researcher (Obesity)	EXP11
Medic (Paediatric Endocrinologist)	EXP12
Researcher (Psychology)	EXP13
Researcher (Medicine)	EXP14

The six-stage framework reported by Braun and Clarke (2006) was adopted to structure the thematic analysis of the data (familiarisation with the data; generation of initial codes; searching for themes; reviewing themes; defining and naming themes; writing up the analysis and results). Although presented as six discrete stages, Braun and Clarke (2006) report that the stages overlap, and the analysis is

an iterative process to ensure the results provide an accurate reflection of the data. The six stages are discussed in the next section.

6.5.1 Six-Stage Framework of Thematic Analysis

Stage 1: Familiarisation with data

The first stage is to immerse oneself in the data to get a sense of the overall discussion (Patton 2002; Braun and Clarke 2006). As mentioned in chapter 5, I first conducted verbatim transcription of the audio recordings from both expert panel consensus meetings. Transcription commenced the day after the meetings using InqScribe (version 2.2.3; https://www.inqscribe.com/) with unique codes assigned to each participant, e.g. EXP01, to allow for anonymity, but also allow me to track the views of participants over the course of the meetings (Pope, Zieband and Mays 2000; Bailey 2008)(see Table 41). To assist with transcription and subsequent analysis I adopted recommendations from Bailey (2008) to include some non-verbal features, such as pauses, where noticeable in the recordings. Non-verbal features can shape how the transcriptions are interpreted and thus the meaning applied. Bailey (2008) recommended that transcriptions need to be sufficiently detailed to capture "features of talk such as emphasis, speed, tone of voice, timing and pauses" (Robinson, Heritage and Maynard 2006).

Upon completion of the transcripts, I re-listened to the audio recording, to balance accuracy and readability of the transcripts prior to commencing analysis (Bailey 2008). Additionally, after transcription I read and re-read the transcripts multiple times prior to starting the next stage of the analysis to further facilitate familiarisation with the data. As part of this stage I also made a note of initial comments and ideas that came to mind during the readings to aid subsequent identification of codes and themes. This process of transcription and re-reading of the transcripts assisted me in becoming familiar with the data and also addressed the criterion of prolonged engagement, increasing the credibility of the results.

Stage 2: Coding data

The second stage involves going through the data and highlighting elements of the text and coding them as a means of organising the data in a meaningful and systematic manner (Tuckett 2005; Maguire and Delahunt 2017). This is a key part of the analysis as it establishes which elements of the text are related to the aim of the analysis, as such some elements of text may be coded multiple times for

multiple codes and others may not be coded at all (Miles and Huberman 1994; Mackieson, Shlonsky and Connolly 2018). This was accomplished by re-reading the transcripts multiple times, which further aided familiarisation with the data and strengthened my involvement with and understanding of the data (see Table 42, page 194) (Pope, Zieband and Mays 2000).

Stage 3: Search for themes

The third stage involves grouping the identified codes to create themes (Maguire and Delahunt 2017). A theme is a group of codes that are similar to each other and capture something significant or interesting about the data (Braun and Clarke 2006; Buetow 2010; Maguire and Delahunt 2017). There are no specific rules or criteria for what constitutes a theme, or how big a theme should be (Braun and Clarke 2006). To assist in identifying themes Buetow (2010) proposed the concept of "saliency analysis", which assesses the frequency (the recurrence of codes) and/or the importance of the themes. Importance is defined as themes that "advance understanding or are useful in addressing real world problems, or both" (Yardley 2000). Frequency alone does not indicate importance, and an important theme or code may not recur often; therefore it is the researcher's interpretation of the data which determines which codes and themes are important to the research question and which are not (Braun and Clarke 2006). For this stage I wrote codes on to post-it notes. This provided a flexible approach to organise and re-organise codes into themes, and enabled me to assess the saliency and frequency of codes to assess their overall importance to the research aim. Additionally potential themes were discussed as part of peer debriefing to enhance credibility and also transferability of the results, by encouraging me to provide thick descriptions of the context and content of the discussions.

Stage 4: Review themes:

The fourth stage involves a process of theme refinement and begins once the initial phase of theme development is completed (Braun and Clarke 2006). During this stage some themes may be excluded due to a lack of frequency and/or importance based on Buetow (2010) idea of saliency analysis. Some themes may be merged with others, whilst some themes may be separated into two or more. The criteria for assessing if a themes are merged or separated is based on Patton's concept of "internal homogeneity" and "external heterogeneity" (Patton 2002). Data within themes should be sufficiently related to one another (internal homogeneity), whilst maintaining a clear and identifiable distinction from other themes (external

heterogeneity) (Braun and Clarke 2006). To assist with this, initial themes were discussed with a supervisor and with peers to offer a different perspective. As part of this stage an initial thematic map was developed. This is a visual representation of how the themes are structured, and which themes are related to other themes (see Figure 63, page 196) (Braun and Clarke 2006).

Stage 5: Define Themes

The fifth stage begins after the development of a satisfactory thematic map, and involves defining and naming the themes – which is identifying the "essence" of each theme and the structure of the themes overall (Braun and Clarke 2006). Guidance on naming a theme dictates that names are clear, concise and easily understood (Boyatzis 1998; Auerbach and Silverstein 2003). Additionally during this process sub-themes are identified, these are themes-within-a-theme, and provide structure to large and/or complex themes and provide a "hierarchy of meaning" within the data (Braun and Clarke 2006). For this, I developed created a paper version of the initial map, alongside a definition of each theme and sub-theme. One by one, the themes and sub-themes were reviewed and a description developed. This increased my understanding of the themes, and allowed for similarities and differences between themes to emerge. This in turn enabled me to re-organise the initial themes to develop the final thematic map (see Figure 64 on page 197).

Stage 6: Producing the report

The sixth and final stage involves the presentation of the findings and themes, which is supported by extracts from the transcript and relevant literature (Braun and Clarke 2006).

6.6 Analysis

As described in 6.4.5, my approach consisted of identifying semantic themes, via inductive analysis whilst adopting an essentialist paradigm (Braun and Clarke 2006; Maguire and Delahunt 2017). The benefits of this approach are that the data would be coded without trying to fit the data to an existing framework (inductive), or attempting to identify underlying assumptions or beliefs of the participants (semantic), which could be biased by the researcher's personal views and knowledge (essentialist paradigm). Given the variation in obesity across the country, the possible variation in observed co-morbidities, and differing service

models it was important to understand the potential local implications of a nationally proposed screening programme.

Once transcription was completed, I conducted line-by-line inductive coding of the text. This was to identify salient segments of text related to the aim of the thematic analysis; identifying factors experts perceived to be important when considering the feasibility of a co-morbidity screening programme for children (Braun and Clarke 2006; Stuckey 2015) (Table 42). During the process some codes were renamed and new codes added to ensure they accurately reflected the text (Pope, Zieband and Mays 2000).

Table 42: Example of the codes identified in the text

Example Text	Example Coding
"…there are a lot of children from really high risk ethnic backgrounds as well(.)if they have IGT within a few months they get diabetes(,)I don't discharge them I keep them under the same category" (EXP09)	Ethnic differences in co-morbidity presence and progression
"its not just referral(,) now you've got your GP saying it's so great that you're doing watchit(,)next time you come in to see me I'm gonna check out how you're doing" – EXP11	GP role as motivator, to encourage weight-loss

Using the aim of the analysis as a guide, the codes from the previous stage were then grouped and organised into potential themes based on potential similarities. Focusing on the aim avoided digression into irrelevant areas. As part of this stage I revisited earlier stages to ensure the codes and initial themes were appropriate and congruent with the research question as well as re-reading the data to ensure I had grasped an understanding of the overall discussion (Boyatzis 1998). Braun and Clarke (2006) noted that the process of re-coding and re-theming may continue ad infinitum. Thus the process was concluded when the refinements were no longer adding anything substantial.

To support the reporting of the themes, extracts from the data are presented, an approach recommended by Braun and Clarke (2006) to "capture the essence" of the discussion (p.23), including elements from the National Screening Committee's criteria (Public Health England 2015a), and other research to support the results.

6.7 Summary of the National Screening Committee Criteria

As detailed in Chapter 2, in the UK the National Screening Committee (NSC) are tasked with assessing a suitability of implementing a screening programme (Kitchener *et al.* 2014; Public Health England 2015a). The NSC's criteria can be separated into four groups; the condition; the test; the intervention; the screening programme. A summary of the criteria can be found in Table 43, with detailed information provided in Chapter 2 (Public Health England 2015a). Upon completion of the inductive analysis, the identified themes were compared with the NSC's criteria, to identify commonalities and differences between them.

Table 43: Summary of the National Screening Committee's criteria

Criterion	Description
Co-morbidities	Should be an important health problem. The epidemiology and natural history of the condition should be adequately understood.
Test	Should be simple, safe, precise, and validated screening test, which is acceptable to the target population with an agreed plan of action for individuals with a positive result.
Intervention	Should be an effective treatment or intervention supported by evidence indicating that early intervention leads to better outcomes.
Screening Programme	Should be evidence from high-quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity.

6.8 Results

Initially the coding and development of the themes identified four main themes and 22 sub-themes. These themes were organised to provide the initial thematic map (Figure 63). I then went back over the transcripts and codes to reconsider the structure of the themes. As part of this process I identified that in my initial map I had separated the screening programme, conducted by the weight management service, from the potential impacts of the screening programme on stakeholders such as the GPs and service users. However the sub-theme *Practicalities of Screening* – which I associated with the weight management service in terms of training and cost - were in fact impacts experienced by the weight management service. These impacts were experienced before and during the screening

programme, whereas the *Screening Impact* theme was focused on the post-screening impacts to other stakeholders. Therefore there was an overlap which could be avoided by combining the themes *Screening Impact* and *Screening Programme.* Related to this, although mentioned a few times, the identified theme *Screening Priorities* did not meet the saliency criteria put forward by Buetow (2010), and did not need to be addressed within the thematic map, therefore was removed. Another sub-theme, practicalities of screening, was also removed from the final thematic map. The reasoning for this was that practical aspects of implementing a screening programme were considered to underlie all the themes. Thus separating it out did not seem appropriate, given the links to all stages of the screening programme. Based on further refinement of the themes, the final thematic map of three themes with six sub-themes was developed (Figure 64). The map highlights the key components of the discussion undertaken by the participants.
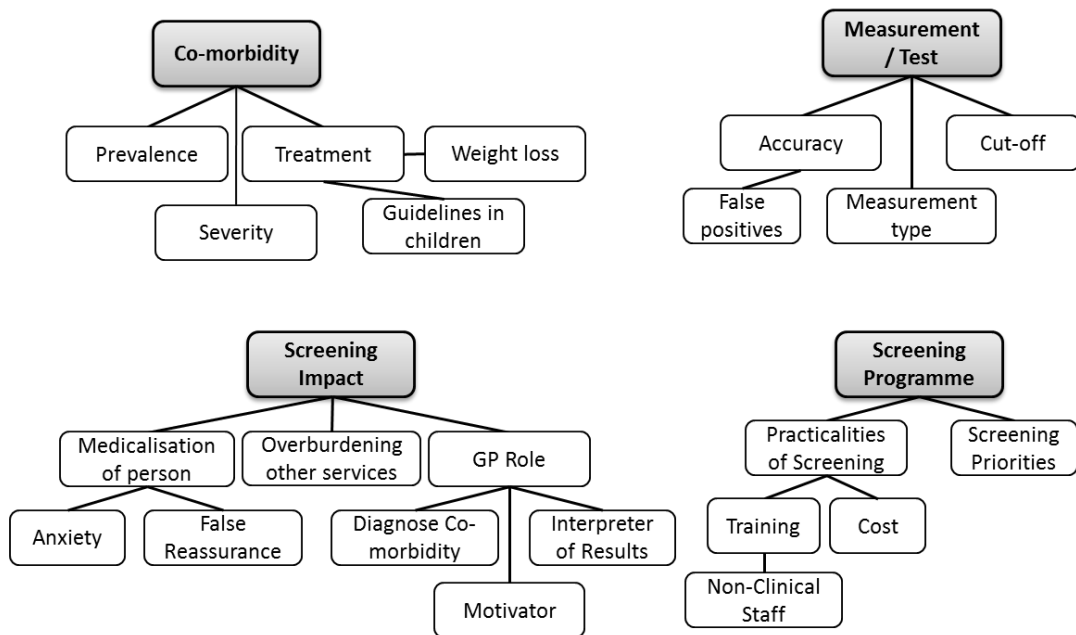

Figure 63: Initial Thematic Map
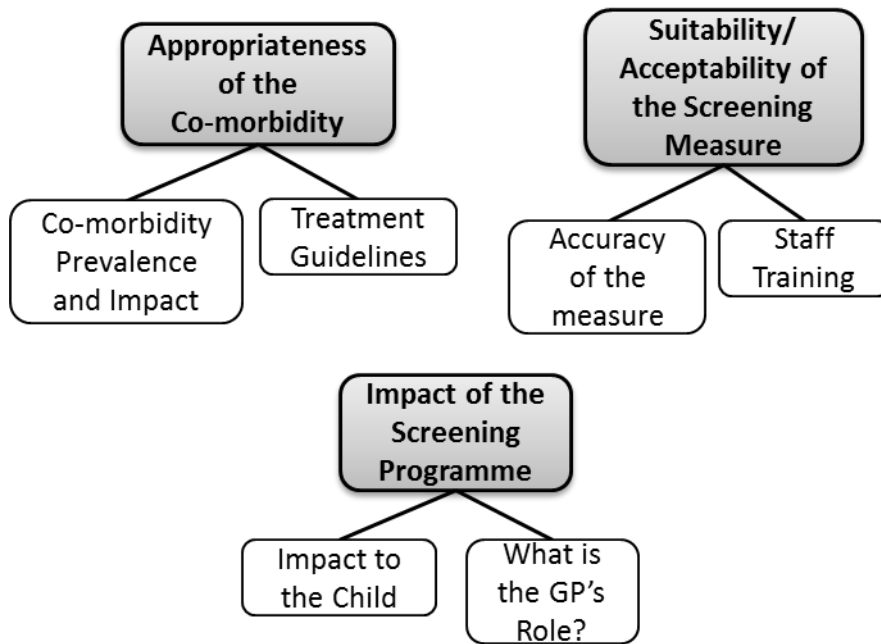
Figure 64: Final Thematic Map

The following sections provide an overview of the three themes and six sub-themes identified as part of the thematic analysis.

Theme 1: Appropriateness of the Co-morbidity

The first theme related to the appropriateness of the co-morbidity for inclusion in the screening programme. Appropriateness was defined as the prevalence, severity and impact of the co-morbidity to the health and well-being of the child. This definition was based on the criteria provided by the NSC for appraising the viability, effectiveness and appropriateness of a screening programme (Public Health England 2015a). Although the NICE recommendation was to screen for obesity-related co-morbidities, the panel noted that currently there is not a universal drive for screening in children, particularly as "*most of the children will be borderline or will have a negative result(.)so why bother screening for them(?)*" – EXP05. The theme itself has two sub-themes, one related to the prevalence of the co-morbidity and the other to existing treatment guidelines for the co-morbidity.

Sub-Theme 1.1: Co-morbidity Prevalence vs Impact

There was concern over the lack of prevalence evidence of some of the co-morbidities from the UK; "*It is very rare, even rarer in children in the UK*" (EXP14 talking about hyperglycaemia), this was perpetuated by the limited number of studies in the systematic review based on a UK population. This suggesting that

screening for the co-morbidities may not be seen as a priority, from both a clinical and commissioning perspective:

> *"I'm trying to imagine all the local commissioners getting feedback from their services…if you were going to try to challenge and override evidence you'd need harder evidence as to why the things that we're choosing are better and might need more evidence than a consensus process based on a discussion about what is imperfectly generalisable evidence to this population"* - EXP05

However some of the panel were less concerned with the prevalence, and viewed the screening as an opportunity to identify those who may require further support:

> *"I took it that this is a point of contact isn't it(?)It's an opportunity(,)so actually it doesn't matter whether it's more prevalent in obese and overweight children it's just that they happen to come to you for their weight"* – EXP07

> *"They [children] have a much more rapid course of deterioration(,)so I think picking up those with IGT is really important cos those kids are not like adults…The literature strongly supports that children deteriorate very quickly"* – EXP09

> *"I still think it's really important even if it's not more prevalence in obese and overweight children, it is quite prevalent and it has an impact(,)so if we have an opportunity to identify it then that's a good thing"* – EXP07

The NSC criteria states the condition should be an "important health problem as judged by its frequency and/or severity" (Public Health England 2015a), therefore a lower prevalence alone may not be sufficient to disregard a co-morbidity, it's severity and impact on the individual should also be taken into consideration. Which support the views of EXP07 and EXP09, the screening is an opportunity, and irrespective of prevalence it is the impact to one's health from the condition which matters. For some of the current population screening programmes the reported UK prevalence is rather low, for example, the reported a prevalence for the breast cancer screening programme between 2015-2016 was 0.85% (18320/2161268) in the targeted population of women aged 50 to 70 (NHS Digital 2017a). For abdominal aortic aneurysm the prevalence in the targeted population of men aged 65 and above was 1.5% (NICE 2009; Gov.UK 2017), suggesting that the impact to one's health and

having a suitable treatment are important factors in deciding if a screening programme is implemented. For instance for abdominal aortic aneurysm, Public Health England reports that screening should prevent approximately 2,000 premature deaths every year in men aged 65 and over (Public Health England n.d.). However, despite the low prevalence of some co-morbidities there was a general view that certain co-morbidities could not be ignored due to the impact they have on one's health and wellbeing:

> *"I'd feel very uncomfortable ultimately ignoring this and not including it"* – EXP04 in relation to diabetes

The discussion regarding prevalence led to the consideration of other factors that may influence prevalence, for example for hyperglycaemia, it was stated that some sub-populations may be at a higher risk, based on factors such as ethnicity:

> *"If a patient in my cohort(,)cos I work in London(,)so there are a lot of children from really high risk ethnic background as well(.)if they have impaired glucose tolerance within a few months they get diabetes(.)I don't discharge them I keep them under the same category"* – EXP09

This view is supported by research that certain ethnic groups, South Asians and Black Caribbean, do have a higher prevalence of type 2 diabetes than other groups, e.g. White European (Riste, Khan and Cruickshank 2001; Oldroyd *et al.* 2005; Reinehr 2013). Furthermore, the potential regional differences in prevalence, due to population variation, prompted some of the panel to suggest the programme may have to accommodate for local factors:

> *"So you're not just giving them a screening tool, you're giving them a screening tool and the steps to identify what needs to happen after the screening, and what they can tap into in their local area"* – EXP07 (discussing how the programme may have to accommodate for local variation)

This would suggest that a standard, national screening programme may not be appropriate or feasible.

Despite initial discussions suggesting prevalence may not matter, subsequently the panel did highlight co-morbidities where the data did not "*suggest that it's worse in obese as opposed to normal weight kids*" – EXP11 in relation to depression,

suggesting a change in perspective, from initially focusing on impact, regardless of prevalence, to, potentially, a more balanced consideration of both.

In summary the panel discussion identified a conflict between co-morbidity prevalence (based on both their experience and the prevalence data provided in the systematic review) and impact of some co-morbidities. Furthermore, the data provided did not report prevalence by ethnicity, however based on their experience some of the panel identified ethnicity as an important consideration when developing the screening programme (Riste, Khan and Cruickshank 2001; Oldroyd *et al.* 2005; Reinehr 2013).

Sub-Theme 1.2: Treatment Guidelines

Related to the prevalence and severity of the co-morbidity was the availability of treatment options for children. Discussion indicated that the first line of treatment for the majority of the co-morbidities in children with obesity would be to lose weight:

> "*Isn't the treatment for even if you find so called pre-diabetes going to be the same anyway(?)so therefore why pick-up if ultimately you're trying to help people reduce weight(?)*" – EXP04

> "*That's what I'm thinking(,)is it referral back into the [weight management] service(?)*" – EXP05

> "*What would you do about that actually other than encourage healthy lifestyle(?)so it's again(,)that's going back into weight management*" – EXP07

> "*Fatty liver is one of those things that will really respond well to just a little bit of weight loss and just a little bit of exercise*" – EXP11

From the panel's perspective there would be a cyclical pathway, from the weight management service who conducted the screening, to the GP to respond to positive screening results, and then refer the child back to the weight management service for the intervention. Thus the benefit of screening, if only to be re-referred to the weight management service, was questioned.

For the majority of the co-morbidities the first-line of treatment would be some form of lifestyle and diet intervention to encourage weight-loss, however there is inconsistent evidence regarding the effectiveness of weight management interventions, particularly in children (Whitlock *et al.* 2010; Hersey *et al.* 2012; Adab *et al.* 2018) (see section 2.4.3, page 20 in Chapter 2) . Although, based on some

panel member's experience, awareness of a co-morbidity may encourage the child to lose weight:

> *"We found that in terms of eating behaviour and lifestyle change*
> *and everything is that the children with a diagnosis of liver disease*
> *were much more motivated and ambitious in their making changes"*
> *– EXP11*

This is supported by other research which suggested that awareness of obesity-associated co-morbidities positively influences weight-loss. Calderón-Larrañaga *et al.* (2015) conducted a longitudinal cohort study of patients in primary care. They reported that patients with obesity who were aware they had at least one co-morbidity were more likely to lose weight. However the type of co-morbidity played a role, with weight loss more likely in patients with a diagnosis of diabetes or dementia and less likely amongst those with hypertension or anxiety. Furthermore weight loss was independently associated with males, higher age and more severe obesity. Although Calderón-Larrañaga *et al.* (2015) included participants aged 15 and above, 43% were aged over 64 and data was not provided for the younger age group limiting understanding of the association between knowledge of co-morbidity and weight loss. This is supported by Gillison *et al.* (2015), who reported that in people at high risk of diabetes or heart disease behaviour change relating to dietary changes, but not physical activity, was associated with weight loss at four and 12 months. Although as noted by another participant that in some cases telling someone to lose weight can have negative consequences:

> *"Saying 'why don't you lose some weight' actually can*
> *rebound(,)people think it's very patronising and will be told where*
> *to go" – EXP04*

The NSC criteria states that there should be an "effective intervention", with evidence that intervention leads to better outcomes (Public Health England 2015a). However at present the guidelines for treatment in children vary based on the co-morbidity. For instance for hyperglycaemia the focus is on initially providing a tailored education programme along with dietary management, with the option of prescribing metformin at a later stage (NICE n.d.-b). For co-morbidities such as depression the pathway would include referral to tier one (GP) or two (CAMHS) mental health services for mild, and tiers two to four (specialist CAMHS services) for moderate to severe depression (NICE n.d.-a; NICE Clinical Guidelines 2005). In

terms of treatment for psychological co-morbidities (depression and anxiety) there were concerns regarding the availability of treatment services for children:

"*have we got adequate mental health services to cope(?)*" – EXP04

Although there are child mental health services, the panel were concerned that the current services did not have the capacity to cope with the potential influx of referrals that the screening programme may result in:

"*but what you do with them is an absolute nightmare(,)because there are just not the services to do it(.)I think that's where it just comes unstuck*" – EXP07

This raised the question of whether it was appropriate to identify a co-morbidity if the services to manage and treat the co-morbidity were not available. As one member stated:

"*If you're gonna identify(,) you've got a responsibility to then do something about it*" – EXP07

This relates to criteria from the NSC regarding the ethics of a screening programme, and if a treatment was not available a positive result may result in greater harm than benefit (Public Health England 2015a). The discussion suggested identification may not be the issue, but the management of the co-morbidities once identified was (Kessler, Sharp and Lewis 2005).

The theme highlighted that for the majority of the co-morbidities the primary treatment is weight loss, typically through diet and exercise, which according to current evidence is of varying effectiveness. For other co-morbidities, the treatment may include referral to other services, which may not be easily accessible, e.g. treatment for obstructive sleep apnoea, or the services are unlikely to be able to cope with the potential increase in referrals, such as child mental health services.

Theme 2: Suitability/Acceptability of the Screening Measure

A key theme evident throughout the transcripts related to the screening measure and its suitability for use within a weight management setting, particularly taking blood samples:

"*Given the constraints of the weight management setting I really can't see how you could go to blood*" – EXP11

> "*it's not in a medical setting is it where they are running the
> session*" – EXP10

As well concern over taking blood samples at the weight management service, there was concern over the experience for the child:

> "*you kind of want to get them started on the programme while that
> motivation's there rather than spend hours and hours talking about
> what's wrong with them.*" – EXP08

The theme had two sub-themes, one related to the accuracy of the measure and the other to staff training.

Sub-Theme 2.1: Accuracy of the measure

The accuracy (sensitivity and specificity) of the screening measures varied within and between co-morbidities, this was of particular concern for the panel in terms of the total number of children (true positive and false positives) that may be sent to the GP:

> "*I rated it highly as well but then I thought what are you going to do
> with all these false positives and intermediate tests you get*" –
> EXP04

The reported prevalence of a co-morbidity is contingent on the sensitivity and specificity of the measure used. Although it was noted that many factors can influence the results of the screening, for instance for blood pressure, "*the most common would be anxiety or the wrong size cuff or a measurement issue*" – EXP02. Which is supported by research indicating that blood pressure can be influence by many factors, such as if forearm or upper arm measurement is adopted, time of day and gender of patient, and appropriate size and placement of the cuff (Karnath 2002; Bauldry, Bollen and Adair 2015; Leblanc, Cloutier and Poirier 2015). Other impracticalities associated with other screening methods were also observed, for instance the nature of current weight management services meant fasting glucose tests would not be feasible:

> "*We mainly do our programmes mainly after school so you can't
> expect the kids to go to school all day*" – EXP08

Thus potentially limiting the option to a random glucose test, however other factors such as age, time of day, physical activity levels can impact glucose levels and influence the proportion of false positives, this is prior to factors associated with the

test itself, such as the testing strip used, or the physical factors of the environment (Ginsberg 2009; Moebus *et al.* 2011).

Issues surrounding the accuracy of the measure are further exacerbated with the use of self-report questionnaires, such as those used for psychological co-morbidities or obstructive sleep apnoea, due to the subjectivity and bias that maybe introduced from parent or child completed questionnaires, further limiting test accuracy. The NSC criteria does not go into detail regarding the test criteria, other than stating that it should be simple, safe, precise and validated (Public Health England 2015a). They have previously not recommended screening for conditions such as alcohol misuse as the test would be a questionnaire, which was considered unsuitable for a population screening programme due to the risk of false positives, which would potentially overwhelm services and reduce access to those who could benefit (Public Health England 2015a; UK National Screening Committee 2017). Furthermore with regards to self-report questionnaires there were concerns regarding how they would be administered. EXP13 talked about how self-report questionnaires are thought of as things you could "just throw at people" and get useful data, when in reality:

> *"Unless it's administered properly I would question what the value*
> *is(.) It really is a training need"* – EXP13

The group discussion highlighted the many factors which might influence the result of the measure, such as the psychometric properties of the tools themselves and others related to the individual differences in the children. A key element which was consistently brought up was the skills and expertise of the weight management staff, and the level of training required to correctly administer the screening measures. Given the frequency with which this was discussed, it was considered as a sub-theme.

Sub-Theme 2.2: Staff Training

It was repeatedly mentioned, by one participant in particular, that weight management service staff are not clinicians, and are not trained in conducting screening physical and psychological co-morbidities:

> *"you've got issues with blood and like being careful(,)with all due*
> *respect they're not nurses running these clinics"* – EXP10

> *"I think that's quite complex to expect [a] worker in like the weight*
> *clinic to make that call really"* – EXP10

Additionally, the staff would have to interpret the results to some extent to ascertain if the child required referral:

> "*you couldn't expect the weight management team to interpret those results(.)So even if they actually took the finger prick test you've got to involve another medical person to do the interpretation*" – EXP10

Though this may indicate a lack understanding by the participants; current clinical tests, such as blood tests for glucose and blood pressure measurements are routinely conducted by trained, but non-clinical staff, including personal trainers at health and fitness centres (McNaughton *et al.* 2011; Nuffield Health 2016; Kielly *et al.* 2018). Furthermore, patients are being trained to self-monitor blood glucose and blood pressure; indicating the concerns raised by the panel are not insurmountable, and with the right training the screening could be conducted by weight management staff (Benjamin 2002; Stergiou *et al.* 2004; Kirk and Stegner 2010).

Concerns were also raised regarding the use of self-report questionnaires, in particular going beyond the initial response to truly understand what the service user/parent is saying:

> "*You have to be careful asking questions like that(,)it's like in my clinic if I saw a child(,)(')do you get headaches?(')they would go (')er yeah I think I do(')(.)I was like now what do I do(?)Do I screen the head, do a scan or do I ignore it(?)I think asking something like that(,)(')do you get joint pain(')(?)(')yes I do(')(.)Sometimes you walk too far(,)I wouldn't need this question and I think you would get a lot of yeses when it's not significant(.)*" – EXP09 in relation to the use of questions to assess for joint pain

> "*I used to pick these up and you've got a child who's got true all the way down(,)and you say (') ok let's have another look at this alright(?)(')(.)So you go through all the items or you talk through all the items(,)and you don't have to get the child to justify everyone but you can with a bit of with just a bit of discussion(.)So ok(,)so you say (')I felt so tired I just sat around and did nothing('), and they put true for that and you say (')Oh(,)so what was the most recent time you did that(?)(') or you can get them to talk through a bit and understand it(.)There are(,)questionnaires are only as good*

> *as their administration and the person's understanding of the*
> *content*" – EXP13 regarding the administration of questionnaires

Both examples articulated the complexity of administering self-report questionnaires for screening, particularly with children, and the importance of having appropriately trained staff to interpret the response, which may require further questioning. The potential concern was that even with training some staff may struggle asking additional questions or correctly interpreting the data. EXP04 gave an example where practice nurses had to ask patients two screening questionnaires and a lot of nurses "felt quite anxious…they didn't feel trained to ask two questions".

The NSC criteria do not explicitly state staff training, however it is implied that the individuals conducting the test are appropriately trained in the administration of the test and interpretation of the result (Public Health England 2015a). However, it was frequently mentioned by the panel, highlighting the importance of having appropriately trained staff to ensure successful implementation of the screening programme. Previous research has also highlighted the importance of having appropriately trained staff and the impact the screening may have by changing the relationship dynamic between the staff and children (Liles *et al.* 2015; Schneider *et al.* 2016).

Theme 3: Impact of the Screening Programme

The final theme focused on the impact of the screening particularly on the child and the GP. For the child the focus was on intended and unintended consequences *"such as labelling, where people now say my kid is depressed"* (EXP04), potentially without a substantial change in management. In contrast, for GPs the discussion focused on the number of referrals to the GP and the multiple roles the GP may potentially have to adopt as part of the screening programme.

Sub-Theme 3.1: Impact to Child

For the child the discussion focused on the negative consequences from the screening, the impact of knowing you have a co-morbidity. EXP04 mentioned a previous programme where patients with diabetes and heart disease were screened for depression; *"well-intended screening programme medicalised people's lives further and probably made them more miserable"*. This prompted further discussion on potential unintended negative effects of the screening programme, such as medicalising the weight management service setting:

> *"I think I feel we have to be careful about medicalising things(,)particularly as a lot of treatments(,)services certainly in our area(,)it's really community lifestyle focused(,)they don't go near the GP…if you're then bringing in quite invasive tests for a child(,)I mean nobody likes a blood test anyway(,)do they(?)but for children you're making it a clinical medical scenario right there and then(.)And I feel quite uncomfortable with that"* – EXP06

And medicalising the lives of the child, and the consequences this might have:

> *"The danger is we're gonna be medicalising everybody with a condition they don't have"* – EXP09

> *"How much harm do we do by inducing a whole load of anxiety which could possibly hang around for years"* – EXP04

> *"Whatever we do here is going to potentially going to have unintended consequences and assumptions about people's readiness to change might vary by both condition and obviously interpersonally"* – EXP04

The potential harmful effects of screening have been previously reported. Feldman (1990) discussed the impacts of false-positive and false-negative results. False-positive results may lead to further discomfort, costs, and risks as the child would have to undergo additional procedures to confirm the diagnosis; for some co-morbidities this may include additional invasive procedures. False-negative results also have negative effects, as they may lead to a false sense of security and the child may be less likely to take their weight management seriously:

> *"You've got to be worried about the ones you say don't have IGT or something(,)cos then you're reassuring them"* – EXP07

> *"Cos they feel supported in the way they are(.)you're almost telling them it's ok"* – EXP03

Overall it seemed there was a lack of clarity as to how children would possibly respond to the screening test results, a negative result may give false reassurance that being overweight/obese is ok, whereas a positive screening result may cause additional harm, due to the child being "labelled" as depressed, or hypertensive, and

potentially medicalising their lives due to the need for treatment and lifestyle changes to manage the co-morbidity (Feldman 1990).

Sub-Theme 3.2: What is the GP's Role?

During the discussion the feasibility and practicality of conducting the screening seemed to repeatedly go back to the impact on GP practices, and also the role the GP may play in the screening programme, this varied from being the diagnosis maker:

> *"for diabetes I would need to see them(,)get them in for a test(,)test them again if it's positive(,)meet them to talk about the test results(,)meeting them again(,)referral"* – EXP04

> *"These are what's required to address the results"* – EXP02

> *"Yeah otherwise you're negligent"* – EXP04

To interpreter of results:

> *"It could be case of everyone who is screened for those conditions, all the result regardless are sent to the GP and the GP then has to do that medical review to see whether or not it is something that requires attention"* – EXP15

This was considered impractical:

> *"I don't think he'd [the GP] be very happy having a whole load of results land on his desk that he [the GP] hasn't requested"* – EXP10

To motivator; encouraging the child to continue with the weight management programme:

> *"It's not just the referral(.)Now you've got your GP saying it's so great that you're doing Watch It next time you come in to see me(,)I'm gonna check out how you're doing(.)You really need to do it"* – EXP11

This all would have to occur in addition to their current workload:

> *"I'm thinking of the 12-hour days and the 10 minute appointments and all of the things that [are] against a GP picking up childhood obesity in the first instance"* – EXP11

> *"It's not just going to be absorbed necessarily by the general day-*
> *to-day role of the GPs"* – EXP07

This may have negative consequences for the practice, and thus the care of the patients. Going back to the screening of depression in patients with diabetes/heart disease it was noted that the *"practice couldn't deal with it(,)they didn't have clear systems for picking up case positive findings and da-de-da(.)so depression was a bit of a nightmare"* – EXP04. The general view was that the current system would not be able to cope with the potential influx of referrals:

> *"We can't throw it into a system that's going to buckle under it you*
> *know if there's lots of tests that come from it so erm it's getting that*
> *balance isn't it"* – EXP06

Overall the discussion considered the potential negative impacts from the screening on the child, such as labelling a child with a condition, which can have negative consequences (Feldman 1990). Additionally there was concern over the impact of false positives and negatives on the well-being of the children (Feldman 1990; Deutekom *et al.* 2011). With regards to GPs, the discussion highlighted the potential variety of roles they may have to undertake; however this would all be without additional resources. Similar concerns, regarding the impact to GPs and the potential for false positives were raised by a group of GPs when assessing the impact of a bowel cancer screening programme (Woodrow *et al.* 2006). However, the results of a pilot study indicated that many of the concerns raised by GPs did not manifest, in particular there was not a substantial increase in their workload; instead their role was limited to more administrative tasks for a small number of patients (Jepson *et al.* 2005).

6.9 Discussion

6.9.1 Summary of Findings

The purpose of the thematic analysis was to understand the factors perceived by the experts as important when thinking about the feasibility of a screening programme for children with obesity attending a community weight management service. The thematic analysis provided a descriptive level analysis of the expert panels' views.

The thematic analysis overall identified three themes and six sub-themes; of which five sub-themes align to various stages of the screening programme and the criteria reported by the National Screening Committee (NSC). The discussion over the two

consensus meetings highlighted the complexity of implementing a screening programme and the importance of carefully considering the various aspects of the entire screening programme. This includes the selection of co-morbidities and screening tests, and the potential consequences resulting from the programme, in terms of the impact to GPs and other health services, such as mental health services. In order to implement a screening programme, particularly for something as complex as obesity-related co-morbidities, stakeholder buy-in at all levels is needed. The number of uncertainties suggested that at present screening for co-morbidities may not be feasible.

When comparing the findings of the expert panel with research assessing the feasibility of developing a screening programme in other areas, similarities were identified. Liles *et al.* (2015) utilised semi-structured interviews and focus groups with health plan leaders, primary care providers, program managers, and endoscopy specialists to explore internal and external barriers for colorectal cancer (CRC) screening. Results of a thematic analysis indicated a number of barriers, such as a lack of consensus on a suitable screening test and the practicalities of training staff.

As part of the consensus study, there were concerns raised about the impact of the screening programme on GPs and other health professionals. This is not considered by NSC criteria, other than stating that "adequate staffing and facilities for testing, diagnosis, treatment and programme management should be available" (Public Health England 2015a). According to Schneider *et al.* (2016), who considered universal lynch syndrome screening, there appeared to be a general resistance to the perceived increase in workload, which mirrors what was noted during the consensus panel discussion. Related to this, Schneider et al. reported concerns related to the role of everyone involved in the screening; in particular where a specific department's responsibility started and ended and how additional time would be created to conduct the screening.

The impact on GP's from the implementation of a screening programme has been raised previously. Prior to the introduction of the bowel cancer screening programme in the UK, a qualitative study of GPs indicated concerns regarding an increased work load for primary care due to the potential number of false positives (Woodrow *et al.* 2006). However a pilot study indicated that such concerns did not manifest, and the GP role was concerned with administrative duties and providing information to patients (UK Colorectal Cancer Screening Pilot Group 2004; Jepson *et al.* 2005). It is unclear at present what the impact on GPs would be from the

proposed co-morbidity screening programme, thus further research in this area is needed.

Given the data from the meta-analyses, the panel questioned the low prevalence of some of the co-morbidities; specifically EXP14 queried the prevalence of hyperglycaemia. However, in the UK every new-born infant is screened for several conditions as part of the newborn blood spot screening programme (Gov.UK 2013b). However, some of the conditions included in the screening programme have a very low prevalence, for example maple syrup urine disease affects only one in 185,000 infants worldwide, isovaleric disease approximately one in 250,000 people in the USA, and homocystinuria only one in 200,000-335,000 people worldwide (US Department of Health and Human Services 2007; US Department of Health and Human Services 2016; US Department of Health and Human Services 2017a). These are considerably lower than the reported prevalence of type 2 diabetes in children, suggesting prevalence should be considered along with the impact of the condition to the individual and the wider society (Diabetes UK 2014; International Diabetes Federation 2017; US Department of Health and Human Services 2017b).

There were also concerns raised by the panel, with regards to the benefits of screening and the accuracy of some of the proposed screening tests. Similar concerns have previously been raised by the NSC when reviewing the feasibility of screening for other conditions (UK National Screening Committee 2017). In their recent report the NSC provide justifications for not screening for specific conditions, and these correspond to the reasons provided by the panel during the face-to-face meeting. For instance, the NSC do not recommend screening for Fetomaternal alloimmune thrombocytopenia and Adolescent idiopathic scoliosis due to uncertain benefits to the patients, for Duchenne muscular dystrophy due to a lack of accurate screening tests, and Asymptomatic bacteriuria due to uncertainty regarding prevalence and suitable interventions (UK National Screening Committee 2017). In terms of the benefits of screening, a systematic review and meta-analysis of 21 trials indicated that early intervention has shown to prevent or delay the progression of impaired glucose tolerance to type 2 diabetes (Gillies *et al.* 2007). Although, the included studies were based on adults, thus generalisability to children is limited. Additionally, the authors reported considerable variation in study quality. Overall, the similarities in the barriers identified in the current analysis and previous results support the trustworthiness of the results (Lincoln and Guba 1985).

6.9.2 Strengths and Limitations

There are a number of strengths identified with the thematic analysis. As I had facilitated both consensus panels I had already been exposed to the data first-hand, which assisted in the transcription in terms of understanding the subtle nuances of language and understanding the context of the comments (Bailey 2008). This was particularly important for the second meeting's recording which was disrupted multiple times due poor recording. Additionally analysis was driven by the data and themes were identified at a semantic level. This reduced bias that may arise from forcing data to fit a particular framework/model, which is evident from subthemes that do not map on the NSC criteria directly, such as the sub-themes related to the GP's Role and impact to the child. Related to this, there is sufficient consistency in the discussions to support the identified themes, as evidenced by the quotes (Lewis and Ritchie 2003). Additionally, as mentioned, the identified themes overlap with themes from other research (Liles *et al.* 2015; Schneider *et al.* 2016; Hossain *et al.* 2017), which indicate that when assessing screening programmes there may be a standard/logical thought process to determine the programme's suitability. Based on research for other screening programmes, this is likely to include a pilot or feasibility study to assess the impact of the screening on the children, GPs and other health services, and identify if the concerns raised in the meeting actually manifest, e.g. the high number of false positives and impact on GPs (Woodrow *et al.* 2006). This is discussed in Section 6.9.4 (page 215).

With regards to the analytical considerations discussed in Section 6.4, the criterion of credibility was met through facilitation of the consensus meetings, which contributed towards prolonged engagement and persistent observations. Triangulation was achieved by identifying supporting research related to other screening programmes, to corroborate participant views. With regards to peer-debriefing, the contents of meeting discussion were discussed with individuals not involved with the project to obtain a different perspective and understanding of the data. As part of the consensus study, the post-meeting questionnaire included a summary of the meeting discussion, allowing the panel to ensure interpretations are correct, meeting the criterion of member checks. This summary included a preliminary thematic analysis; however the panel were not able to review the final thematic analysis. Therefore, member checks of this were not possible, and it is possible some of the panel may not agree with the interpretation of the transcripts. The criterion of transferability was met by providing of the participants' roles in Table 41 (page 190), to allow the reader to assess the participant's level of expertise. As

part of the analysis, the discussion highlights implications for future research, with Chapter 5, in combination with this chapter, providing details of the context and discussion, thereby providing a thick description. To address the criterion of dependability, chapter 5 provides a detailed account of the methods undertaken as part of the consensus study and this chapter provides a detailed account of the thematic analysis. Allowing others to critique the steps taken, and ensure the methods are logical and traceable. Confirmability was met by providing a detailed account of the themes, which were supported by extracts from the discussion to demonstrate the themes are derived from the data. Furthermore, the themes were also supported by other research which has reported similar findings to corroborate the results and theme structure.

Although the thematic analysis did identify important themes relating to the implementation of a screening programme, there are still some limitations worth noting. Firstly, although inductive analysis methods were adopted, the themes are identified based on my knowledge and experience, which may explain why the themes map fairly well on to NSC criteria and may be subject to some bias and may indicate a lack of reflexivity (Palaganas *et al.* 2017). Thus someone without that prior knowledge may develop different codes and themes, resulting in a different insight into the discussion. Secondly, the analysis was based on two face-to-face meetings between 12 and 11 participants. Therefore it is likely that a different panel composition would lead to different results. Additionally, review of the transcripts indicated that a handful of participants appeared to dominate the discussions, as evidenced by the quotations provided. This would suggest poor facilitation on my part in ensuring all attendees were able to express their views. Related to this, not all participants were able to attend the face-to-face meetings, for instance EXP17 is a proponent of the screening programme, however due to other commitments was unable to attend. A further limitation relates to the purposive sample of health professionals and researchers which limits the generalisability of the results. This is particularly evident in the second meeting where the addition of two members resulted in the subsequent exclusion of dyslipidaemia and low self-esteem. This has also been observed in other research (Shekelle *et al.* 1998; Raine *et al.* 2015). Finally, the concerns raised by the panel were not explored in real-life. For example, the concerns raised by GPs with regards to the bowel cancer screening programme were assessed via a pilot study (Jepson *et al.* 2005; Woodrow *et al.* 2006). This enabled an assessment of whether the perceived barriers would manifest in reality.

6.9.3 Implications of the Results

The results of the thematic analysis do have implications for policy, implementation, and future research. Based on the discussion, there was concern over the impact the screening results would have on the current and future health and well-being of the child. Of the current 14 screening programmes in the UK, three are focused on new-borns and one is for children aged 12 and over (diabetic eye screening). As such, there is limited evidence of the impact of false and true positive screening results on the health and well-being of children. Research into false-positives in breast cancer screening reported that patients were less likely to attend future screenings if a false positive was occurred in previous screening, this is despite this group of participants being at greater of a true positive in the future (Álamo-Junquera *et al.* 2012; Nightingale *et al.* 2015). With regards to false-negatives, a systematic review concluded that false-negatives have the potential to delay the detection of conditions, but there was little evidence assessing the psychological consequences on patients (Petticrew *et al.* 2000).

Furthermore, based on the experiences of the participants, there appears to be local variation in the prevalence of certain co-morbidities based on the characteristics of the population. For instance EXP09 reported a higher prevalence of type 2 diabetes than EXP14. Part of this may be due to the populations seen by the two clinicians; EXP09 moved from Birmingham to London and stated there was an increase in the number of patients from an ethnic minority community. In contrast EXP14 is based in Liverpool with a predominantly White population. The variation in the populations seen by services may have an impact on the how the screening programme is implemented, for instance participant ethnicity may influence which co-morbidities are screened. Related to this, the discussion also highlighted variation in the treatment options available across the country, particularly for psychological co-morbidities. Therefore local guidance may be required based on the services available.

Additionally, the use of self-report questionnaires as a method of screening was questioned. Historically the NSC has opted to not screen for particular conditions due to the lack of an objective screening tool, such as for alcohol misuse (UK National Screening Committee 2017). Therefore, the inclusion of co-morbidities such as depression, anxiety and sleep apnoea – which can only be screened for via self-report questionnaire – may require further research to assess the impact of false positives and negatives, and the potential increased burden this may place on

services. The sensitivity and specificity of self-report questionnaires are typically calculated in relation to a specific high risk population. Therefore, when the tool is used in a more general population its sensitivity and specificity are likely to change, i.e. there may be greater false positives when comparing results from a general population to those from a clinical population due to differences in actual prevalence (Leeflang *et al.* 2013). This also applies to clinical tests; for instance Geleijnse *et al.* (2009) concluded that the reported sensitivity of dobutamine stress echocardiography is likely higher and the specificity lower than expected in routine clinical practice due to patient specific factors (Detrano *et al.* 1988; Kriege *et al.* 2006). Both Liles et al. (2015) and Schneider et al. (2016) concluded that awareness of the challenges at local and national levels would assist in the development of screening programmes. Thus, awareness of these potential barriers provides a foundation for future research in the area.

6.9.4 Future Research

Based on the results and implications there are areas for future work. Firstly, the available interventions for children with a positive screening result need to be identified in all local areas. This builds on the participants views that there may be local variations in how the screening programme is implemented. Furthermore, NSC state that it would be unethical to screen for a co-morbidity if a suitable intervention is not available (Public Health England 2015a).

Secondly, the development of new screening tests or improvements in existing tests to limit the number of false positives. Research has indicated that false positives at screening can result in long-term psychosocial harm for up to three years, though this is based on research in breast cancer screening in the UK (Bond *et al.* 2013; Brodersen and Siersma 2013). Furthermore, a false positive result at the first screening has been shown to negatively affect attendance at subsequent screenings (Álamo-Junquera *et al.* 2012). Thus increased test accuracy is likely to reduce the proportion of false-positives and reduce the psychological harm caused to patients. In terms of false-negatives, a systematic review concluded that they have the potential to delay the detection of disease and are likely to lead to legal action being taken by those individuals affected, and potentially may reduce public confidence in screening (Petticrew *et al.* 2000).

Related to this, research could assess the reliability of utilising self-report screening tests, such as questionnaires, in screening programmes. Many co-morbidities can only be screened via self-report measures, such psychological co-morbidities and

obstructive sleep apnoea; however the NSC have historically not supported self-report measures, which means co-morbidities may not be included in the screening programme. Self-report measures are seen as problematic in terms of the potential number of false positives that may result and the subsequent burden on health services and the impact to the child from being labelled, for example, as depressed (Angermeyer and Matschinger 2003; UK National Screening Committee 2017). Although more recent research has suggested that being labelled with depression did not impact peer response (Dolphin and Henness 2017). Additionally as part of the consensus study EXP13 stated that administration of questionnaires is key to ensuring accurate information is captured, which requires additional training for staff. However, when comparing self-reports with clinical tests, Ning, Zhang and Yang (2016) reported for hypertension and diabetes self-reports led to an underestimation of prevalence, though in more developed communities, self-reported data can be a reliable estimate of prevalence. With regards to mental health conditions, a systematic review concluded that there was insufficient evidence regarding the accuracy of any depression screening tool and cut-off for major depressive disorder in children and adolescents (Roseman *et al.* 2016). Furthermore, Roseman *et al.* (2016) stated that screening could lead to over diagnosis and the consumption of scarce health care resources, in line with the views of the NSC (UK National Screening Committee 2017).

Thirdly, there were concerns about the impact to GP's from the screening programme. This was based on the NICE recommendation suggesting the pathway for treatment went via the GP. The panel discussion highlighted many potential roles for GPs, such as motivator and test result interpreter; however it is unclear if GPs have the capacity to fulfil these roles to meet the currently unknown demands on their already limited time. Therefore, a feasibility study could assess the impact the screening programme would have on GPs. A previous study assessing GP attitudes towards a bowel cancer screening programme indicated similar concerns to those identified in the consensus study, with regards to the impact on GPs (Jepson *et al.* 2005; Woodrow *et al.* 2006). However, after taking part in a pilot study the concerns had gone, as they had not manifested. In fact the GP role, although varying by area of the country was typically concerned with administrative duties and providing information to patients about the screening process (UK Colorectal Cancer Screening Pilot Group 2004; Jepson *et al.* 2005).

Taking all this into consideration, a pilot study implementing a co-morbidity screening programme would be beneficial. As with previous pilot screening

programmes, the pilot would allow for a clearer understanding of screening uptake by service users, which previous studies have indicated as ranging from 10% to 61% - though this is with adults and required patients to come to the screening, whereas in the proposed programme the screening would be brought to the patients (Webb *et al.* 2011; Goyder *et al.* 2008; Eborall *et al.* 2012). Furthermore the pilot would assess the validity of the concerns raised by the expert consensus panel, regarding the effectiveness of the screening programme, the accuracy of the screening tests (the number of false positives and false negatives), and the subsequent impact on primary services (Jepson *et al.* 2005; Woodrow *et al.* 2006).

6.10 Conclusion

This thematic analysis of discussions between a varied group of health professionals and researchers regarding the implementation of a screening programme yielded valuable information regarding the potential challenges that would need to be overcome when implementing a screening programme in community weight management settings. The challenges related to various stages of the screening programme and highlighted areas where additional work is required prior to implementation, in particular, suitable and more accurate screening methods for use within weight management services, and adequately funded and resourced services for referral.

Overall there was support for screening as evidenced by consensus on which co-morbidities to include in the screening programme, yet based on the screening programmes the NSC chose not to recommend, as additional data is required to i) assess the prevalence of the co-morbidities in the defined population (e.g. to ascertain the potential number of referrals to the GP); ii) better understand the views of children undergoing the screening, in terms of positive and negative consequences; and iii) ensure an appropriate intervention exits for those with a positive screening result. This additional data would enable a more in-depth assessment of the feasibility of developing a co-morbidity screening programme for children attending UK weight management services.

Chapter 7: General Discussion and Conclusion

7.1 Introduction

Over the past four decades, there has been a steady rise in the global prevalence of childhood obesity (Han, Lawlor and Kimm 2010; OECD 2014). In England, according to recent data, approximately 22% of 5-6 year olds are overweight/obese, increasing to over 34% at age 10-11 and this has received widespread media attention (BBC News 2018; Public Health England 2018a; Public Health England 2018b). Being overweight/obese can have severe implications for the individual, ranging from physical co-morbidities such as hyperglycaemia and hypertension to psychological co-morbidities such as depression and anxiety (Daniels 2009; Han, Lawlor and Kimm 2010; Güngör 2014; Yoon 2014; Parker *et al.* 2016; Reuter *et al.* 2016; Brady 2017).

Given the prevalence of obesity and increasing evidence indicating a higher prevalence of co-morbidities in children with obesity, The National Institute for Health and Care Excellence (NICE; a UK body which provides national guidance and advice to improve health and social care) recommended that community weight management services assess each child for obesity-associated comorbidities and refer them to their GP if any concerns are identified (NICE 2013). However the guidance did not indicate specific co-morbidities required screening or how screening could be implemented.

In the UK, the National Screening Committee (NSC) was established to assess the viability, effectiveness and appropriateness of a screening programme, for which criteria was developed. The NSC's criteria consider the condition, the test, the intervention, and the screening programme. Using the National Screening Committee's (NSC) criteria, this PhD project was undertaken to assess the feasibility of developing a co-morbidity screening programme for children attending weight management services in the UK (Public Health England 2015a).

The programme of work consisted of three distinct phases; systematic review and meta-analyses (chapters 3 and 4); consensus study (chapter 5); and thematic analyses (chapter 6). A detailed account of these phases is provided in the aforementioned chapters. This chapter summarises the key findings of each phase and how they relate to the NSC's criteria (Section 7.2), and goes on to discuss strengths and limitations (Section 7.3), implications for research (Section 7.4), and

implications for practice (Section 7.5), before a final conclusion is presented (Section 7.6).

7.2 Summary of Findings

7.2.1 Phase 1: Systematic Review and Meta-Analyses

The systematic review and meta-analyses provided the foundation of the PhD, identified co-morbidities present in children who are overweight/obese, and reported the co-morbidities prevalence and prevalence ratio relative to children of a healthy weight, and identified screening tests for the co-morbidities. This relates to the NSC's criteria related to condition and test. According to the NSC criteria, the condition (co-morbidities) should be "an important health problem as judged by its frequency and/or severity", and the test should be "simple, safe, precise and validated" (Public Health England 2015a).

The systematic review enabled the identification of co-morbidities and screening tests, and the meta-analysis enabled an estimation of the co-morbidities population prevalence ratio for the overweight and obese groups relative to the healthy weight group. One hundred and sixty-two studies were included in the systematic review which identified 22 co-morbidities, including hyperglycaemia, hypertension, depression, and anxiety. Of these 162 studies, 45 general population studies reporting prevalence for all three weight categories (healthy weight, overweight, and obese) were included in the meta-analysis. The meta-analyses indicated that the prevalence of the co-morbidities was greater in overweight and obese groups than the healthy weight groups for the majority of the co-morbidities. However, there was heterogeneity between co-morbidities for the prevalence ratios which ranged from 1.4 (diabetes) to 58.0 (metabolic syndrome) for children/adolescents with obesity, and 1.2 to 15.8 for children/adolescents with overweight relative to those of a healthy weight. Part of this variation in prevalence ratios may be the result of between study heterogeneity resulting from study methodology and definitions of co-morbidities and test cut-offs, age, ethnicity, and country-specific factors. Furthermore, only a limited number of studies were conducted in the UK, which restricted the generalisability of the results to the UK population; the target population for the proposed screening programme. The high number of co-morbidities identified highlighted the importance of developing a prioritised list of which co-morbidities to include in the screening programme.

Results from previous systematic reviews supported the current findings, that children who were overweight or obese were more likely to have a number of co-morbidities when compared to those of a healthy weight (Guh *et al.* 2009; Pulgarón 2013; Sanders *et al.* 2015). Only Guh *et al.* (2009) conducted a meta-analyses, calculating the relative risk in the overweight and obese groups relative to the healthy weight groups. Statistically significant associations with obesity were found with the incidence of type II diabetes, all cancers except oesophageal and prostate cancer, all cardiovascular diseases, asthma, gallbladder disease, osteoarthritis and chronic back pain. However the studies included participants predominantly aged 40 years and above, thus the results were not appropriate for the current meta-analyses.

Overall the results indicated that children with obesity had a higher prevalence of a number of physical and psychological co-morbidities, which met the NSC's criterion in relation to increased frequency in the target population for the screening programme. However, due to the number of co-morbidities identified it would be impractical to screen for them all. Thus, the number would need to be reduced in line with NSC criteria relating to severity of the co-morbidity.

7.2.2 Phase 2: Consensus Study

This study developed upon the results of the systematic review and meta-analyses, using a modified version of the RAND/UCLA appropriateness method over two cycles, to obtain consensus on which co-morbidities (cycle 1) and screening tests (cycle 2) to include in the proposed screening programme. Participants consisted of two groups; the expert panel consisted of health professionals and researchers, and the separate panel of service users were recruited from a local weight management service.

Results indicated there was consensus for the inclusion of five co-morbidities in the proposed screening programme (hyperglycaemia, hypertension, obstructive sleep apnoea, depression and anxiety), despite some having little evidence in the systematic review and meta-analysis of increased prevalence in children with obesity relative to those of a healthy weight (e.g. depression). However, consensus was not achieved on suitable screening methods for use within a weight management setting. Review of the transcripts from the expert panel's face-to-face meetings indicated concerns regarding specific aspects of the proposed programme, such as the impact to child from a positive or negative result and the lack of prevalence data from a UK population.

Results also indicated that the expert panel, over the course of the two meetings, went back and forth regarding the inclusion/exclusion of certain co-morbidities; based on the discrepancy between the reported prevalence data and the known severity of the co-morbidity.

A previous consensus study with the aim of developing a co-morbidities screening programme for children with obesity was not identified. However, a screening algorithm was, although details of its development were not available (Patel n.d.). The algorithm recommended screening for hypertension, diabetes (hyperglycaemia), and obstructive sleep apnoea; in line with the consensus study. Additionally, the algorithm recommended screening for high cholesterol and non-alcoholic steatohepatitis (NASH), and did not recommend screening for any psychological co-morbidities. As part of the pre-meeting questionnaire both high cholesterol and NASH were included, but face-to-face discussion resulted in their exclusion. With regards to psychological co-morbidities, their exclusion may be due to the use of questionnaires, which are regarded as subjective and may result in a high number of false positives. The NSC has previously recommended not to screen for specific conditions for this exact reason, the risk of false positives taking away resources from those who genuinely need them (UK National Screening Committee 2017).

The consensus study was able to reach agreement on the inclusion of five co-morbidities in the proposed screening programme (hyperglycaemia, hypertension, obstructive sleep apnoea, anxiety and depression); however consensus on screening methods was not achieved. The discussion during the expert panel's face-to-face meetings indicated there were a number of potential concerns that needed further investigation. This was achieved in the third stage, via thematic analysis.

7.2.3 Phase 3: Thematic Analyses

As the consensus study was not able to achieve consensus on appropriate screening methods for the five co-morbidities. Discussions highlighted areas of concern regarding the implementation of a screening programme; an in-depth thematic analysis was conducted to help understand the possible reasons for this. Transcripts from the expert panel's consensus meetings were reviewed to identify themes to better understand their concerns and potential barriers to implementing a screening programme.

After adopting an inductive approach to identify semantic themes within an essentialist paradigm, three themes and six sub-themes were identified. The themes highlighted a number of potential barriers in implementing a screening programme, such as the lack of suitable screening tools for the proposed setting, the impact on the weight management service in terms of staff training and costs, and the impact on GPs from potentially increased referrals.

Previous qualitative research, which assessed the feasibility of screening programmes, identified similar barriers and facilitators using thematic analyses (Liles *et al.* 2015; Schneider *et al.* 2016). For instance, Liles *et al.* (2015) identified barriers to colorectal cancer (CRC) screening, such as the lack of consensus on a suitable screening test and the practicalities of training staff. Furthermore, previous reviews of potential screening programmes by the NSC has indicated concerns over the lack of objective screening tests and increased false positive results (alcohol misuse), and uncertainty over the benefits from screening (asymptomatic bacteriuria) (UK National Screening Committee 2017). In addition, the thematic analysis also identified barriers related to the impact to GP practices from increased referrals and the consequences to the weight management services, which are not currently set up to conduct screening programmes.

The results of the PhD overall indicated that children who are overweight/obese have a higher prevalence of many co-morbidities, of which five were deemed important based on prevalence and/or impact on health by a multi-disciplinary panel of experts in a structured consensus study. However, there were concerns highlighted regarding implementation of a screening programme which warrants further research.

7.3 Strengths and Limitations

The results of this PhD project are strengthened by a structured and transparent approach adopted for each phase. This allows for critique of the methods and future replication to assess the reliability of the results. With regards to the systematic review and meta-analyses, a transparent process means new research can be integrated into the results as and when they become available.

Additionally this is the first known piece of work that attempted to assess the feasibility of a co-morbidity screening programme for children attending a weight management service. It is also the first known systematic review and meta-analysis estimating population prevalence ratio of co-morbidities associated with childhood

obesity, relative to those of a healthy weight. The systematic review and meta-analysis identified a list of co-morbidities associated with childhood obesity and identified limitations which future research can address.

Recruiting a multi-disciplinary group of experts for the consensus panel, with wide-ranging knowledge and expertise increased the face validity of the results. Furthermore consensus on co-morbidities deemed appropriate for inclusion in a screening programme provides a starting point for future work in the area.

The systematic approach to the thematic analysis increased the credibility, transferability, dependability and confirmability of the results. Additionally, the results highlight areas for future work/research to overcome the concerns regarding the current lack of suitable screening tools, the impact to services and service users, as well as the role of the GP in the screening programme.

In addition to these strengths, some limitations are noteworthy. Firstly, there is a lack of prevalence data from general population studies that provide prevalence for all three weight categories, by age, gender and ethnicity, and conducted in the UK. These limitations prevented an accurate estimation of UK prevalence of co-morbidities. Furthermore, some co-morbidities were assessed using self-report methods, which made them vulnerable to misclassification and over-estimation of prevalence. Additionally, there was significant heterogeneity between studies in some prevalence ratio estimates, therefore results from the meta-analysis should be interpreted with caution.

Although the consensus study was strengthened by the inclusion of a multi-disciplinary group of experts, not all members could attend all face-to-face meetings, particularly a key proponent of the proposed screening programme. Therefore they were unable to hear discussions and offer their views face-to-face which may have influenced results of the post-meeting questionnaire. To mitigate this, the discussion was summarised in the post-meeting questionnaires and non-attendees were offered the opportunity to talk to the researcher to assist with knowledge transfer. However it is unclear if this was sufficient, and many did not contact the researcher for additional information. Related to this, the sample consisted of experts who were available and consented to take part in the study. Previous research as reported that results of a consensus study are contingent, partly, on the composition of the group; thus a different panel may have reached a different decision (Shekelle *et al.* 1998). Furthermore, there was variation in the service user panels, with different service users attending each focus group. This may have

influenced the discussion and their views, particularly in the second meeting as the service users did not have the prior knowledge from the first meeting, although attempts were made to fill any potential knowledge gaps by providing a detailed overview of the study, its purpose and the methods being employed.

This project was heavily focussed on the prevalence of co-morbidities in overweight and obese children/adolescents relative to children of a healthy weight. According to NSC criteria, prevalence alone is not sufficient to justify a screening programme, and other factors should be considered. These include assessment of the cost and clinical effectiveness, and availability and acceptability of the proposed screening programme. For instance, some of the existing screening programmes have a much lower prevalence, but are screened for because of the impact to one's health and the availability of effective treatment (US Department of Health and Human Services 2007; US Department of Health and Human Services 2016; US Department of Health and Human Services 2017a).

As mentioned in Chapter 6, the methods adopted by the PhD did not allow for exploration of the perceived barriers identified by the panel of health professionals and researchers, to assess whether they would manifest if the screening programme was actually implemented. Thus one of the recommendations for future research is to conduct a pilot study to assess whether the concerns are genuine, and identify the impact of a co-morbidities screening programme. Furthermore, the pilot study would provide additional evidence regarding the feasibility of implementing the proposed screening programme; this would be in line with NSC criteria that highly quality evidence is required to assess the impact of the screening programme (Public Health England 2015a).

A final, potential limitation is that the PhD considered multiple co-morbidities at once, whereas typically NSC screening programmes focus on one disease. The proposed screening programme is similar to the fetal anomaly screening programme and the new born blood spot screening programme, which screen for 11 and nine health conditions, respectively (Gov.UK 2013a; Gov.UK 2013b). Considering one co-morbidity, e.g. type 2 diabetes, at a time may have been more beneficial, as it would have allowed for a greater depth of discussion between the expert panel and a more detailed assessment of its feasibility.

7.4 Implications for Research

This PhD project provides the first step in developing a screening programme for co-morbidities associated with childhood obesity. Firstly the results of the systematic review and meta-analysis provide the basis for future research in terms of assessing the prevalence of co-morbidities in children by weight status category, and ethnicity, and by country. Furthermore, it highlights the need for prevalence studies in a UK population, particularly if the programme is to be implemented in the UK.

The results of the consensus study identified five co-morbidities (hyperglycaemia, hypertension, obstructive sleep apnoea, depression and anxiety), deemed important by a panel of health professionals and researchers and service users for inclusion in a screening programme. Future research can build on this by confirming the results through replication studies. The lack of consensus on screening methods, indicates a need for better screening methods, in terms of accuracy, which can be used within non-clinical settings, whilst still maintaining a degree of objectivity in line the NSC criteria (Public Health England 2015a).

The thematic analysis indicated potential barriers to implementing a screening programme, such as inappropriate screening methods, cost implications, and impacts to services, service users and GPs. Future research may be able to test the validity of these barriers through feasibility studies in a local area. This would enable a detailed assessment of actual and perceived barriers. Furthermore, the results of the thematic analysis act as a starting point for the factors that future research would need to consider when developing a feasibility study to assess the practicalities of implementing a co-morbidity screening programme.

7.5 Implications for Practice

The results provide a starting point for services that wish to screen for co-morbidities, by identifying key co-morbidities to include, with flexibility in screening method based on their local practices and limitations. This can be supported by results from the thematic analysis, which indicated the screening programme may have to account for local differences in terms of available services and treatments for children. Related to this, the results highlight that the NICE guidance cannot be implemented without further work to assess the feasibility, practical constraints and implications of implementing a co-morbidity screening programme for children with obesity attending a weight management service.

7.6 Conclusions

The aim of the PhD was to assess the feasibility of developing a co-morbidity screening programme for children with obesity attending community weight management services. The results of the PhD indicated that although there is agreement to screen for specific co-morbidities at a young age, there are currently practical limitations which mean at present a co-morbidity screening programme is not feasible within UK weight management services.

Despite this, the PhD makes a contribution to the field by conducting the first known global systematic review and meta-analysis of the prevalence of co-morbidities in children who are overweight/obese relative to those of a healthy weight; providing a consolidated view of literature and prevalence of obesity-associated co-morbidities in children. Additionally, via a structured and transparent consensus study, the 22 identified comorbidities were reduced to five, which were deemed by a panel of experts to be appropriate for inclusion in a screening programme. This was the first known consensus study, which along with the thematic analysis explored the feasibility of the NICE recommendation to develop a screening programme. The thematic analysis also identified potential barriers to the successful implementation of the proposed screening programme, some which have not been identified by prior research, and which future research can investigate and identify means to overcome these barriers. The PhD also highlighted areas for future research to build on the results, and implications for practice with regards to the screening of obesity-associated co-morbidities in children.

References

ABDULLAH, A., A. PEETERS, M. DE COURTEN and J. STOELWINDER. 2010. The magnitude of association between overweight and obesity and the risk of diabetes: A meta-analysis of prospective cohort studies. *Diabetes Research & Clinical Practice,* **89**, pp.309-319.

ADAB, P., M. J. PALLAN, E. R. LANCASHIRE, K. HEMMING, E. FREW, T. BARRETT, R. BHOPAL, J. E. CADE, A. CANAWAY, J. L. CLARKE, A. DALEY, J. J. DEEKS, J. L. DUDA, U. EKELUND, P. GILL, T. GRIFFIN, E. MCGEE, K. HURLEY, J. MARTIN, J. PARRY, S. PASSMORE and K. K. CHENG. 2018. Effectiveness of a childhood obesity prevention programme delivered through schools, targeting 6 and 7 year olds: cluster randomised controlled trial (WAVES study). *BMI,* **360**, pK211.

ADIBI, A., R. KELISHADI, A. BEIHAGHI, H. SALEHI and M. TALAEI. 2009. Sonographic fatty liver in overweight and obese children, a cross sectional study in Isfahan. *Endokrynologia Polska,* **60**(1), pp.14-19.

ADRIAANSE, M., F. SNOEK, J. DEKKER, H. VAN DER PLOEG and R. HEINE. 2002. Screening for Type 2 diabetes: an exploration of subjects' perceptions regarding diagnosis and procedure. *Diabet Med.,* **19**(5), pp.405-411.

ADRIAANSE, M. C. and F. J. SNOEK. 2006. The psychological impact of screening for type 2 diabetes. *Diabetes Metab Res Rev.,* **22**(1), pp.20-25.

AINSWORTH, N. P., E. J. VARGO and A. PETRÓCZI. 2018. Being in control? A thematic content analysis of 14 in-depth interviews with 2,4-dinitrophenol users. *International Journal of Drug Policy,* **52**, pp.106-114.

AKINS, R. B., H. TOLSON and B. R. COLE. 2005. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. *BMC Med Res Methodol,* **5**(37).

AL-AGHA, A., A. OCHELTREE and N. SHATA. 2012. Prevalence of hyperinsulinism, type 2 diabetes mellitus and metabolic syndrome among Saudi overweight and obese pediatric patients. *Minerva Pediatrica,* **64**(6), pp.623-631.

AL-BAJJALI, T. T. and L. D. RAJAB. 2014. Traumatic dental injuries among 12-year-old Jordanian schoolchildren: an investigation on obesity and other risk factors. *BMC Oral Health,* **14**, p101.

AL-KHUDAIRY, L., E. LOVEMAN, J. L. COLQUITT, E. MEAD, R. E. JOHNSON, H. FRASER, J. OLAJIDE, M. MURPHY, R. VELHO, C. O'MALLEY, L. B. AZEVEDO, L. J. ELLS, M. METZENDORF and K. REES. 2017. Diet, physical activity and behavioural interventions for the treatment of overweight or obese adolescents aged 12 to 17 years. *Cochrane Database Syst Rev.,* **6**(CD012691).

ÁLAMO-JUNQUERA, D., C. MURTA-NASCIMENTO, F. MACIÀ, M. B. J. GALCERÁN, N. ASCUNCE, R. ZUBIZARRETA, D. SALAS, R. ROMÁN, X. CASTELLS, M. SALA and THE CUMULATIVE FALSE-POSITIVE RISK GROUP. 2012. Effect of false-positive results on reattendance at breast cancer screening programmes in Spain. *European Journal of Public Health,* **22**(3), pp.404-408.

ALAVIAN, S. M., A. H. MOHAMMAD-ALIZADEH, F. ESNA-ASHARI, G. ARDALAN and B. HAJARIZADEH. 2009. Non-alcoholic fatty liver disease prevalence among school-aged children and adolescents in Iran and its association with biochemical and anthropometric measures. *Liver International,* **29**(2), pp.159-63.

ALHOJAILAN, M. I. 2012. Thematic Analysis: A critical review of its process and evaluation. *In: WEI International European: Academic Conference Proceedings,* Zagreb, Croatia. WEI International European, pp.8-21.

ALLEN, J., J. DYAS and M. JONES. 2004. Building consensus in health care: a guide to using the nominal group technique. *Br J Community Nurs,* **9**(3), pp.110-114.

ANDERSON, P. M. and K. E. BUTCHER. 2006. Childhood obesity: trends and potential causes. *Future Child,* **16**(1), pp.19-45.

ANGERMEYER, M. C. and H. MATSCHINGER. 2003. The stigma of mental illness: effects of labelling on public attitudes towards people with mental disorder. *Acta Psychiatr Scand,* **108**(4), pp.304-309.

ANGLIN, R. E. S., Z. SAMAAN, S. D. WALTER and S. D. MCDONALD. 2013. Vitamin D deficiency and depression in adults: systematic review and meta-analysis. *The British Journal of Psychiatry,* **202**, pp.100-107.

ANGOLD, A., E. J. COSTELLO and A. ERKANLI. 1999. Comorbidity. *J Child Psychol Psychiatry,* **40**(1), pp.57-87.

ANTON, S. D., R. L. NEWTON JR, M. SOTHERN, C. K. MARTIN, T. M. STEWART and D. A. WILLIAMSON. 2006. Association of depression with Body Mass Index, sedentary behavior, and maladaptive eating attitudes and behaviors in 11 to 13-year old children. *Eat Weight Disord,* **11**(3), pp.102-108.

ARDITI, C., I. PEYTREMANN-BRIDEVAUX, B. BURNAND, V. F. ECKARDT, P. BYTZER, L. AGRÉUS, R. W. DUBOIS, J. P. VADER, F. FROEHLICH, V. PITTET, S. SCHUSSELÉ FILLIETTAZ, P. JUILLERAT, J. J. GONVERS and THE EPAGE II STUDY GROUP. 2009. Appropriateness of colonoscopy in Europe (EPAGE II): Screening for colorectal cancer. *Endoscopy,* **41**, pp.200-208.

ARTERBURN, D. E., M. L. MACIEJEWSKI and J. TSEVAT. 2005. Impact of morbid obesity on medical expenditures in adults. *International Journal of Obesity,* **29**(3), pp.334-339.

ASIF, I. M., S. JOHNSON, J. SCHMIEG, T. SMITH, A. L. RAO, K. G. HARMON, J. C. SALERNO and J. A. DREZNER. 2014. The psychological impact of cardiovascular screening: the athlete's perspective. *Br J Sports Med,* **48**, pp.1162-1166.

AUERBACH, C. and L. B. SILVERSTEIN. 2003. *Qualitative Data: An Introduction to Coding and Analysis.* New York, USA: New York University Press.

AUGUST, G. P., S. CAPRIO, I. FENNOY, M. FREEMARK, F. R. KAUFMAN, R. H. LUSTIG, J. H. SILVERSTEIN, P. W. SPEISER, D. M. STYNE, V. M. MONTORI and ENDOCRINE SOCIETY. 2008. Prevention and treatment of pediatric obesity: an endocrine society clinical practice guideline based on expert opinion. *J Clin Endocrinol Metab.,* **93**(12), pp.4576-99.

AYONRINDE, O. T., J. K. OLYNYK, L. J. BEILIN, T. A. MORI, C. E. PENNELL, N. DE KLERK, W. H. ODDY, P. SHIPMAN and L. A. ADAMS. 2011. Gender-Specific Differences in Adipose Distribution and Adipocytokines Influence Adolescent Nonalcoholic Fatty Liver Disease. *Hepatology,* **53**(3), pp.800-809.

BAILEY, J. 2008. First steps in qualitative data analysis: transcribing. *Fam Pract,* **25**(2), pp.127-131.

BALÁN, I., T. FRASCA, M. IBITOYE, C. DOLEZAL and A. CARBALLO-DIÉGUEZ. 2016. Fingerprick Versus Oral Swab: Acceptability of Blood-Based Testing Increases If Other STIs Can Be Detected. *AIDS and Behavior,* **21**(2), pp.501-504.

BANDURA, A. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review,* **84**(2), pp.191-215.

BAR DAYAN, Y., K. ELISHKEVITS, I. GROTTO, L. GOLDSTEIN, A. GOLDBERG, S. SHVARTS, A. LEVIN, N. OHANA, E. ONN, Y. LEVI and Y. BAR DAYAN. 2005. The prevalence of obesity and associated morbidity among 17-year-old Israeli conscripts. *Public Health,* **119**(5), pp.385-389.

BARATLOO, A., M. HOSSEINI, A. NEGIDA and G. EL ASHAL. 2015. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emerg (Tehran),* **3**(2), pp.48-39.

BASEN-ENGQUIST, K. and M. CHANG. 2011. Obesity and Cancer Risk: Recent Review and Evidence. *Curr Oncol Rep.,* **13**(1), pp.71-76.

BASGER, B. J., T. F. CHEN and R. J. MOLES. 2012. Validation of prescribing appropriateness criteria for older Australians using the RAND/UCLA appropriateness method. *BMJ Open,* **2**, pp.doi:10.1136/bmjopen-2012-001431.

BAULDRY, S., K. A. BOLLEN and L. S. ADAIR. 2015. Evaluating measurement error in readings of blood pressure for adolescents and young adults. *Blood Pressure,* **24**(2), pp.96-102.

BAUM, C. L. N. and W. F. FORD. 2004. The wage effects of obesity: a longitudinal study. *Health Economics,* **13**(9), pp.885-899.

BAUMAN, A. E., R. S. REIS, J. F. SALLIS, J. C. WELLS, R. J. F. LOOS and B. W. MARTIN. 2012. Correlates of physical activity: why are some people physically active and others not? *Lancet* **380**(9838), pp.258-271.

BBC NEWS. 2018. *One in 25 children aged 10 or 11 'severely obese'* [online]. [Accessed 29th May]. Available from: https://www.bbc.co.uk/news/health-44261065.

BEATTIE, E. and K. MACKWAY-JONES. 2004. A Delphi study to identify performance indicators for emergency medicine. *Emerg Med J,* **21**, pp.47-50.

BELL, B. G., R. SPENCER, A. J. AVERY and S. M. CAMPBELL. 2014. Tools for measuring patient safety in primary care settings using the RAND/UCLA appropriateness method. *BMC Fam Pract,* **15**, p110.

BELL, L. M., S. BYRNE, A. THOMPSON, N. RATNAM, E. BLAIR, M. BULSARA, T. W. JONES and E. A. DAVIS. 2007. Increasing body mass index z-score is continuously associated with complications of overweight in children, even in the healthy weight range. *J Clin Endocrinol Metab.,* **92**(2), pp.517-522.

BELL, L. M., J. A. CURRAN, S. BYRNE, H. ROBY, K. SURIANO, T. W. JONES and E. A. DAVIS. 2011. High incidence of obesity co-morbidities in young children: a cross-sectional study. *J Paediatr Child Health,* **47**(12), pp.911-917.

BENJAMIN, E. M. 2002. Self-Monitoring of Blood Glucose: The Basics. *Clinical Diabetes,* **20**(1), pp.45-47.

BENSON, L. J., H. J. BAER and D. C. KAELBER. 2012. Screening for Obesity-Related Complications Among Obese Children and Adolescents: 1999–2008. *Obesity,* **19**(5), pp.1077-82.

BERGE, J. M. 2010. A Review of Familial Correlates of Child and Adolescent Obesity: What has the 21st Century Taught us so Far? *Int J Adolesc Med Health,* **21**(4), pp.457-483.

BERMAN, L., J. JACKSON, K. MILLER, R. KOWALSKI, P. KOLM and F. I. LUKS. 2017. Expert surgical consensus for prenatal counseling using the Delphi method. *Journal of Pediatric Surgery*.

BHATTACHERJEE, A., N. CHAU, C. O. SIERRA, B. LEGRAS, L. BENAMGHAR, J. P. MICHAELY, A. K. GHOSH, F. GUILLEMIN, J. F. RAVAUD, J. M. MUR and L. GROUP. 2003. Relationships of job and some individual characteristics to occupational injuries in employed people: a community-based study. *Journal of Occupational Health,* **45**(6), pp.382-91.

BIRKO, S., E. S. DOVE and V. ÖZDEMIR. 2015. Evaluation of Nine Consensus Indices in Delphi Foresight Research and Their Dependency on Delphi Survey Characteristics: A Simulation Study and Debate on Delphi Design and Interpretation. *PLos One,* **10**(9), pe0135162.

BJØRGE, T., A. ENGELAND, A. TVERDAL and G. D. SMITH. 2008. Body mass index in adolescence in relation to cause-specific mortality: a follow-up of 230,000 Norwegian adolescents. *Am J Epidemiol.,* **168**(1), pp.30-7.

BLACK, J. A., B. WHITE, R. M. VINER and R. K. SIMMONS. 2013. Bariatric surgery for obese children and adolescents: a systematic review and meta-analysis. *Obes Rev.,* **14**(8), pp.634-644.

BLACK, N., M. MURPHY, D. LAMPING, M. MCKEE, C. SANDERSON, J. ASKHAM and T. MARTEAU. 1999a. Consensus Development Methods: A Review of Best Practice in

Creating Clinical Guidelines. *Journal of Health Services Research & Policy,* **4**(4), pp.236-248.

BLACK, N., M. MURPHY, D. LAMPING, M. MCKEE, C. SANDERSON, J. ASKHAM and T. MARTEAU. 1999b. Consensus Development methods: a review of best practice in creating clinical guidelines. *Journal of Health Services and Research Policy,* **4**(4), pp.236-248.

BOND, M., T. PAVEY, K. WELCH, C. COOPER, R. GARSIDE, S. DEAN and C. J. HYDE. 2013. Psychological consequences of false-positive screening mammograms in the UK. *Evid Based Med,* **18**(2), pp.54-61.

BOOTH, M. L., J. GEORGE, E. DENNEY-WILSON, A. D. OKELY, L. L. HARDY, R. AITKEN and T. DOBBINS. 2008. The population prevalence of adverse concentrations and associations with adiposity of liver tests among Australian adolescents. *Journal of Paediatrics and Child Health,* **44**(12), pp.686-691.

BORENSTEIN, M., L. V. HEDGES, J. P. T. HIGGINS and H. R. ROTHSTEIN. 2009. *Introduction to Meta-Analysis.* West Sussex, UK: John Wiley & Sons Ltd.

BORENSTEIN, M., L. V. HEDGES, J. P. T. HIGGINS and H. R. ROTHSTEIN. 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods,* **1**(2), pp.97-111.

BORENSTEIN, M., J. P. T. HIGGINS, L. HEDGES and H. ROTHSTEIN. 2017. Basics of Meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods,* **8**, pp.5-18.

BORJESON, M. 1976. The Aetiology of Obesity in Children: A Study of 101 Twin Pairs. *Acta Paediatrica,* **65**(3), pp.279-287.

BOULKEDID, R., H. ABDOUL, M. LOUSTAU, O. SIBONY, C. ALBERTI and J. M. WRIGHT. 2011. Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. *PLoS One,* **6**(6), pe20476.

BOURREE, F., P. MICHEL and L. R. SALMI. 2008. Consensus methods: Review of original methods and their main alternatives used in public health. *Revue d'Epidemiologie et de Sante Publique,* **56**, pp.13-21.

BOWLING, A. 2005. Mode of questionnaire administration can have serious effects on data quality. *J Public Health* **27**(3), pp.281-291.

BOWN, M. J. and A. J. SUTTON. 2010. Quality Control in Systematic Reviews and Meta-analyses. *European Journal of Vascular and Endovascular Surgery,* **40**(5), pp.669-677.

BOYATZIS, R. E. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development.* London: Sage Publishing.

BRADY, T. M. 2017. Obesity-Related Hypertension in Children. *Front Pediatr.,* **5**(197).

BRAR, S. S., A. L. MAHAR, L. K. HELYER, C. SWALLOW, C. LAW, L. PASZAT, R. SEEVARATNAM, R. CARDOSO, R. MCLEOD, M. DIXON, L. YOHANATHAN, L. G. LOURENCO, A. BOCICARIU, T. BEKAII-SAAB, I. CHAU, N. CHURCH, D. COIT, C. H. CRANE, C. EARLE, P. MANSFIELD, N. MARCON, T. MINER, S. H. NOH, G. PORTER, M. C. POSNER, V. PRACHAND, T. SANO, C. VAN DE VELDE, S. WONG and N. G. COBURN. 2014. Processes of Care in the Multidisciplinary Treatment of Gastric Cancer: Results of a RAND/UCLA Expert Panel. *JAMA Surg,* **149**(1), pp.18-25.

BRAUN, V. and V. CLARKE. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology,* **3**(2), pp.77-101.

BRAY, G. A., K. K. KIM and J. P. H. WILDING. 2017. Obesity: a chronic relapsing progressive disease process. A position statement of the World Obesity Federation. *Obesity Reviews,* **18**(7), pp.715-723.

BRESLOW, L. and D. W. ROBERTS. 1955. Screening for Asymptomatic Disease. *J. Chron Dis.,* **2**, pp.363-490.

BRETT, J., S. STANISZEWSKA, I. SIMERA, K. SEERS, C. MOCKFORD, S. GOODLAD, D. ALTMAN, D. MOHER, R. BARBER, S. DENEGRI, A. R. ENTWISTLE, P. LITTLEJOHNS, C. MORRIS, R. SULEMAN, V. THOMAS and C. TYSALL. 2017. Reaching consensus on reporting patient and public involvement (PPI) in research: methods and lessons learned from the development of reporting guidelines. *BMJ Open,* **7**(10), pe016948.

BRITISH HEART FOUNDATION. 2012. *Coronary Heart Disease Statisitcs: A compendium of health statistics* [online]. [Accessed 3rd May]. Available from: https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja &uact=8&ved=0ahUKEwjD-8ny8NPTAhWLPxQKHbEpAKwQFgguMAE&url=https%3A%2F%2Fwww.bhf.org.uk%2 F-%2Fmedia%2Ffiles%2Fpublications%2Fresearch%2F2012_chd_statistics_compendi um.pdf&usg=AFQjCNGpcjJZeSKbo476DUvdVaSxkNS6hg&sig2=4kwQC3VHQfyYpJpU RcTvig.

BRODERSEN, J. and V. D. SIERSMA. 2013. Long-Term Psychosocial Consequences of False-Positive Screening Mammography. *Ann Fam Med,* **11**(2), pp.106-115.

BROOK, R. H. 1995. *The RAND/UCLA Appropriateness Method* [online]. [Accessed]. Available from: http://www.rand.org/pubs/reprints/RP395.html.

BROOMFIELD, D. and G. M. HUMPHRIS. 2001. Using the Delphi technique to identify the cancer educationrequirements of general practitioners. *Medical Education,* **35**, pp.928-937.

BUDD, G. M. and L. L. HAYMAN. 2008. Addressing the childhood obesity crisis. *Am J Matern Child Nurs,* **33**, pp.113-117.

BUETOW, S. 2010. Thematic analysis and its reconceptualization as 'saliency analysis'. *J Health Serv Res Policy,* **15**(2), pp.123-125.

BUICK, S., M. JOFFE, T. TAYLOR and J. CONLY. 2014. A Consensus Development Conference Model for Establishing Health Policy for Surveillance and Screening of Antimicrobial-Resistant Organisms. *Clinical Infectious Diseases*.

BURTON, W. N., C. Y. CHEN, A. B. SCHULTZ and D. W. EDINGTON. 1998. The economic costs associated with body mass index in a workplace. *Journal of Occupational and Environmental Medicine,* **40**(9), pp.786-792.

BUTLER, M. M., T. C. MEEHAN, M. KEMPLE, J. DRENNAN, M. TREACY and M. JOHNSON. 2009. Identifying research priorities for midwifery in Ireland. *Midwifery,* **25**, pp.576-587.

CALDERÓN-LARRAÑAGA, A., P. HERNÁNDEZ-OLIVAN, F. GONZÁLEZ-RUBIO, L. A. GIMENO-FELIU, B. POBLADOR-PLOU and A. PRADOS-TORRES. 2015. Multimorbidity and weight loss in obese primary care patients: longitudinal study based on electronic healthcare records. *BMJ Open,* **5**(3), pe006227.

CAMPBELL, M. K. 2016. Biological, environmental, and social influences on childhood obesity. *Pediatric Research,* **79**, pp.205-211.

CAMPBELL, S. E., A. AZUARA-BLANCO, M. K. CAMPBELL, J. J. FRANCIS, A. C. GREENE, C. R. RAMSAY and J. M. BURR. 2012. Developing the specifications of an Open Angle Glaucoma screening intervention in the United Kingdom: a Delphi approach. *BMC Health Services Research,* **12**(447).

CAMPBELL, S. M., T. SHIELD, A. ROGERS and L. GASK. 2004. How do stakeholder groups vary in a Delphi technique about primary mental health care and what factors influence their ratings? *BMJ Quality and Safety,* **13**, pp.428-434.

CAREY, T. S. and S. D. BODEN. 2003. A Critical Guide to Case Series Reports. *SPINE,* **28**(15), pp.1631-1634.

CARVER, A., A. F. TIMPERIO, K. D. HESKETH, N. D. RIDGERS, J. L. SALMON and D. A. CRAWFORD. 2011. How is active transport associated with children's and

adolescents' physical activity over time? *International Journal of Behavioral Nutrition and Physical Activity,* **8**(1), pp.126-131.

CASERTA, C. A., G. M. PENDINO, S. ALICANTE, A. AMANTE, F. AMATO, M. FIORILLO, A. MESSINEO, I. POLITO, M. SURACE, P. SURACE, C. VACALEBRE, M. ZUIN, R. COTICHINI, F. MARCUCCI, F. ROSMINI, A. MELE and M. S. GROUP. 2010. Body mass index, cardiovascular risk factors, and carotid intima-media thickness in a pediatric population in southern Italy. *Journal of Pediatric Gastroenterology & Nutrition,* **51**(2), pp.216-20.

CASSAR FLORES, A., S. MARSHALL and M. CORDINA. 2014. Use of the Delphi technique to determine safety features to be included in a neonatal and paediatric prescription chart. *Int J Clin Pharm,* **36**(6), pp.1179-1189.

CAWLEY, J. and C. MEYERHOEFER. 2012. The medical care costs of obesity: An instrumental variables approach. *Journal of Health Economics,* **31**, pp.219-230.

CDC. 2016. *Global School-based Student Health Survey (GSHS)* [online]. [Accessed 2nd Jan 2018]. Available from: https://www.cdc.gov/gshs/.

CENTERS FOR DISEASE CONTROL AND PREVENTION. 2005. Reducing Childhood Asthma Through Community-Based Service Delivery - New York City, 2001-2004. *Morbidity and Mortality Weekly Report,* **54**(1), pp.11-14.

CENTERS FOR DISEASE CONTROL AND PREVENTION. 2015. *High Cholesterol Facts* [online]. [Accessed 14th September]. Available from: https://www.cdc.gov/cholesterol/facts.htm.

CENTRE FOR REVIEWS AND DISSEMINATION. 2009. *Systematic Reviews: CRD's guidance for undertaking reviews in health care* [online]*.* York: CRD, University of York. [Accessed 3rd March 2017]. Available from: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.

CHAN, K. and M. BHANDARI. 2011. Three-minute critical appraisal of a case series article. *Indian J Orthop,* **45**(2), pp.103-104.

CHAPMAN, A. L., M. HADFIELD and C. J. CHAPMAN. 2015. Qualitative research in healthcare: an introduction to grounded theory using thematic analysis. *J R Coll Physicians Edinb,* **45**, pp.201-205.

CHARAP, M. H. 1981. The periodic health examination: genesis of a myth. *Ann Intern Med,* **95**(6), pp.733-735.

CHARLESWORTH, L. W. and M. K. RODWELL. 1997. Focus Groups with Children: A Resource for Sexual Abuse Prevention Program Evaluation. *Child Abuse and Neglect,* **21**, pp.1205-1216.

CHAU, N., J. M. MUR, C. TOURON, L. BENAMGHAR and D. DEHAENE. 2004. Correlates of occupational injuries for various jobs in railway workers: a case-control study. *Journal of Occupational Health,* **46**(4), pp.272-80.

CHEN, F., Y. WANG, X. SHAN, H. CHENG, D. HOU, X. ZHAO, T. WANG, D. ZHAO and J. MI. 2012. Association between childhood obesity and metabolic syndrome: evidence from a large sample of Chinese children and adolescents. *PLoS ONE [Electronic Resource],* **7**(10), pe47380.

CHEN, J. P., M. J. CHUNG and M. J. WANG. 2009. Flatfoot prevalence and foot dimensions of 5- to 13-year-old children in Taiwan. *Foot & Ankle International,* **30**(4), pp.326-32.

CHENG, E. M., C. J. CRANDALL, C. T. BEVER JR, B. GIESSER, J. K. HASELKORN, R. D. HAYS, P. SHEKELLE and B. G. VICKREY. 2010. Quality indicators for multiple sclerosis. *Multiple Sclerosis,* **16**(8), pp.970-980.

CHENG, J., J. GAO, X. SHUAI, G. WANG and K. TAO. 2016. The comprehensive summary of surgical versus non-surgical treatment for obesity: a systematic review and meta-analysis of randomized controlled trials. *Oncotarget.,* **7**(26), pp.39216-39230.

CHENG, J. P. EAST, E. BLANCO, E.K. SIM, M. CASTILLO, B. LOZOFF, S. GAHAHAN. 2016a. Obesity leads to declines in motor skills across childhood. *Child Care Health Dev., 42(3), pp.343-350.*

CHOPRA, M., S. GALBRAITH and I. DARNTON-HILL. 2002. A global response to a global problem: the epidemic of overnutrition. *Bulletin of the World Health Organization,* **80**(12), pp.952-958.

CHU, N. F. and W. H. PAN. 2007. Prevalence of obesity and its comorbidities among schoolchildren in Taiwan. *Asia Pacific Journal of Clinical Nutrition,* **16 Suppl 2**, pp.601-7.

CLARK, C. D. 1996. Interviewing Children in Qualitative Research: a show and tell. *Canadian Journal of Market Research,* **15**, pp.74-79.

CLARKE, V. and V. BRAUN. 2013. Teaching Thematic Analysis: Overcoming challenges and developing strategies for effective learning. *The Psychologist,* **26**(2), pp.120-123.

COHEN, A. K., M. RAI, D. H. REHKOPF and B. ABRAMS. 2013. Educational attainment and obesity: A systematic review. *Obes Rev,* **14**(12), pp.989-1005.

COLEMAN, S., E. A. NELSON, J. KEEN, L. WILSON, E. MCGINNIS, C. DEALEY, N. STUBBS, D. MUIR, A. FARRIN, D. DOWDING, J. M. G. A. SCHOLS, J. CUDDIGAN, D. BERLOWITZ, E. JUDE, P. VOWDEN, D. L. BADER, A. GEFEN, C. W. J. OOMENS, L. SCHOONHOVEN and J. NIXON. 2014. Developing a pressure ulcer risk factor minimum data set and risk assessment framework. *Journal of Advanced Nursing,* **70**(10), pp.2339-2352.

COLLINS, R. E., L. M. LOPEZ and T. M. MARTEAU. 2011. Emotional impact of screening: a systematic review and meta-analysis. *BMC Public Health.,* **11**(603).

COMMISSION ON CHRONIC ILLNESS. 1957. *Chronic Illness in the United States. Volume 1 - Prevention of Chronic Illness.* Cambridge, MA: Harvard University Press.

CONNELLY, J., J. COOPER, A. MANN and T. W. MEADE. 1998. The psychological impact of screening for risk of coronary heart disease in primary care settings. *J Cardiovasc Risk.,* **5**(3), pp.185-191.

Consensus Development Conference on Antipsychotic Drugs and Obesity and Diabetes. 2004. *Diabetes Care,* **27**(2), pp.596-601.

Consensus development conference: prophylaxis and treatment of osteoporosis. 1987. *British Medical Journal (Clinical research ed.),* **295**(6603), pp.914-915.

Consensus development conference: prophylaxis and treatment of osteoporosis. 1991. *Am J Med,* **90**(1), pp.107-110.

COOK, D. J., C. D. MULROW and R. B. HAYNES. 1997. Systematic reviews: Synthesis of best evidence for clinical decisions. *Ann Intern Med,* **126**, pp.376-380.

COOK, S., M. WEITZMAN, P. AUINGER, M. NGUYEN and W. H. DIETZ. 2003. Prevalence of a metabolic syndrome phenotype in adolescents: findings from the third National Health and Nutrition Examination Survey, 1988-1994. *Archives of Pediatrics and Adolescent Medicine,* **157**(8), pp.821-827.

CORTÉS-TÉLLES, A., L. TORRE-BOUSCOULET, M. SILVA-CERÓN, R. MEJÍA-ALFARO, N. SYED, G. S. ZAVORSKY and J. A. GUENETTE. 2015. Combined effects of mild-to-moderate obesity and asthma on physiological and sensory responses to exercise. *Respiratory Medicine,* **109**(11), pp.1394-1403.

CRANE, D., E. J. HENDERSON and D. R. CHADWICK. 2016. Exploring the acceptability of a 'limited patient consent procedure' for a proposed blood-borne virus screening programme: a Delphi consensus building technique. *BMJ Open,* **7**(5), pe015373.

CUTTING, C. C. 1985. *HISTORY OF THE KAISER PERMANENTE MEDICAL CARE PROGRAM In:* M. CHALL (Ed.).

D'AMORE, P. J. 2006. Childhood Obesity. *Laboratory Medicine,* **37**(5), pp.298-305.

DABELEA, D., J. M. STAFFORD, E. J. MAYER-DAVIS, R. D'AGOSTINO JR, L. DOLAN, G. IMPERATORE, B. LINDER, J. M. LAWRENCE, S. M. MARCOVINA, A. K. MOTTL, M. H. BLACK, R. POP-BUSUI, S. SAYDAH, R. F. HAMMAN and C. PIHOKER. 2017.

Association of Type 1 Diabetes vs Type 2 Diabetes Diagnosed During Childhood and Adolescence With Complications During Teenage Years and Young Adulthood. *JAMA,* **317**(8), pp.825-835.

DALKEY, N. 1969. An experimental study of group opinion: The Delphi method. *Futures,* **1**(5), pp.408-426.

DALKEY, N. and O. HELMER. 1962. *An Experimental Application of the DELPHI Method to the Use of Experts.* Santa Monica, California: RAND Corporation.

DANIELS, S. R. 2009. Complications of obesity in children and adolescents. *Int J Obes,* **33**, pp.S60-65.

DAVIS, C. L., B. FLICKINGER, D. MOORE, R. BASSALI, S. D. BAXTER and Z. YIN. 2005. Prevalence of cardiovascular risk factors in schoolchildren in a rural Georgia community. *American Journal of the Medical Sciences,* **330**(2), pp.53-59.

DAVISON, K. K. and L. L. BIRCH. 2001. Childhood overweight: a contextual model and recommendations for future research. *Obes Rev.,* **2**(3), pp.159-171.

DE COSMI, V., S. SCAGLIONI and C. AGOSTONI. 2017. Early Taste Experiences and Later Food Choices. *Nutrients.,* **9**(2), pp.107-115.

DE FERRANTI, S. D., K. GAUVREAU, D. S. LUDWIG, E. J. NEUFELD, J. W. NEWBURGER and N. RIFAI. 2004. Prevalence of the metabolic syndrome in American adolescents: findings from the Third National Health and Nutrition Examination Survey. *Circulation,* **110**(16), pp.2494-2497.

DE GONZALEZ, A. B., P. HARTGE, J. R. CERHAN, A. J. FLINT, L. HANNAN, R. J. MACINNIS, S. C. MOORE, G. S. TOBIAS, H. ANTON-CULVER, L. B. FREEMAN, W. L. BEESON, S. L. L. CLIPP, D. R. ENGLISH, A. R. FOLSOM, D. M. FREEDMAN, G. GILES, N. HAKANSSON, K. D. HENDERSON, J. HOFFMAN-BOLTON, J. A. HOPPIN and E. AL. 2010. Body-Mass Index and Mortality among 1.46 Million White Adults. *N Engl J Med,* **363**, pp.2211-2219.

DE SCHREYE, R., D. HOUTTEKIER, L. DELIENS and J. COHEN. 2017. Developing indicators of appropriate and inappropriate end-of-life care in people with Alzheimer's disease, cancer or chronic obstructive pulmonary disease for population-level administrative databases: A RAND/UCLA appropriateness study. *Palliat Med,* **31**(10), pp.932-945.

DEAN, H. J. and E. A. SELLERS. 2007. Comorbidities and microvascular complications of type 2 diabetes in children and adolescents. *Pediatr Diabetes,* **8**(S9), pp.35-41.

DEEKS, J. J. and J. P. T. HIGGINS. 2010. Statistical algorithms in Review Manager 5 [online]. p11. Available from: https://www.researchgate.net/profile/Jonathan_Deeks2/publication/252093205_Statistical_Algorithms_in_Review_Manager_5/links/54d159b70cf28370d0e07f9e/Statistical-Algorithms-in-Review-Manager-5.pdf?origin=publication_detail&ev=pub_int_prw_xdl&msrp=egicQkOEoUa0ee7E12M1zFhzgpfpQQSOYCNwJACRm_ucgp66XIsfFUDbubBZG9dGn0b0QrJ2cBrKPCID6TbSUowAO-1qKpAqygQvZ7dEwoFfnzSxgOtma-sw.d5iy1zbArLFmK-0A4su4gWInPzY2-qfoNX-2PwlmG07CM0ZQfMiKwYoPvQYBgAwLZ6LLhfaLFdpkrqtwJeRLTtVtVT_3Yk_G-CluRg.oGYr915GwfdVjVtVw80i9Mx5dt8gGcVNEzswZXoXfq4YeiEgl8zEohOKm73engDFXbsskS9KCPI0YucUNx_51ejnHQR8CTqTRpjw3A.uiMOyfGtVwLmF4NX7g_CUaxSxlSVTUbPpGO5LOddKRCrcmP7n9xSpwt826reJ70hJOCovMA3RShTLHGVh0h6mT038O9cWlW0qzWuNg.

DEGLI ESPOSTI, E., A. STURANI, G. VALPIANI, M. DI MARTINO, F. ZICCARDI, A. RITA CASSANI, C. BARACCANI, M. GENTILE, M. G. PUGLIA and L. DEGLI ESPOSTI. 2006. The relationship between body weight and drug costs: An Italian population-based study. *Clinical Therapeutics,* **28**(9), pp.1472-81.

DEL-RIO-NAVARRO, B. E., O. VELAZQUEZ-MONROY, A. LARA-ESQUEDA, R. VIOLANTE-ORTIZ, G. FANGHANEL, L. PEREZ-SANCHEZ and A. BERBER. 2008. Obesity and Metabolic Risks in Children. *Archives of Medical Research,* **39**(2), pp.215-221.

DELBECQ, A. L. and A. H. VAN DE VEN. 1971. A Group Process Model For Problem Identification and Program Planning. *The Journal of Applied Behavioral Science,* **7**(4), pp.466-492.

DEPARTMENT OF HEALTH. 2004. *Choosing Health: Making Healthy Choices Easier.* London: Department of Health.

DEPARTMENT OF HEALTH. 2005. *Research Governance Framework for Health and Social Care.* London: Department of Health.

DEPARTMENT OF HEALTH. 2006. *Our health, our care, our say: a new direction for community services* London: Department of Health.

DEPARTMENT OF HEALTH. 2010. *Equity and excellence: Liberating the NHS* London: Department of Health.

DEPARTMENT OF HEALTH. 2013. *Developing a specification for lifestyle weight management services: Best practice guidance for tier 2 services* [online]. [Accessed 18th September]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/142723/Weight_Management_Service_Spec_FINAL_with_IRB.pdf.

DETRANO, R., A. JANOSI, K. P. LYONS, G. MARCONDES, N. ABBASSI and V. F. FROELICHER. 1988. Factors affecting sensitivity and specificity of a diagnostic test: the exercise thallium scintigram. *Am J Med,* **84**(4), pp.699-710.

DEUTEKOM, M., F. VANSENNE, K. MCCAFFERY, M. ESSINK-BOT, K. STRONKS and P. M. M. BOSSUYT. 2011. The effects of screening on health behaviour: a summary of the results of randomized controlled trials. *Journal of Public Health,* **33**(1), pp.71-79.

DEVAUX, M. and F. SASSI. 2013. Social inequalities in obesity and overweight in 11 OECD countries. *Eur J Public Health,* **23**(3), pp.464-469.

DEVEREAUX, P. J., P. T. L. CHOI, S. EL-DIKA, M. BHANDARI, V. M. MONTORI, H. J. SCHÜNEMANN, A. X. GARG, J. W. BUSSE, D. HEELS-ANSDELL, W. A. GHALI, B. J. MANNS and G. H. GUYATT. 2004. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. *J Clin Epidemiol,* **57**, pp.1232-1236.

DHILLON, J. K. and N. C. GILL. 2014. Deciphering the system of a systematic review. *Dent Res J (Isfahan),* **11**(5), pp.531-536.

DIABETES UK. 2014. *Diabetes: Facts and Stats* [online]. [Accessed 5th May]. Available from: https://www.diabetes.org.uk/resources-s3/2017-11/diabetes-key-stats-guidelines-april2014.pdf.

DIABETES UK. 2016. *Oral Glucose Tolerance Test* [online]. [Accessed 21st June]. Available from: https://www.diabetes.co.uk/oral-glucose-tolerance-test.html.

DIABETES UK. 2018a. *Diabetes and Hyperglycemia* [online]. [Accessed 13th Feb 2018]. Available from: https://www.diabetes.co.uk/Diabetes-and-Hyperglycaemia.html.

DIABETES UK. 2018b. *Diabetes Tests* [online]. [Accessed 13th May]. Available from: https://www.diabetes.co.uk/Diabetes-screening-test.html.

DIABETES UK. n.d. *Cost of Diabetes* [online]. [Accessed 5th January]. Available from: https://www.diabetes.co.uk/cost-of-diabetes.html.

DIAMOND, I. R., R. C. GRANT, B. M. FELDMAN, P. B. PENCHARZ, S. C. LING, A. M. MOORE and P. W. WALES. 2014. Defining Consensus: A systematic review recommends methodologic criteria for reporting Delphi studies. *Journal of Clinical Epidemiology,* **67**, pp.401-409.

DINSDALE, H., C. RIDLER and L. ELLS. 2011. *A simple guide to classifying body mass index in children.* Oxford: National Obesity Observatory.

DIXON-WOODS, M., R. L. SHAW, S. AGARWAL and J. A. SMITH. 2004. The problem of appraising qualitative research. *Qual Saf Health Care,* **13**, pp.223-225.

DOBBIE, A., M. RHODES, J. W. TYSINGER and J. FREEMAN. 2004. Using a modified nominal group technique as a curriculum evaluation tool. *Core Concepts in Family Medicine Evaluation,* **36**(6), pp.402-406.

DOBBS, R., C. SAWERS, F. THOMPSON, J. MANYIKA, J. WOETZEL, P. CHILD, S. MCKENNA and A. SPATHAROU. 2014. *Overcoming Obesity: An initial economic analysis* [online]. [Accessed 13th December]. Available from: http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights/Economic%20Studies/How%20the%20world%20could%20better%20fight%20obesity/MGI_Overcoming_obesity_Full_report.ashx.

DOLPHIN, L. and E. HENNESS. 2017. Labelling effects and adolescent responses to peers with depression: an experimental investigation. *BMC Psychiatry,* **17**(228).

DOO, E. C., J. H. HOOFNAGLE and G. P. RODGERS. 2009. National Institutes of Health (NIH) Consensus Development Conference on the Management of Hepatitis B. *Hepatology,* **49**(5), pp.S1-S3.

DOVER, D. C. and D. P. SCHOPFLOCHER. 2011. Using funnel plots in public health surveillance. *Population Health Metrics,* **9**(1), pp.58-68.

DREYER, N. A., P. VELENTGAS, K. WESTRICH and R. DUBOIS. 2014. The GRACE Checklist for Rating the Quality of Observational Studies of Comparative Effectiveness: A Tale of Hope and Caution. *J Manag Care Pharm,* **20**(3), pp.301-308.

DROST, E. A. 2011. Validity and Reliability in Social Science Research. *Education Research and Perspectives,* **38**(1), pp.105-123.

DUZOVA, A., F. YALCINKAYA, E. BASKIN, A. BAKKALOGLU and O. SOYLEMEZOGLU. 2013. Prevalence of hypertension and decreased glomerular filtration rate in obese children: results of a population-based field study. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association,* **28 Suppl 4**, pp.iv166-71.

DYSON, P. A., D. ANTHONY, B. FENTON, D. R. MATTHEWS, D. E. STEVENS and COMMUNITY INTERVENTIONS FOR HEALTH COLLABORATION. 2014. High rates of child hypertension associated with obesity: a community survey in China, India and Mexico. *Paediatrics & international Child Health,* **34**(1), pp.43-9.

EBORALL, H., M. STONE, N. AUJLA and N. TAUB. 2012. Influences on the uptake of diabetes screening: a qualitative study in primary care. *Br J Gen Pract.,* **62**(596), pp.204-211.

EBORALL, H. C., S. J. GRIFFIN, T. PREVOST, A. KINMONTH, D. P. FRENCH and S. SUTTON. 2007. Psychological impact of screening for type 2 diabetes: controlled trial and comparative study embedded in the ADDITION (Cambridge) randomised controlled trial. *BMJ,* **335**, p486.

ECCLES, M., Z. CLAPP, J. GRIMSHAW, P. C. ADAMS, B. HIGGINS, I. PURVES and I. RUSSELL. 1996. Developing valid guidlines: methodological and procedural issues from the North of England Evidence Based Guideline Development Project. *Quality in Health Care,* **5**, pp.44-50.

EGGER, M., G. D. SMITH and D. G. ALTMAN. 2001. *Systematic Reviews in Health Care: Meta-Analysis in Context.* 2nd Edition ed. Tavistock Square, London: BMJ Publishing Group.

EGGER, M., G. D. SMITH and A. N. PHILLIPS. 1997. Meta-analysis: principles and procedures. *BMJ,* **315**(7121), pp.1533-1537.

ELIZONDO-MONTEMAYOR, L., P. A. UGALDE-CASAS, L. LAM-FRANCO, H. BUSTAMANTE-CAREAGA, M. SERRANO-GONZALEZ, N. G. GUTIERREZ and U. MARTINEZ. 2014. Association of ALT and the metabolic syndrome among Mexican children. *Obesity Research & Clinical Practice,* **8**(1), pp.e79-87.

ELLIOTT, E., C. T. FISCHER and D. L. RENNIE. 1999. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology,* **38**(3), pp.215-229.

ELLS, L. J., C. HANCOCK, V. R. COPLEY, E. MEAD, H. DINSDALE, S. KINRA, R. M. VINER and H. RUTTER. 2015. Prevalence of severe childhood obesity in England: 2006–2013. *Arch Dis Child.,* **100**, pp.631-636.

ELVSAAS, I. K. Ø., L. GISKE, B. FURE and L. K. JUVET. 2017. Multicomponent Lifestyle Interventions for Treating Overweight and Obesity in Children and Adolescents: A Systematic Review and Meta-Analyses. *Journal of Obesity,* **2017**(5021902).

ENES, C. C., P. M. F. FERNANDEZ, S. M. VOCI, N. TORAL, A. ROMERO and B. SLATER. 2009. Validity and reliability of self-reported weight and height measures for the diagnoses of adolescent's nutritional status. *Rev Bras Epidemiol,* **12**(4), pp.627-35.

ENZI, G. 1994. Socioeconomic Consequences of Obesity. *PharmacoEconomics,* **5**(Supplement 1), pp.54-57.

EPP EVALUATION TEAM. 2005. *Process Evaluation of the EPP: Report II*. Manchester: National Primary Care Research and Development Centre, University of Manchester.

EPSTEIN, L. H., S. RAJA, T. O. DANIEL, R. A. PALUCH, D. E. WILFLEY, B. E. SAELENS and J. N. ROEMMICH. 2012. The built environment moderates effects of family-based childhood obesity treatment over 2 years. *Ann Behav Med.,* **44**(2), pp.248-258.

ERKELENZ, N., S. KOBEL, S. KETTNER, C. DRENOWATZ and J. M. STEINACKER. 2014. Parental Activity as Influence on Children`s BMI Percentiles and Physical Activity. *J Sports Sci Med,* **13**(3), pp.645-650.

ESPOSITO, M., B. GALLAI, M. ROCCELLA, R. MAROTTA, F. LAVANO, S. M. LAVANO, G. MAZZOTTA, D. BOVE, M. SORRENTINO, F. PRECENZANO and M. CAROTENUTO. 2014. Anxiety and depression levels in prepubertal obese children: a case-control study. *Neuropsychiatric Disease and Treatment,* **10**, pp.1897-1902.

EUBANK, B. H., N. G. MOHTADI, M. R. LAFAVE, J. P. WILEY, A. J. BOIS, R. S. BOORMAN and D. M. SHEPS. 2016. Using the modified Delphi method to establish clinical consensus for the diagnosis and treatment of patients with rotator cuff pathology. *BMC Medical Research Methodology,* **16**(1), p56.

EUROPEAN ASSOCIATION FOR THE STUDY OF OBESITY. 2016. *European Obesity Day announces parliamentary initiative* [online]. [Accessed 30th April]. Available from: http://easo.org/european-obesity-day-announces-parliamentary-initiative/.

EVANS, N., B. HANNIGAN, S. PRYJMACHUK, E. GILLEN, D. EDWARDS, M. LONGO, G. TRAINOR and F. HATHWAY. 2017. Using the nominal group technique to involve young people in an evidence synthesis which explored 'risk' in inpatient mental healthcare. *Research Involvement and Engagement,* **3**(16).

EWALD, H., J. KIRBY, K. REES and W. ROBERTSON. 2014. Parent-only interventions in the treatment of childhood obesity: a systematic review of randomized controlled trials. *Journal of Public Health,* **36**(3), pp.476-489.

EXPERT PANEL ON DETECTION, E., AND TREATMENT OF HIGH BLOOD CHOLESTEROL IN ADULTS,. 2001. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA,* **285**(19), pp.2486-2497.

EXPERT PANEL ON WOMEN'S IMAGING, P. V. PANDHARIPANDE, K. P. LOWRY, C. REINHOLD, M. ATRI, C. B. BENSON, P. R. BHOSALE, E. D. GREEN, S. K. KANG, Y. LAKHMAN, K. E. MATUREN, R. NICOLA, G. M. SALAZAR, T. D. SHIPP, L. SIMPSON, B. L. SUSSMAN, J. UYEDA, D. J. WALL, B. WHITCOMB, C. M. ZELOP and P. GLANC. 2017. ACR Appropriateness Criteria: Ovarian Cancer Screening. *Journal of the American College of Radiology,* **14**(11), pp.S490-S499.

EYLES, C., M. MOORE, N. SHERON, P. RODERICK, W. O'BRIEN and G. M. LEYDON. 2013. Acceptability of screening for early detection of liver disease in hazardous/harmful drinkers in primary care. *Br J Gen Pract.,* **33**(513), pp.e516-e522.

FAROOQI, I. and S. O'RAHILLY. 2000. Recent advances in the genetics of severe childhood obesity. *Arch. Dis. Child,* **83**(1), pp.31-34.

FAROOQI, I. S., J. M. KEOGH, G. S. YEO, E. J. LANK, T. CHEETHAM and S. O'RAHILLY. 2003. Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med.,* **348**(12), pp.1085-1095.

FAWCETT, K. A. and I. BARROSO. 2010. The genetics of obesity: FTO leads the way. *Trends in Genetics,* **26**(6), pp.266-274.

FEINSTEIN, A. R. 1970. The pre-therapeutic classification of comorbidity in chronic disease. *Journal of Chronic Diseases,* **23**, pp.455-468.

FELDMAN, W. 1990. How serious are the adverse effects of screening? *J Gen Intern Med.,* **5**(5 Suppl), pp.S50-53.

FEREDAY, J. and E. MUIR-COCHRANE. 2006. Demonstrating rigour using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods,* **5**(1), pp.80-92.

FIELD, M. J. and K. N. LOHR. 1992. *Guidlines for Clinical Practice: From Development to Use.* Washington, D.C.: National Academy Press.

FIELDEN, A. L., E. SILLENCE and L. LITTLE. 2011. Children's Understandings' of obesity, a thematic analysis. *Int J Qualitative Stud Health Well-being,* **6**, p7170.

FINEGOOD, D. T., T. D. N. MERTH and H. RUTTER. 2010. Implications of the Foresight Obesity System Map for Solutions to Childhood Obesity. *Obesity,* **18**(1), pp.13-16.

FINK, A., J. KOSECOFF, M. CHASSIN and R. H. BROOK. 1984. Consensus methods: characteristics and guidelines for use. *American Journal of Public Health,* **74**(9), pp.979-983.

FISHER, J., M. CABREL DE MELLO, V. PATEL, A. RAHMAN, T. TRAN, S. HOLTON and W. HOLMES. 2012. Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle income countries: a systematic review. *Bulletin of the World Health Organization,* **90**, pp.139-149.

FITCH, K., S. J. BERNSTEIN, M. D. AGUILAR, B. BURNAND, J. R. LACALLE, P. LAZARO, M. VAN HET LOO, J. MCDONNELL, J. P. VADER and J. P. KAHAAN. 2001. *The RAND/UCLA Appropriateness Method User's Manual. RAND Corporation*. Santa Monica, CA.

FLEGAL, K. M., B. K. KIT, H. ORPANA and E. AL. 2013. Association of All-Cause Mortality With Overweight and Obesity Using Standard Body Mass Index Categories: A Systematic Review and Meta-analysis. *JAMA,* **309**(1), pp.71-72.

FLETCHER, R. W. and S. W. FLETCHER. 2005. *Clinical Epidemiology: The Essentials.* 4th Edition ed. Baltimore, MD: Lippincott Williams & Williams.

FORD, D. A. 1975. Shang inquiry as an alternative to Delphi: Some experimental findings. *Technological Forecasting and Social Change,* **7**(2), pp.139-164.

FOX, M. D., A. AFROZE, I. J. STUDEBAKER, T. WEI and C. M. HELLMAN. 2010. The prevalence of elevated blood pressure among obese adolescents in a pediatric resident continuity clinic. *Journal - Oklahoma State Medical Association,* **103**(4-5), pp.111-4.

FRANKLIN, J., G. DENYER, K. S. STEINBECK, I. D. CATERSON and A. J. HILL. 2006. Obesity and risk of low self-esteem: a statewide survey of Australian children. *Pediatrics,* **118**(6), pp.2481-7.

FRASER, G. M., D. PILPEL, J. KOSECOFF and R. H. BROOK. 1994. Effect of Panel Composition on Appropriateness Ratings. *International Journal for Quality in Healthcare,* **6**(3), pp.251-255.

FRAYLING, T. M., N. J. TIMPSON, M. N. WEEDON, E. ZEGGINI, R. M. FREATHY, C. M. LINDGREN, J. R. PERRY, K. S. ELLIOTT, H. LANGO, N. W. RAYNER, B. SHIELDS, L. W. HARRIES, J. C. BARRETT, S. ELLARD, C. J. GROVES, B. KNIGHT, A. M. PATCH, A. R.

NESS, S. EBRAHIM, D. A. LAWLOR, S. M. RING, Y. BEN-SHLOMO, M. R. JARVELIN, U. SOVIO, A. J. BENNETT, D. MELZER, L. FERRUCCI, R. J. LOOS, I. BARROSO, N. J. WAREHAM, F. KARPE, K. R. OWEN, L. R. CARDON, M. WALKER, G. A. HITMAN, C. N. PALMER, A. S. DONEY, A. D. MORRIS, G. D. SMITH, A. T. HATTERSLEY and M. I. MCCARTHY. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science,* **316**(5826), pp.889-94.

FROOD, S., L. M. JOHNSTON, C. L. MATTESON and D. T. FINEGOOD. 2013. Obesity, Complexity and the role of the Health System. *Curr Obes Rep,* **2**, pp.320-326.

GABRIEL, D. 2013. *Inductive and Deductive approaches to research* [online]. [Accessed 6th March]. Available from: http://deborahgabriel.com/2013/03/17/inductive-and-deductive-approaches-to-research/.

GALLAGHER, M., T. HARES, J. SPENCER, C. BRADSHAW and I. WEBB. 1993. The nominal group technique: a research tool for general practice? *Fam Pract,* **10**(1), pp.76-81.

GARG, A. X., D. HACKAM and M. TONELLI. 2008. Systematic Review and Meta-anlysis: When one study is just not enough. *Clin J Am Soc Nephrol,* **3**, pp.253-260.

GARG, M. K., N. TANDON, R. K. MARWAHA and Y. SINGH. 2014. Evaluation of surrogate markers for insulin resistance for defining metabolic syndrome in urban Indian adolescents. *Indian Pediatrics,* **51**(4), pp.279-284.

GARRIGUET, D., R. COLLEY and T. BUSHNIK. 2017. Parent-Child association in physical activity and sedentary behaviour. *Health Rep,* **28**(6), pp.3-11.

GATES, D. M., P. SUCCOP, B. J. BREHM, G. L. GILLESPIE and B. D. SOMMERS. 2008. Obesity and Presenteeism: The Impact of Body Mass Index on Workplace Productivity. *Journal of Occupational and Environmental Medicine,* **50**(1), pp.39-45.

GATINEAU, M. and S. MATHRANI. 2011. *Obesity and ethnicity.* Oxford: National Obesity Observatory.

GEERTZ, C. 1978. Thick Description: Toward an Interpretive Theory of Culture. *In:* C. GEERTZ, ed. *The Interpretation of Cultures: Selected Essays.* New York, USA: Basic Books, Inc.

GELEIJNSE, M. L., B. J. KRENNING, B. M. VAN DALEN, A. NEMES, O. I. SOLIMAN, J. G. BOSCH, T. W. GALEMA, F. J. TEN CATE and E. BOERSMA. 2009. Factors affecting sensitivity and specificity of diagnostic testing: dobutamine stress echocardiography. *J Am Soc Echocardiogr.,* **22**(11), pp.1199-1208.

GEMMILL, A. W., B. LEIGH, J. ERICKSEN and J. MILGROM. 2006. A survey of the clinical acceptability of screening for postnatal depression in depressed and non-depressed women. *BMC Public Health,* **6**(211).

GENOVESI, S., L. ANTOLINI, M. GIUSSANI, P. BRAMBILLA, V. BARBIERI, S. GALBIATI, S. MASTRIANI, V. SALA, M. G. VALSECCHI and A. STELLA. 2010. Hypertension, Prehypertension, and Transient Elevated Blood Pressure in Children: Association With Weight Excess and Waist Circumference. *American Journal of Hypertension,* **23**(7), pp.756-761.

GIBSON, L. Y., S. M. BYRNE, E. BLAIR, E. A. DAVIS, P. JACOBY and Z. S.R. 2008. Clustering of psychosocial symptoms in overweight children. *Aust N Z J Psychiatry,* **42**(2), pp.118-125.

GILLIES, C. L., K. R. ABRAMS, P. C. LAMBERT, N. J. COOPER, A. J. SUTTON, R. T. HSU and K. KHUNTI. 2007. Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis. *BMJ,* **334**(299).

GILLISON, F., A. STATHI, P. REDDY, R. PERRY, G. TAYLOR, P. BENNETT, J. DUNBAR and C. GREAVES. 2015. Processes of behavior change and weight loss in a theory-based weight loss intervention program: a test of the process model for lifestyle behavior change. *Int J Behav Nutr Phys Act,* **12**(2).

GINSBERG, B. H. 2009. Factors affecting blood glucose monitoring: source of error in measurement. *Journal of Diabetes Science and Technology,* **3**(4), pp.903-913.

GLOBAL MEDIA INSIGHTS. 2018. *United Arab Emirates: Population Statistics (2018)* [online]. [Accessed 21th May 2018]. Available from: https://www.globalmediainsight.com/blog/uae-population-statistics/.

GODBER, G. E. 1975. *The Health Service Past, Present and Future.* London, UK: Athlone Press.

GOKLER, M. E., N. BUGRUL, A. O. SARı and S. METINTAS. 2018. The validity of self-reported vs. measured body weight and height and the effect of self-perception. *Arch Med Sci,* **14**(1), pp.174-181.

GOLAFSHANI, N. 2003. Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report,* **8**(4), pp.597-607.

GOLDSTEIN, B. I., B. BIRMAHER, D. A. AXELSON, T. R. GOLDSTEIN, C. ESPOSITO-SMYTHERS, M. A. STROBER, J. HUNT, H. LEONARD, M. K. GILL, S. IYENGAR, C. GRIMM, M. YANG, N. D. RYAN and M. B. KELLER. 2008. Preliminary findings regarding overweight and obesity in pediatric bipolar disorder. *J Clin Psychiatry,* **69**(12), pp.1953-1959.

GONG, C. D., Q. L. WU, Z. CHEN, D. ZHANG, Z. Y. ZHAO and Y. M. PENG. 2013. Glycolipid metabolic status of overweight/obese adolescents aged 9- to 15-year-old and the BMI-SDS/BMI cut-off value of predicting dyslipidemiain boys, Shanghai, China: a cross-sectional study. *Lipids in Health & Disease,* **12**, p129.

GONZÁLEZ-SERRANO, J., J. SERRANO, R. M. LÓPEZ-PINTOR, V. M. PAREDES, E. CASAÑAS and G. HERNÁNDEZ. 2016. Prevalence of Oral Mucosal Disorders in Diabetes Mellitus Patients Compared with a Control Group. *Journal of Diabetes Research,* **2016**, pp.1-11.

GORTMAKER, S. L., A. MUST, J. M. PERRIN, A. M. SOBOL and W. H. DIETZ. 1993. Social and Economic Consequences of Overweight in Adolescence and Young Adulthood. *N Engl J Med,* **329**, pp.1008-1012.

GOV. UK. n.d. *Population screening programmes* [online]. [Accessed 24th March]. Available from: https://www.gov.uk/topic/population-screening-programmes.

GOV.UK. 2013a. *Fetal anomaly screening: programme overview* [online]. [Accessed 29th April]. Available from: https://www.gov.uk/guidance/fetal-anomaly-screening-programme-overview.

GOV.UK. 2013b. *Newborn blood spot screening: programme overview* [online]. [Accessed 30th April]. Available from: https://www.gov.uk/guidance/newborn-blood-spot-screening-programme-overview.

GOV.UK. 2017. *Abdominal aortic aneurysm screening across the UK* [online]. [Accessed 12th March]. Available from: https://phescreening.blog.gov.uk/2017/01/20/abdominal-aortic-aneurysm-screening-across-the-uk/.

GOYDER, E., S. WILD, C. FISCHBACHER, J. CARLISLE and J. PETERS. 2008. Evaluating the impact of a national pilot screening programme for type 2 diabetes in deprived areas of England. *Fam Pract.,* **25**(5), pp.370-375.

GRAHAM, T., D. ROSE, J. MURRAY, M. ASHWORTH and A. TYLEE. 2014. User-generated quality standards for youth mental health in primary care: a participatory research design using mixed methods. *BMJ Quality and Safety,* **23**(10), pp.857-866.

GRANDORI, F. and M. E. LUTMAN. 1998. European Consensus Statement on Neonatal Hearing Screening Finalised at the European Consensus Development Conference on Neonatal Hearing Screening 15–16 May 1998, Milan, Italy. *Scandinavian Audiology,* **27**(4), pp.259-260.

GRAY, R. E., M. FITCH, C. DAVIS and C. PHILLIPS. 2000. Challenges of participatory research: reflections on a study with breast cancer self-help groups. *Health Expectations,* **3**, pp.243-252.

GREENBAUM, T. L. 1988. *The Practical Handbook and Guide to Focus Group Research.* Lexignton, MA: Lexington Books.

GREENLAND, S. and K. O'ROURKE. 2001. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics,* **2**(4), pp.463-471.

GRÉGOIRE, G., F. DERDERIAN and J. LE LORIER. 1995. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *Journal of Clinical Epidemiology,* **48**(1), pp.159-163.

GRIMM, P. 2010. Social Desirability Bias. *Wiley International Encyclopedia of Marketing.* John Wiley & Sons, Ltd.

GUEST, G., K. M. MACQUEEN and E. E. NAMEY. 2012. *Applied thematic analysis.* Thousand Oaks, California: Sage Publications Inc.

GUH, D., W. ZHANG, N. BANSBACK, Z. AMARSI, C. L. BIRMINGHAM and A. ANIS. 2009. The incidence of co-morbidities related to obesity and overweight: A systematic review and meta-analysis. *BMC Public Health,* **9**(1), p88.

GUNGOR, N., T. THOMPSON, K. SUTTON-TYRRELL, J. JANOSKY and S. ARSLANIAN. 2005. Early Signs of Cardiovascular Disease in Youth With Obesity and Type 2 Diabetes. *Diabetes Care,* **28**(5), pp.1219-1221.

GÜNGÖR, N. K. 2014. Overweight and Obesity in Children and Adolescents. *J Clin Res Pediatr Endocrinol,* **6**(3), pp.129-143.

GUNNARSDOTTIR, T., U. NJARDVIK, A. S. OLAFSDOTTIR, L. CRAIGHEAD and R. BJARNASON. 2012. Childhood obesity and co-morbid problems: effects of Epstein's family-based behavioural treatment in an Icelandic sample. *Journal of Evaluation in Clinical Practice,* **18**(2), pp.465-472.

GUO, X., Y. LI, G. SUN, Y. YANG, L. ZHENG, X. ZHANG, Z. SUN, H. MA, N. WANG, M. JIANG, J. LI and Y. SUN. 2012. Prehypertension in children and adolescents: association with body weight and neck circumference. *Internal Medicine,* **51**(1), pp.23-7.

GUPTA, R., M. SHARMA, N. K. GOYAL, P. BANSAL, S. LODHA and K. K. SHARMA. 2016. Gender differences in 7 years trends in cholesterol lipoproteins and lipids in India: Insights from a hospital database. *Indian J Endocrinol Metab,* **20**(2), pp.211-218.

GUSTAFSON, S. L. and R. E. RHODES. 2006. Parental correlates of physical activity in children and early adolescents. *Sports Medicine,* **36**(1), pp.79-97.

HAIDICH, A. B. 2010. Meta-analysis in medical research. *Hippokratia,* **14**(1), pp.29-37.

HAN, E., E. C. NORTON and S. C. STEARNS. 2009. Weight and wages: fat versus lean paychecks. *Health Economics,* **18**(5), pp.535-548.

HAN, J. C., D. A. LAWLOR and S. Y. KIMM. 2010. Childhood obesity – 2010: Progress and Challenges. *Lancet,* **375**(9727), pp.1737-1348.

HANEKOM, S., H. VAN ASWEGEN, N. PLANI and S. PATMAN. 2015. Developing minimum clinical standards for physiotherapy in South African intensive care units: the nominal group technique in action. *J Eval Clin Pract.,* **21**(1), pp.118-127.

HANNON, T. S., G. RAO and S. A. ARSLANIAN. 2005. Childhood obesity and type 2 diabetes mellitus. *Pediatrics,* **116**, pp.473-480.

HARRIS, J. M. and R. SHIPTSOVA. 2007. Consumer Demand for Convenience Foods: Demographics and Expenditures. *Journal of Food Distribution Research,* **38**(3), pp.22-36.

HARRIS, R., G. F. SAWAYA, V. A. MOYER and N. CALONGE. 2011. Reconsidering the Criteria for Evaluating Proposed Screening Programs: Reflections From 4 Current and Former Members of the U.S. Preventive Services Task Force. *Epidemiologic Reviews,* **33**, pp.20-35.

HARTLING, L., A. MILNE, M. P. HAMM, B. VANDERMEER, M. ANSARI, A. TSERTSVADZE and D. M. DRYDEN. 2013. Testing the Newcastle Ottawa Scale showed low reliability between individual reviewers. *J Clin Epidemiol,* **66**, pp.982-993.

HAWTHORNE, V. M. 1964. Mass radiography in an ecological approach to respiratory disease. *Scott Med J.,* **9**, pp.115-124.

HEALTH & SOCIAL CARE INFORMATION CENTRE. 2011. *Statistics on Obesity, Physical Activity and Diet: England, 2011* [online]. [Accessed 1st December]. Available from: http://www.ic.nhs.uk/pubs/opad11.

HEALY, M. and C. PERRY. 2000. Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research: An International Journal,* **3**(3), p118/126.

HECKMAN, J. J. 2010. Selection Bias and Self-Selection. *In:* S. N. DURLAUF and L. E. BLUME, eds. *Microeconometrics.* London: Palgrave Macmillan UK, pp.242-266.

HEDGES, L. V. and J. L. VEVEA. 1998. Fixed- and Random-Effects Models in Meta-Analysis. *Psychological Methods,* **3**(4), pp.486-504.

HERMONT, A. P., P. A. D. OLIVEIRA, C. C. MARTINS, S. M. PAIVA, I. A. PORDEUS and S. M. AUAD. 2014. Tooth Erosion and Eating Disorders: A Systematic Review and Meta-Analysis. *PLos One,* **9**(11), pe111123.

HERSEY, J. C., O. KHAVJOU, L. B. STRANGE, R. L. ATKINSON, S. N. BLAIR, S. CAMPBELL, C. L. HOBBS, B. KELLY, T. M. FITZGERALD, J. KISH-DOTO, M. A. KOCH, B. MUNOZ, E. PEELE, J. STOCKDALE, C. AUGUSTINE, GLENDAMITCHELL, D. ARDAY, J. KUGLER and M. BRITT. 2012. The efficacy and cost-effectiveness of a community weight management intervention: A randomized controlled trial of the health weight management demonstration. *Preventive Medicine,* **54**(1), pp.42-49.

HICKS, N. R. 1994. Some observations on attempts to measure appropriateness of care. *BMJ,* **309**, p730.

HIGGINS, J. P. and S. GREEN. 2011. *Cochrane Handbook for Systematic Reviews of Interventions* [online]. [Accessed 22nd February]. Available from: http://handbook.cochrane.org/.

HIGGINS, J. P. T., D. G. ALTMAN, P. C. GØTZSCHE, P. JÜNI, D. MOHER, A. D. OXMAN, J. SAVOVIĆ, K. F. SCHULZ, L. WEEKS and J. A. C. STERNE. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ,* **343**(d5928).

HIGGINS, J. P. T., S. G. THOMPSON, J. J. DEEKS and D. ALTMAN, G,. 2003. Measuring inconsistency in meta-analyses. *BMJ,* **327**(7414), pp.557-560.

HIGGINS, J. P. T., S. G. THOMPSON and D. J. SPIEGELHALTER. 2009. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society. Series A, (Statistics in Society,* **172**(1), pp.137-159.

HILIGSMANN, M., C. VAN DURME, P. GEUSENS, B. G. DELLAERT, C. D. DIRKSEN, T. VAN DER WEIJDEN, J. Y. REGINSTER and B. A. 2013. Nominal group technique to select attributes for discrete choice experiments: an example for drug treatment choice in osteoporosis. *Patient Preference and Adherence,* **7**, pp.133-139.

HINDORFF, L. A., P. SETHUPATHY, H. A. JUNKINS, E. M. RAMOS, J. P. MEHTA, F. S. COLLINS and T. A. MANOLIO. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A,* **106**(23), pp.9362-9367.

HINGLE, M. and D. KUNKEL. 2012. Childhood obesity and the media. *Pediatr Clin North Am.,* **59**(3), pp.677-692.

HOLLAND, W. W. and S. STEWART. 2005. *Screening in Disease Prevention: What Works?* Oxon, UK: Radcliffe Publishing Ltd.

HOLLOWAY, I. and L. TODRES. 2003. The Status of Method: Flexibility, Consistency and Coherence. *Qualitative Research,* **3**(2), pp.345-357.

HOLTON, G. 1973. *Thematic Origins of Scientific Thought: Kepler to Einstein.* Cambridge, MA: Harvard University Press.

HOOTMAN, J. M., J. B. DRIBAN, M. R. SITLER, K. P. HARRIS and N. M. CATTANO. 2011. Reliability and validity of three quality rating instruments for systematic reviews of observational studies. *Research Synthesis Methods,* **2**(2), pp.110-118.

HOPKINS, A. 1993. What do we mean by appropriate health care. *Quality in Health Care,* **2**, pp.117-123.

HOSSAIN, L. N., F. FERNANDEZ-LLIMOS, T. LUCKETT, J. C. MOULLIN, D. DURKS, L. FRANCO-TRIGO, S. I. BENRIMOJ and D. SABATER-HERNÁNDEZ. 2017. Qualitative meta-synthesis of barriers and facilitators that influence the implementation of community pharmacy services: perspectives of patients, nurses and general medical practitioners. *BMJ Open* **7**(9), pe015471.

HRUBY, A. and F. B. HU. 2015. The Epidemiology of Obesity: A Big Picture. *Pharmacoeconomic,* **33**(7), pp.673-689.

HSIA, D. S., S. C. FALLON and M. L. BRANDT. 2012. Adolescent Bariatric Surgery. *JAMA Pediatrics,* **166**(8), pp.757-766.

HSIEH, H. F. and S. E. SHANNON. 2005. Three approaches to qualitative content analysis. *Qualitative Health Research,* **15**(9), pp.1277-1288.

HSU, C. C. and B. A. SANDFORD. 2007. The Delphi Technique: Making Sense of Consensus. *Practical Assessment, Research and Evaluation,* **12**(10).

HUDDA, M. T., C. M. NIGHTINGALE, A. S. DONIN, C. G. OWEN, A. R. RUDNICKA, J. C. K. WELLS, H. RUTTER, D. G. COOK and P. H. WHINCUP. 2017. Reassessing patterns of childhood body-mass index, overweight, and obesity in South Asian and black participants in the English National Child Measurement Programme: use of ethnicity-specific body-mass index adjustments. *Lancet,* **390**(S10).

HUEDO-MEDINA, T. B., J. SÁNCHEZ-MECA, F. MARÍN-MARTÍNEZ and J. BOTELLA. 2006. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods,* **11**(2), pp.193-206.

HUMPHERY-MURTO, S., L. VARPIO, T. J. WOOD, C. GONSAVES, L. A. UFHOLZ, K. MASCIOLI, C. WANG and T. FOTH. 2017. The Use of the Delphi and Other Consensus Group Methods in Medical Education Research: A Review. *Academic Medicine*.

HUMPHREY-MURTO, S., L. VARPIO, C. GONSALVES and T. J. WOOD. 2017. Using consensus group methods such as Delphi and Nominal Group in medical education research. *Medical Teacher,* **39**(1), pp.14-19.

HUNTER, D. J., C. M. MCKEE, C. F. SANDERSON and N. A. BLACK. 1994. Appropriate indications for prostatectomy in the UK--results of a consensus panel. *J Epidemiol Community Health,* **48**(1), pp.58-64.

HUTCHINGS, A. and R. RAINE. 2006. A systematic review of factors affecting the judgments produced by formal consensus development methods in health care. *Journal of Health Services & Research Policy,* **11**(3), pp.172-179.

HUTCHINGS, A., R. RAINE, C. SANDERSON and N. BLACK. 2006. A comparison of formal consensus methods used for developing clinical guidelines. *Journal of Health Services & Research Policy,* **11**(4), pp.218-224.

INGE, T. H., W. C. KING, T. M. JENKINS, A. P. COURCOULAS, M. MITSNEFES, D. R. FLUM, B. M. WOLFE, A. POMP, G. F. DAKIN, S. KHANDELWAL, M. H. ZELLER, M. HORLICK, J. R. PENDER, J. CHEN and S. R. DANIELS. 2013. The Effect of Obesity in Adolescence on Adult Health Status. *Pediatrics,* **132**(6), pp.1098-104.

INTERNATIONAL DIABETES FEDERATION. 2017. *Diabetes Atlas* [online]. [Accessed 7th May 2018]. Available from: http://diabetesatlas.org/resources/2017-atlas.html.

IQBAL, S. and L. PIPON-YOUNG. 2009. The Delphi method. *The Psychologist,* **22**, pp.598-601.

JAGO, R., E. SOLOMON-MOORE, C. MACDONALD-WALLIS, J. L. THOMPSON, D. A. LAWLOR and S. J. SEBIRE. 2017. Association of parents' and children's physical activity and sedentary time in Year 4 (8–9) and change between Year 1 (5–6) and Year 4: a

longitudinal study. *International Journal of Behavioral Nutrition and Physical Activity,* **14**(110).

JAMES, P. T., R. LEACH, E. KALAMARA and M. SHAYEGHI. 2001. The worldwide obesity epidemic. *Obesity Research,* **9**(5), pp.228-233.

JANSSEN, X., K. D. MANN, L. BASTERFIELD, K. N. PARKINSON, M. S. PEARCE, J. K. REILLY, A. J. ADAMSON and J. J. REILLY. 2016. Development of sedentary behavior across childhood and adolescence: longitudinal analysis of the Gateshead Millennium Study. *Int J Behav Nutr Phys Act.,* **13**(88).

JEBB, S. 2017. Dusting off Foresight's Obesity Report [online]. [Accessed 2nd Feb 2019]. Available from: https://foresightprojects.blog.gov.uk/2017/10/04/dusting-off-foresights-obesity-report/

JEPSON, R., D. WELLER, F. ALEXANDER and J. WALKER. 2005. Impact of UK colorectal cancer screening pilot on primary care. *Br J Gen Pract.,* **55**(510), pp.20-25.

JONES, H. and B. C. TWISS. 1978. *Forecasting technology for planning decisions.* London: The MacMillan Press, Ltd.

JONES, J. and D. HUNTER. 1995. Qualitative Research: Consensus Methods for Medical and Health Services Research. *BMJ,* **311**, p376.

JONES, J. E., L. L. JONES, T. J. H. KEELEY, M. J. CALVERT and M. J. 2017. A review of patient and carer participation and the use of qualitative research in the development of core outcome sets. *Plos ONE,* **12**(3), pe0172937.

JONES, S. C. 2004. Using the Nominal Group Technique to Select the Most Appropriate Topics for Postgraduate Research Students' Seminars. *Journal of University Teaching and Learning Practice,* **1**(1), pp.20-34.

JUNI, P., F. HOLENSTEIN, J. STERNE, C. BARTLETT and M. EGGER. 2002. Direction and impact of language bias in meta-analyses of controlled trials: Empirical Study. *Int J Epidemiol,* **31**, pp.115-123.

JÜNI, P., A. WITSCHI, R. BLOCH and M. EGGER. 1999. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA,* **282**(11), pp.1054-1060.

JUONALA, M., C. G. MAGNUSSEN, G. S. BERENSON, A. VENN, T. L. BURNS, M. A. SABIN, S. R. SRINIVASAN, S. R. DANIELS, P. H. DAVIS, W. CHEN, C. SUN, M. CHEUNG, J. S. A. VIIKARI, T. DWYER and O. T. RAITAKARI. 2011. Childhood Adiposity, Adult Adiposity, and Cardiovascular Risk Factors. *New England Journal of Medicine,* **365**, pp.1876-1885.

JUSOT, F., M. KHLAT, T. ROCHEREAU and C. SERME. 2008. Job loss from poor health, smoking and obesity: a national prospective survey in France. *J Epidemiol Community Health,* **62**, pp.332-337.

KADAM, U. T., K. JORDAN and P. R. CROFT. 2006. A comparison of two consensus methods for classifying morbidities in a single professional group showed the same outcomes. *Journal of Clinical Epidemiology,* **59**, pp.1169-1173.

KAJBAF, T. Z., S. ASAR and M. R. ALIPOOR. 2011. Relationship between obesity and asthma symptoms among children in Ahvaz, Iran: a cross sectional study. *Italian Journal of Pediatrics,* **37**, p1.

KARK, M. and N. KARNEHED. 2012. Weight status at age 18 influences marriage prospects. A population-based study of Swedish men. *BMC Public Health,* **12**, p833.

KARNATH, B. M. 2002. Sources of Error in Blood Pressure Measurement. *Hospital Physician,* **March**, pp.33-37.

KARRA, E., O. G. O'DALY, A. I. CHOUDHURY, A. YOUSSEIF, S. MILLERSHIP, M. T. NEARY, W. R. SCOTT, K. CHANDARANA, S. MANNING, M. E. HESS, H. IWAKURA, T. AKAMIZU, Q. MILLET, C. GELEGEN, M. E. DREW, S. RAHMAN, J. J. EMMANUEL, S. C. WILLIAMS, U. U. RÜTHER, J. C. BRÜNING, D. J. WITHERS, F. O. ZELAYA and R. L. BATTERHAM. 2013. A link between FTO, ghrelin, and impaired brain food-cue responsivity. *J. Clin. Invest.,* **123**(8), pp.3539-51.

KASSI, E., P. PERVANIDOU, G. KALTSAS and G. CHROUSOS. 2011. Metabolic syndrome: definitions and controversies. *BMC Medicine,* **9**, pp.48-60.

KATIKIREDDI, S. V., M. EGAN and M. PETTICREW. 2015. How do sysmtematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. *J Epidemiol Community Health,* **69**, pp.189-195.

KEARNEY-MITCHELL, P. I., K. M. MILSOM, A. S. BLINKHORN and M. TICKLE. 2006. Referral criteria for school dental screening. *British Dental Journal,* **200**(9), p501.

KELLY, T., W. YANG, C. S. CHEN, K. REYNOLDS and J. HE. 2008. Global burden of obesity in 2005 and projections to 2030. *Int J Obes,* **32**(9), pp.1431-1437.

KENNEDY, C., S. KOOLS and R. KRUEGER. 2001. Methodological Considerations in Children's Focus Groups. *Nursing Research,* **50**(3), pp.184-187.

KESSLER, D., D. SHARP and G. LEWIS. 2005. Screening for depression in primary care. *Br J Gen Pract.,* **55**(518), pp.659-660.

KHADER, Y. S., A. BATIEHA, H. JADDOU, Z. BATIEHA, M. EL-KHATEEB and K. AJLOUNI. 2011. Metabolic abnormalities associated with obesity in children and adolescents in Jordan. *International Journal of Pediatric Obesity,* **6**(3-4), pp.215-22.

KHODYAKOV, D., S. HEMPEL, L. RUBENSTEIN, P. SHEKELLE, R. FOY, S. SALEM-SCHATZ, S. O'NEILL, M. DANZ and S. DALAL. 2011. Conducting Online Expert panels: a feasibility and experimental replicability study. *BMC Medical Research Methodology,* **11**, p174.

KIELLY, J., D. V. KELLY, C. HUGHES, K. DAY, S. HANCOCK, S. ASGHARI, J. GAHAGAN, C. MARRA and H. NGUYEN. 2018. Adaptation of POCT for pharmacies to reduce risk and optimize access to care in HIV, the APPROACH study protocol: examining acceptability and feasibility. *Pilot and Feasibility Studies,* **4**(59).

KING, N. 2004. Using templates in the thematic analysis of text. *In:* C. CASSELL and G. SYMON, eds. *Essential guide to qualitative methods in organisational research.* London, UK: Sage., pp.256-270.

KIRK, J. K. and J. STEGNER. 2010. Self-Monitoring of Blood Glucose: Practical Aspects. *Journal of Diabetes Science and Technology,* **4**(2), pp.435-439.

KIRKØEN, B., P. BERSTAD, E. BOTTERI, E. DALÉN, J. A. NILSEN, G. HOFF, T. DE LANGE and T. BERNKLEV. 2017. Acceptability of two colorectal cancer screening tests: pain as a key determinant in sigmoidoscopy. *Endoscopy.,* **49**(11), pp.1075-1086.

KITCHENER, H. C., K. CANFELL, C. GILHAM, A. SARGENT, C. ROBERTS, M. DESAI and J. PETO. 2014. The clinical effectiveness and cost-effectiveness of primary human papillomavirus cervical screening in England: extended follow-up of the ARTISTIC randomised trial cohort through three screening rounds. *Health Technology Assessment, No. 18.23,* (Appendix 4).

KLASSEN, T. P., A. R. JADAD and D. MOHER. 1998. Guides for Reading and Interpreting Systematic Reviews. 1. Getting Started. *Arch Pediatr Adolesc Med,* **152**, pp.700-704.

KODAMA, S., C. HORIKAWA, K. FUJIHARA, S. YOSHIZAWA, Y. YACHI, S. TANAKA, N. OHARA, S. MATSUNAGA, T. YAMADA, O. HANYU and H. SONE. 2014. Quantitative relationship between body weight gain in adulthood and incident type 2 diabetes: a meta-analysis. *Obes Rev,* **15**(3), pp.202-214.

KOEBNICK, C., K. J. COLEMAN, M. H. BLACK, N. SMITH, J. K. DER-SARKISSIAN, S. J. JACOBSEN and A. H. PORTER. 2012a. Cohort Profile: The KPSC Children's Health Study, a population-based study of 920 000 children and adolescents in southern California. *Int J Epidemiol,* **41**(3), pp.627-633.

KOEBNICK, C., N. SMITH, M. H. BLACK, A. H. PORTER, B. A. RICHIE, S. HUDSON, D. GILILLAND, S. J. JACOBSEN and G. F. LONGSTRETH. 2012b. Pediatric Obesity and Gallstone Disease. *Journal of Pediatric Gastroenterology and Nutrition,* **55**(3), pp.328-333.

KOSLOFF, S., S. IRISH, L. PERREAULT, G. ANDERSON and A. NOTTBOHM. 2017. Assessing relationships between conformity and meta-traits in an Asch-like paradigm. *Social Influence,* **12**(2-3), pp.90-100.

KRAL, T. V. E. and M. S. FAITH. 2009. Influences on Child Eating and Weight Development from a Behavioral Genetics Perspective. *Journal of Pediatric Psychology,* **34**(6), pp.596-605.

KREFTING, L. 1991. Rigor in Qualitative Research: The assessment of Trustworthiness. *The American Journal of Occupational Therapy,* **45**(3), pp.214-222.

KRIEGE, M., C. T. BREKELMANS, I. M. OBDEIJN, C. BOETES, H. M. ZONDERLAND, S. H. MULLER, T. KOK, R. A. MANOLIU, A. P. BESNARD, M. M. TILANUS-LINTHORST, C. SEYNAEVE, C. C. BARTELS, R. KAAS, S. MEIJER, J. C. OOSTERWIJK, N. HOOGERBRUGGE, R. A. TOLLENAAR, E. J. RUTGERS, H. J. DE KONING and J. G. KLIJN. 2006. Factors affecting sensitivity and specificity of screening mammography and MRI in women with an inherited risk for breast cancer. *Breast Cancer Res Treat,* **100**(1), pp.109-19.

KUMANYIKA, S. K. 2008. Environmental influences on childhood obesity: ethnic and cultural influences in context. *Physiol Behav. ,* **94**(1), pp.61-70.

KUNTZ, B. and T. LAMPERT. 2010. Socioeconomic Factors and Obesity. *Dtsch Arztebl Int.,* **107**(30), pp.517-522.

KWON, H. L., B. ORTIZ, R. SWANER, K. SHOEMAKER, B. JEAN-LOUIS, M. E. NORTHRIDGE, R. D. VAUGHAN, T. MARX, A. GOODMAN, L. N. BORRELL and S. W. NICHOLAS. 2006. Childhood asthma and extreme values of body mass index: The Harlem Children's Zone Asthma Initiative. *Journal of Urban Health,* **83**(3), pp.421-433.

LALKHEN, A. G. and A. MCCLUSKEY. 2008. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain,* **8**(61), pp.221-223.

LAMBERT, S. and E. O'HALLORAN. 2008. Deductive thematic analysis of a female paedophilia website. *Psychiatry, Psychology and Law,* **15**(2), pp.284-300.

LANDIS, J. R. and G. G. KOCH. 1977. The measurement of observer agreement for categorical data. *Biometrics,* **33**, pp.159-174.

LANG, J. E., J. HOSSAIN, K. SMITH and J. J. LIMA. 2012. Asthma severity, exacerbation risk, and controller treatment burden in underweight and obese children. *Journal of Asthma,* **49**(5), pp.456-463.

LAWLOR, D. A., M. CALLAWAY, C. MACDONALD-WALLIS, E. ANDERSON, A. FRASER, L. D. HOWE, C. DAY and N. SATTAR. 2014. Nonalcoholic Fatty Liver Disease, Liver Fibrosis, and Cardiometabolic Risk Factors in Adolescence: A Cross-Sectional Study of 1874 General Population Adolescents. *Journal of Clinical Endocrinology & Metabolism,* **99**(3), pp.E410-E417.

LAWSON, E. H., M. M. GIBBONS, C. Y. KO and P. G. SHEKELLE. 2012. The appropriateness method has acceptable reliability and validity for assessing overuse and underuse of surgical procedures. *Journal of Clinical Endocrinology,* **65**, pp.1133-1143.

LEBLANC, M. È., L. CLOUTIER and P. POIRIER. 2015. Sensitivity, specificity, and predictive values of a forearm blood pressure measurement method in severe obesity. *Blood Press Monit,* **20**(2), pp.79-82.

LEEFLANG, M. M. G., A. W. S. RUTJES, J. B. REITSMA, L. HOOFT and P. M. M. BOSSUYT. 2013. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ,* **185**(11), pp.E537–E544.

LEWIS, J. and J. RITCHIE. 2003. Generalising from qualitative research. *In:* J. RITCHIE and J. LEWIS, eds. *Qualitative research in practice: A guide for social science students and researchers.* London, UK.: Sage.

LI, Y., X. YANG, F. ZHAI, F. J. KOK, W. ZHAO, J. PIAO, J. ZHANG, Z. CUI and G. MA. 2008. Prevalence of the metabolic syndrome in Chinese adolescents. *British Journal of Nutrition,* **99**(3), pp.565-570.

LILES, E. G., J. L. SCHNEIDER, A. C. FELDSTEIN, D. M. MOSEN, N. PERRIN, A. G. ROSALES and D. H. SMITH. 2015. Implementation challenges and successes of a population-based colorectal cancer screening program: a qualitative study of stakeholder perspectives. *Implement Sci,* **10**(41).

LINCOLN, Y. S. and E. G. GUBA. 1985. *Naturalistic Inquiry.* Thousand Oaks, CA: Sage Publications.

LINSTONE, H. A. and M. TUROFF. 1975. *The Delphi Method: Techniques and applications.* Reading, MA: Addison-Wesley Pub Co.

LO, C. K., D. MERTZ and M. LOEB. 2014. Newcastle-Ottawa Scale: Comparing reviewers' to authors' assessments. *BMC Med Res Methodol,* **14**, pp.45-49.

LOCKE, J. E., H. QU, R. SHEWCHUK, R. B. MANNON, R. GASTON, D. L. SEGEV, E. C. MANNON and M. Y. MARTIN. 2015. Identification of strategies to facilitate organ donation among African Americans using the nominal group technique. *Clin J Am Soc Nephrol.,* **10**(2), pp.286-293.

LOTH, K. A., J. MOND, M. WALL and D. NEUMARK-SZTAINER. 2011. Weight Status and Emotional Well-Being. *Journal of Pediatric Psychology,* **36**, pp.216-225.

LOUTHAN, M. V., J. A. THERIOT, E. ZIMMERMAN, J. T. STUTTS and C. J. MCCLAIN. 2005. Decreased prevalence of nonalcoholic fatty liver disease in black obese children. *Journal of Pediatric Gastroenterology & Nutrition,* **41**(4), pp.426-9.

LOVEMAN, E., L. AL-KHUDAIRY, R. E. JOHNSON, W. ROBERTSON, J. L. COLQUITT, E. L. MEAD, L. J. ELLS, M. I. METZENDORF and K. REES. 2015. Parent-only interventions for childhood overweight or obesity in children aged 5 to 11 years. *Cochrane Database Syst Rev. ,* **12**(CD012008).

LUBRANO-BERTHELIER, C., M. CAVAZOS, B. DUBERN, A. SHAPIRO, C. L. STUNFF, S. ZHANG, F. PICART, C. GOVAERTS, P. FROGUEL, P. BOUGNERES, K. CLEMENT and C. VAISSE. 2003. Molecular genetics of human obesity-associated MC4R mutations. *Ann N Y Acad Sci,* **994**, pp.49-57.

LUPPINO, F. S., L. M. DE WIT, P. F. BOUVY, T. STIJNEN, P. CUIJPERS, B. W. J. H. PENNINX and F. G. ZITMAN. 2010. Overweight, Obesity, and Depression: A Systematic Review and Meta-analysis of Longitudinal Studies. *Arch Gen Psychiatry,* **67**(3), pp.220-229.

MACCOBY, E. E. 1998. *The two sexes: growing apart, coming together.* Cambridge, MA: Belknap Press.

MACKIESON, P., A. SHLONSKY and M. CONNOLLY. 2018. Increasing rigor and reducing bias in qualitative research: A document analysis of parliamentary debates using applied thematic analysis. *Qualitative Social Work,* **0**(0), pp.1-16.

MACKINNON III, G. E. and L. SWANSON. 2002. Treatment Options for Adult Obesity. *Internet Continuing Education* **6**(1).

MACLEAN, A., H. SWEETING, L. WALKER, C. PATTERSON, U. RÄISÄNEN and K. HUNT. 2015. "It's not healthy and it's decidedly not masculine": a media analysis of UK newspaper representations of eating disorders in males. *BMJOpen,* **5**(e007468).

MACMAHON, B. and D. TRICHOPOULOS. 1996. *Epidemiology: Principles and Methods.* 2nd Edition ed. Boston, MA: Little, Brown and Company.

MADILL, A. J., A. and C. SHIRLEY. 2000. Objectivity and reliability in qualitative analysis: Realist, contextualist and radical constructionist epistemologies. *British Journal of Psychology,* **91**(1), pp.1-20.

MAGGIO, A. B. R., X. E. MARTIN, C. S. GASSER, C. GAL-DUDING, M. BEGHETTI, N. J. FARPOUR-LAMBERT and C. CHAMAY-WEBER. 2014. Medical and non-medical complications among children and adolescents with excessive body weight. *BMC Pediatrics,* **14**, p232.

MAGUIRE, M. and B. DELAHUNT. 2017. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Teaching and Learning in Higher Education,* (3), pp.3351-33514.

MAINE, A., A. DICKSONB, M. TRUESDALEA and B. BROWN. 2017. An application of Bandura's 'Four Sources of Self-Efficacy' to the self-management of type 2 diabetes in people with intellectual disability: An inductive and deductive thematic analysis. *Research in Developmental Disabilities,* **70**, pp.75-84.

MALTERUD, K. 2001. Qualitative research: standards, challenges, and guidelines. *The Lancet,* **358**(9290), pp.483-488.

MANIOS, Y., G. MOSCHONIS, G. KOURLABA, Z. BOULOUBASI, E. GRAMMATIKAKI, A. SPYRIDAKI, C. HATZIS, A. KAFATOS and G. A. FRAGIADAKIS. 2008. Prevalence and independent predictors of insulin resistance in children from Crete, Greece: the Children Study. *Diabetic Medicine,* **25**(1), pp.65-72.

MANN, C. J. 2003. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J,* **20**, pp.54-60.

MANSOOR, K., S. SABA SHAHNAWAZ, M. RASOOL, H. CHAUDHRY, G. AHUJA and S. SHAHNAWAZ. 2016. Automated Versus Manual Blood Pressure Measurement: A Randomized Crossover Trial in the Emergency Department of a Tertiary Care Hospital in Karachi, Pakistan: Are Third World Countries Ready for the Change? *J Med Sci,* **4**(3), pp.404-409.

MARKER, K. 2016. *Sugar-Sweetened Beverages Increase Risk for Hyperuricemia* [online]. [Accessed 21st December]. Available from: https://transforming-science.com/featured-research/sugar-sweetened-beverages-hyperuricemia/.

MARSHALL, C. and G. B. ROSSMAN. 1999. *Designing Qualitative Research.* 3rd ed. Thousand Oaks, CA: Sage.

MARTIN, J. L., V. PÉREZ, M. SACRISTÁN and E. ALVAREZ. 2005. Is grey literature essential for a better control of publication bias in psychiatry? An example from three meta-analyses of schizophrenia. *European Psychiatry,* **20**(8), pp.550-553.

MARTIN, P. 2003. The epidemiology of anxiety disorders: a review. *Dialogues Clin Neurosci. ,* **5**(3), pp.281-298.

MAXIM, L. D., R. NIEBO and M. J. UTELL. 2014. Screening tests: a review with examples. *Inhal Toxicol.,* **26**(13), pp.811-828.

MCGLYNN, E. A., J. KOSECOFF and R. H. BROOK. 1990. Format and Conduct of Consensus Development Conferences: A Multination Comparison. *In:* C. GOODMAN and S. R. BARATZ, eds. *Improving Consensus Development for Health Technology Assessment: An International Perspective.* Washington D.C.: National Academies Press.

MCKEOWN, T. 1968. *Screening in Medical Care: reviewing the evidence.* Oxford: Oxford University Press for hte Nuffield Provincial Hospitals Trust.

MCLAUGHLIN, H. 2010. Keeping Service User Involvement in Research Honest. *The British Journal of Social Work,* **40**(5), pp.1591-1608.

MCMANUS, S., P. BEBBINGTON, R. JENKINS and T. BRUGHA. 2016. *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014.* [online]. [Accessed 2nd January]. Available from: http://digital.nhs.uk/media/32987/APMS-2014-Full-Report/pdf/Mental_health_and_wellbeing_in_England_full_report.

MCMILLAN, S. S., F. KELLY, A. SAV, E. KENDALL, M. A. KING, J. A. WHITTY and A. J. WHEELER. 2014. Using the Nominal Group Technique: how to analyse across multiple groups. *Health Services and Outcomes Research Methodology,* **14**(3), pp.92-108.

MCMILLAN, S. S., M. KING and M. P. TULLY. 2016. How to use the nominal group and Delphi techniques. *Int J Clin Pharm,* **38**, pp.655-662.

MCNAUGHTON, R. J., N. T. A. OSWALD, J. S. SHUCKSMITH, P. J. HEYWOOD and P. S. WATSON. 2011. Making a success of providing NHS Health Checks in community pharmacies across the Tees Valley: a qualitative study. *BMC Health Services Research,* **11**(222).

MEAD, E., G. ATKINSON, B. RICHTER, M. METZENDORF, L. BAUR, N. FINER, E. CORPELEIJN, C. O'MALLEY and L. J. ELLS. 2016. Drug interventions for the treatment of obesity in children and adolescents. *Cochrane Database Syst Rev.,* **11**(CD012436).

MEDCALC SOFTWARE. 2017. *MedCalc - easy-to-use statistical software* [online]. [Accessed 19th October]. Available from: https://www.medcalc.org/index.php.

MEHAIRI, A. E., A. A. KHOURI, M. M. NAQBI, S. J. MUHAIRI, F. A. MASKARI, N. NAGELKERKE and S. M. SHAH. 2013. Metabolic syndrome among Emirati adolescents: a school-based study. *PLoS ONE [Electronic Resource],* **8**(2), pe56159.

MENTAL HEALTH FOUNDATION. 2016. *Fundamental Facts About Mental Health* [online]. [Accessed 2nd January]. Available from: https://www.mentalhealth.org.uk/file/2518/download?token=RU6lzcSK.

MERGEN, M., H. MERGEN, M. OZATA, R. ONER and C. ONER. 2001. A novel melanocortin 4 receptor (MC4R) gene mutation associated with morbid obesity. *J Clin Endocrinol Metab.,* **86**(7), pp.3448-51.

MERTON, R. K. 1975. Thematic Analysis on Science: Notes on Holton's Concept. *Science,* **188**(4186), pp.335-338.

MESSIAH, S. E., K. L. ARHEART, B. LUKE, S. E. LIPSHULTZ and T. L. MILLER. 2008. Relationship between body mass index and metabolic syndrome risk factors among US 8- to 14-year-olds, 1999 to 2002. *Journal of Pediatrics,* **153**(2), pp.215-21.

MILES, M. B. and A. M. HUBERMAN. 1994. *An Expanded Sourcebook: Qualitative Data Analysis.* Thousand Oaks, California: Sage Publications.

MINOGUE, V., J. BONESS, A. BROWN and J. GIRDLESTONE. 2005. The impact of service user involvement in research. *International Journal of health Care Quality Assurance,* **18**(2), pp.103-112.

MOEBUS, S., L. GÖRES, C. LÖSCH and K. JÖCKEL. 2011. Impact of time since last caloric intake on blood glucose levels. *Eur J Epidemiol,* **26**(9), pp.719-728.

MONTEIRO, C. A., W. L. CONDE and B. M. POPKIN. 2004. The Burden of Disease From Undernutrition and Overnutrition in Countries Undergoing Rapid Nutrition Transition: A View From Brazil. *Am J Public Health,* **94**(3), pp.433-434.

MORANDI, A., E. M. DEL GIUDICE, F. MARTINO, E. MARTINO, M. BOZZOLA and C. MAFFEIS. 2014. Anthropometric Indices are not satisfactory predictors of Metabolic Comorbidities in Obese Children and Adolescents. *The Journal of Pediatrics,* **165**(6), pp.1178-1183.

MORENO, M. A. 2011. Reducing Screen Time for Children. *Arch Pediatr Adolesc Med. ,* **165**(11), p1056.

MORRISON, A., J. POLISENA, D. HUSEREAU, K. MOULTON, M. CLARK, M. FIANDER, M. MIERZINSKI-URBAN, T. CLIFFORD, B. HUTTON and D. RABB. 2012. The Effect of English-Language Restriciton on Systematic Review-Based Meta-Analyses: A Systematic Review of Empirical Studies. *International Journal of Technology Assessment in Health Care,* **28**(2), pp.138-144.

MORSE, J. M. 1999. Qualitative Generalizability. *Qualitative Health Research,* **9**(1), pp.5-6.

MORTENSEN, A., K. LENZ, H. ABILDSTRØM and T. B. LAURITSEN. 2011. Anesthetizing the obese child. *Pediatric Anesthesia,* **21**(6), pp.623-629.

MORTON, R., M. SAYMA and M. S. SURA. 2017. Economic analysis of the breast cancer screening program used by the UK NHS: should the program be maintained? *Breast Cancer (Dove Med Press),* **9**, pp.217-225.

MOURA, A. A., M. A. M. SILVA, M. R. M. T. FERRAZ and I. R. RIVERA. 2004. Prevalence of high blood pressure in children and adolescents from the city of Maceio, Brazil. [Portuguese]Prevalencia de pressao arterial elevada em escolares e adolescentes de Maceio. *Jornal de Pediatria,* **80**(1), pp.35-40.

MUNN, Z., S. MOOLA, D. RIITANO and K. LISY. 2014. The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence. *International Journal of Health Policy and Management,* **3**(3), pp.123-128.

MURPHY, M. K., N. A. BLACK, D. L. LAMPING, C. M. MCKEE, C. F. SANDERSON, J. ASKHAM and T. MARTEAU. 1998. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess,* **2**(3), pp.1-88.

MURRY JR., J. W. and J. O. HAMMONS. 1995. Delphi: A Versatile Methodology for Conducting Qualitative Research. *The Review of Higher Education,* **18**(4), pp.423-436.

MUST, A., P. F. JACQUES, G. E. DALLAL, C. J. BAJEMA and W. H. DIETZ. 1992. Long-term morbidity and mortality of overweight adolescents. A follow-up of the Harvard Growth Study of 1922 to 1935. *N Engl J Med.,* **327**(19), pp.1350-1355.

MUST, A., S. M. PHILLIPS and E. N. NAUMOVA. 2012. Occurrence and timing of childhood overweight and mortality: findings from the Third Harvard Growth Study. *J Pediatr. ,* **160**(5), pp.743-50.

MWAIKAMBO, S. A., G. H. LEYNA, J. KILLEWO, A. SIMBA and T. PUOANE. 2015. Why are primary school children overweight and obese? A cross sectional study undertaken in Kinondoni district, Dar-es-salaam. *BMC Public Health.,* **15**(1269).

NABBE, P., J. Y. LE RESTE, M. GUILLOU-LANDREAT, E. BECK-ROBERT, R. ASSENOVA, D. LAZIC, S. CZACHOWSKI, S. STOJANOVIĆ-ŠPEHAR, M. HASANAGIC, H. LINGNER, A. CLAVERÍA, M. I. FERNANDEZ SAN MARTIN, A. SOWINSKA, S. ARGYRIADOU, C. LYGIDAKIS, B. LE FLOCH, C. DOERR, T. MONTIER, H. VAN MARWIJK and P. VAN ROYEN. 2016. One consensual depression diagnosis tool to serve many countries: a challenge! A RAND/UCLA methodology. *BMC Res Notes,* **11**(4).

NADLER, D. L. and I. G. ZURBENKO. 2014. Estimating Cancer Latency Times Using a Weibull Model. *Advances in Epidemiology*, pp.1-8.

NAIR, R., R. AGGARWAL and D. KHANNA. 2011. Methods of formal consensus in classification/diagnositic criteria and guideline development. *Semin Arthritis Rheum.,* **41**(2), pp.95-105.

NARAYAN, K. M. V., J. P. BOYLE, T. J. THOMPSON, E. W. GREGG and D. F. WILLIAMSON. 2007. Effect of BMI on Lifetime Risk for Diabetes in the U.S. *Diabet Med.,* **30**(6), pp.1562-1566.

NATIONAL HEART LUNG AND BLOOD INSTITUTE. 2005. *High Blood Cholesterol: What You Need To Know* [online]. [Accessed 2nd May]. Available from: https://www.nhlbi.nih.gov/health/resources/heart/heart-cholesterol-hbc-what-html.

NATIONAL INSTITUTE OF MENTAL HEALTH. 2017. *What is prevalence?* [online]. [Accessed 13th June]. Available from: https://www.nimh.nih.gov/health/statistics/prevalence/index.shtml.

NEWTON, S., D. BRAITHWAITE and T. F. AKINYEMIJU. 2017. Socio-economic status over the life course and obesity: Systematic review and meta-analysis. *Plos ONE*.

NG, M., T. FLEMING, M. ROBINSON, B. THOMSON, N. GRAETZ, C. MARGONO, E. C. MULLANY, S. BIRYUKOV, C. ABBAFATI, S. F. ABERA, J. P. ABRAHAM, N. M. ABU-RMEILEH, T. ACHOKI, F. S. ALBUHAIRAN, Z. A. ALEMU, R. ALFONSO, M. K. ALI, R. ALI, N. A. GUZMAN, W. AMMAR, P. ANWARI, A. BANERJEE, S. BARQUERA, S. BASU, D. A. BENNETT, Z. BHUTTA, J. BLORE, N. CABRAL, I. C. NONATO, J. C. CHANG, R. CHOWDHURY, K. J. COURVILLE, M. H. CRIQUI, D. K. CUNDIFF, K. C. DABHADKAR, L. DANDONA, A. DAVIS, A. DAYAMA, S. D. DHARMARATNE, E. L. DING, A. M. DURRANI, A. ESTEGHAMATI, F. FARZADFAR, D. F. FAY, V. L. FEIGIN, A. FLAXMAN, M. H. FOROUZANFAR, A. GOTO and M. A. G. GREEN, R.; HAFEZI-NEJAD, N.; HANKEY, G.J.; HAREWOOD, H.C.; HAVMOELLER, R.; HAY, S.; HERNANDEZ, L.; HUSSEINI, A.; IDRISOV, B.T.; IKEDA, N.; ISLAMI, F.; JAHANGIR, E.; JASSAL, S.K.; JEE, S.H.; JEFFREYS,

M.; JONAS, J.B.; KABAGAMBE, E.K.; KHALIFA, S.E.; KENGNE, A.P.; KHADER, Y.S.; KHANG, Y.H.; KIM, D.; KIMOKOTI, R.W.; KINGE, J.M.; KOKUBO, Y.; KOSEN, S.; KWAN, G.; LAI, T.; LEINSALU, M.; LI, Y.; LIANG, X.; LIU, S.; LOGROSCINO, G.; LOTUFO, P.A.; LU, Y.; MA, J.; MAINOO, N.K.; MENSAH, G.A.; MERRIMAN, T.R.; MOKDAD, A.H.; MOSCHANDREAS, J.; NAGHAVI, M.; NAHEED, A.; NAND, D.; NARAYAN, K.M.; NELSON, E.L.; NEUHOUSER, M.L.; NISAR, M.I.; OHKUBO, T.; OTI, S.O.; PEDROZA, A.; PRABHAKARAN, D.; ROY, N.; SAMPSON, U.;.SEO, H.; SEPANLOU, S.G.; SHIBUYA, K.; SHIRI, R.; SHIUE, I.; SINGH, G.M.; SINGH, J.A.; SKIRBEKK, V.; STAPELBERG, N.J.; STURUA, L.; SYKES, B.L.; TOBIAS, M.; TRAN, B.X.; TRASANDE, L.; TOYOSHIMA, H.; VAN DE VIJVER, S.; VASANKARI, T.J.; VEERMAN, J.L.; VELASQUEZ-MELENDEZ, G.; VLASSOV, V.V.; VOLLSET, S.E.; VOS, T.; WANG, C.; WANG, X.; WEIDERPASS, E.; WERDECKER, A.; WRIGHT, J.L.; YANG, Y.C.; YATSUYA, H.; YOON, J.; YOON, S.J.; ZHAO, Y.; ZHOU, M.; ZHU, S.; LOPEZ, A.D.; MURRAY, C.J.; GAKIDOU, E.;. 2014. *Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet,* **384**(9945), pp.766-781.

NHS. 2013. *High cholesterol* [online]. [Accessed 19th June]. Available from: http://www.nhs.uk/Conditions/Cholesterol/Pages/Introduction.aspx.

NHS. 2015a. *Cervical Screening* [online]. [Accessed 27th January]. Available from: http://www.nhs.uk/conditions/Cervical-screening-test/Pages/Introduction.aspx.

NHS. 2015b. *Child obesity rates are 'stabilising'* [online]. [Accessed 22nd February]. Available from: https://www.nhs.uk/news/obesity/child-obesity-rates-are-stabilising/.

NHS. 2017a. *Flatfoot* [online]. [Accessed 19th December]. Available from: http://www.ouh.nhs.uk/paediatricorthopaedics/information/conditions/flat-foot.aspx.

NHS. 2017b. *Weight loss surgery* [online]. [Accessed 23rd April]. Available from: https://www.nhs.uk/conditions/weight-loss-surgery/types/.

NHS CHOICES. 2015. *Gallstones* [online]. [Accessed 27th December]. Available from: https://www.nhs.uk/conditions/gallstones/.

NHS CHOICES. 2016. *Blood Tests - C-Reactive Protein* [online]. [Accessed 27th December]. Available from: https://www.nhs.uk/conditions/blood-tests/types/#c-reactive-protein-crp-test.

NHS CHOICES. 2018. *NHS screening* [online]. [Accessed 4th June]. Available from: https://www.nhs.uk/conditions/nhs-screening/.

NHS DIGITAL. 2016. *National Child Measurement Programme - England, 2015-16* [online]. [Accessed 17th January]. Available from: http://content.digital.nhs.uk/searchcatalogue?productid=23381&q=national+child+measurement+programme&sort=Relevance&size=10&page=1#top.

NHS DIGITAL. 2017a. *Breast Screening Programme: England, 2015-16* [online]. [Accessed 12th March]. Available from: https://digital.nhs.uk/media/30533/Breast-Screening-Programme-England-2015-16-Report-/Any/bres-scre-prog-eng-2015-16-rep.

NHS DIGITAL. 2017b. *National Child Measurement Programme - England, 2016-17* [online]. [Accessed 5th May]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/national-child-measurement-programme/2016-17-school-year.

NHS DIGITAL. 2017c. *Statistics on Obesity, Physical Activity and Diet - England, 2017* [online]. [Accessed 1st February]. Available from: http://webarchive.nationalarchives.gov.uk/20180328133908/http://digital.nhs.uk/catalogue/PUB23742.

NHS DIGITAL. n.d. *National Child Measurement Programme* [online]. [Accessed 17th January]. Available from: http://content.digital.nhs.uk/ncmp.

NHS ENGLAND. 2015. *Service User Involvement* [online]. [Accessed 25th June]. Available from: https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2015/10/ohc-paper-06.pdf.

NICE. 2009. *Endovascular stent–grafts for the treatment of abdominal aortic aneurysms* [online]. [Accessed 13th March]. Available from: https://www.nice.org.uk/guidance/ta167/chapter/2-Clinical-need-and-practice.

NICE. 2013. *Weight management: lifestyle services for overweight or obese children and young people [PH47]* [online]. [Accessed 10th October]. Available from: https://www.nice.org.uk/guidance/ph47/chapter/1-Recommendations.

NICE. 2014a. *News and Features: Offer weight loss surgery to obese people with diabetes* [online]. [Accessed 23rd March]. Available from: https://www.nice.org.uk/news/article/offer-weight-loss-surgery-to-diabetics.

NICE. 2014b. *Obesity* [online]. [Accessed 18th April]. Available from: https://bnf.nice.org.uk/treatment-summary/obesity.html.

NICE. 2014c. *Obesity: identification, assessment and management* [online]. [Accessed 20th April]. Available from: https://www.nice.org.uk/guidance/cg189/chapter/1-recommendations#lifestyle-interventions.

NICE. 2017. *Obese, overweight with risk factors: liraglutide (Saxenda)* [online]. [Accessed 23rd April]. Available from: https://www.nice.org.uk/advice/es14/chapter/Key-points.

NICE. n.d.-a. *Care for children and young people with depression* [online]. [Accessed 15th March]. Available from: https://pathways.nice.org.uk/pathways/depression#path=view%3A/pathways/depression/care-for-children-and-young-people-with-depression.xml&content=view-node%3Anodes-assessment-and-treatment-considerations.

NICE. n.d.-b. *Type 2 diabetes in children and young people* [online]. [Accessed 13th March]. Available from: https://pathways.nice.org.uk/pathways/diabetes-in-children-and-young-people#path=view%3A/pathways/diabetes-in-children-and-young-people/type-2-diabetes-in-children-and-young-people.xml&content=view-node%3Anodes-metformin.

NICE CLINICAL GUIDELINES. 2005. *Depression in Children and Young People: Identification and Management in Primary, Community and Secondary Care*. Leicester, UK: British Psychological Society.

NIGHTINGALE, J. M., R. BORGEN, L. PORTER-BENNETT and K. SZCZEPURA. 2015. An audit to investigate the impact of false positive breast screening results and diagnostic work-up on re-engagement with subsequent routine screening. *International Journal of Diagnostic Imaging and Radiation Therapy,* **21**(1), pp.7-10.

NIH CONSENSUS STATEMENT. 2000. Phenylketonuria (PKU): Screening and Management. *National Institutes of Health Consensus Development Conference Statement,* **17**(3), pp.1-27.

NING, M., Q. ZHANG and M. YANG. 2016. Comparison of self-reported and biomedical data on hypertension and diabetes: findings from the China Health and Retirement Longitudinal Study (CHARLS). *BMJ Open,* **6**(1), pe009836.

NOO. 2010. *The economic burden of obesity* [online]. [Accessed 1st October]. Available from: www.noo.org.uk/uploads/doc/vid_8575_Burdenofobesity151110MG.pdf.

NOO. 2011. *A simple guide to classifying body mass index in children* [online]. [Accessed 17th January]. Available from: http://www.noo.org.uk/uploads/doc/vid_11601_A_simple_guide_to_classifying_BMI_in_children.pdf.

NOONAN, C. W., B. D. BROWN, B. BENTLEY, K. CONWAY, M. CORCORAN, K. FOURSTAR, P. FREIDE, B. HEMLOCK, S. WAGNER and T. WILSON. 2010. Variability in childhood

asthma and body mass index across Northern Plains American Indian communities. *Journal of Asthma,* **47**(5), pp.496-500.

NOWELL, L. S., J. M. NORRIS, D. E. WHITE and N. J. MOULES. 2017. Thematic Analysis: Striving to meet the Trustworthiness Criteria. *International Journal of Qualitative Methods,* **16**, pp.1-13.

NUFFIELD HEALTH. 2016. *10 questions to ask during your Health MOT* [online]. [Accessed 6th February].

NUR, N., S. CETINKAYA, A. YILMAZ, A. AYVAZ, M. O. BULUT and H. SUMER. 2008. Prevalence of hypertension among high school students in a middle Anatolian province of Turkey. *Journal of Health, Population & Nutrition,* **26**(1), pp.88-94.

O'BRIEN, P. 2016. *Surgical Treatment of Obesity [Updated 2016 Jan 19]* [online]. [Accessed 23rd April]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK279090/.

O'BRIEN, P. E., S. M. SAWYER, C. LAURIE, W. A. BROWN, S. SKINNER, F. VEIT, E. PAUL, P. R. BURTON, M. MCGRICE, M. ANDERSON and J. B. DIXON. 2010. Laparoscopic adjustable gastric banding in severely obese adolescents: a randomized trial. *JAMA,* **303**(6), pp.519-526.

OCHOA, A. and J. M. BERGE. 2017. Home Environmental Influences on Childhood Obesity in the Latino Population: A Decade Review of Literature. *J Immigr Minor Health.,* **19**(2), pp.430-447.

OECD. 2014. *Obesity Update* [online]. [Accessed 19th December]. Available from: http://www.oecd.org/health/Obesity-Update-2014.pdf.

OGDEN, C. L. and K. M. FLEGAL. 2010. Changes in Terminology for Childhood Overweight and Obesity. *National Health Statistics Reports,* (25), pp.1-8.

OLDROYD, J., M. BANERJEE, A. HEALD and K. CRUICKSHANK. 2005. Diabetes and ethnic minorities. *Postgraduate Medical Journal* **81**, pp.486-490.

OREMUS, M., C. OREMUS, G. B. C. HALL, M. C. MCKINNON and ECT & COGNITION SYSTEMATIC REVIEW TEAM. 2012. Inter-rater and test–retest reliability of quality assessments by novice student raters using the Jadad and Newcastle–Ottawa Scales. *BMJ Open,* **2**, pe001368.

OWEN, N., E. LESLIE, J. SALMON and M. J. FOTHERINGHAM. 2000. Environmental Determinants of Physical Activity and Sedentary Behavior. *Exercise and Sport Sciences Reviews,* **28**(4), pp.153-158.

OWEN, N., P. B. SPARLING, G. N. HEALY, D. W. DUNSTAN and C. E. MATTHEWS. 2010. Sedentary Behavior: Emerging Evidence for a New Health Risk. *Mayo Clin Proc.,* **85**(12), pp.1138-1141.

OXMAN, A. D., D. J. COOK and G. H. GUYATT. 1994. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *JAMA,* **272**, pp.1367-1371.

ÖZTÜRK, Y. 2017. Etiology and comorbidities of childhood obesity. *Turk J Gastroenterol.,* **28**(2), pp.149-151.

PALAGANAS, E. C., M. C. SANCHEZ, M. V. P. MOLINTAS and R. D. CARICATIVO. 2017. Reflexivity in Qualitative Research: A Journey of Learning. *The Qualitative Report,* **22**(2), pp.426-438.

PANTALONE, K. M., T. M. HOBBS, K. M. CHAGIN, S. X. KONG, B. J. WELLS, M. W. KATTAN, J. BOUCHARD, B. SAKURADA, A. MILINOVICH, W. WENG, J. BAUMAN, A. D. MISRA-HEBERT, R. S. ZIMMERMAN and B. BURGUERA. 2017. Prevalence and recognition of obesity and its associated comorbidities: cross-sectional analysis of electronic health record data from a large US integrated health system. *BMJ Open,* **7**(11), pe017583.

PAPOUTSAKIS, C., M. YANNAKOULIA, I. NTALLA and G. V. DEDOUSSIS. 2012. Metabolic syndrome in a Mediterranean pediatric cohort: Prevalence using International

Diabetes Federation-derived criteria and associations with adiponectin and leptin. *Metabolism: Clinical and Experimental,* **61**(2), pp.140-145.

PARK, R. E., A. FINK, R. H. BROOK, M. R. CHASSIN, K. L. KAHN, N. J. MERRICK, J. KOSECOFF and D. H. SOLOMON. 1986. Physician Ratings of ApprQpriate Indications for Six Medical and Surgical Procedures. *American Journal of Public Health,* **76**(7), pp.766-772.

PARK, R. E., A. FINK, R. H. BROOK, M. R. CHASSIN, K. L. KAHN, N. J. MERRICK, J. KOSECOFF and D. H. SOLOMON. 1989. Physician Ratings of Appropriate Indications for Three Procedures: Theoretical Indications vs Indications Used in Practice. *American Journal of Public Health,* **79**(4), pp.445-447.

PARKER, E. D., A. R. SINAIKO, E. O. KHARBANDA, K. L. MARGOLIS, M. F. DALEY, N. K. TROWER, N. E. SHERWOOD, L. C. GREENSPAN, J. C. LO, D. J. MAGID and P. J. O'CONNOR. 2016. Change in Weight Status and Development of Hypertension. *Pediatrics,* **137**(3), pe20151662.

PARNELL, K. J., N. A. STANTON and K. L. PLANT. 2018. Creating the environment for driver distraction: A thematic framework of sociotechnical factors. *Applied Ergonomics,* **68**, pp.213-228.

PARTHASARATHY, S. and S. SHARMA. 2014. Determining ERP customization choices using nominal group technique and analytical hierarchy process. *Computers in Industry,* **65**(6), pp.1009-1017.

PATAKY, Z., E. BOBBIONI-HARSCH and A. GOLAY. 2010. Obesity: a complex growing challenge. *Experimental and clinical endocrinology and diabetes,* **118**(7), pp.427-433.

PATEL, S. n.d. *Overweight and obese comorbidity screening algorithm* [online]. [Accessed 4th July]. Available from: https://idph.iowa.gov/Portals/1/Files/WIC/Comorbidity%20Algorithm%20final-Sonali%202013.pdf.

PATRICK, H. and T. A. NICKLAS. 2005. A review of family and social determinants of children's eating patterns and diet quality. *J Am Coll Nutr.,* **24**(2), pp.83-92.

PATTON, M. Q. 2002. *Qualitative Research and Evaluation Methods.* Thousand Oaks, CA: SAge Publications Inc.

PECIN, I., R. SAMOVOJSKA, B. HEINRICH, T. ZELJKOVIC-VRKIC, M. LAGANOVIC and B. JELAKOVIC. 2013. Hypertension, overweight and obesity in adolescents: the CRO-KOP study. *Collegium Antropologicum,* **37**(3), pp.761-4.

PEREIRA-MIRANDA, E., P. R. F. COSTA, V. A. O. QUEIROZ, M. PEREIRA-SANTOS and M. L. P. SANTANA. 2017. Overweight and Obesity Associated with Higher Depression Prevalence in Adults: A Systematic Review and Meta-Analysis. *J Am Coll Nutr.,* **36**(3), pp.223-233.

PEREIRA, A., A. D. GUEDES, I. T. VERRESCHI, R. D. SANTOS and T. L. MARTINEZ. 2009. Obesity and its association with other cardiovascular risk factors in school children in Itapetininga, Brazil. *Arquivos Brasileiros de Cardiologia,* **93**(3), pp.253-60.

PETTICREW, M. P., A. J. SOWDEN, D. LISTER-SHARP and K. WRIGHT. 2000. False-negative results in screening programmes: systematic review of impact and implications. *Health Technol Assess.,* **4**(5), pp.1-120.

PHAROAH, P. D. P., B. SEWELL, D. FITZSIMMONS, H. S. BENNETT and N. PASHAYAN. 2013. Cost effectiveness of the NHS breast screening programme: life table model. *BMJ,* **346**(f2618).

PHELPS, C. E. 1993. The Methodologic Foundations of Studies of the Appropriateness of Medical Care. *N Engl J Med,* **329**, pp.1241-1245.

POLLACK, A. 2013. *A.M.A. Recognizes Obesity as a Disease* [online]. [Accessed 30th April]. Available from: https://www.nytimes.com/2013/06/19/business/ama-recognizes-obesity-as-a-disease.html?smid=pl-share.

POPE, C., S. ZIEBAND and N. MAYS. 2000. Qualitative research in health care: Analysing qualitative data. *BMJ,* **320**, pp.114-116.

PORTER, J. 2012. Methodological Issues in Collecting Children's View Part 2: Using Nominal Group Technique to Explore Children's Views of the Difficulties Encountered in School. *Procedia - Social and Behavioral Sciences,* **47**, pp.1627-1631.

PROSTATE CANCER UK. 2016. *Consensus statements on PSA testing in asymptomatic men in the UK - information for health professionals* [online]. [Accessed 1st June]. Available from: https://prostatecanceruk.org/for-health-professionals/our-projects/psa-consensus-for-health-professionals.

PUBLIC HEALTH ENGLAND. 2015a. *Criteria for appraising the viability, effectiveness and appropriateness of a screening programme* [online]. [Accessed 6th January]. Available from: https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme.

PUBLIC HEALTH ENGLAND. 2015b. *National Child Measurement Programme* [online]. [Accessed 12th October]. Available from: https://www.noo.org.uk/data_sources/child/national_child_measurement_programme.

PUBLIC HEALTH ENGLAND. 2018a. *Child obesity and excess weight: small area level data* [online]. [Accessed 28th April]. Available from: https://www.gov.uk/government/statistics/child-obesity-and-excess-weight-small-area-level-data.

PUBLIC HEALTH ENGLAND. 2018b. *NCMP Local Authority Profile - January 2018 update* [online]. [Accessed 5th May]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673318/Official_statistics_short_commentary_NCMP_LA_Profile_update_January_2018.pdf.

PUBLIC HEALTH ENGLAND. n.d. *NHS Abdominal Aortic Aneurysm Screening Programme: deciding whether to be screened* [online]. [Accessed 12th March]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/617871/AAA_draft_short_form_decision_aid.pdf.

PUBMED HEALTH. 2016. *Benefits and risks of screening tests* [online]. [Accessed 6th June]. Available from: https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0072602/.

PUHL, R. and K. D. BROWNELL. 2012. Bias, Discrimination, and Obesity. *Obesity,* **9**(12), pp.788-805.

PUHL, R. M. and C. A. HEUER. 2010. Obesity Stigma: Important Considerations for Public Health. *Am J Public Health. ,* **100**(6), pp.1019–1028.

PULGARÓN, E. R. 2013. Childhood Obesity: A Review of Increased Risk for Physical and Psychological Co-morbidities. *Clinical therapeutics,* **35**(1), pp.A18-A32.

RACETTE, S. B., S. S. DEUSINGER and R. H. DESUSINGER. 2003. Obesity: overview of prevalence, etiology, and treatment. *Physical Therapy,* **83**(3), pp.276-288.

RADOMSKI, M. V., M. FINKELSTEIN, I. LLANOS, M. SCHEIMAN and S. G. WAGENER. 2014. Composition of a Vision Screen for Servicemembers With Traumatic Brain Injury: Consensus Using a Modified Nominal Group Technique. *Am J Occup Ther.,* **68**(4), pp.422-429.

RAFFLE, A. 2017a. *Chapter 5 - Measures of test performance* [online]. [Accessed 3rd March]. Available from: https://www.healthknowledge.org.uk/interactive-learning/screening/chapter5.

RAFFLE, A. 2017b. *Chapter 8 - Screening prior to evidence* [online]. [Accessed 31st May]. Available from: https://www.healthknowledge.org.uk/interactive-learning/screening/chapter8.

RAFRAF, M., B. POURGHASSEM GARGARI and A. SAFAIYAN. 2010. Prevalence of prehypertension and hypertension among adolescent high school girls in Tabriz, Iran. *Food and Nutrition Bulletin,* **31**(3), pp.461-465.

RAINE, R., N. C. BHÁIRD, P. XANTHOPOULOU, I. WALLACE, D. ARDRON, M. HARRIS, J. BARBER, A. PRENTICE, S. GIBBS, M. KING, J. M. BLAZEBY, S. MICHIE, A. LANCELEY, A. CLARKE and G. LIVINGSTON. 2015. Use of a formal consensus development technique to produce recommendations for improving the effectiveness of adult mental health multidisciplinary team meetings. *BMC Psychiatry,* **15**, pp.143-154.

RAINE, R., C. SANDERSON and N. BLACK. 2005. Developing clinical guidelines: a challenge to current methods. *BMJ,* **331**, pp.631-633.

RAINE, R., C. SANDERSON, A. HUTCHINGS, S. CARTER, K. LARKIN and N. BLACK. 2004. An experimental study of determinants of group judgments in clinical guideline development. *Lancet,* **364**, pp.429-437.

RAMIREZ, F. D., Y. CHEN, P. DI SANTO, T. SIMARD, P. MOTAZEDIAN and B. HIBBERT. 2017. Association Between Self-Reported Potentially Modifiable Cardiac Risk Factors and Perceived Need to Improve Physical Health: A Population-Based Study. *Journal of the American Heart Association,* **6**(5), pe005491.

RAMOS, E. and H. BARROS. 2005. Prevalence of hypertension in 13-year-old adolescents in Porto, Portugal. *Revista Portuguesa de Cardiologia,* **24**(9), pp.1075-87.

RANKIN, J., L. MATTHEWS, S. COBLEY, A. HAN, R. SANDERS, H. D. WILTSHIRE and J. S. BAKER. 2016. Psychological consequences of childhood obesity: psychiatric comorbidity and prevention. *Asolesc Health Med Ther,* **7**, pp.125-146.

REDMAN, S., S. CARRICK, J. COCKBURN and S. HIRST. 1997. Consulting about priorities for the NHMRC National Breast Cancer Centre: how good is the nominal group technique. *Australian and New Zealand Journal of Public Health,* **21**(3), pp.250-256.

REEVES, G. M., T. T. POSTOLACHE and S. SNITKER. 2008. Childhood Obesity and Depression: Connection between these Growing Problems in Growing Children. *Int J Child Health Hum Dev,* **1**(2), pp.103-114.

REGITZ-ZAGROSEK, V. 2012. Sex and gender differences in health. *EMBO Reports,* **13**(7), pp.596-603.

REILLY, J. J. and J. KELLY. 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**, pp.891-898.

REILLY, J. J., E. METHVEN, Z. C. MCDOWELL, B. HACKING, D. ALEXANDER, L. STEWART and C. J. H. KELNAR. 2003. Health Consequences of Obesity. *Arch Dis Child.,* **88**, pp.748-752.

REINEHR, T. 2013. Type 2 diabetes mellitus in children and adolescents. *World J Diabetes,* **4**(6), pp.270-281.

REISER, S. J. 1978. The emergence of the concept of screening for disease. *Milbank Mem Fund Q Health Soc,* **56**(4), pp.403-425.

REMES, O., C. BRAYNE, R. VAN DER LINDE and L. LAFORTUNE. 2016. A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain Behav.,* **6**(7), pe00497.

REUTER, C. P., P. T. DA SILVA, J. D. P. RENNER, E. D. DE MELLO, A. R. DE MOURA VALIM, L. PASA, R. DA SILVA and M. S. BURGOS. 2016. Dyslipidemia is Associated with Unfit and Overweight-Obese Children and Adolescents. *Arq Bras Cardiol.,* **106**(3), pp.188-193.

RIBEIRO-SILVA, R. D. C., A. M. OLIVEIRA-ASSIS, S. B. JUNQUEIRA, R. L. FIACCONE, S. M. C. DOS SANTOS, M. L. BARRETO, E. D. J. PINTO, L. A. DA SILVA, L. C. RODRIGUES and N. M. ALCANTARA-NEVES. 2013. Food and nutrition insecurity: a marker of vulnerability to asthma symptoms. *Public health nutrition,* **17**(1), pp.14-19.

RIGGS, W. E. 1983. The Delphi technique: An experimental evaluation. *Technological Forecasting and Social Change,* **23**(1), pp.89-94.

RISTE, L., F. KHAN and K. CRUICKSHANK. 2001. High prevalence of type 2 diabetes in all ethnic groups, including Europeans, in a British inner city: relative poverty, history, inactivity, or 21st century Europe? *Diabetes Care,* **24**(8), pp.1377-1383.

ROBINSON, J., J. HERITAGE and D. W. MAYNARD. 2006. *Soliciting patients' presenting concerns, Communication in Medical Care: Interaction between Primary Care Physicians and Patients. 1st edn. .* Cambridge: Cambridge University Press.

ROSEMAN, M., L. A. KLODA, N. SAADAT, K. E. RIEHM, A. ICKOWICZ, F. BALTZER, L. Y. KATZ, S. B. PATTEN, C. ROUSSEAU and B. D. THOMBS. 2016. Accuracy of Depression Screening Tools to Detect Major Depression in Children and Adolescents: A Systematic Review. *Can J Psychiatry,* **61**(12), pp.746-757\.

ROSENMAN, R., V. TENNEKOON and L. G. HILL. 2011. Measuring bias in self-reported data. *Int J Behav Healthc Res,* **2**(4), pp.320-332.

RUDISILL, C., J. CHARLTON, H. P. BOOTH and M. C. GULLIFORD. 2016. Are healthcare costs from obesity associated with body mass index, comorbidity or depression? Cohort study using electronic health records. *Clin Obes.,* **6**(3), pp.225-31.

RUF, M. and O. MORGAN. 2008. *Principles, methods, applications and organisation of screening for early detection, prevention, treatment and control of disease* [online]. [Accessed 22nd December]. Available from: http://www.healthknowledge.org.uk/public-health-textbook/disease-causation-diagnostic/2c-diagnosis-screening/principles-methods-applications.

RUMIŃSKA, M., E. WITKOWSKA–SĘDEK, A. MAJCHER, M. BRZEWSKI and A. CZERWONOGRODZKA–SENCZYNA. 2017. Carotid Intima-Media Thickness and Metabolic Syndrome Components in Obese Children and Adolescents. *In:* P. M., ed. *Pulmonary Care and Clinical Medicine. Advances in Experimental Medicine and Biology.* Cham: Springer.

RYCROFT-MALONE, J. 2001. Formal consensus: the development of a national clinical guideline. *Quality in Health Care,* **10**(4), pp.238-244.

SABIN, M., E. CROWNE and J. P. SHIELD. 2002. Childhood obesity and type 2 diabetes. *Nursing Times,* **98**(19), p49.

SAC. 2002. Assessing Health Status and Quality of Life Instruments: Attributes and Review Criteria. *Quality of Life Research,* **11**(3), pp.193-205.

GREAT BRITAIN. U. S. A. F. P. RAND. 1974. *Delphi Assessment: Expert Opinion, Forecasting, and Group Process*. RAND Corporation: Santa Monica, CA.

SADOVSKY, R. 1998. Value of C-Reactive Protein Measurements in Children. *Am Fam Physician,* **57**(6), pp.1384-1386.

SAHOO, K., B. SAHOO, A. K. CHOUDHURY, N. Y. SOFI, R. KUMAR and A. S. BHADORIA. 2015. Childhood obesity: causes and consequences. *J Family Med Prim Care,* **4**(2), pp.187-192.

SALANCIK, J. R. 1973. Assimilation of aggregated inputs into Delphi forecasts: A regression analysis. *Technological Forecasting and Social Change,* **5**(3), pp.243-247.

SALVY, S. J. and J. C. BOWKER. 2014. Peers and Obesity during Childhood and Adolescence: A Review of the Empirical Research on Peers, Eating, and Physical Activity. *J Obes Weight Loss Ther.,* **4**(1), p207.

SALVY, S. J., J. C. BOWKER, L. GERMEROTH and J. BARKLEY. 2012. Influence of peers and friends on overweight/obese youths' physical activity. *Exerc Sport Sci Rev. ,* **40**(3), pp.127-32.

SAMPLE, J. A. 1984. Nominal Group Technique: An Alternative to Brainstorming. *Journal of Extension,* **22**(2).

SANDELOWSKI, M. 1986. The problem of rigor in qualitative research. *Advances in Nursing Science,* **8**(3), pp.27-37.

SANDERS, R. H., A. HAN, J. S. BAKER and S. COBLEY. 2015. Childhood obesity and its physical and psychological co-morbidities: a systematic review of Australian children and adolescents. *European Journal of Pediatrics,* **174**(6), pp.715-746.

SANDERSON, S., I. D. TATT and J. P. T. HIGGINS. 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.666-679.

SANGHERA, S., J. COAST, R. M. MARTIN, J. L. DONOVAN and S. MOHIUDDIN. 2018. Cost-effectiveness of prostate cancer screening: a systematic review of decisionanalytical models. *BMC Cancer*.

SARGENT, J. D. 1994. Obesity and stature in adolescence and earnings in young adulthood. *Arch Pediatr Adolesc Med,* **148**, pp.681-7.

SCHEELE, D. S. 1975. Consumerism comes to Delphi: Comments on Delphi assessment, expert opinion, forecasting, and group process by H. Sackman. *Technological Forecasting and Social Change,* **7**(2), pp.215-219.

SCHIENKIEWITZ, A., G. B. M. MENSINK and C. SCHEIDT-NAVE. 2012. Comorbidity of overweight and obesity in a nationally representative sample of German adults aged 18-79 years. *BMC Public Health,* **12**, p658.

SCHMIDT, M. E., J. HAINES, A. O'BRIEN, J. MCDONALD, S. PRICE, B. SHERRY and E. M. TAVERAS. 2012. Systematic Review of Effective Strategies for Reducing Screen Time Among Young Children. *Obesity (Silver Spring),* **20**(7), pp.1338-1354.

SCHNEIDER, J. L., J. DAVIS, T. L. KAUFFMAN, J. A. REISS, C. MCGINLEY, K. ARNOLD, J. ZEPP, M. GILMORE, K. R. MUESSIG, S. SYNGAL, L. ACHESON, G. L. WIESNER, S. K. PETERSON and K. A. B. GODDARD. 2016. Stakeholder perspectives on implementing a universal lynch syndrome screening program: A qualitative study of early barriers and facilitators. *Genet Med,* **18**(2), pp.152-161.

SCHWANDT, T. A. 2001. *Dictionary of Qualitative Inquiry.* 2nd ed. Thousand Oaks, CA: Sage Publications Inc.

SCIENTIFIC ADVISORY COMMITTEE ON NUTRITION. 2016. *Vitamin D and Health*.

SCOTT, E. A. and N. BLACK. 1991. Appropriateness of cholecystectomy in the United Kingdom: A consensus panel approach. *Gut,* **32**(9), pp.1066-1070.

SEDGWICK, P. 2014. Retrospective cohort studies: advantages and disadvantages. *BMJ,* **348**.

SEGULA, D. 2014. Complications of obesity in adults: A short review of the literature. *Malawi Med J.,* **26**(1), pp.20-24.

SEIBERT, T. S., D. B. ALLEN and A. L. CARREL. 2014. Adolescent Obesity and Its Risks: How to Screen and When to Refer. *J Clin Outcomes Manag,* **21**(2), pp.87-96.

SEKI, M., T. MATSUO and A. J. FARIA CARRILHO. 2008. Prevalence of metabolic syndrome and associated risk factors in Brazilian schoolchildren. *Public Health Nutrition,* **12**(7), pp.947-952.

SEMPLE, A. 1953, 1960, 1966). *Annual Reports for the City of Liverpool*.

SHADISH, W. R., T. D. COOK and D. T. CAMPBELL. 1979. *Experimental and Quasi-Experimental Designs for Generalized Casual Inference.* New York: Houghton Mifflin Company.

SHAHBABU, B., A. DASGUPTA, K. SARKAR and S. K. SAHOO. 2016. Which is More Accurate in Measuring the Blood Pressure? A Digital or an Aneroid Sphygmomanometer. *J Clin Diagn Res,* **10**(3), pp.11-14.

SHARIFF, N. J. 2015. Utilizing the Delphi Survey Approach: A Review. *Nurs Car,* **4**, p246.

SHARMA, V. 2014. *Conversation with Sarah Meanwell, Senior Watch It Trainer.* 18th December 2014.

SHEEHY, A. M., D. B. COURSIN and R. A. GABBAY. 2009. Back to Wilson and Jungner: 10 Good Reasons to Screen for Type 2 Diabetes Mellitus. *Mayo Clin Proc.,* **84**(1), pp.38-42.

SHEKELLE, P. G., J. P. KAHAN, S. J. BERNSTEIN, L. L. LEAPE, C. J. KAMBERG and R. E. PARK. 1998. The Reproducibility of a Method to Identify the Overuse and Underuse of Medical Procedures. *The New England Journal of Medicine,* **336**, pp.1888-1895.

SHERRY, B., M. E. JEFFERDS and L. M. GRUMMER-STRAWN. 2007. Accuracy of adolescent self-report of height and weight in assessing overweight status: a literature review. *Arch Pediatr Adolesc Med.,* **161**(12), pp.1154-61.

SICKLES, E. A. 1997. Breast cancer screening outcomes in women ages 40-49: clinical experience with service screening using modern mammography. *J Natl Cancer Inst Monogr.,* **22**, pp.99-104.

SINGH, A. K., S. K. SINGH, N. SINGH, N. AGRAWAL and K. GOPAL. 2011. Obesity and dyslipidemia. *International Journal of Biological & Medical Research,* **2**(3), pp.824-828.

SINGH, G. M., R. MICHAM, S. KHATIBZADEH, P. SHI, S. LIM, K. G. ANDREWS, R. E. ENGELL, M. EZZATI, D. MOZAFFARIAN and GLOBAL BURDEN OF DISEASES NUTRITION AND CHRONIC DISEASES EXPERT GROUP (NUTRICODE). 2015. Global, Regional, and National Consumption of Sugar-Sweetened Beverages, Fruit Juices, and Milk: A Systematic Assessment of Beverage Intake in 187 Countries. *PLoS ONE [Electronic Resource],* **10**(8), pe0124845.

SKINNER, A. C. and J. A. SKELTON. 2014. Prevalence and trends in obesity and severe obesity among children in the United States, 1999-2012. *JAMA Pediatrics,* **168**(6), pp.561-566.

SKINNER, J. D., B. R. CARRUTH, W. BOUNDS, P. ZIEGLER and K. REIDY. 2002a. Do food-related experiences in the first 2 years of life predict dietary variety in school-aged children? *J Nutr Educ Behav.,* **34**(6), pp.310-315.

SKINNER, J. D., B. R. CARRUTH, B. WENDY and P. J. ZIEGLER. 2002b. Children's food preferences: a longitudinal analysis. *J Am Diet Assoc,* **102**(11), pp.1638-47.

SMEDLEY, R. M. and N. S. COULSON. 2017. A thematic analysis of messages posted by moderators within health-related asynchronous online support forums. *Patient Education and Counseling,* **100**(9), pp.1688-1693.

SMITH, L. 2011. Obesity [online]. [Accessed 2nd Feb 2019]. Available from: http://researchbriefings.files.parliament.uk/documents/SN03724/SN03724.pdf.

SOARES, H. P., S. DANIELS, A. KUMAR, M. CLARKE, C. SCOTT, S. SWANN and B. DJULBEGOVIC. 2004. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ,* **328**, pp.22-24.

SOLOMON, M. 2015. *Making Medical Knowledge.* Oxford, UK: Oxford University Press.

SORRELL, M. F., E. A. BELONGIA, J. COSTA, I. F. GAREEN, J. L. GREM, J. M. INADOMI, E. R. KERN, J. A. MCHUGH, G. M. PETERSEN, M. F. REIN, D. B. STRADER and H. T. TROTTER. 2009. National Institutes of Health Consensus Development Conference Statement: Management of Hepatitis B. *Annals of Internal Medicine,* **150**(2), pp.104-110.

SPELIOTES, E. K., C. J. WILLER, S. I. BERNDT, K. L. MONDA, G. J. THORLEIFSSON, A.U., H. LANGO ALLEN, C. M. LINDGREN, J. LUAN, R. R. MÄGI, J,C., S. VEDANTAM, T. W. WINKLER, L. QI and ET AL. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.,* **42**(11), pp.932-948.

SPENCER, D. M. 2010. Facilitating public participation in tourism planning on American Indian reservations: A case study involving the Nominal Group Technique. *Tourism Management,* **31**(5), pp.684-690.

SPETHMAN, B. 1992. Focus groups key to reaching kids. *Advertising Age,* (S1 and S24), p10.

SPIEGELHALTER, D. J. 2005. Funnel plots for comparing institutional performance. *Statistics in Medicine,* **24**, pp.1185-1202.

SPRINGER, S. C., J. SILVERSTEIN, K. COPELAND, K. R. MOORE, G. E. PRAZAR, T. RAYMER, R. N. SHIFFMAN, V. V. THAKER, M. ANDERSON, S. J. SPANN and S. K. FLINN. 2013. Management of Type 2 Diabetes Mellitus in Children and Adolescents. *Pediatrics,* **131**(2), pp.648-664.

STAIMEZ, L. R., M. B. WEBER, K. M. NARAYAN and R. OZA-FRANK. 2013. A systematic review of overweight, obesity, and type 2 diabetes among Asian American subgroups. *Curr Diabetes Rev,* **9**(4), pp.312-331.

STAMATAKIS, E., J. WARDLE and T. J. COLE. 2013. Childhood obesity and overweight prevalence trends in England: evidence for growing socio-economic disparities. *Int J Obes,* **34**(1).

STANG, A. 2010. Critical Evaluation of the Newcastle-Ottawa Scale for the Assessment of the Quality of Non-Randomized Studies in Meta-Analyses. *Eur J Epidemiol,* **25**, pp.603-605.

STEGBAUER, C., G. WILLMS, K. KLEINE-BUDDE, A. BRAMESFELD, C. STAMMANNA and J. SZECSENY. 2017. Development of indicators for a nationwide cross-sectoral quality assurance procedure for mental health care of patients with schizophrenia, schizotypal and delusional disorders in Germany. *Z Evid Fortbild Qual Gesundhwes,* (126), pp.13-22.

STENBACKA, C. 2001. Qualitative research requires quality concepts of its own. *Management Decision,* **39**(7), pp.551-556.

STERGIOU, G., T. MENGDEN, P. L. PADFIELD, G. PARATI and E. O'BRIEN. 2004. Self monitoring of blood pressure at home. *BMJ,* **329**(7471), pp.870-871.

STITT-GOHDES, W. L. and T. B. CREWS 2004. The Delphi Technique: A research strategy for career and technical education. *Journal of Career and Technical Education,* **20**(2), pp.55-67.

STORY, M., D. NEUMARK-SZTAINER and S. FRENCH. 2002. Individual and environmental influences on adolescent eating behaviors. *J Am Diet Assoc,* **102**(3), pp.S40-51.

STRACHAN, J., G. YELLOWLEES and A. QUIGLEY. 2015. General practitioners' assessment of, and treatment decisions regarding, common mental disorder in older adults: thematic analysis of interview data. *Ageing and Society,* **35**(1), pp.150-168.

STRAUSS, R. S. 2000. Childhood Obesity and Self-Esteem. *Pediatrics,* **105**(1), pp.1-5.

STRAUSS, R. S. and H. A. POLLACK. 2003. Social marginalisation of overweight children. *Archives of Pediatrics and Adolescent Medicine,* **157**, pp.746-752.

STROUP, D. F., J. A. BERLIN, S. C. MORTON, I. OLKIN, G. D. WILLIAMSON, D. RENNIE, D. MOHER, B. J. BECKER, T. A. SIPE and S. B. THACKER. 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA,* **283**(15), pp.2008-12.

STUCKEY, H. L. 2015. The second step in data analysis: coding qualitative research data. *Journal of Social Health and Diabetes,* **3**(1), pp.7-10.

STUNKARD, A. and V. BURT. 1967. Obesity and the body image. II. Age at onset of disturbances in the body image. *Am J. Psychiatry,* **123**, pp.1443-1447.

STUNKARD, A. J., T. T. FOCH and Z. HRUBEC. 1986. A twin study of human obesity. *JAMA,* **256**(1), pp.51-54.

SUDORE, R. L., H. D. LUM, J. J. YOU, L. C. HANSON, D. E. MEIER, S. Z. PANTILAT, D. D. MATLOCK, J. A. C. RIETJENS, I. J. KORFAGE, C. S. RITCHIE, J. S. KUTNER, J. M. TENO, J. THOMAS, R. D. MCMAHAN and D. K. HEYLAND. 2017. Defining Advance Care Planning for Adults: A Consensus Definition from a Multidisciplinary Delphi Panel. *J Pain Symptom Manage,* **53**(5), pp.821-832.

TAM, V. C. and S. J. HOTTE. 2008. Consistency of phase III clinical trial abstracts presented at an annual meeting of the American Society of Clinical Oncology compared with their subsequent full-text publications. *Journal of Clinical Oncology,* **26**(13), pp.2205-2211.

TANDON, N., M. K. GARG, Y. SINGH and R. K. MARWAHA. 2013. Prevalence of metabolic syndrome among urban Indian adolescents and its relation with insulin resistance (HOMA-IR). *Journal of Pediatric Endocrinology,* **26**(11-12), pp.1123-30.

TANOFSKY-KRAFF, M., J. C. HAN, K. ANANDALINGAM, L. B. SHOMAKER, K. M. COLUMBO, L. E. WOLKOFF, M. KOZLOSKY, C. ELLIOTT, L. M. RANZENHOFER, C. A. ROZA, S. Z. YANOVSKI and J. A. YANOVSKI. 2009. The FTO gene rs9939609 obesity-risk allele and loss of control over eating. *Am J Clin Nutr.,* **90**(6), pp.1483-1488.

TENENBAUM, S., O. HERSHKOVICH, B. GORDON, N. .BRUCK, R. THEIN, E. DERAZNE, D. TZUR, A. SHAMISS and A. AFEK. 2013. Flexible pes planus in adolescents: body mass index, body height, and gender--an epidemiological study. *Foot & Ankle International,* **34**(6), pp.811-7.

THANGARATINAM, S. and C. W. E. REDMAN. 2005. The Delphi Technique. *The Obstetrician and Gynaecologist,* **7**(2), pp.120-125.

THE EXPERT COMMITTEE ON THE DIAGNOSIS AND CLASSIFICATION OF DIABETES MELLITUS. 2003. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care,* **26**, pp.3160-3167.

THE JOANNA BRIGGS INSTITUTE. 2014. *The Joanna Briggs Institute Reviewers' Manual 2014: The Systematic Review of Prevalence and Incidence Data* [online]. [Accessed 12th January]. Available from: http://www.joannabriggs.org/assets/docs/sumari/ReviewersManual_2014-The-Systematic-Review-of-Prevalence-and-Incidence-Data_v2.pdf.

THE JOANNA BRIGGS INSTITUTE. 2016. *The Joanna Briggs Institute Critial Appraisal Tools for use in JBI Systematic Reviews: Checklist for Case Series* [online]. [Accessed 21st February]. Available from: http://joannabriggs.org/assets/docs/critical-appraisal-tools/JBI_Critical_Appraisal-Checklist_for_Case_Series.pdf.

THE, N. S., C. SUCHINRAN, K. E. NORTH, B. M. POPKIN and P. GORDON-LARSEN. 2010. The association of adolescent obesity with risk of severe obesity in adulthood. *JAMA,* **304**(18), pp.2042-2047.

TOBIN, G. A. and C. M. BEGLEY. 2004. Methodological rigour within a qualitative framework. *J Adv Nurs,* **48**(4), pp.388-396.

TOUBOUL, J. P. and J. R. I. CROUSE. 1997. *Intima-media thickness and atherosclerosis: predicting the risk?* New York: Parthenon Publishing Group.

TRICKEY, H., I. HARVEY, G. WILCOCK and D. SHARP. 1998. Formal consensus and consultation: a qualitative method for development of a guideline for dementia. *Quality in Health Care,* **7**, pp.192-199.

TU, S. H., Y. T. HUNG, H. Y. CHANG, C. M. HANG, N. S. SHAW, W. LIN, Y. C. LIN, S. W. HU, Y. H. YANG, T. C. WU, Y. H. CHANG, S. C. SU, H. C. HSU, K. S. TSAI, S. Y. CHEN, C. J. YEH and W. H. PAN. 2007. Nutrition and Health Survey of Taiwan Elementary School Children 2001-2002: research design, methods and scope. *Asia Pac J Clin Nutr,* **16**(S2), pp.507-517.

TUCKETT, A. G. 2005. Applying thematic analysis theory to practice: a researcher's experience. *Contemp Nurse,* **19**(1-2), pp.75-87.

TUNCELI, K., K. LI and W. L.K. 2006. Long-Term Effects of Obesity on Employment and Work Limitations Among U.S. Adults, 1986 to 1999. *Obesity,* **14**(9), pp.1637-46.

TZOU, I. L. and N. F. CHU. 2012. Parental influence on childhood obesity: A review *Health,* **4**, pp.1464-1470.

UK COLORECTAL CANCER SCREENING PILOT GROUP. 2004. Results of the first round of a demonstration pilot of screening for colorectal cancer in the United Kingdom. *BMJ,* **329**(133).

UK GOUT SOCIETY. 2014. *All about gout and diet* [online]. [Accessed 21st December]. Available from: http://www.ukgoutsociety.org/docs/goutsociety-allaboutgoutanddiet-0113.pdf.

UK NATIONAL SCREENING COMMITTEE. 1998. *First Report of the UK National Screening Committee.* London: Health Departments of the United Kingdom.

UK NATIONAL SCREENING COMMITTEE. 2000. *Second Report of the UK National Screening Committee.* London: Health Departments of the United Kingdom.

UK NATIONAL SCREENING COMMITTEE. 2013. *NHS population screening explained* [online]. [Accessed 6th January]. Available from: https://www.gov.uk/guidance/nhs-population-screening-explained.

UK NATIONAL SCREENING COMMITTEE. 2017. *Screening in the UK: Making effective recommendation (1 April 2016 to 31 March 2017)* [online]. [Accessed 5th May]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/649986/Screening_in_the_UK_making_effective_recommendations_2016_to_2017.pdf.

US DEPARTMENT OF HEALTH AND HUMAN SERVICES. 2007. *Isovaleric acidemia* [online]. [Accessed 2nd May]. Available from: https://ghr.nlm.nih.gov/condition/isovaleric-acidemia#statistics.

US DEPARTMENT OF HEALTH AND HUMAN SERVICES. 2016. *Homocystinuria* [online]. [Accessed 2nd May]. Available from: https://ghr.nlm.nih.gov/condition/homocystinuria#statistics.

US DEPARTMENT OF HEALTH AND HUMAN SERVICES. 2017a. *Maple syrup urine disease* [online]. [Accessed 2nd May]. Available from: https://ghr.nlm.nih.gov/condition/maple-syrup-urine-disease#statistics.

US DEPARTMENT OF HEALTH AND HUMAN SERVICES. 2017b. *Type 2 diabetes* [online]. [Accessed 2nd May]. Available from: https://ghr.nlm.nih.gov/condition/type-2-diabetes#statistics.

USDHHS, FDA, CDER, CBER and CDRH. 2009. *Guidance for industry: patient-reported outcome measures:use in medical product development to support labeling claims.*

VADON, R. 2007. Cost of Obesity 'over estimated' [online]. [Accessed 2nd Feb 2019]. Available from: http://news.bbc.co.uk/1/hi/health/7106219.stm.

VAISMORADI, M., H. TURUNEN and T. BONDAS. 2013. Content analysis and thematic analysis: Implications forconducting a qualitative descriptive study. *Nursing and Health Sciences,* **15**, pp.398-405.

VALDÉS, S., P. BOTAS, E. DELGADO, F. ÁLVAREZ and F. D. CADÓRNIGA. 2008. Does the new American Diabetes Association definition for impaired fasting glucose improve its ability to predict type 2 diabetes mellitus in Spanish persons? The Asturias Study. *Metabolism,* **57**(3), pp.399-403.

VAN DAM, R. M., W. C. WILLETT, J. E. MANSON and F. B. HU. 2006. The relationship between overweight in adolescence and premature death in women. *Ann Intern Med. ,* **145**(2), pp.91-7.

VAN DE VEN, A. H. and A. L. DELBECQ. 1971. Nominal versus Interacting Group Processes for Committee Decision-Making Effectiveness. *The Academy of Management Journal,* **14**(2), pp.203-212.

VAN DE VEN, A. H. and A. L. DELBECQ. 1972. The Nominal Group as a Research Instrument for Exploratory Health Studies. *American Journal of Public Health,* **62**(3), pp.337-342.

VAN JAARSVELD, C. H. M. and M. C. GULLIFORD. 2015. Childhood obesity trends from primary care electronic health records in England between 1994 and 2013: population-based cohort study. *Arch Dis Child.,* **0**, pp.1-6.

VAN RIJSSEN, L. B., A. GERRITSEN, I. HENSELMANS, M. A. SPRANGERS, M. JACOBS, C. BASSI, O. R. BUSCH, C. FERNÁNDEZ-DEL CASTILLO, Z. V. FONG, J. HE, J. Y. JANG, A. A. JAVED, S. W. KIM, L. MAGGINO, A. MITRA, V. OSTWAL, S. PELLEGRINI, S. V. SHRIKHANDE, J. W. WILMINK, C. L. WOLFGANG, H. W. VAN LAARHOVEN, M. G.

BESSELINK and C. S. GROUP. 2017. Core Set of Patient-reported Outcomes in Pancreatic Cancer (COPRAC): An International Delphi Study Among Patients and Health Care Providers. *Ann Surg.*

VAN TEIJLINGEN, E., E. PITCHFORTH, C. BISHOP and R. E. 2006. Delphi method and nominal group technique in family planning and reproductive health research. *J Fam Plann Reprod Health Care,* **32**(4), pp.249-252.

VAN WERSCH, A. and M. ECCLES. 2001. Involvement of consumers in the development of evidence based clinical guidelines: practical experiences from the North of England evidence based guideline development programme. *BMJ Quality and Safety,* **10**(1), pp.10-16.

VANDENBROECK, P., J. GOOSSENS and M. CLEMENS. 2007. *Tackling Obestities: Future Choices - Obesity System Atlas.* Department of Innovation Universities and Skills: Crown.

VAUGHN, S., J. S. SCHUMM and J. SINAGUB. 1996. *Focus group interviews in education and psychology.* London: Sage.

VELDWIJK, J., M. S. LAMBOOIJ, F. G. J. KALLENBERG, H. J. VAN KRANEN, A. L. BREDENOORD, E. DEKKER, H. A. SMIT and G. A. DE WIT. 2016. Preferences for genetic testing for colorectal cancer within a population-based screening program: a discrete choice experiment. *European Journal of Human Genetics,* **24**, pp.361-366.

VELLA, K., C. GOLDFRAD, K. ROWAN, J. BION and N. BLACK. 2000. Use of consensus development to establish national research priorities in critical care. *BMJ,* **320**(7240), pp.976-980.

VIERA, A. J. and J. M. GARRETT. 2005. Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine,* **37**(5), pp.360-363.

VILCHIS-GIL, J., M. GALVÁN-PORTILLO, M. KLÜNDER-KLÜNDER, M. CRUZ and S. FLORES-HUERTA. 2015. Food habits, physical activities and sedentary lifestyles of eutrophic and obese school children: a case–control study. *BMC Public Health,* **15**, p124.

WAGGONER, J., J. CARLINE and S. J. DURNING. 2016. Is There a Consensus on Consensus Methodology? Descriptions and Recommendations for Future Consensus Research. *Academic Medicine,* **91**(5), pp.663-668.

WAINWRIGHT, D., C. BOICHAT and L. M. MCCRACKEN. 2013. Using the nominal group technique to engage people with chronic pain in health service development. *The Iinternational Journal of Health Planning and Management,* **29**(1), pp.52-69.

WAKE, M., L. CANTERFORD, G. C. PATTON, K. HESKETH, P. HARDY, J. WILLIAMS, E. WATERS and J. B. CARLIN. 2010. Comorbidities of overweight/obesity experienced in adolescence: longitudinal study. *Arch Dis Child.,* **95**(3), pp.162-168.

WALES, M. E. 2009. *Understanding the role of convenience in consumer food choices: a review article* [online]. [Accessed 20th April]. Available from: https://journal.lib.uoguelph.ca/index.php/surg/article/view/983/1431.

WANG, Q., J. YIN, L. XU, H. CHENG, X. ZHAO, H. XIANG, H. S. LAM, J. MI and M. LI. 2013. Prevalence of metabolic syndrome in a cohort of Chinese schoolchildren: comparison of two definitions and assessment of adipokines as components by factor analysis. *BMC Public Health,* **13**, p249.

WANG, Y. and T. LOBSTEIN. 2006. Worldwide trends in childhood overweight and obesity. *Int J Pediatr Obes.,* **1**(1), pp.11-25.

WANG, Y. C., K. MCPHERSON, T. MARSH, S. L. GORTMAKER and M. BROWN. 2011. Health and economic burden of the projected obesity trends in the USA and the UK. *The Lancet,* **378**(9793), pp.815-825.

WANG, Z., C. M. PATTERSON and A. P. HILLS. 2002. A Comparison of Self-Reported and Measured Height, Weight and BMI in Australian Adolescents. *Australian and New Zealand Journal of Public Health,* **26**(5), pp.473-478.

WARDLE, J. and L. COOKE. 2005. The Impact of Obesity on Psychological Well-Being. *Best Practice & Research,* **19**(3), pp.421-440.

WARDLE, J., C. GUTHRIE, S. SANDERSON, L. BIRCH and R. PLOMIN. 2001. Food and activity preferences in children of lean and obese parents. *International Journal of Obesity,* **25**(7), pp.971-977.

WASHINGTON, D. L., S. J. BERNSTEIN, J. P. KAHAN, L. L. LEAPE, C. J. KAMBERG and P. G. SHEKELLE. 2003. Reliability of Clinical Guideline Development Using Mail-Only versus In-Person Expert Panels. *Medical Care,* **41**(12), pp.1374-1381.

WATMOUGH, D. J. and K. KUMAR. 1994. Screening programme not cost effective. *BMJ,* **308**, p202.

WEBB, D. R., L. J. GRAY, K. KHUNTI, B. SRINIVASAN, N. TAUB, S. CAMPBELL, J. BARNETT, A. FAROOQI, J. ECHOUFFO-TCHEUGUI, B., S. GRIFFIN, J., N. J. WAREHAM and M. J. DAVIES. 2011. Screening for diabetes using an oral glucose tolerance test within a western multi-ethnic population identifies modifiable cardiovascular risk: the ADDITION-Leicester study. *Diabetologia.,* **54**(9), pp.2237-46.

WELLS, G. A., B. SHEA, D. O'CONNELL, J. PETERSON, V. WELCH, M. LOSOS and P. TUGWELL. 2014. *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses* [online]. [Accessed 25th April]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.

WELSH, J. A., S. KARPEN and M. B. VOS. 2013. Increasing prevalence of nonalcoholic fatty liver disease among United States adolescents, 1988-1994 to 2007-2010. *J Pediatr. ,* **162**(3), pp.496-500.

WHITAKER, R. C., J. A. WRIGHT and M. S. PEPE. 1997. Predicting obesity in young adulthood from childhood and parental obesity. *New England Journal of Medicine,* **337**, pp.869-873.

WHITLOCK, E. P., E. A. O'CONNOR, S. B. WILLIAMS, T. L. BEIL and K. W. LUTZ. 2010. Effectiveness of Weight Management Interventions in Children: A Targeted Systematic Review for the USPSTF. *Pediatrics,* **125**(2), pp.396-418.

WHO. 2000. WHO Technical Report Series 894. *Obesity: Preventing and Managing the Global Epidemic*. Geneva, Switzerland: World Health Organisation.

WHO. 2003. *Screening for Type 2 Diabetes*. Geneva, Switzerland: World Health Organisation.

WHO. 2014a. *Childhood overweight and obesity* [online]. [Accessed 13th December]. Available from: http://www.who.int/dietphysicalactivity/childhood/en/.

WHO. 2014b. *WHO Handbook for Guideline Development (2nd edition).* Geneva, Switzerland: WHO Press.

WHO. 2016. *Factsheets: Obesity and overweight* [online]. [Accessed 19th December]. Available from: http://www.who.int/mediacentre/factsheets/fs311/en/.

WIEGAND, S., K. M. KELLER, M. ROBL, D. L'ALLEMAND, T. REINEHR, K. WIDHALM and R. W. HOLL. 2010. Obese boys at increased risk for nonalcoholic liver disease: Evaluation of 16 390 overweight or obese children and adolescents. *International Journal of Obesity,* **34**(10), pp.1468-1474.

WILKIE, H. J., M. STANDAGE, F. B. GILLISON, S. P. CUMMING and P. T. KATZMARZYK. 2016. Multiple lifestyle behaviours and overweight and obesity among children aged 9–11 years: results from the UK site of the International Study of Childhood Obesity, Lifestyle and the Environment. *BMJ Open,* **6**(e010677).

WILSON, J. M. G. and G. JUNGNER. 1968. *Principles and Practice of Screening for Disease.* 34. Geneva: World Health Organisation.

WISE, J. 2016. Weight loss drugs are compared for effectiveness. *BMJ,* **353**(i3330).

WISKIN, A. E., J. H. DAVIES, S. A. WOOTTON and R. M. BEATTIE. 2011. Energy expenditure, nutrition and growth. *Arch Dis Child.,* **96**(6), pp.567-72.

WOODROW, C., L. ROZMOVITS, P. HEWITSON, P. ROSE, J. AUSTOKER and E. WATSON. 2006. Bowel Cancer screening in England: a qualitative study of GPs' attitudes and information needs. *BMC Fam Pract,* **7**(53).

WOOLF, S. H. 1992. Practice Guidelines, a new reality in medicine: Methods of Developing Guidelines. *Arch Intern Med,* **152**, pp.946-952.

WOOLFORD, S. J., S. J. CLARK, A. GEBREMARIAM, M. M. DAVIS and G. L. FREED. 2010. To Cut or Not to Cut: Physicians' Perspectives on Referring Adolescents for Bariatric Surgery. *Obes Surg.,* **20**(7), pp.937-942.

WORLD HEALTH ORGANISATION. 2010. *WHO Guidelines on Drawing Blood: Best Practices in Phlebotomy* [online]. [Accessed 9th April]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK138647/.

WORLD HEALTH ORGANISATION. 2014. *Ambient Air Pollution Database* [online]. [Accessed 18th December]. Available from: http://www.who.int/quantifying_ehimpacts/national/countryprofile/AAP_PM_database_May2014.xls.

WORLD HEALTH ORGANISATION. 2017. *Obesity and overweight* [online]. [Accessed 17th May]. Available from: http://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight.

WORLD HEALTH ORGANISATION. 2018. *Global school-based student health survey* [online]. [Accessed 2nd January]. Available from: http://www.who.int/ncds/surveillance/global-school-student-survey/datasets/en/.

WORLD HEALTH ORGANISATION. n.d. *What is the double burden of malnutrition?* [online]. [Accessed 6th October]. Available from: http://www.who.int/childgrowth/4_double_burden.pdf.

WORRALL-DAVIES, A. and F. MARINO-FRANCIS. 2008. Eliciting Children's and Young People's Views of Child and Adolescent Mental Health Services: A Systematic Review of Best Practice. . *Child and Adolescent Mental Health,* **13**(1), pp.9-15.

XU, H., A. LIU, Q. ZHANG, X. HU, H. FANG, T. LI, H. GUO, Y. LI, G. XU, J. MA, L. DU and G. MA. 2012. Prevalence of the metabolic syndrome among children from six cities of China. *BMC public health,* **12**, p13.

XU, Y. and C. JI. 2008. Prevalence of the metabolic syndrome in secondary school adolescents in Beijing, China. *Acta Paediatrica, International Journal of Paediatrics,* **97**(3), pp.348-353.

YANG, C. L., E. SIMONS , R. FOTY, P. SUBBARAO, T. TO and S. D. DELL. 2017. Misdiagnosis of asthma in schoolchildren. *Pediatric Pulmonology,* **52**(3), pp.293-302.

YARDLEY, L. 2000. Dilemmas in qualitative health research. *Psychology and Health,* **15**(2), pp.215-228.

YOON, J. M. 2014. Dyslipidemia in Children and Adolescents: When and How to Diagnose and Treat? *Pediatr Gastroenterol Hepatol Nutr.,* **17**(2), pp.85-92.

YOUSUF, M. I. 2007. Using Experts' Opinions Through Delphi Technique. *Practical Assessment, Research and Evaluation,* **12**(4), pp.1-8.

ZAKERI, M., M. SEDAGHAT, M. E. MOTLAGH, R. TAYARI ASHTIANI and G. ARDALAN. 2012. BMI correlation with psychiatric problems among 10-18 years Iranian students. *Acta Medica Iranica,* **50**(3), pp.177-84.

ZGAGA, L., E. THEODORATOU, J. KYLE, S. M. FARRINGTON, F. AGAKOV, A. TENESA, M. WALKER, G. MCNEILL, A. F. WRIGHT, I. RUDAN, M. G. DUNLOP and H. CAMPBELL. 2012. The Association of Dietary Intake of Purine-Rich Vegetables, Sugar-Sweetened Beverages and Dairy with Plasma Urate, in a Cross-Sectional Study. *PLoS One,* **7**(6), pe38123.

ZHENG, M., A. RANGAN, B. MEERTENS and J. H. Y. WU. 2017. Changes in Typical Portion Sizes of Commonly Consumed Discretionary Foods among Australian Adults from 1995 to 2011–2012. *Nutrients,* **9**(6), pp.577-586.

ZIEMBA, S. E., F. A. HUBBELL, M. J. FINE and M. J. BURNS. 1991. Resting electrocardiograms as baseline tests: impact on the management of elderly patients. *Am J Med,* **91**(6), pp.576-583.

ZIMMET, P., G. ALBERTI, J. SHAW and S. M. GRUNDY. 2007. *The IDF consensus worldwide definition of the Metabolic Syndrome* [online]. [Accessed 28th September]. Available from: https://www.idf.org/webdata/docs/IDF_Meta_def_final.pdf.

Appendices

Appendix 1: Medline Search Strategy

Below is the search criteria used for Medline followed by the first 5 search results.

Database: Ovid MEDLINE® <1946 to February Week 4 2015>

Search Strategy:

Prevalence/ (197910)

Cross-Sectional Studies/ (186239)

prevalen*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (495489)

4   cross-sectional*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (243011)

5   1 or 3 (495489)

6   2 or 4 (243011)

7   5 or 6 (676636)

8   Obesity/ or Obesity, Morbid/ or Obesity Hypoventilation Syndrome/ or Obesity, Abdominal/ or Pediatric Obesity/ (143072)

9   body constitution/ or "body weights and measures"/ or body fat distribution/ or adiposity/ or body mass index/ or body size/ or body weight/ or overweight/ or obesity/ or obesity, abdominal/ or obesity, morbid/ or sagittal abdominal diameter/ or waist circumference/ or waist-height ratio/ or body surface area/ or skinfold thickness/ or waist-hip ratio/ (360030)

10   body weight/ or weight gain/ or overweight/ or obesity/ or obesity, morbid/ (308427)

11   obes*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (207223)

12   adiposi*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (16976)

13   overweight*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (38307)

14   BMI*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (73604)

15   waist circumference*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (14417)

16   neck circumference*.mp. (563)

17   body mass index*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (135630)

18   body mass index/ (83547)

19   8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 (480959)

20   causality/ or risk factors/ or comorbidity/ (640870)

21   illness.mp. (345281)

22   Disease/ or disease.mp. (2954606)

23   comorb*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (117136)

24   co-morb*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (13993)

25   depression.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (265182)

26   Depression/ep, me, pc, px [Epidemiology, Metabolism, Prevention & Control, Psychology] (36237)

27   self esteem.mp. or Self Concept/ (51419)

28   self-esteem.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (13143)

29   diabetes.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (397183)

30   Diabetes Mellitus, Type 2/co, dh, ep, me, pc, px [Complications, Diet Therapy, Epidemiology, Metabolism, Prevention & Control, Psychology] (48984)

31   Diabetes Mellitus, Type 1/ (61783)

32   29 or 30 (397183)

33   32 not 31 (335400)

34   Sleep Apnea, Obstructive/ep, pc, px [Epidemiology, Prevention & Control, Psychology] (2355)

35   Sleep Disorders/ep, pc, px [Epidemiology, Prevention & Control, Psychology] (4823)

36   Sleep Apnea Syndromes/ or sleep apnea*.mp. (26949)

37   20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 33 or 34 or 35 or 36 (4008830)

38   Child/ (1377338)

39   Pediatrics/ (41366)

40   Adolescent/ (1633164)

41   Students/ (35858)

42   boy*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (107551)

43   girl*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (102623)

44   youth*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (43553)

45   teen*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (20841)

46   child*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (1865059)

47   school-age*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (13577)

48   juvenil*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (65544)

49   student*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (192613)

50   preteen*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (217)

51   pedia*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (211675)

52   paedi*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (38419)

53   adoles*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (1657230)

54   38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 (2941874)

55   7 and 19 and 37 and 54 (13201)

56   limit 55 to animals (184)

57   55 not 56 (13017)

58   limit 57 to ("all infant (birth to 23 months)" or "all adult (19 plus years)" or "newborn

infant (birth to 1 month)" or "infant (1 to 23 months)") (8173)

59   57 not 58 (4844)

60   Anorexia Nervosa/ or Anorexia/ (14831)

61   Bulimia Nervosa/ or Bulimia/ (6373)

62   bulimi*.mp. [mp=title, abstract, original title, name of substance word, subject heading

word, keyword heading word, protocol supplementary concept word, rare disease

supplementary concept word, unique identifier] (8412)

63   anorex*.mp. [mp=title, abstract, original title, name of substance word, subject heading

word, keyword heading word, protocol supplementary concept word, rare disease

supplementary concept word, unique identifier] (27773)

64   60 or 61 or 62 or 63 (32023)

65   59 not 64 (4754)

Appendix 2: Data Extraction Form

Extraction Details:

Endnote Identifier The unique identifier given to the article by Endnote.	
Lead Author	
Year	
Data Extractor The person extracting the data	☐ VS ☐ MB ☐ JN ☐ LS ☐ SC ☐ Other: _____
Date of Extraction	

Study Details

Details of the study, e.g. design, country etc.

Country - Location The country where the study was conducted	
Setting Where was the study conducted?	
Study Design	☐ Cross Sectional (prospective) ☐ Cross Sectional (retrospective) ☐ Cohort ☐ Case Series ☐ Case Control ☐ Other:_____
Number of time points The number of time points included in the article.	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5+, please state _____
Inclusion criteria:	
Exclusion criteria:	

Participant Details

Population/Sample description Summary of population and sample	
How were participants recruited? Any indication the sample was representative of the target population?	
Was a sample size calculation done?	☐ Yes, sufficient details given for calculation (e.g. power, error, effect size, etc.) ☐ Yes, lack of calculation details ☐ Not required, e.g. large national survey ☐ No/Not stated
Calculated sample size:	
Participant characteristics reported? E.g. mean age, gender split, ethnicity split (where applicable), age split (where applicable), SES etc.	☐ Yes ☐ No Comments:
Initial sample size The total number of people recruited	Overall:_____ Male:_____ Female:_____ ☐ Not reported
Final sample size The total number of people included in the analysis, after drop-outs, non-responders, missing data etc.	Overall:_____ Male:_____ Female:_____ ☐ Not reported
Reasons for drop-outs, missing data provided?	☐ Yes ☐ No ☐ Not applicable (everyone took part)
☐ Mean ☐ Median age of the participants	Overall: _____ Male: _____ Female: _____ ☐ Not reported
Minimum age Lowest age of the participants	Overall: _____ Male: _____ Female: _____ ☐ Not reported
Maximum age: Highest age of the participants	Overall: _____ Male: _____ Female: _____ ☐ Not reported
Ethnicity (ET) List how the participants were grouped by ethnicity in the analysis, e.g. "white" and "non-white", if "other" is stated, provide	ET1. ET2. ET3.

details of what this includes.	ET4. ☐ Not stated

Categorisation of Weight and cut-offs

Weight Categorisation method used for analysis Some studies may consider more than one method, if so, enter the method used to categorise participants for the analysis	☐ BMI ☐ Standardised BMI ☐ BMI Percentile ☐ Body Fat ☐ Waist circumference ☐ Other: _____
Weight Category (WC) cut-offs Provide the names for the categories and the cut-offs, e.g. if the study used BMI percentiles (ticked in the previous question), for this question you might enter: 0. Healthy weight; $5^{th} \leq$ BMI $<85^{th}$ 1. Overweight; $85^{th} \leq$ BMI $<95^{th}$ 2. Obese; $\geq 95^{th}$ 3. N/A Enter "X" if the cut-offs are not provided And provide a reference for the cut-offs if provided.	WC0: Cut-off: WC1: Cut-off: WC2: Cut-off: WC3: Cut-off: Reference:

Co-morbidity Details

There may be multiple co-morbidities considered in the article, list details for each one.

Co-morbidity 1 Name	
Co-morbidity Definition State if the study uses a national or international definition, e.g. "International Diabetes Federation definition for metabolic syndrome" and reference.	

Assessment(s) conducted and cut-off(s)
If more than one assessment pertaining to the co-morbidity is conducted list them all.
Confirm if self-reported (SR), conducted for study (CS) or taken from case notes (CN)

Name/Type & Reference	Cut-off(s)	SR	CS	CN

Co-morbidity N Name	
Co-morbidity Definition State if the study uses a national or international definition, e.g. "International Diabetes Federation definition for metabolic syndrome" and reference.	

Assessment(s) conducted and cut-off(s)
If more than one assessment pertaining to the co-morbidity is conducted list them all.
Confirm if self-reported (SR), conducted for study (CS) or taken from case notes (CN)

Name/Type & Reference	Cut-off(s)	SR	CS	CN

Prevalence Data

Year(s) data collected	Gender 0. male 1. female 2. mixed	Age Group/ Pubertal status	Ethnicity	Nationality	Weight Category 0. H/W 1. O/W 2. Ob	Co-morbidity	Co-morbidity definition (if multiple compared in study, e.g. MS) List reference, e.g. IDF 2007	Co-morbidity severity, e.g. 0. Mild 1. Moderate 2. Severe 3. Overall	n	N

Appendix 3: Blood Test Acceptability Questionnaires

Weight Management Staff Questionnaire

As part of some research looking into screening children for weight-related illnesses, such as diabetes, we would like to ask you some questions about taking a sample of blood from a child/adolescent

1) Would you be comfortable taking a blood sample from a child/adolescent? (training would be provided)

○ Yes

○ No, because

_____

If yes:

2) Would you prefer taking a blood sample by:

○ A needle in the arm to take up to 4 teaspoons of blood

○ A needle-prick to the finger to take a drop of blood

○ Either

Any further comments?

_____

_____

3) What is your job title/role?

_____

4) What age of children/adolescents do you work with?

_____

Thank You

Parent and Child Questionnaire

Introduction: As part of some research looking into screening children for weight-related illnesses, such as diabetes, we would like to ask you and your child some questions about having a sample of blood taken by a trained member of staff.

To start, I'd like to get some information about your child

a) How old is your child? _____

b) What is your child's gender? ◯ Boy   ◯ Girl

Questions for parent:

1) Would you be happy for staff to take a blood sample from your child?

◯ Yes - please go to question 2

◯ No,

because_____

2) Which would you find acceptable to be done?

◯ A needle in the arm to take up to 4 teaspoons of blood

◯ A needle-prick to the finger to take a drop of blood

◯ Either

Questions for child:

1) Would you be happy to have a sample of blood taken from you?

◯ Yes – please go to question 2

◯ No, because

_____

2) How would you like the blood sample to be taken?

◯ A needle in the arm to take up to 4 teaspoons of blood

◯ A needle-prick to the finger to take a drop of blood

◯ Either

Any other comments?

_____

_____

_____

Thank you for your time