EXPLORE AND DEVELOP METHODS FOR THE ECONOMIC EVALUATION OF SCHOOL-BASED INTERVENTIONS TO PREVENT CHILDHOOD OBESITY IN LOW AND MIDDLE INCOME COUNTRIES

by

MANDANA ZANGANEH

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Health Economics Unit

Institute of Applied Health Research

College of Medical and Dental Sciences

University of Birmingham

September 2019

ABSTRACT

Childhood obesity is a major global public health challenge with associated health, social, and emotional consequences, leading to long term direct and indirect costs. However, there are few published economic evaluations of interventions and only one from a Chinese setting. This thesis aims to explore and develop methods for the economic evaluation of school-based interventions to prevent obesity in children in low and middle income countries, thus making a methodological contribution to the literature.

The methods for the economic evaluation were derived from a combination of published literature and guidelines for conducting economic evaluation. The systematic review undertaken within this thesis discovered heterogeneity regarding methods applied. The evaluation, conducted alongside the CHIRPY DRAGON trial, reported the intervention to be highly cost-effective. A number of methodological issues were explored: measuring household cost and outcome data and the construct validity of the CHU-9D in a Chinese sample. Including societal costs and effects increased the incremental cost-effectiveness ratio, however the intervention remained cost-effective using conventional decision making rules and throughout a series of sensitivity analyses. Furthermore, the thesis findings provide support for the construct validity of the CHU-9D within this population.

To Ebrahim, Maryam, Mahshid and Mahtab Zanganeh and Scott Paton

“Education is the most powerful weapon which you can use to change the world.”

Nelson Mandela

“When faced with a challenge, look for a way, not a way out.”

David Weatherford

ACKNOWLEDGEMENTS

I would like to thank my supervisors Emma Frew, Peymane Adab and Bai Li for their constant supervision, guidance, support and encouragement throughout the three years. I am very grateful to them for making time to answer my questions, read all sections of my thesis chapters and papers, and providing invaluable feedback despite their busy schedules.

I would like to thank all people contributing to the data collection within the CHIRPY DRAGON trial, the project management and the participating families. Special thanks to the investigators of the trial. I gratefully acknowledge the PhD funding from the College of Medical and Dental Sciences.

I extend my gratitude to people at the Obesity Prevention and Management Research Group and the Health Economics Unit (HEU) at Birmingham; the Association for the Study of Obesity, Newcastle; the Health Economics Study Group, Norwich; and the 13th International Health Economics Association, Basel for their helpful and valuable comments and suggestions as they have helped to shape the ideas and have provided thought provoking discussions on the direction of the research.

I thank the HEU staff for the supportive environment: Raymond Oppong and Hareth Al-Janabi who shared tips on economic evaluation methods and analysis; Louise Jackson for being my internal assessor and also providing valuable feedback on my Annual Progress Report. I also thank the Institute of Applied Health Research statisticians at Birmingham, James Martin and Sayeed Haque, for their statistical advice for chapters 5 and 6. In addition, I thank Sue Bayliss and Jon Andrews for providing guidance in determining the search strategy for the systematic review. I thank my fellow PhD students for keeping my spirits up and making the whole experience so enjoyable, especially James Hall who was the 2nd reviewer for the systematic review.

I thank my friends and family for their encouragement. I am very grateful to my parents (Ebrahim, Maryam) and sisters (Mahshid, Mahtab). I thank them for their love and support, despite the distance we are apart, and for taking an interest in my research. I especially thank my mother for her endless encouragement.

This thesis was copy edited for conventions of language, spelling and grammar by my husband Scott. I am very grateful to him for this and for his love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ATP	Ability-to-Pay
BHIS	Basic Health Insurance Scheme
BMI	Body Mass Index
BN	Bulimia Nervosa
CBA	Cost-Benefit Analysis
CCA	Cost-Consequence Analysis
CDC	Centre for Disease Control and Prevention
CDSR	Cochrane Database of Systematic Reviews
CEA	Cost-Effectiveness Analysis
CEAC	Cost-Effectiveness Acceptability Curve
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CENTRAL	Cochrane Central Register of Controlled Trials
CHU-9D	Child Health Utility Nine Dimension
CI	Confidence Interval
CNHDRC	China National Health Development Research Centre
CMA	Cost-Minimisation Analysis
CRD	Centre for Reviews and Dissemination
CUA	Cost-Utility Analysis
DAC	Development Assistance Committee
DALYs	Disability-Adjusted Life Years
DARE	Database of Abstracts of Reviews of Effects

DSA	Deterministic Sensitivity Analysis
DWCB	Disordered Weight Control Behaviours
EQ-5D	EuroQoL Five Dimension
GDP PPPs	Gross Domestic Product Purchasing Power Parities
GP	General Practitioner
GNP	Gross National Product
HMIC	Healthcare Management Information Consortium
HRQoL	Health-Related Quality of Life
HTA	Health Technology Assessment
HUI	Health Utility Index
ICER	Incremental Cost-Effectiveness Ratio
IOTF	International Obesity Task Force
IQR	Inter-Quartile Range
ITT	Intention-To-Treat
LEAP	Live, Eat and Play
LMICs	Low and Middle Income Countries
MD	Mean Difference
MeSH	Medical Subject Headings
MET	Metabolic Equivalents
MI	Motivational Interviewing
MRC	UK Medical Research Council
MVPA	Moderate to Vigorous Physical Activity
NB	Net Benefit
NBR	Net Benefit Regression

NCMS	New Cooperative Medical System
NE	North East
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NW	North West
OECD	Organisation for Economic Co-operation and Development
PA	Physical Activity
PBAC	Pharmaceutical Benefits Advisory Committee
PE	Physical Education
PedsQL	Paediatric Quality of Life Inventory
PSA	Probabilistic Sensitivity Analysis
PSS	Personal Social Services
QALYs	Quality-Adjusted Life Years
QoL	Quality of Life
RCT	Randomised-Controlled Trial
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SCT	Social Cognitive Theory
SD	Standard Deviation
SE	South East
SES	Socio-Economic Status
SSB	Sugar Sweetened Beverage
SF-6D	Short Form Six Dimension
SF-36/12	Short Form 36/12
SW	South West

UK	United Kingdom
USA	United States of America
WC	Waist Circumference
WHO	World Health Organization
WHtR	Waist-to-Height Ratio
WTA	Willingness-to-Accept
WTP	Willingness-to-Pay

DISSEMINATION

LIST OF SCIENTIFIC PUBLICATIONS

1. **Zanganeh M**, Adab P, Li B, Frew E. Protocol for a systematic review of methods and cost-effectiveness findings of economic evaluations of obesity prevention and/or treatment interventions in children and adolescents. *BMC Systematic Reviews*. 2018;54(7):1-7.
2. **Zanganeh M**, Adab P, Li B, Frew E. A systematic review of methods, study quality, and results of economic evaluation for childhood and adolescent obesity intervention. *International Journal of Environmental Research and Public Health*. 2019;16(3):485-499.
3. Li B, Pallan M, Liu WJ, Hemming K, Frew E, Lin R, Liu W, Martin J, **Zanganeh M**, Hurley K, Cheng KK, Adab P. The CHIRPY DRAGON intervention in preventing obesity in Chinese primary school-aged children: A cluster-randomised controlled trial. *Plos Medicine*. 2019;16(11):1-20.

CONFERENCE PAPERS, ORAL & POSTER PRESENTATIONS

<u>Conference Papers</u>

1. **Zanganeh M**, Adab P, Li B, Frew E. An assessment of the construct validity of the CHU-9D in school-aged children: Evidence from a Chinese trial. Health Economists' Study Group (HESG) conference, University of East Anglia, Norwich, UK. 3rd July 2019.
2. **Zanganeh M**, Adab P, Li B, Frew E. Cost-effectiveness of the CHIRPY DRAGON obesity prevention intervention in Chinese primary school-aged children: A cluster-

randomised controlled trial. International Health Economics Association (iHEA) pre-congress session (Economics of Obesity), University of Basel, Switzerland. 14th July 2019.

Oral Presentations as Part of Organised Sessions

1. **Zanganeh M**, Adab P, Li B, Frew E. Cost-effectiveness of the CHIRPY DRAGON obesity prevention intervention in Chinese primary school-aged children: A cluster-randomised controlled trial. iHEA Congress, University of Basel, Switzerland. 15th July 2019.
2. **Zanganeh M**, Adab P, Li B, Frew E. The association between weight and health-related quality of life in children: Evidence from a Chinese trial. iHEA Congress, University of Basel, Switzerland. 17th July 2019.

The above were also presented at the Health Economics Unit (HEU) at the University of Birmingham in June 2019.

Poster Presentations

1. **Zanganeh M**, Adab P, Li B, Frew E. A systematic review of methods, study quality, and results of economic evaluation for childhood and adolescent obesity intervention. Association for the Study of Obesity (ASO) conference, University of Newcastle, UK. 5th – 7th September 2018.
2. **Zanganeh M**, Adab P, Li B, Frew E. Cost-effectiveness of the CHIRPY DRAGON obesity prevention intervention in Chinese primary school-aged children: A cluster-randomised controlled trial. ASO conference, University of Newcastle, UK. 5th – 7th

September 2018.

These were also presented at the Postgraduate Research Festival within the College of Medical and Dental Sciences at the University of Birmingham, the first in May 2018 & the second in May 2019. In addition, the systematic review was presented as an oral presentation at the Obesity Prevention and Management Research Group (OPMRG) meeting in November 2017 and at HEU in March 2018, both at the University of Birmingham.

Additionally, after the submission of this thesis, the findings presented in chapters 5 and 6 will be disseminated through publications.

CHAPTER 1. INTRODUCTION

This thesis aims to contribute towards reducing the burden of childhood obesity and helping public health decision making in low and middle income countries (LMICs). This has been done by exploring and developing methods for the economic evaluation of obesity interventions and was facilitated by estimating the cost-effectiveness of a childhood obesity programme in a school setting, using evidence from the CHIRPY DRAGON trial. This introductory chapter summarises the background behind the research question. It also describes the aims and objectives, and provides a summary of each of the remaining six chapters within the thesis.

1.1 Childhood Obesity

Childhood obesity, defined as "abnormal or excessive fat accumulation", is one of the biggest public health challenges of this time with associated health, social, and emotional consequences, as well as long term direct and indirect costs [1-4]. Rapid socioeconomic and nutritional transitions in urban Chinese populations over a relatively short period have contributed to the rising prevalence of overweight/obesity among children [5-8]. In some populations, this prevalence is approaching the level of high-income countries [9, 10] and unlike Western countries, which are at a more advanced stage of the obesity epidemic, obesity prevalence, in China, is positively associated with socio-economic status (SES), particularly in boys [10, 11]. Most recent national data report that in China, 42% of adults and around one-fifth of children are overweight or obese [12]. A Chinese case study found that the indirect effects of obesity and obesity-related dietary and physical activity patterns reached 3.58% of gross national product (GNP) in 2000 and was projected to reach 8.73% in 2025 [13]. Therefore, childhood obesity is a cause of concern for several stakeholders (including school teachers and local authorities) within China.

Because of the health consequences of overweight and obesity and the lack of sustainable treatment options, prevention is likely to be the most (cost-) effective approach to address the childhood obesity problem in society. In 2016, the Chinese Government launched the "Healthy China 2030" policy, which seeks to improve health standards in China to be on a par with that in developed countries, through a range of initiatives including health promotion and improvement of public health services [14]. To achieve this grand vision there is an urgent need for effective preventive interventions to address the rapid increase of obesity prevalence. It is vital to develop effective, culturally appropriate, prevention interventions in China to control the obesity epidemic in children.

A few Chinese studies have demonstrated that comprehensive school-based interventions, which targeted diet and physical activity, were effective [15-17]. However, there has been little research applying rigorous methods and established theoretical tools/framework to develop and evaluate prevention interventions for such a population [18-20], and only one study reported on the cost-effectiveness of a childhood obesity intervention (this study was not conducted from a societal perspective and only included clinical outcome measures) [17].

In China, obesity prevalence has been increasing year on year, and childhood obesity is a growing concern for health professionals and policy makers alike. The Centre for Disease Control and Prevention (CDC) has called for research on the costs and benefits of strategies to prevent childhood obesity. The CHIRPY DRAGON obesity prevention intervention study commenced in 2015. This intervention began development in 2009, using guidelines from the UK Medical Research Council (MRC) framework for complex interventions [21, 22], in consultation with parents, grandparents, teachers and school staff. It targets children and their families, encouraging healthy eating and physical activity behaviours and is delivered in a school setting.

1.2 Economic Evaluation and Outcome Measures

Public health priorities vary from country to country, and also from region to region. Like many other countries, China suffers from a scarcity of public health resources and decision makers need to prioritise spending towards policies that offer the greatest value for money [23, 24]. Economic evaluation is a means to aid decisions about public resource allocation [24-26] and as obesity prevention and treatment often involves lifestyle interventions which have costs and consequences that fall outside the health care sector, a societal perspective for evaluation is usually recommended [26]. This means that all relevant resource use/costs and consequences are measured, outlining how these fit within a given sector, such as health, education or the wider community [27]. However, when incorporating costs and outcomes that span multiple sectors, it is not always clear how much society is willing to pay for a ‘health’ effect caused by an intervention funded from a ‘non-health care budget’ [28]. Also, the valuation of resources for which no market exists, such as informal care, or patient time costs (e.g travel to appointments), requires specific methods [23].

Within economic evaluations of clinical interventions, outcomes are often measured in natural units or quality-adjusted life years (QALYs). For the economic evaluation of public health interventions, other outcomes might be relevant including effects on individuals not directly targeted by the intervention and other non-health related effects such as educational or wider wellbeing outcomes. Some of these effects can be incorporated into QALYs, some not [29]. Costs and benefits may fall on parts of the public sector not confined to health alone, such as the judicial system, education and housing. There is a consensus within the economic guidance that wider social and environmental costs and benefits should be looked at due to the complex nature of public health [29].

Implementation of a particular intervention is not recommended without evidence of both the effectiveness and cost effectiveness relative to usual practice [30]. To conduct an economic evaluation, information on the costs and benefits of competing interventions are considered. The findings of cost-effectiveness analyses will help to inform decisions on whether to implement multi-component interventions such as CHIRPY DRAGON within primary schools as so far, little is known about the costs of school-based childhood obesity prevention interventions in China.

Health Technology Assessment (HTA) is a field of scientific policy research that adopts multidisciplinary approaches to undertake systematic evaluation of health technologies and inform policy and clinical decision-making [31]. There has been rapid development of HTA in China over the past two decades, and with the introduction of universal health care coverage and intensive collaborations of the China National Health Development Research Centre (CNHDRC) with NICE in the UK, the remit of HTA in China is continually expanding to support and guide decision-making for policy makers [32, 33]. HTA now includes a multitude of topics, such as drug resource allocation, medical devices, procedures, and vaccines [32]. However, despite these developments, since HTA in China remains fragmented and is not yet formally integrated into health policy as a mandatory component, health policy and decision-making in China still largely relies on experience rather than research evidence [31].

In health economics there are two main alternative frameworks to conducting economic evaluations. These are welfarism and extra-welfarism. The theoretical and methodological basis for carrying out economic evaluations is different within these frameworks, as each uses different value judgements to determine various states of resource allocation. Briefly, the aim of welfarism is to maximise social welfare whilst the aim of extra-welfarism is to maximize the total health of a population and allows the consideration of outcomes beyond utility. Extra-welfarism goes beyond welfarism in four ways: 'it permits the use of outcomes other than utility, it permits source of valuation other than the affected individuals, it permits the weighting of outcomes other than preference-based, and it permits interpersonal comparisons of well-being' [34].

Currently, decision-makers, such as the National Institute for Health and Care Excellence (NICE), recommend the use of extra-welfarism as a framework for conducting economic evaluations, particularly cost-utility analysis (CUA) [35]. In a CUA, the costs are valued in monetary units and the effects are valued as a multi-dimensional unit (e.g. QALYs) [36]. The QALY incorporates length and quality of life (QoL) in a single metric. QALYs are used as the unit of assessment to make judgements about the relative cost-effectiveness of competing interventions [36] and require an understanding of the relationship between weight and health related QoL when measured in utility terms. QALYs are also used to inform resource allocation decisions in other country-settings [37].

Ideally, utility-based health-related quality of life (HRQoL) in children should be measured using an instrument specifically designed for them. Although there is no gold standard for measuring utility-based HRQoL in primary school-aged children, previous research has shown the Child Health Utility Nine Dimension (CHU-9D), a recently developed instrument, is the most appropriate choice [38]. It is not specific to any one condition or disease. Originally tested for children aged 7–11 years [39, 40], it has more recently demonstrated good construct validity in 11–17 year olds [41]. The tool has successfully been applied to wider populations, from six years old up to 17 years old [42, 43]. Although there is emerging evidence regarding the psychometric properties of the CHU-9D instrument [41, 42, 44], more evidence is required with respect to its validity for use in different age groups and country settings considering different tariffs. This is important because the measure may have different construct validity in different populations.

The acknowledgement that behaviour change interventions have spillover effects on family members has led to an increased interest on how to adapt methods for capturing these broader effects to maximise population health, rather than just the target participant's health [45]. The choice of evaluative space of health (e.g. EuroQol Five Dimension (EQ-5D)), which is designed to measure a generic related HRQoL and is not specific to any one condition or disease, for economic evaluation is an important value judgment which can have a large impact on resource allocation decisions [30]. Currently EQ-5D is recommended by decision-makers to generate QALYs for adults [30]. However, when EQ-5D is not considered to be suitable for a condition, decision-makers will accept the QALY outcome derived from another HRQoL measure [30].

For economic evaluation, the value that individuals place on healthcare is usually assessed using measures that assess preferences for possible outcomes (preference-based measures). This basis for measuring outcomes has been used in the case of the extra-welfarist (CUA measures: e.g. CHU-9D and EQ-5D) [36].

When costs are valued in monetary units and benefits are measured in natural units (which are specific to the condition under analysis, e.g. Body Mass Index (BMI) z-score change), the economic evaluation is termed a cost-effectiveness analysis (CEA) [36]. CEA is useful to compare interventions, which target the same health condition, and is particularly useful in a clinical setting. Although a CEA is an extra-welfarist evaluation, a major limitation of it for decision makers is its inability to directly compare cost-effectiveness of interventions across various areas of health conditions, or sectors of the economy, due to the disease-specific nature of the outcome measure used [46].

1.3 Aim and Objectives

To address the evidence gap, in China, of what interventions to implement to prevent childhood obesity and to address the methodological challenges of conducting an economic evaluation within this setting, the CHIRPY DRAGON study was developed. Using data from this study, the aim of this thesis is to (1) explore and develop methods for the economic evaluation of school-based interventions to prevent obesity in children in LMICs, and (2) consider the suitability of economic outcome measures for interventions in this age group and country setting. This will be useful, more broadly, to inform the design of future economic evaluations with the aim of generating economic evidence to assist decision makers in LMICs.

The thesis has the following objectives:

- Systematically review the literature to identify the current evidence regarding methods, study quality, and results of economic evaluation for childhood and adolescent obesity interventions. There is a lack of synthesised evidence available on appropriate methodology for the economic evaluation of obesity interventions. Undertaking this review will add to the evidence available.
- Discuss the methodology of the CHIRPY DRAGON cluster-randomised controlled trial regarding its economic evaluation.
- Explore the methodological challenges of conducting an economic evaluation within a

Chinese setting and including spillover effects. The findings of this objective will develop methods for the economic evaluation of school-based interventions to prevent obesity in children in low and middle income countries as, so far, little is known about this topic [47].

- Estimate the costs and health impacts associated with the implementation of the CHIRPY DRAGON prevention intervention programme by conducting trial-based cost-effectiveness and cost-utility analyses from both public sector and societal perspectives. The findings of these analyses will help to inform decisions on whether to implement multi-component interventions such as CHIRPY DRAGON through primary schools as, so far, little is known about the costs of childhood obesity prevention interventions in China.
- Estimate how weight status relates to HRQoL in children from a Chinese setting, and assess the construct validity of the CHU-9D in school-aged Chinese children. This will help to assess the suitability of economic outcome measures such as the CHU-9D for assessing qualify of life within an economic framework in this age group and country setting as, so far, the information regarding this topic is scarce [47]. These findings will be useful to inform the design of future economic evaluations both within a childhood population and within a LMIC setting.

1.4 Thesis Outline

This thesis is divided into three parts: background research; empirical work; and discussion.

The first part, background research, comprises: Chapter 2, background regarding the epidemiology and public health context; and Chapter 3, the methodological foundations for the work developed within this thesis.

Chapter Two explores the epidemiology and public health context of obesity in China by describing the measurement and definitions, the prevalence and trends, the risk factors, and costs and consequences of obesity in this setting. It also summarises the evidence on prevention interventions. Finally, it provides a description of the CHIRPY DRAGON trial, within which this PhD is nested. **Chapter Three** explores the theoretical foundations of economic evaluation in healthcare, and describes the different types of economic evaluation in light of methodological considerations, presenting the possible applications and limitations of each method. It goes on to describe the other vehicles for economic evaluation and the HRQoL measurements for economic evaluation. It concludes by discussing decision-making beyond economic evaluation.

The second part of the thesis describes the empirical work which comprises: Chapter 4, the background and methodological guidance from the previously published systematic review of economic evaluations; Chapter 5, reporting the economic evaluation alongside the CHIRPY DRAGON trial; and Chapter 6, consideration of the suitability of economic outcome measure for interventions in school-aged children.

Chapter Four presents a systematic review of the literature with the aim of exploring the available evidence regarding the methods of economic evaluations used to estimate the costs and benefits of intervention strategies, including both trial-based and model-based evaluations, to prevent/treat obesity in children and adolescents. The review also provides a narrative synthesis of cost-effectiveness evidence and assesses the quality of the included studies. **Chapter Five** describes the intervention and its delivery. It presents the methods to estimate the costs and benefits of the CHIRPY DRAGON trial. It describes the sources for the resource use data, as well as the costs and effectiveness associated with implementation of the CHIRPY

DRAGON prevention intervention programme by conducting a trial-based cost-effectiveness and cost-utility analysis from both public sector and societal perspectives using both clinical and economic outcome measures. This is to provide robust evidence on the cost-effectiveness of this intervention programme compared to usual practice. It explains the economic evaluation structure and its main characteristics. It also discusses the methods and results for the base-case analysis, the uncertainty analyses and the sensitivity analyses. Furthermore, it explores the methodological challenges of conducting an economic evaluation within a Chinese setting when including spillover effects. Where possible, it reports the results using Consolidated Health Economic Evaluation Reporting Standards (CHEERS) guidelines [48]. **Chapter Six** considers the suitability of economic outcome measures for interventions in school-aged children in China. For this purpose, it firstly explores how weight status relates to HRQoL; then it assesses the construct validity of the CHU-9D. To facilitate this assessment, the CHU-9D is compared to the Paediatric Quality of Life Inventory (PedsQL) instrument which is a widely used, validated generic HRQoL measure in children. This chapter uses the baseline data from the CHIRPY DRAGON trial.

Finally, **Chapter Seven** starts by revisiting the aims of this thesis; and provides a summary of the key findings from the entire thesis perspective and discusses them within the context of the wider literature. This is followed by discussing the applied findings to inform policy development and reflections on the methods used for conducting the economic evaluation, highlighting the main strengths and limitations of the approach. The final sections discuss the implications for current policy making, and suggest future research recommendations.

CHAPTER 2. CHILDHOOD OBESITY

2.1 Introduction

The overall aim of this thesis is to explore and develop methods for the economic evaluation of school-based interventions to prevent obesity in children in LMICs. This chapter begins with a discussion on the epidemiology of childhood obesity with a specific focus on China, followed by a summary of the evidence-based decision-making, the evidence on prevention interventions and, finally, an introduction into the design of the CHIRPY DRAGON trial.

2.2 Epidemiology of Childhood Obesity

2.2.1 Measurement and Definitions

The World Health Organization (WHO) defines overweight and obesity as "abnormal or excessive fat accumulation that may impair health" [49]. Although not as sensitive as some other measurements, the most simple and frequently used measure to classify overweight and obesity is the BMI [50]. This is defined as an individual's weight in kilograms divided by their height in meters squared (kg/m2) [49]. Based on observational studies that report major morbidity and mortality risks associated with different BMI levels, the WHO have defined thresholds for overweight (BMI >=25<30) and obesity (BMI>=30) [49].

These BMI-cut offs are, however, not used for children and adolescents age 2 to 18 years and assessing the BMI of children and adolescents is more complicated than for adults. This is because children's BMI normally changes considerably as they mature over time, with a fast increase in the first year, followed by a decrease between 2-4 years of age and then a gradual increase from ages 5-6 until age 18 [51]. Also, these patterns of growth differ between boys and girls. In addition, the association between different thresholds and health outcomes is not straightforward. Therefore, a statistical approach is used for assigning thresholds for high or low BMI in children, based on the child's gender and age [52].

The most widely used reference standard to classify weight status has been established by the International Obesity Task Force (IOTF). Thresholds are derived to line up with the adult BMI thresholds for obesity and overweight at age 18 years. However, many countries have their own population-specific reference standards for assessing BMI in children. For example, in England the British 1990 growth reference (UK90) is recommended for population monitoring and clinical assessment in children aged 4 years and over. Other BMI standards are sometimes used, particularly for international comparison of obesity prevalence [53]. An alternative to IOTF is the WHO standard, which is used for children up to 5 years of age and based on a sample of healthy breastfed children from six countries (Brazil, Ghana, India, Norway, Oman, and the USA). The WHO standard denotes overweight as > 2 standard deviations above the WHO growth standard median [54]. Using different approaches might lead to different classification.

2.2.2 Prevalence and Trends

Childhood obesity is a growing problem worldwide. Recent reports state that it has increased ten-fold from 1975 to 2016, affecting 41 million children under the age of 5 years [55]. Estimated age-standardised prevalence of obesity in children and adolescents in 2016 ranged from higher than 30% in, for example, Nauru, the Cook Islands, and Palau to lower than 2% in other countries, including Ethiopia [55]. Obesity is as much an issue in developing as in developed countries. Although the prevalence of childhood obesity may be higher in developed countries, the rate of increase over the last decade is steeper in many developing countries [55].

In the past three decades, there have been rapid socioeconomic and nutritional transitions in many urban Chinese populations (e.g. Guangzhou), which are important centres of China's trade and economic power [5-7]. Along with these life-style changes, the prevalence of overweight and obesity among Chinese school-aged children has increased more rapidly and over a shorter period of time in comparison with other countries [8, 56]. In some populations, this prevalence is approaching the level of high-income countries [9, 10] and unlike Western countries, which are at a more advanced stage of the obesity epidemic, obesity prevalence, in China, is positively associated with socio-economic status (e.g. maternal education level), particularly in boys [10, 11]. Studies across China have also highlighted increasing childhood obesity rates in rural areas [57].

In the 1980s, childhood obesity was not a public health problem in China (the prevalence of being overweight or obese in school children was approximately 1% (in both genders) in 1985) [58]. It increased to 8.8% in 2000 and then to 17.1% in 2011 [59]. After three decades (in 2015), this prevalence has increased to 28.2% in boys and 16.4% in girls [58]. Most recent national data reports that in China, 42% of adults and around one-fifth of children are overweight or obese [12]. In major cities, around one-third of boys are overweight or obese [60]. Without comprehensive effective interventions, it is predicted that more than one in four Chinese school children (around 50 million) will be overweight/obese by 2030 [58].

In response to this, two nation-wide Public Health interventions introduced by the Chinese state are:

1) Annual measurement of height, weight and fitness of 7-18 year old registered students. This is for everybody in all year groups and is conducted by either local education (Education Bureau at city and district levels) or health (Health Commission at city and district levels) authorities.

The data is analysed locally (at city and province levels) and sent to central government. The analysed data is published every 5 years, which forms the basis for national monitoring of trends.

2) A requirement for children to have one hour of physical activity on campus every school day. However, local implementation has been poor due to contextual issues and the exam-oriented education system [61]. This was the reason why one of the CHIRPY DRAGON components targeted this problem. The local authorities have responsibility of monitoring the implementation of this national policy, however, this is complicated by the fact schools often manipulate behaviour which makes it difficult for local authorities to accurately track adherence [61].

In 2016, the Chinese Government launched the “Healthy China 2030” policy, which seeks to improve health standards in China to be on a par with that in developed countries, through a range of initiatives including health promotion and improvement of public health services [62]. This recent Heathy China 2030 national action plan includes elements relating to nutrition and physical activity interventions in schools but these actions predominantly focus on educating individuals (e.g. providing information and encouragement to help them to eat and live more healthily) [63].

2.2.3 Risk Factors

Individual health behaviours (e.g. physical activity, sedentary behaviour, dietary intake and eating behaviours), environmental factors (e.g. green space availability, healthy food access), policies (e.g. food marketing, nutritional labelling, transport policies), culture, other factors (e.g. maternal pre-pregnancy weight) and interactions between these play an important role in the aetiology of obesity [64].

Individual health behaviours

A lack of physical activity (PA) and an increase in time spent sedentary are known to be associated with various chronic diseases (e.g. obesity). It is recommended by WHO that children and youths aged 5-17 years spend at least one hour per day in moderate to vigorous physical activity (MVPA) [65]. Most of these daily physical activities should be aerobic. Physical activity for all healthy children at this age, to increase their energy output, includes active play, sports, games, chores, recreation, planned exercise, physical education, in the context of family, school, and community activities [65].

Any waking behaviour characterised by low energy expenditure (e.g. resting metabolic rate, typically $\leq$ 1.5 metabolic equivalents (METs)) while in a sitting, reclining or lying posture is defined as sedentary behaviour [66]. This includes sitting at school, or sitting at home while watching TV or playing with electronic devices [67]. Based on expert opinion, WHO recommend, for example, children aged 1-4 years should not be sedentary for more than 1 hour at a time [68]. Over the past few years, there has been a large increase in sedentary behaviour in various societies. This is largely associated with the workplace demands and increase in screen time [69]. Similar trends have been seen among children. A recent systematic review indicated that there is limited available evidence for an association between sedentary behaviour and health outcomes in children and adolescents when accounting for MVPA [70]. However, other studies have shown contrasting results [71, 72]. According to the NICE guidelines, five reviews in children and youths found a positive association between amount of screen time in childhood and weight related outcomes, although the association was not statistically significant in one of the reviews [73]. As physical inactivity and sedentary behaviour can coexist, they may need to be targeted separately [74].

Obesity is due to energy intake being greater than energy expenditure and the importance of dietary intake for obesity development is undebatable [64]. Nevertheless, the evidence base regarding energy intake, diet composition and intake of particular food items; and children being overweight or obese is not fully clarified. This can be partly because an accurate assessment of dietary intake and eating behaviours is notoriously challenging to obtain as individuals vary greatly in their interpretation of how much they eat [64]. The only sound evidence regarding a risk of overweight or obesity has been noted with consumption of sugar-sweetened beverages [75]. Higher consumption of fruits and vegetables has been significantly associated with a lower risk of some diseases (e.g. cancer) [76]. However, an inverse relationship between fruit and vegetable intake and obesity among children is unclear [77].

Parents can negatively impact their children's food preferences and eating patterns [78]. This may include, for example, intake of certain unhealthy food items as part of regular family meals, pressuring them to eat (e.g. finishing the plate) or giving food as a reward [79], all of which might possibly lead them to develop obesity [80]. On the other hand, good parental practices, feeding styles, nutritional knowledge and health behaviours are of major importance for children's development of healthy eating and physical-activity habits [81].

Culture

Cultural factors may have specific influence on habits and development of obesogenic behaviours. For example, most Chinese parents and, moreover, grandparents, aspire for children to be overweight, as (in common with many cultures where there has been a recent history of poverty and famine) this is taken to be a sign of health, growth and prosperity [82].

A major reason for higher rates of obesity among Chinese boys may be because of a cultural preference for boys to be overweight [83], leading to overfeeding.

In China, there is a high prevalence of grandparents taking an active role in child care. Additionally, the one-child policy means that often 4 grandparents as well as the parents are involved in the care of a single child. A mixed methods study based on the families of primary school children in Guangzhou, suggested that Chinese grandparents contribute to childhood obesity through their misperceptions (e.g. children with obesity are healthy), lower levels of knowledge about healthy eating and harms of obesity (e.g. beliefs that obesity related diseases do not affect children; or that foods with higher dietary energy/fat content are more nutritious) and promotion of unhealthy behaviours out of their desire to please and protect their grandchildren (e.g. overfeeding and indulging through excusing children from household chores) [84].

The popularity of cross-generation living, conflicting child care beliefs and practices between parents and grandparents, and between grandparents and school teachers are also important factors contributing to childhood obesity in China, and can undermine efforts to promote healthy behaviours in children [84].

In a recent qualitative study to explore how children and parents make eating and physical activity decisions in China, three main themes were identified [12]. Firstly, children chose food based on flavour: commonly consuming high-calorie snacks rather than fruits and vegetables. Secondly, there were inconsistent standards and practices regarding lunch services across schools: children and parents' perceptions of school lunch services differed among schools. Thirdly, children spent limited time on physical activity because of study burdens (e.g. excessive homework and weekend studies for academic attainment).

Other Factors

A large number of other genetic, lifestyle, environmental and social factors have also been implicated. Prenatal and the early postnatal periods are considered to be "critical" for the development of obesity [85]. Some of the more commonly studied and consistent risk factors include maternal pre-pregnancy weight [86], smoking during pregnancy [87], high gestational weight gain, maternal socio-economic status [88], gestational diabetes [87], high birth weight [89], no or short duration of breastfeeding [90], rapid growth during the first year of life [91], short sleep duration [92], and early introduction of solid foods [93].

2.2.4 Costs and Consequences

Childhood overweight/obesity is a major global public health problem for three main reasons.

Firstly, obesogenic behaviours can persist from childhood to adulthood, and children who are overweight or obese run an increased risk of becoming obese in adulthood [94, 95].

Secondly, overweight and obesity is linked to serious physiological, psychological and social consequences in both children and adults [3, 4, 96]. Once people develop overweight or obesity, it is more likely that they will develop chronic disease conditions, such as diabetes [97] cardiovascular diseases [98], musculoskeletal disorders [99], and some types of cancer [100]. Even during childhood, they are more likely to develop early symptoms and signs of co-morbidities, hypertension and insulin resistance [101]. There is growing evidence that obesity in childhood has a detrimental effect on HRQoL, as children living with severe obesity have reported HRQoL that is comparable with cancer [102]. In addition, childhood obesity is associated with low self-esteem and, because of stigmatisation and weight-related teasing, it can lead to symptoms of depression and (perceived) social rejection [103, 104].

Thirdly, the obesity epidemic leads to significant economic and societal consequences via both direct and indirect costs [1, 2]. Direct costs relate to the healthcare needs arising from associated health problems (both childhood related diseases and those in adulthood), whilst indirect costs, which are estimated to exceed the direct costs, result from productivity losses associated with overweight and obesity (e.g. sick leave, disability pension, death before retirement) and other types of exclusion from the labour market (e.g. stigmatisation) [2, 105-107]. Furthermore, lower academic achievement among children with overweight and obesity could hinder future employment prospects [101].

The Chinese government established the Basic Health Insurance Scheme (BHIS) for urban residents in 1997. The BHIS is operated by the local municipal governments and has implemented extensive cost-containment measures. Only drugs and services approved by the BHIS can be reimbursed. The BHIS provides limited coverage. Its worst limitation is that it does not cover dependents. Commercial health insurance plans are only available in some cities, and premiums are high. As a consequence, according to a survey in 2003, about 45% of China's urban residents did not have any access to health insurance coverage [108]. China launched the New Cooperative Medical System (NCMS) for the rural population in 2003. As of 2006, the system had reached about 50% of the rural population [108]. Overall, China is still in the early stages of building a health safety net for China's citizens. The Chinese low insurance coverage makes patient out-of-pocket expenses the major financing source for health care (58% in 2002) [108].

According to a recent systematic review, which included studies from high income country settings, the mean total lifetime costs due to obesity in childhood and adolescence were estimated to be €149,206 (range: €129,410 to €178,933) for a boy and €148,196 (range:

€136,576 to €173,842) for a girl, with the vast majority of the cost being due to productivity losses and, more specifically, income penalties [109]. A Chinese case study found that the indirect effects of obesity and obesity-related dietary and physical activity patterns reached 3.58% of GNP in 2000 and was projected to reach 8.73% in 2025 [13]. However, there is a lack of evidence regarding the economics of childhood obesity and the long-term economic consequences of childhood obesity in China.

2.3 Evidence-Based Decision Making

Randomised Controlled Trials

Randomised controlled trials (RCTs) are considered the gold standard for evaluating the efficacy of interventions [110]. Randomly assigning participants to either an intervention or control group minimises the risk of bias, allowing causal interpretation of the findings [111]. RCTs are known to provide the most rigorous method regarding whether a relationship exists between the treatment and outcome. This means that the results of economic evaluations (to identify the cost-effectiveness of a new intervention), alongside these trials, are one of the most robust forms of evidence [112]. Evidence on effectiveness and cost-effectiveness is required by decision-makers when making decisions regarding resource allocation.

However, the use of RCT in a public health setting, particularly in the case of behavioural interventions, faces some difficulties mainly for two reasons [113]. Firstly, the complex causal processes in public health interventions makes RCT results subject to effect modification in various populations. Secondly, there are issues concerning ethics and feasibility.

Lack of perfect blinding, losses to follow-up, and cross-over between groups are some of the reasons that internal validity sometimes cannot be assured. Both the internal and external validity of RCT findings can be largely enhanced by observational studies using adequacy or plausibility designs [113].

Cluster Randomised Controlled Trials

As explained above, RCT is the most efficient design for allowing causal interpretation of the findings. However, in some cases randomisation at the individual level is inadequate or impractical. In these cases a cluster RCT, which is a sub-type of RCT, is chosen in which groups or clusters of individuals, rather than individuals themselves, are randomized [114]. Indeed, for the evaluation of certain types of intervention (e.g. those used in health promotion and educational interventions) a cluster RCT is the only valid approach [114]. Cluster RCTs are preferred in two study situations:

- The intervention is delivered at an organisational level or unit (e.g. a surgery unit).
- The intervention cannot be directly targeted at individuals (e.g. school, class room).

However, the cluster RCT design has the disadvantage that there is a need for more study participants to reach the same statistical power.

2.4 Evidence on Prevention Interventions

It is vital to develop effective, culturally appropriate, prevention interventions in order to control the obesity epidemic in children in society.

Schools are considered an ideal place to implement prevention interventions [115]. International research including the updated published Cochrane review, which included trials, approximately 90% of which were conducted in high income countries, has shown that well-designed and well-implemented school obesity prevention interventions were effective in the reduction of BMI in children [116, 117]. However, for instance, findings of three large, well conducted childhood obesity prevention trials in the UK found no evidence for the effectiveness of school based prevention interventions [118-120]. This highlights the importance of 'context' in determining intervention effectiveness. Also, some systematic reviews demonstrate that some short-term interventions (less than 12 months), which focused on combining dietary and physical activity initiatives, did not significantly decrease BMI [121, 122].

It is unknown whether rigorous development frameworks and research methods established in the West can be applied in a low/middle income country setting to develop effective childhood obesity prevention interventions.

Relatively few intervention studies have been conducted in low and middle income countries (LMICs) [123] but these suggest that combining dietary and physical activity initiatives were effective in the reduction of BMI in children [124].

A few Chinese studies have demonstrated that comprehensive school-based interventions, which combined diet and physical activity, were effective [15-17]. However, according to previous [18, 19] and the most recent systematic reviews of childhood obesity intervention studies conducted in China [20], there has been little research applying rigorous methods and established theoretical tools/framework to develop and evaluate prevention interventions for such a population. Shortcomings (methodological flaws) included: poor reporting of process and implementation measures; lack of sub-group analyses (e.g. gender and socio-economic status); short-term follow-up; lack of a control group, randomisation, blinding; failure to report dropouts; insufficient adjustment of confounders; no intention to treat (ITT) analysis and little information on potentially harmful effects in Chinese studies. In addition, the success of the interventions when scaled up remains unclear. Also, most prevention interventions in China have focused on physical activity promotion and most previous intervention studies were treatment focused [20].

As previously mentioned, in light of the limited public resources available, interventions must not only be effective but also cost-effective. With regard to the number of intervention studies in the field of childhood obesity, the number of economic evaluations is relatively small [26]. Only one Chinese study has reported on the cost-effectiveness of a childhood obesity prevention intervention in China [17, 47]. Economic evaluation is important as a means to aid decisions about public resource allocation and it provides information on the relative costs and effects of competing interventions (see chapter 3) [36].

China is in the early stage of the childhood obesity epidemic. The epidemic only became a public health concern from the early 20th century. The range and quality of public health interventions are generally behind those tried in high income countries where the epidemic has been a longer-lasting issue. In fact CHIRPY DRAGON is believed to be the first example of rigorous development and evaluation of a childhood obesity prevention intervention programme not only in China but in LMICs, in general. To date, government/environmental interventions (i.e. interventions beyond school and home settings - e.g. banning unhealthy food adverts during children's peak TV viewing time, sugar taxation) have not been implemented in China. Two nation-wide public health interventions introduced by the Chinese state are: (1) annual measurement of height, weight and fitness of 7-18 year old registered students; and (2) a requirement for children to have one hour of physical activity on campus every school day. The recent Heathy China 2030 national action plan includes elements relating to nutrition and physical activity interventions in schools but these actions predominantly focus on educating individuals [63]. Government provision of effective policy measures such as control of advertising, use of taxation of certain food and drinks, introducing healthy food policies in workplaces/schools; cross-sector collaboration; increasing corporate social responsibility; and reforming health insurance policies could be introduced to curb the trend toward overweight and obesity in China but as yet there is no evidence to show that is happening [63].

2.5 The CHIRPY DRAGON Trial

In summary, China is a useful case study because:

It is a large middle income country which shares several features with other LMIC settings and other South East Asian countries. These include a rapidly increasing prevalence of childhood obesity within the context of limited evidence on how to prevent this. Though there have been previous trials, economic analysis has been limited and not comprehensively undertaken.

China is in a similar stage of the nutrition transition as many other LMICs, with similar social and cultural influences, as well as its economic situation, to these settings [7, 56]. Thus learning from China is potentially more transferrable to these settings than those from the West.

To address the evidence gap of what interventions to implement to prevent childhood obesity in China, and to address the methodological challenges of conducting an economic evaluation within this setting, the CHIRPY DRAGON study was developed using guidelines from the UK MRC Framework for complex interventions [21, 22]. Intervention development was led by researchers from the University of Birmingham in partnership with China Guangzhou CDC, in consultation with parents, grandparents, teachers and school staff [21, 22].

A range of behaviour change techniques and social marketing principles were incorporated in designing the CHIRPY DRAGON (CHInese pRimary school children PhYsical activity and DietaRy behAviour chanGes InterventiON) intervention and a feasibility study was conducted to test and refine the intervention prior to the main cluster randomised controlled trial (cRCT) reported in this thesis. This study provides one of the first examples of a rigorous development and evaluation of a childhood obesity prevention programme outside of high income countries. This study also provides one of the first examples of rigorous economic evaluation of a childhood obesity prevention programme in a low/middle income country (non-western population). The study was funded through a philanthropic donation from Zhejiang Yong Ning Pharmaceutical Ltd Co from 2014 to 2018 and was evaluated from 2015 to 2017 (It was developed from 2009 to 2015 [84, 125]).

My role in the trial was the development of the analysis plan for the economic evaluation; undertaking the statistical and economic analysis including 'bottom-up' costing of the intervention; analysis of outcomes; expanding the evaluation to include spillover effects; undertaking sensitivity analysis, and helping interpret the clinical and cost-effectiveness results.

2.5.1 Trial Design

Study Design and Setting

The CHIRPY DRAGON trial protocol has been published previously [126]. The protocol was implemented without changes. The parallel, two-arm cRCT evaluated the CHIRPY DRAGON obesity prevention intervention, designed for boys and girls aged 6-7 years at baseline. This was implemented within 40 non-boarding, state-funded primary schools located in the largest Southern Chinese city of Guangzhou, a socio-economically advanced city with an urban population of 12.9 million [5].

Ethics Statement

Ethical approval was received from the Life and Health Sciences Ethical Review Committee at the University of Birmingham on 2nd March, 2015 (ERN_14-1440). In addition, local ethical approval was obtained by the Ethical Committee of Guangzhou CDC on 1st December, 2014. The trial registration number is ISRCTN11867516.

School/Participant Identification/Recruitment

All year-one schools (n=353) from non-boarding, state-funded (residents) public primary schools (clusters) located in the largest Southern Chinese city of Guangzhou were eligible for inclusion. The majority of Chinese children attend this type of school [11, 127]. There are also a few private schools, mainly for foreign children from a lower socio-economic background [11, 127], which were not eligible. Basic education in China includes pre-school (usually three years), primary-school (six years, usually starting at the age of six), middle-school (three years) and high-school (three years) periods with the primary- and middle-school education being compulsory [128]. The academic year is divided into two terms for all educational institutions: February to mid-July (six weeks summer vacation) and September to mid/late-January (three weeks winter vacation). There are no half-terms. On average, a primary school student spends about six to seven hours per day at school including a lunch break and mid-day nap (at home or school), while for a middle- or high-school student this rises to about ten hours (excluding a lunch break and mid-day nap but including evening self-study time in classroom) [128].

A research team member randomly selected 40 schools, using a random number generator. From the first week of September 2015, the selected schools were invited to take part in the trial. Each school was sent an official support letter from the relevant local education and health authority/bureau. They additionally received personal visits (with written information sheet/consent form) or telephone communication from the research team members. All 40 schools agreed to take part in the trial within 1-2 weeks.

In China, primary schools have an average of four (range two to eight) year-one classes in each school. Each class consists of approximately 45 children. There is a senior teacher (known as a class-level head teacher) for each class. In the participating schools, all children from year-one classes (age 6-7), along with their family members, were eligible for inclusion and were offered the opportunity to take part in the prevention programme. One year-one class was randomly selected (by a research member during a personal visit to the school) from each of the participating schools to have outcome measurements taken. The research team members distributed invitation letters, information sheets and consent forms to the head teacher of each selected class. These were then given to each child to take home and pass to their parents/other family members. Those who returned a completed/signed consent form were allowed to take part in outcome measures. Parents (or family members) of children were advised to inform the research team if they believed that there were any medical reasons why a child should not take part in any outcome measures and/or intervention activities; and/or if they would like to withdraw from the study at any time for other reasons.

Participant/Cluster Baseline/Primary Follow-Up Assessment

Baseline assessments were completed in the first/autumn term (September to December 2015) when participating children were in year one (aged 6-7 years), followed by the delivery of the 12-month intervention programme which started in the second/spring term after the school winter holiday. The primary (first) follow- up measurements were taken upon completion of the 12 months intervention (April to July 2017) when all participating children had reached year two (aged 7-8 years). A second follow up measurement was taken after a further 12 months, but is not reported here.

Sample Size

1,641 children from 40 schools were needed to have 80% power at a 5% significance level to detect a difference of 0.17 units in the mean BMI z-score between the treatment groups, assuming an average of 45 children per cluster (school). Loss to follow up was anticipated to be 10%.

Randomisation and Blinding/Masking

Randomisation took place after consent was obtained from the 40 schools/clusters and all participating children, and after baseline measurements (January 2016). A UK-based trial statistician used a computer generated sequence (ralloc function in STATA version 14) to allocate schools to the intervention and control groups, and performed stratified randomisation based on whether the school provided mid-morning snacks or had an indoor activity space. These two factors were selected based on the advice received from the trial's lead statistician and local education authority who hold good knowledge of local primary schools. These factors were most likely to impact on the delivery of the intervention programme (potentially influence dietary and physical-activity behaviours of children in local primary schools). Schools were randomly allocated to either the usual practice (n=20) or the intervention arm (n=20). Participating schools were then informed of their allocation.

Because of the nature of the intervention, school staff, children, parents and other live-in adult family members of the participating children and trained project staff who delivered the intervention could not be masked to group allocation during the intervention period. Therefore, independent and trained external assessors (research staff), who were blinded to study allocation and who were not involved in any part of the intervention delivery, were employed to undertake all outcome measures.

2.5.2 Intervention and Comparator

The intervention and its delivery are described in full in chapter 5. Briefly, the CHIRPY DRAGON programme was a 12-month multi-component intervention which was implemented from March 2016 to March 2017. It consisted of four components targeting diet and physical activity behaviours, inside and outside of school, through nine interactive workshops, daily family home activities, and supporting school physical-activity and food provision. It aimed to facilitate the development and/or maintenance of a healthy weight through improving diet and promoting physical activity in children. The intervention was delivered by five full-time Chinese trained project staff (known as CHIRPY DRAGON teachers/researchers). For equity and practical reasons, in the intervention schools, the workshops and activities were delivered to every child in the target year, and the school meal component was a whole school intervention. However, for research practical reasons and to stay within resources, study measurements were limited to children in one class per school.

Schools allocated to the comparator/control arm continued with their ongoing standard provision (usual practice) during the full trial period with no access to any of the CHIRPY DRAGON intervention activities and resources.

2.5.3 Outcome Measures

The outcome measures are described in detail in chapters 3 and 5. Briefly, the clinical measure of effectiveness was BMI z-score. The economic outcome measure was QALYs. Utility-data was collected using the CHU-9D for children and EQ-5D-3L for parents/carers.

2.5.4 Participant Flow during the Trial

1,799 children from the 40 consented schools were eligible in September 2015. 158 of them did not reply to the recruitment letters. The remaining 1,641 (91.2%) consented and participated in study measurements. 20 schools (832 consented children) were randomly allocated to the intervention programme and 20 schools (809 consented children) to the control group. No schools dropped out of the trial. Loss to follow up was lower (3·4% overall, 3·3% in the intervention, and 3·5% in the control arms) than the estimated level used (10%) in the sample size determination. 794 children (95.4%) from the intervention group and 768 (95.0%) from the control group were included in the primary outcome analysis. A CONSORT flow diagram of the CHIRPY DRAGON trial participants is presented in Figure 1.1. The results of the CHIRPY DRAGON trial and its economic evaluation based on cost per QALY and BMI z-score change are reported in chapter 5.

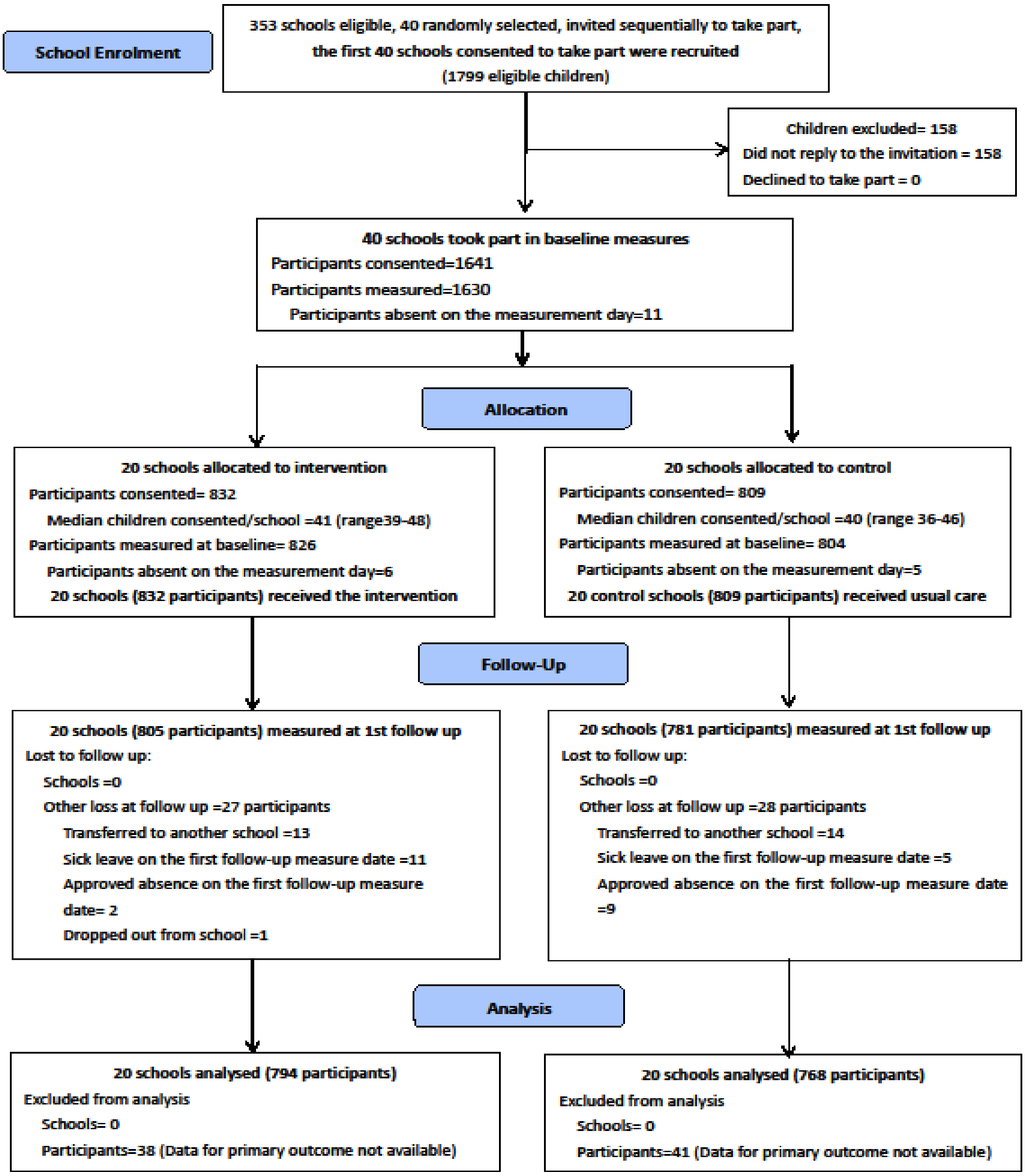

Figure 2.1 - CONSORT flow diagram of the CHIRPY DRAGON trial participants

Li et al. (2019) [129]

2.6 Conclusion

Approximately 42% of adults and around one-fifth of children in China are overweight or living with obesity. Overweight and obesity is linked to serious consequences and costs. Despite an increasing number of behavioural obesity prevention intervention studies, there are relatively few published economic evaluations. That is why the CHIRPY DRAGON study was developed.

In the following chapter, a background to the methods for economic evaluation is provided to help build the context for the subsequent chapters reporting the economic evaluation alongside the CHIRPY DRAGON trial.

CHAPTER 3. THEORETICAL FOUNDATIONS AND APPLICATION OF ECONOMIC EVALUATION

3.1 Introduction

Providing better health and well-being for all individuals in society is the ultimate goal of health care and public health. A key aspect of this is organizing health services around individuals' needs and expectations [130]. However, as discussed in chapter 1, the resources used to provide healthcare services – facilities, equipment, people, time and knowledge – are scarce in relation to the demand for them [36, 131]. Public health goes beyond the health care sector, acting across several fronts such as environment, housing and education. Public health programmes typically aim to reduce the risk of illness or premature death.

Because provision of healthcare and public health is seen as fundamental to individuals' lives, and due to the scarcity of resources, decisions regarding how to allocate these resources are difficult. Therefore, assessing an intervention's relative costs and benefits using economic evaluation is important to enable public health decision-makers to maximise the efficient use of resources. Implementation of a particular intervention is not recommended without evidence of both the effectiveness and cost effectiveness relative to a comparator which is, commonly, usual practice [30].

Some policy decision-making bodies have published methodological guidelines for submissions of economic evaluations to adhere to. These include NICE, the American CDC, the WHO, along with many others [132].

According to Drummond et al (2015) economic evaluation is important for three main reasons [36]. Firstly, it minimizes the chances of an important alternative being excluded from consideration. Secondly, it considers different viewpoints, for example: the patient, the particular institution, the target group for particular services, the budget of the decision-makers, the overall budget of the government, and the societal or community viewpoint. Thirdly, it seeks to estimate the opportunity cost and to compare it with the programme benefits.

Economic evaluation has two characterizing features. Firstly, the linkage between inputs and outputs (sometimes called costs and consequences) of activities is established. Secondly, it is concerned with choices. Many criteria, both implicit and explicit, have to be taken into consideration in making these choices [36]. Economic evaluation attempts to identify and make explicit a set of criteria which might help in deciding between various possible uses of scarce healthcare or public health resources [36].

These characteristics lead to economic evaluation being defined as the comparative analysis of alternative courses of action considering both their cost and consequences [133]. Hence, identification, measurement, valuation, and comparison of the costs and consequences of the alternatives being considered are the fundamental tasks of any economic evaluation [36].

Shiell et al (2002) suggest that efficiency is a criterion which can be used to help decision-makers choose between alternative courses of action [134]. Economics considers two key types of efficiency: allocative and technical. Allocative efficiency concerns whether or not particular resources should be allocated to an intervention. Technical efficiency relates to how best to allocate resources, either minimising costs to achieve a desired level of output, or maximising a particular output relative to its input [134]. These efficiency concerns are addressed to varying degrees by the different methods within economic evaluation.

It is important to always specify within an economic evaluation the perspective for measuring and valuing the costs and benefits included [36]. This matters because an intervention may be cost-effective from one point of view (e.g. healthcare), but not from another (e.g. societal). As discussed in chapter 1, a more comprehensive economic evaluation should consider both the healthcare and societal perspectives. However, the likelihood of studies using a societal perspective is usually based on the decision-makers recommendations or policies (e.g. in the UK, the NICE reference case specifies a (Personal Social Services/National Health Service (PSS/NHS) perspective). Clearly stating the perspective adopted is accordingly a required task for the researcher(s) and is consistently recommended in guidelines for economic evaluations.

This chapter describes the two main theoretical foundations of economic evaluation in the context of health care and public health: welfarism and extra-welfarism. Then, it discusses the five types of economic evaluation. These are: cost-benefit analysis (CBA), cost-effectiveness analysis (CEA), cost-utility analysis (CUA), cost-minimisation analysis (CMA) and cost-consequence analysis (CCA). It continues by describing the other vehicles for economic evaluation and the HRQoL measurements for economic evaluation; and conclude by discussing decision-making beyond economic evaluation.

3.2 Theoretical Foundations

Economics is divided into two main streams: positive and normative economics. Positive economics answers objective and verifiable questions [131]. For instance, it would determine the relationship between user charges and the demand for care. Normative economics is a subjective approach that involves addressing decisions regarding how economic systems should work. For example, how the demand for healthcare should be addressed [135]. Normative economics assess the most desirable resource allocation in accordance with value judgements.

Black, Hashimzade and Myles (2012) state that normative economics is required for making decisions regarding the production and distribution of scarce healthcare resources [135]. Its value judgements being informed by efficiency and equity considerations. Economic evaluations are normative in their nature [131]. This is because there is a fundamental normative stage of deciding what costs and benefits should be considered within the analysis [36], along with how benefits are valued and what is regarded as the optimal option [131].

There are two frameworks, within health economics, which can be used to conduct economic evaluations: welfarism and extra-welfarism. According to NICE, extra-welfarism is widely implemented in health economics and is recommended by decision-makers [35]. The methodological basis for conducting economic evaluations is different within the welfarist and extra-welfarist framework, as each uses different value judgements to determine various states of resource allocation. In the following sections these two different approaches to economic evaluation are discussed.

3.2.1 Welfarism

Welfarism or the neo-classical framework is the traditional theoretical basis for economic evaluation [136]. Welfarism tries to maximize individuals' utilities subject to a budget constraint [131]. A utility is described as the representation of the preference of an individual for a particular good, regarding the satisfaction and/or happiness which they gain from it [34]. In a welfare economic approach, the best judges of their own welfare are the individuals themselves [137].

According to consumer theory, a welfare economic approach considers only the outcomes and resulting utilities for people with regard to their consumption of particular goods or services

[131]. In theory the aim of a welfare economic approach is to generate improvements in social welfare. Other people's judgements, such as those of health care professionals or decision makers, are irrelevant [137].

In order to identify efficient resource allocations, economists use the Pareto principle. This determines that a policy change is socially desirable if either everyone is made better off (weak Pareto improvement: formed by a weak value judgement, i.e., one that gathers consensus) or at least some are made better off while no one is made worse off (strong Pareto improvement: based on a strong value judgement, i.e., a less consensual and therefore a more debatable judgement) [131, 138]. A resource allocation is judged to be Pareto optimal if and only if there is no alternative state that would increase one individual's utility without decreasing another's [133].

Although considered relatively uncontroversial, this approach is not particularly useful for decision-makers because, for any health care or public health decision that involves providing resources to one group at the expense of another, this value judgement offers little indication as to the appropriate course of action [137]. Also, measuring social welfare only with regard to the utility that individuals gain from consumption is too restrictive [139, 140] as firstly, social welfare is a function of only utility and judgements regarding the superiority of one state of the world over another are made irrespective of the non-utility aspects of each state and secondly, the utilities are a function only of goods and services consumed by the individual himself [34, 141].

Policy decisions which are able to satisfy every individual and/or guarantee no losers are very few [142]. The practical drawbacks related to the Pareto optimality criterion resulted in the adoption of the Potential Pareto Improvement concept. This concept expects the losers to be compensated and, if after compensation, the gainers are still better-off, the change is desirable [143]. This is more easily understood if gains and losses are expressed in monetary terms. However, the Potential Pareto Improvement concept might be viewed as representing a serious compromising of Pareto optimality. It makes explicit interpersonal comparisons of uncompensated gains or losses (by allocating an equal unitary shadow weight to each one) and, at the same time, it ignores all other types of uncompensated change and any initial differences between individuals [34].

From a practical perspective and on the basis of the use of the Potential Pareto improvement criterion, welfare economics is criticised because decisions can result in an inequitable distribution of healthcare resources [136, 144]. The extra-welfarist approach was developed in response to these criticisms.

3.2.2 Extra-Welfarism

Extra-welfarism tries to maximise the total health of a population and allows the consideration of outcomes beyond utility. Healthcare or public health resources ought to be allocated on the basis of need. As the need for healthcare implies a deprivation in health, this deprivation can be reduced by healthcare provision [141]. The extra-welfarist approach takes people’s characteristics into account in making its decisions. These characteristics, such as whether or not a person is happy, free of pain, physically mobile, describe and provide a comparison of an individual’s health [34, 141].

The extra-welfarist approach adds something ‘extra’ to welfare economics. Extra-welfarism goes beyond welfarism in four ways. That is: ‘it permits the use of outcomes other than utility, it permits source of valuation other than the affected individuals, it permits the weighting of outcomes other than preference-based, and it permits interpersonal comparisons of well-being' [34].

Health measures such as the quality adjusted life year (QALY) allow individuals to be compared within a health domain [141]. QALYs might be interpreted as a principal outcome of effective health care that reflects the principal dimensions of QoL deemed to be of significance by the decision-maker involved [34].

Indeed, health economists' major criticism of the commonly applied extra-welfarist approach, in practice, has been that it relies purely on health as the single outcome which is too narrow and restrictive [137]. Within the extra-welfarist framework, an intervention's benefits can only be valued regarding the intervention's ability to produce health. There is a possibility that other, non-health related, benefits such as education and housing, could be generated as a result of a special healthcare programme [34]; but, these cannot be captured using health measures, such as the QALY. However, other measures such as capabilities measures under the extra-welfarist paradigm have been proposed.

Capability wellbeing or wider wellbeing is an evaluative space and assesses well-being based on an individual's ability to do and be the things they value in life. It goes beyond consideration of peoples' health as a simple function. Instead it focuses on their ability to do or be in life e.g. the difference between starving due to lack of food versus starving due to voluntary fasting. Sufficient capability is an egalitarian approach to decision making that aims to ensure everyone in society achieves a normatively sufficient level of capability wellbeing [145]. There are cases where researchers have applied this approach [145, 146]. For example, a study showed that capability wellbeing can be incorporated into economic evaluation when considering the impact of addiction treatments, in direct comparison to QALYs [146].

Extra-welfarism using QALYs is typically the basis on which economic evaluation in health care is undertaken within the UK and internationally [137, 144]. It offers a pragmatic approach that is seen to fit well with decision makers' goals. A summary of the five main differences between the welfarist and extra-welfarist approaches are presented in Table 3.1.

Table 3.1 - Summary of welfarist and extra-welfarist approaches

Welfarism	**Extra-Welfarism**
Maximise social welfare (utility)	Maximise health, outcomes beyond utility
Allocative efficiency	Technical efficiency
Combine health and non-health benefits	Consider only health benefits
Consider affected individuals' values	Consider societal values
Distribution and equity are considered separate	Distribution and equity are not separable

Adapted from Brouwer et al. (2008) [34]

3.3 Types of Economic Evaluation

The different types of economic evaluation will be described in light of these methodological considerations. It is common to differentiate between five types of economic evaluation:

3.3.1 Cost-Benefit Analysis (CBA)

CBA provides a systematic assessment of the costs and benefits of an intervention [147]. CBA is theoretically grounded in the welfare economic approach and is a method of economic evaluation where both the costs and benefits of a healthcare intervention are quantified in monetary units [147]. This enables the analyst to compare the discounted future streams of incremental benefits of a programme along with its incremental costs [36]. Thus, it can be used in a comparative analysis of alternatives that have various objectives [133]. The difference between the two streams being the net benefit of the programme.

The aim of CBA is to establish if the benefits of a programme exceed its costs. A positive net benefit (NB>0) demonstrates that an intervention is cost-beneficial and that a programme is worthwhile and should be implemented [147]. Using monetary units as a measure of (positive and negative) benefit allows not only a comparison of the cost-benefits of interventions and services within the health sector, but also between various areas of expenditure within different sectors of the economy [148].

The value of a commodity, in a normal market, is determined directly by observing how an individual responds to changes in quantity and price [149]. Thus, whether the individual consumes the commodity is associated with their willingness to pay (WTP) and ability to pay (ATP) [131]. The healthcare market does not resemble a normal market. Thus, it is impossible to measure individual's preferences for healthcare by assessing their response to changes in price because health care is free at the point of use. Therefore, it is necessary to use other measures to assess the benefits of an intervention, which can be measured either directly or indirectly.

Human capital, revealed preferences and stated preferences of WTP (known as contingent valuation) are three general approaches to the monetary valuation of health outcomes [36]. Stated preference is a direct method that can measure the observed change in welfare from the provision of an intervention [149]. The outcome measure used is known as WTP, which is a measure of the maximum amount of money individuals are willing to sacrifice for their treatment [36, 150]. A less commonly used outcome measure is called willingness-to accept

(WTA), which is a measure of the minimum amount of money the individuals are willing to accept to abandon a good [36, 150].

In practice, healthcare decisions are associated with the addition of a new intervention and tend to measure welfare gain (WTP). Revealed preference is an indirect method and is rarely used in health economics because most health care markets are either financed by tax revenue or social insurance and therefore free at the point of use. It was proved by Olsen (1997) that some analysts have attempted to use WTP for comparing health and non-health programmes [151].

CBA is broad in scope and enables both technical and allocative efficiency concerns to be addressed. Relative values are allocated to health and non-health related goals to assess which are worth achieving [36]. CBA is considered to be a tool that can be used by decision makers to assess various alternatives [131]. However, in economic evaluation and policy decision-making, the cost-benefit approach has had limited use due to the difficulty in measuring the value of all health benefits in monetary units. In addition, there might be moral objections regarding the influence of ATP in the process of valuing the effects [152]. Although CBA is still used, other types of analysis, namely CEA and CUA are more frequently used in health care. Nonetheless, CBAs are still the preferred type of evaluation in other sectors such as transport [153] and environment [153, 154]. Although the use of CBA in public health is a relatively recent phenomenon, it has been used in recent years to evaluate a number of accident prevention strategies (such as introduction of roundabouts, speed cameras and 20 mph zones), capturing a wide range of health and non-health impacts such as congestion, time saving and CO2 emissions [35]. CBAs have not to date been used in obesity prevention interventions [35].

3.3.2 Cost-Effectiveness Analysis (CEA)

CEA is theoretically grounded in the extra-welfarist approach. Here, the incremental cost of a programme is compared to its incremental health benefits, measuring the health effects in natural units related to the objective of the programme such as life years saved or improvements in functional status (units of BMI z-score, cholesterol or blood pressure) [36]. Therefore, CEA is useful to compare interventions, which target the same health condition, and is particularly useful in a clinical setting.

Since costs and benefits of a CEA are measured in non-comparable units, their ratio can be used to assess productive or technical efficiency [46]. The results of a CEA can address questions of technical efficiency, by demonstrating if an intervention compares favourably to an alternative. In particular, it can show whether an intervention can minimise costs to achieve a specified level of health benefit [155]. However, a major limitation of CEA for decision makers is its inability to directly compare cost-effectiveness of interventions across various areas of health conditions, or sectors of the economy. Therefore, it cannot directly address allocative efficiency [36, 46]. This limits the CEAs usefulness in informing budget allocation decisions. This is because, at a national and local level, these decisions need to consider a variety of interventions in various areas of public health, health and social care.

For two alternative interventions, an incremental cost-effectiveness ratio (ICER) is calculated by dividing the differences in total costs (incremental cost) by the difference in the chosen measure of health outcome or effect (incremental effect) [152]. This can be interpreted as the extra cost of obtaining an additional unit of health outcome from a particular health intervention, when compared to an alternative, or the welfare loss from removal of an intervention [156].

The cost-effectiveness plane is used to visually represent the differences in costs and health effects between treatment alternatives in two dimensions, by plotting the costs against effects on a graph [157]. Health effects are usually plotted on the x axis and costs on the y axis. Frequently 'current practice' is plotted at the origin, and so the x and y values represent the incremental health effects and incremental costs versus current practice. More than two points can be represented on the plane, with the line connecting cost-effective alternatives being called the cost-effectiveness frontier. Cost-effectiveness planes are also useful to show the uncertainty around cost-effectiveness outcomes, often represented as a cloud of points on the plane corresponding to different iterations of an economic model in a Probabilistic Sensitivity Analysis (PSA). The cost-effectiveness plane is divided into four quadrants [157]:

North West (NW): When a new intervention generates poorer health effects and is also more costly, which indicates that current practice is dominant.

South East (SE): A new intervention is considered to be dominant, when both less expensive and more effective than its alternative. Therefore, it is the most cost-effective option [36].

North East (NE): In many economic evaluations, an intervention might be relatively more effective than usual care, yet also be more costly.

South West (SW): In some cases the intervention generates lower costs but it is less effective.
In these two last cases, in order to judge if the intervention is a cost-effective use of public resources, the ICER should be compared to a cost-effectiveness threshold that is based on the amount society is willing to pay for an extra unit of health benefit [131, 152, 158].

3.3.3 Cost-Utility Analysis (CUA)

CUA is theoretically grounded in the extra-welfarist approach and focuses on the comparison of healthcare preventions or treatments regarding both the quantitative and qualitative aspects of health outcomes (morbidity and mortality) produced [46]. CUA is an adaptation of the CEA approach. The incremental cost of a programme is compared to its incremental health improvement. Health improvement is measured in utility terms which can be framed as either a QALYs (the product of the preference based measure of QoL 'utility' and length of life) [152] or a disability-adjusted life year (DALY). Both are used as a unit of assessment to make judgements about the relative cost-effectiveness of competing interventions [131]. In fact, in a LMIC setting, the DALY is the more popular outcome unit.

In theory, the QALY measure is universal; therefore, various programmes across different health conditions which are evaluated using QALYs can be compared. However, there are limitations to using CUA for the reasons that QALY measures vary by method; QALY results vary by respondent (for example, the patient, family, clinician, or general public) and society might value a QALY for various population groups differently [159, 160]. The use of QALYs in economic evaluation of public health interventions has been criticised. This is partly for considering health as the only relevant outcome, and ignoring the production of wider benefits which are not captured in the QALY.

However, the use of a single measure of health benefit enables comparison of diverse healthcare interventions, so CUA allows both technical and allocative efficiency questions to be addressed [46]. Like CEA, the most cost-effective intervention, from those compared, is considered to be the one which generates the most QALYs for the lowest incremental cost, considering a budget constraint [144]. CUA is more useful to decision-makers with a broad ruling than CEA, because of its broad applicability [36].

There is uncertainty around how the cost-effectiveness threshold should be defined and current cost-effectiveness thresholds are valued in cost-per-QALY terms. For instance, in the UK, using NICE public health guidelines, ICER values under £20,000-30,000 per QALY are deemed a cost-effective use of resources [161]. By contrast in the USA the recommendation is US$50,000 per QALY [162] and in many countries (e.g. China) there are no clearly defined thresholds at all. The uncertainty around the ICER is typically assessed through the use of the net-benefit regression (NBR) framework [163, 164] using non-parametric bootstrapping. Decision uncertainty is presented using cost-effectiveness acceptability curves (CEACs) [164]. CEACs are developed to estimate the probability of the intervention being cost-effective across a range of values of WTP for an extra QALY.

Different healthcare systems have different HTA contexts, systems and priorities. Some countries do not apply specific thresholds. Where they do not, the WHO make recommendations for cost-effectiveness thresholds of 1 to 3 times the Gross Domestic Product (GDP) per capita of that country [165, 166].

The GDP per capita, which calculates all the goods and services produced in a country in one year divided by the total population, is a useful metric for categorizing countries as either developing or developed. According to WHO, countries are considered to have developed economies when they have a GDP per capita of >= US$12,000 [166, 167]. However, some economists believe that US$25,000 is a more realistic threshold [167].

With a GDP/capita of US$19,000 [166, 168], China's status is complex. Despite having the world's second-largest economy and third-largest military, it is generally not classified as a developed country (e.g. it is on the Development Assistance Committee (DAC) list of LMICs, developed by OECD [169]). This is because many economists argue that the country's per capita GDP remains below the accepted minimum thresholds for developed-country status. In addition, it has a low level of technological innovation [167]. However, the status is controversial, as despite the GDP, it shares many social, cultural and economic features of other middle income countries.

It is inevitable that an economic evaluation contains some degree of uncertainty in its assessment. The choice of sensitivity analysis may depend on the methodology applied or the setting in which the intervention was conducted. In order to assess the level of uncertainty, one can apply a deterministic and probabilistic sensitivity analyses. In a deterministic sensitivity analysis (DSA), model parameters are changed individually. To assess the level of uncertainty for analyses, it has become common practice and it is the gold standard to apply a Probabilistic Sensitivity Analysis (PSA), where model parameters are changed simultaneously according to a given distribution [30, 170].

3.3.4 Cost-Minimisation Analysis (CMA)

CMA focuses solely on costs' differences therefore questions such as "Is the extra effectiveness worth the extra cost?" cannot be answered [152]. CMA is the analysis of choice either when the common outcomes are equal or assumed to be equal (owing to outcomes being roughly identical). It identifies the lowest cost alternative [133]. One of the advantages of this analysis is that only cost data needs to be collected.

CMA is inappropriate in cases where there might be a difference in effectiveness, because effectiveness is not measured and the option which is cheaper might be harmful for the patient

[152]. In these situations, researchers should use an economic evaluation that allows the competing interventions or treatments to have different effectiveness; CEA is the simplest of these types. Because the estimations of costs and outcomes are not certain, it is difficult to have equal outcomes unless the alternatives are approximately identical [133]. In a review of economic evaluation approaches [171], CMA was deemed inappropriate because it is very rare to have a set of circumstances whereby intervention effectiveness are known with certainty and are assessed as identical in advance of analysis.

3.3.5 Cost-Consequence Analysis (CCA)

CCA makes the influence of the new treatment as comprehensive as possible. It simply lists out all the cost-incurring events and the consequences in a disaggregated fashion. Therefore, it will enable decision makers to select which components are most relevant to their perspective. In addition, it will give them confidence that the data are valid to use as the basis for resource allocation decisions [172]. However, the decision made by an individual may not be in the best interest of either patients or society.

The details of characteristics of health care evaluation and types of economic evaluation are provided in Tables 3.2 and 3.3.

Table 3.2 - Distinguishing characteristics of healthcare evaluation

Are both costs and consequences of the alternatives examined?

		NO		YES
		Only consequences	**Only costs**	
Is there a comparison of two or more alternatives?	**NO**	Partial economic evaluation Outcome description	Partial economic evaluation Cost description	Partial economic evaluation Cost outcome description
	YES	Partial economic evaluation Efficiency or effectiveness evaluation	Partial economic evaluation Cost analysis	Full economic evaluation

Adapted from Drummond et al. (2015) [36]

Table 3.3 - Economic evaluation types

Method of analysis	**Valuation of costs**	**Valuation of effects**	**Paradigm**
CBA	Monetary units	Monetary units (WTP)	Welfarism
CEA	Monetary units	Natural units: One-dimensional unit (e.g. BMI z-score)	Extra-welfarism
CUA	Monetary units	Multi-dimensional unit (e.g. QALYs)	Extra-welfarism
CMA	Monetary units	Equivalent: None (alternatives are assumed to have equal outcomes)	Extra-welfarism
CCA	Monetary units	Description	-

Adapted from Drummond et al. (2015) [36]

3.4 Other Vehicles for Economic Evaluation

Decision-modelling can be used as a vehicle to conduct an economic evaluation. Modelling can be done following a trial and/or using a framework of evidence synthesis [173]. For example, following a trial, the model-based economic evaluation goes beyond the observed time of the trial and extrapolates the effectiveness and costs over a longer time horizon with using assumptions applied to the trial data and/or combining the trial data with evidence from multiple sources [174]. These methods provide a framework for developing and applying an appropriate structure for the natural progression of the underlying disease which enables an assessment of the long term costs and benefits of relevant options; all existing evidence can be brought to bear and, through PSA, the implication of parameter uncertainty on the recommendation can be quantified [173].

In addition, modelling links intermediate outcomes to final outcomes and it supports the analysis using secondary data (e.g. morbidity and mortality data) [174]. This role of modelling in economic evaluation is especially relevant in the field of obesity prevention or treatment interventions in children and adolescents, where the outcomes may only be realised long after the trial has finished. Model-based economic evaluation can capture the uncertainty linked to any assumptions made and could also improve the generalization of results obtained in one setting to other settings [174].

The most common types of model structures are decision trees and Markov models. Decision trees are commonly used, and are one of the most simplistic model structures in decision analytic modelling [131]. Markov models are a more complex type of model, widely used in economic evaluation, which overcome the limitations of decision trees [36]. They reflect various health states or consequences of treatment, where the probability of transitioning between health states occurs in accordance with the time cycle of the model [36].

To assess the level of uncertainty for model-based analyses, it has become common practice to apply a Probabilistic Sensitivity Analysis (PSA), where model parameters are changed simultaneously according to a given distribution [30, 170]. It is also common to conduct a more comprehensive approach which is a combination of DSA and PSA to assess the uncertainty around the parameters included [170].

Discounting

Discounting is the process that enables a comparison between the value of a good consumed in the future with the value of that same good consumed now. Lipscomb et al (1996) state that in economic evaluation, discounting is based on the principle that, all future costs and health benefits/consequences should be considered in relation to their present value to the decision maker [175]. Discounting enables a comparison between interventions that produce benefits and incur costs over a number of years [36]. Preventive interventions will usually require costs to be incurred in the present for effects to be enjoyed in the later stages of life. For instance, interventions to prevent obesity will result in a lower risk of some types of cancer that will be experienced later in life. Currently, NICE recommend a discount rate of 3.5% per year for both costs and outcomes for economic evaluation of public health interventions that extend beyond one year, with a sensitivity analysis of 1.5% [30]. For childhood obesity prevention interventions, the effect of discounting will cause future health gains to be devalued [26]. However, failure to discount future benefits would consider interventions to be more cost effective than they would otherwise appear [176].

Methodological challenges of conducting economic evaluations

There are four key methodological challenges that face the economic analyst when conducting evaluations of public health interventions [29]:

These comprise firstly, the problems of attributing effects and time scales to a specific public health intervention: most published guidelines, including NICE, recommend the use of RCTs to compare alternatives. There are likely to be fewer controlled trials of public health interventions because of the very large sample size required to power pragmatic trials of public health interventions adequately. Whilst public health interventions could be expected to have an impact over the longer term, follow up in clinical trials is often limited to one or two years at most.

Secondly, measuring and valuing outcomes: in economic evaluations of clinical interventions, outcomes are often measured in natural units or QALYs. In the economic evaluation of public health interventions, other outcomes might be important including effects on individuals not directly targeted by the intervention and other non-health related effects such as education. Some of these effects can be incorporated into QALYs, some not. So alternative measures need to be adopted and offset against costs using different frameworks for economic evaluation.

Thirdly, incorporating equity considerations: in many cases the aim of the public health intervention is to reduce inequalities. The normal assumption in economic evaluation methods is that a QALY is of the same value to everyone who receives it and resources are allocated using a maximisation principle to achieve the greatest amount of QALYs for every unit of resource. It is possible to move to a position in society where inequalities in QALYs are reduced but this would require a trade-off from the most efficient allocation, and requires a change to the framework of economic evaluation.

Fourth, identifying inter-sectoral costs and consequences: the impact of the public health interventions can be wide-ranging. The costs and benefits may fall on parts of the public sector not confined to health alone, such as the judicial system, education and housing. There is a consensus within the economic guidance that wider social and environmental costs and benefits should be looked at due to the complex nature of public health but it is unclear how this is done in practice with the opportunity costs in other non-health settings unknown [29].

There are also some additional technical (methodological) and context specific challenges of conducting economic evaluations in LMICs as below [177]:

Technical (methodological) difficulties include:

- Scarcity of robust local clinical data.
- Lack of data on costing (e.g. unit cost, source of cost).
- Paucity of commonly accepted guidelines (e.g. perspective, discounting) for economic evaluations.
- Lack of local data or appropriate value sets for estimating QALYs or DALYs (in CUA).
- Difficulty in interpreting and using the findings of a CEA and CUA (e.g. lack of an equivalent threshold value for a QALY).

Context specific difficulties include:

- Economic evaluations are not formally linked to the healthcare decision-making process.
- Limited local research capacity to undertake economic evaluation.
- Lack of funding for the necessary research.
- Communication barriers between researchers, academia and end users of the evidence.

- Limited number of published local journals with a standard review process.

3.5 HRQoL Measurements for Economic Evaluation

The concept of HRQoL takes into account several dimensions of health. It has been recognised as an increasingly important concept, because it acknowledges the subjective perception of disease and health [69]. It may be more accurate and appropriate in assessing someone's health status as it allows assessment of health benefits before the consequences of disease are physically manifested.

3.5.1 HRQoL in Children

Weight prevention and management interventions have increasingly targeted primary school-aged children to address the growing problem of childhood obesity [47]. This has implications for choosing the methods of outcome measurement within economic evaluation of these interventions as few instruments exist which are designed to generate utilities in this age group [47].

HRQoL for adults has been widely accepted as an endpoint in a research and clinical setting. In contrast, the assessment of HRQoL among children has only gained interest and awareness in recent years. Assessment of health status in children differs from adults as it requires a different conceptual approach because of rapid rates of development, dependency on parents/caregivers and differences in disease epidemiology [178]. The assessment of each individual's HRQoL relies on their subjective evaluation of functioning in different domains. It was believed that children's subjective health reports were not reliable and were therefore of limited use [179]. But, research demonstrates that primary school-age children, aged 8-10 years [180], and maybe even younger [181] can adequately reflect and report their health state if instruments are adapted to them. HRQoL measurements may either be self-administered or administered by parents, caregivers or researchers. Because the cognitive and language skills of young children are not completely developed, it is essential to rely on parents/caregivers or researchers for the assessment of HRQoL in this age group. Many aspects of QALY measurement in young children are not completely developed and as a result hinder their application in CUA [182].

Generic Measures

Currently, there are a range of generic health related QoL instruments available which can be used in all conditions. These measures include: CHU-9D; EuroQol-5 Dimension Youth (EQ-5D-Y); Health Utility Instrument (HUI-2); and Paediatric Quality of Life Inventory TM (PedsQL).

CHU-9D

As discussed in chapter 1, ideally, utility-based HRQoL in children should be measured using an instrument specifically designed for them. Although there is no gold standard for measuring utility-based HRQoL in primary school-aged children, previous research has shown the CHU-9D, a recently developed instrument, is the most appropriate choice [38]. As a utility-based instrument, it is preference-based and thus generates utility values anchored between the values of 0 (death) and 1 (perfect health), with negative values denoting states worse than death. It is not specific to any one condition or disease. It is designed for application in cost-effectiveness analyses of prevention, treatment and service programmes targeted at young people where the QALY is the desired outcome measure [183]. Originally tested for children aged 7–11 years [39, 40], it has more recently demonstrated good construct validity in 11–17 year olds [41]. The tool has successfully been applied to wider populations, from six years old up to 17 years old [42, 43].

The CHU-9D instrument combines nine dimensions of HRQoL: worried; sad; pain; tired; annoyed; schoolwork/homework; sleep; daily routine; and ability to join in activities [39, 184]. Each dimension comprises five severity levels, resulting in 1,953,125 unique health states associated with the measure. Each child under investigation completes the questionnaire and an analyst (researcher) can attach a utility value for the child's health state by using the appropriate algorithm. QALYs are calculated for each individual, over the specific period, using the standard area under the curve approach [36].

EQ-5D-Y and HUI-2

In 2010, the EQ-5D-Y was developed which is designed specifically for children [185]. The EQ-5D-Y has been tested for feasibility, validity and reliability in children, adolescents and young adults aged 8–19 [186]. This instrument is not applicable for infants and young children younger than 4 years. The EQ-5D-Y was adapted directly from the EQ-5D. The EQ-5D-Y describes five dimensions of HRQoL and wording of the questions in each dimension is modified to make it relevant to a younger age range. For instance, mobility is referred to as 'walking about' [186]. Similar to the 3L-version adult instrument, there are three severity levels for each dimension (no problems, some problems, extreme problems). Until the EuroQol group produces tariff values for the EQ-5D-Y, the use of EQ-5D-Y is not recommended for utility elicitation in young children [38].

The HUI-2 is a generic measure of health status which has been used extensively in both clinical and general populations, in both children and adolescents [187]. But, it has rarely been studied with regard to child and adolescent obesity [188]. This instrument has been valued using a sample of parents from Hamilton, Canada and allows utility estimation. The HUI-2 questionnaire has six dimensions: sensation, mobility, emotion, cognition, self-care and pain. Each of these has between three and five levels. The use of HUI2 has been recommended by its developers for use in children of five years and older.

PedsQL

The PedsQL is a non-preference based instrument and therefore does not apply any explicit weighting between item domains. It therefore cannot be used to generate utility values for the construction of QALYs. However, it would be expected to produce HRQoL values which move in the same direction as the utility measures. This instrument is a widely used HRQoL instrument validated for use with young children over 5 years old. It has good reliability and validity in both sick and healthy populations [189, 190]. It is a 23-item instrument comprising four domains: physical (8 items), emotional (5 items), social (5 items), and school (5 items) functioning [189]. Each item has five response options: never; hardly ever; sometimes; often; almost always. Emerging from the instrument is a score (transformed on to a 0–100 scale) for each domain and a score for total HRQoL. A low score indicates a poor HRQoL, and a high score indicates high HRQoL.

3.5.2 HRQoL in Adult (Spillover Effects)

The acknowledgement that behaviour change interventions have spillover effects on family members has led to an increased interest on how to adapt methods for capturing these broader effects to maximise population health, rather than just the target participant's health [45]. The choice of evaluative space of health (e.g. EQ-5D), which is designed to measure a generic-related HRQoL and is not specific to any one condition or disease, for economic evaluation is an important value judgment which can have a large impact on resource allocation decisions [30].

Generic Measures

Currently, there are a range of generic health related utility-based QoL multi-attribute instruments available which can be used in all conditions. These measures are used in conducting a CUA. These include: EQ-5D (3L and 5L); Short Form 6 Dimension (SF-6D); and Health Utility Instrument (HUI3) [36].

EQ-5D

EQ-5D is recommended by decision-makers (NICE) [30] and it is one of the most commonly used instruments to estimate HRQoL. The EQ-5D is a measure used to capture HRQoL through a combination of five dimensions: mobility; self-care; usual activities; pain and discomfort; and anxiety and depression [191]. Each dimension comprises three (no problems, some problems, extreme problems (unable)) or five (no problems, slight problems, moderate problems, severe problems, extreme problems (unable)) possible levels to describe the extent of the problem associated with each domain. The individual under investigation completes the questionnaire and an analyst (researcher) can attach a utility value for the individual's health state by using the appropriate algorithm derived from a population sample. This accounts for public preferences, to obtain an overall EQ-5D utility index score for the individual [192]. QALYs are calculated for each individual, over the specific period, using the standard area under the curve approach [36].

However, EQ-5D may not be appropriate in all cases as it may not be sufficiently sensitive or relevant for all conditions. It is recognised that some clinicians and researchers do not consider EQ-5D to be a suitable measure because they are usually interested in condition specific outcomes [193].

SF-6D and HUI-3

The SF-6D is a measure used to capture HRQoL through a combination of six dimensions: physical functioning; role limitations; social functioning; pain; mental health; and validity. SF-6D was not originally designed as a preference based measure. It is developed from a non-preference based measure (SF-36) [194] and generated by mapping either the non-utility based SF-12 or SF-36 to SF-6D.

The HUI3 is recommended for primary analysis in adults. Dimensions include vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain. The system exists in two complementary forms: HUI2 and HUI3 and the use of both can result in insightful results.

3.6 Decision-Making beyond Economic Evaluation

As described above and in chapter 2, in public health, decisions should be based on the best available evidence, and provide the best value for money. However, there are other additional factors which decision-makers need to consider when deciding on the implementation of public health interventions [195]. These factors may not be directly quantifiable. These, which were primarily developed for the assessment of childhood obesity interventions, include the following criteria: “strength of evidence”, “feasibility of implementation”, “equity”, “sustainability”, “side-effects”, and “acceptance by other stakeholders” [196]. These criteria are applicable to decision-making regarding other public health interventions as well.

3.7 Conclusion

This chapter has described the theoretical foundations of economic evaluation in healthcare. The different types of economic evaluation have been discussed in light of methodological considerations, presenting the possible applications and limitations of each method (e.g discount rate, perspectives, sensitivity analysis). It has shown that the extra-welfarist approach is inclusive of welfarism; and is typically the basis on which economic evaluation is undertaken particularly in the West as well as internationally. There is no 'gold standard' method by which to evaluate public health interventions, although CUA has been recommended in many countries, for instance, in the UK by NICE. It has highlighted that model based analyses, as a widely used and powerful instrument, are particularly relevant to the evaluation of the cost-effectiveness of public health interventions.

The condition and trends in obesity rates have been described in chapter 2. This chapter has provided a background to economic evaluation to set the sense for later chapters. The next chapter will provide an up-to-date review of methods for generating economic evidence for the treatment and prevention of childhood and adolescent obesity.

CHAPTER 4. A SYSTEMATIC REVIEW OF METHODS, STUDY QUALITY, AND RESULTS OF ECONOMIC EVALUATION FOR CHILDHOOD AND ADOLESCENT OBESITY INTERVENTION

4.1 Introduction

As described in chapter 2, childhood obesity is a major global public health problem. Cost-effective obesity prevention and treatment in children and adolescents is therefore a priority. However, despite an increasing number of intervention studies, there are relatively few published economic evaluations [197-199].

As discussed in chapter 3, economic evaluation is a means to aid decisions about public resource allocation [25, 26] and as obesity prevention and treatment often involves lifestyle interventions that have costs and consequences that fall outside the health care sector, a societal perspective for evaluation is usually recommended [26].

Seven recent reviews [101, 196, 200-204] have summarised the cost-effectiveness of obesity prevention and/or treatment interventions in young people however none were designed to offer a rigorous review of methods applied for economic evaluation. Five reviews had language restrictions [101, 196, 200, 201, 203] and four excluded studies that were conducted in developing countries [101, 196, 201, 203], limiting global interpretation. Only two reviews appraised methods for handling inter-sectoral costs [101, 203]. Just three of the reviews used established criteria e.g. Drummond checklist [36] to assess the quality of the primary studies

[101, 200, 203]. The search strategy was inadequate (e.g. search terms not fully reported) in three reviews [196, 201, 204], and in the remaining four there were omissions of relevant databases, which means that relevant studies could have been missed [101, 200, 202, 203]. Furthermore, the most recent review, which only focused on interventions in pre-school children, included studies reported up to November 2015 and, at least 3 new economic evaluation studies of childhood obesity interventions have been published since then [205-207].

This chapter reports on a systematic review of published economic evaluations of obesity prevention and/or treatment interventions in children and adolescents (0-19 years) with the primary objective of appraising the methods used and assessing the quality of the economic evaluations using the Drummond checklist [36]. More specifically this review provides a systematic overview of the study context; the type of economic evaluation and measures of effectiveness; the evaluation perspective taken; the time horizon considered and type of modelling approach taken; the choice of discount rate; the methods for collecting and estimating resource use/costs; and the type of sensitivity analyses undertaken. A secondary objective was to undertake a narrative synthesis of the evidence of the cost-effectiveness.

A systematic review is a widely used methodology. It is supported by many research groups and policy-makers. In general, it is placed at the top of hierarchy evidence regarding effectiveness studies, and it is robust, replicable and its results are reliable [208].

4.2 Methods

This systematic review follows the reporting guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [209]. The protocol is registered with the international prospective register of systematic reviews (PROSPERO) database ref (CRD42017062236) and has previously been published [210].

4.2.1 Literature Search

The following electronic health economics/biomedical databases were searched: MEDLINE (Ovid); EMBASE (Ovid); Web of Science; CINAHL Plus; EconLit; PsycINFO; Cochrane Database of Systematic Reviews (CDSR); Database of Abstracts of Reviews of Effects (DARE); the National Health Service Economic Evaluation Database (NHS EED); Health Technology Assessment (HTA) and Cost-Effectiveness Analysis (CEA) Registry. The following sources were also used to identify potential additional studies: Google Scholar; relevant NICE guidelines; the reference lists of eligible studies and review articles; and Grey literature such as OpenSIGLE, National Obesity Observatory, NHS Evidence, National Technical Information Service, Healthcare Management Information Consortium (HMIC) and RepEC (Economic Working papers) database. The search was conducted in May 2017 and studies were sought between January 2001 and April 2017. The year 2001 was chosen since the first study evaluating the cost-effectiveness of a childhood obesity treatment intervention was published then followed 2 years later by the first economic evaluation of a childhood obesity prevention intervention [202]. Search strategies included Medical Subject Headings (MeSH) terms and text words of key papers that were identified beforehand. The search terms and text words were adapted for use within other bibliographic databases. The full search strategy is provided in Appendix 1.1.

4.2.2 Inclusion and Exclusion Criteria

Economic evaluations were included or excluded based on the following criteria:

Types of study: Primary full economic evaluations were included (studies in which both the costs and outcomes of the alternatives are examined and in which a comparison of two or more interventions or case alternatives are undertaken) including trial-based and model-based (using trial data) evaluations. Partial economic evaluations; qualitative studies; conference abstracts; and study protocols were excluded.

Participants/ population: Children and adolescents aged 0-19 years at the start of the intervention and/or their parents/guardians were included. Family based interventions were also included when the target participants were the children. Economic evaluations undertaken within any country context were included. Interventions to tackle obesity due to a secondary cause (e.g. Prader-Willi syndrome) were excluded.

Intervention(s), exposure(s): All behavioural (focused on individual behaviour change techniques), environmental (focused on modifying the local environment) or policy (focused on population-wide legislative or fiscal action) interventions for the treatment or prevention of overweight/obesity in children and/or adolescents were included. Pharmacological or surgical interventions were excluded.

Comparator(s)/control: Only studies with a clearly defined comparator were included with no restrictions on the types of comparator(s).

Outcome(s): No restrictions on outcomes measures. Potentially relevant outcomes were: DALYs; QALYs; effectiveness outcomes such as kilogramme weight loss; % Body Fat; BMI z-score; waist circumference; overweight and obesity cases avoided; additional minute of MVPA; increase in overall physical activity level and METs hour gained.

Other criteria: There were no restrictions based on language, evaluation perspective taken, duration of intervention, time horizon for evaluation or setting.

4.2.3 Study Selection Procedure

The review followed a two-stage method. First, the main researcher (Mandana Zanganeh) and an independent researcher (James Hall) individually screened titles and abstracts of identified publications against the selection criteria. If in doubt, the full text version was requested. Second, full-text papers were reviewed by both researchers and a final decision made with respect to the inclusion/exclusion criteria. There was 85% agreement between the 2 reviewers. Any disagreements between the reviewers over the eligibility of specific studies were resolved by discussion between the main reviewer and all supervisors (Emma Frew, Peymane Adab, Bai Li). To aid study selection and analysis of non-English language articles, translation either in part or in whole was undertaken by academic colleagues with the appropriate language skills. The literature search results were managed using EndNote X7 (Thomson Reuters).

4.2.4 Data Extraction

The study characteristics and findings were recorded by the main researcher (Mandana Zanganeh) using a standardised, pre-piloted data extraction form. Extracted information included: authors; publication year; country; study design; setting; target population/age group; N (analytical sample); parent/guardians included; intervention overview/target; aim of the intervention/mode of delivery; comparator; measures of effectiveness; type of modelling approach; study perspective; duration of intervention/follow-up; time horizon; price year; currency unit; discount rate; methods for estimating/collecting resource use; costs categories; largest cost drivers; excluded costs; total/average costs per participant; funding source; ICER/average cost per benefit; uncertainty analysis; sensitivity analysis; and cost-effectiveness results.

This process was checked for completeness and accuracy by an independent researcher (James Hall). Any discrepancies between the reviewers over the data extraction process was identified and resolved by discussion or by consensus with all three supervisors. Missing data was requested from study authors.

4.2.5 Quality Assessment of Included Studies

To allow a comparison of the economic evaluation methods used in the studies, the Drummond checklist was used by the main researcher (Mandana Zanganeh) [36]. Given that none of the available checklists or scoring schemes had been validated to quantify the methodological quality, quality assessment of individual items or an overall summary score was not applied. Instead, the quality assessment provides a systematic and critical descriptive overview of key methodological elements. A full copy of the quality assessment sheet is presented within Appendix 1.2. Quality assessment of the included studies was independently checked for completeness and accuracy by an independent researcher (James Hall) and any discrepancies were resolved by discussion with all three supervisors.

4.3 Results

Figure 4.1 illustrates the flow of papers identified, screened and included in the review. Of the 4,185 studies identified in the initial literature search, 2,607 were screened. From the screened papers, 2,559 were excluded based on titles and abstracts. 48 articles were considered potentially relevant and remained for subsequent detailed assessment. Of these, 39 were in line with the eligibility criteria. Therefore, these articles were included in the analysis and synthesis. The most common reasons for exclusion were the lack of (full) economic evaluations, being a protocol study, or including an ineligible target population.

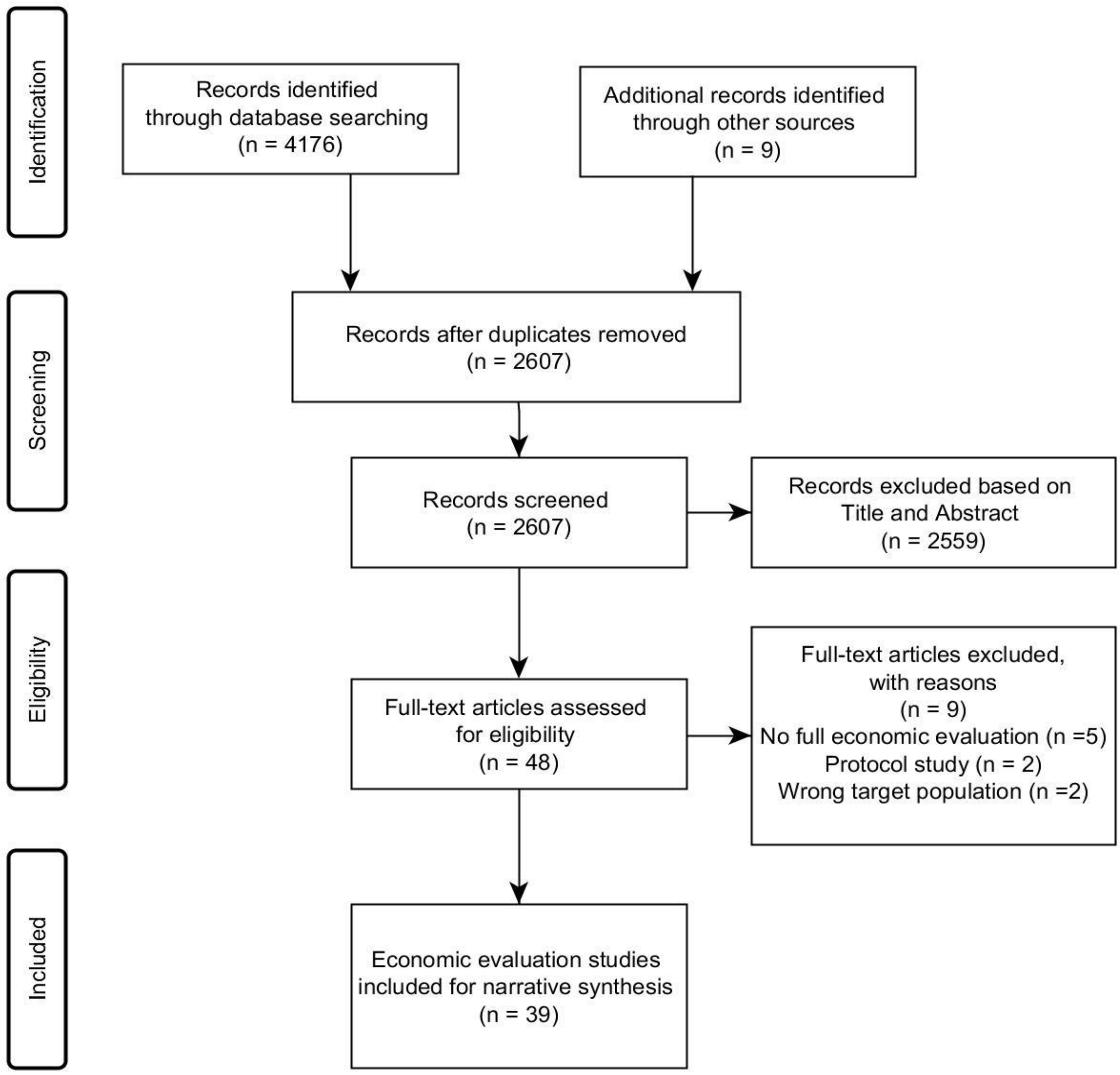

Figure 4.1 - Adaptation of PRISMA 2009 flow diagram

4.3.1 Details about Study Context

Full details about study context are presented in Appendix 1.3 (i) - 1.3 (iv), and summarised in Table 4.1.

Table 4.1 - Summary of general characteristics of the studies

Study characteristics		Number of studies identified (%)
Year of publication		
2001-2009		17 (44)
2010-2017		22 (56)
Study approach		
Trial-based	Prevention	9 (23)
	Treatment	7 (18)
Model-based	Prevention	19 (49)
	Treatment	4 (10)
Comparator selected		
Usual care		33 (85)
Another intervention		6 (15)
Country		
High-income		
Australia		15 (38.5)
New Zealand		2 (5)
The USA		12 (31.5)
Canada		1 (2.5)
The UK		4 (10)
Germany		2 (5)
Finland		1 (2.5)
Spain		1 (2.5)
Low and middle-income		
China		1 (2.5)
Setting		
Prevention		
School		21 (54)
US/Australian state		5 (13)
Community		1 (2.5)
Home		1 (2.5)
Treatment		
Clinical		9 (23)
School		1 (2.5)
Community		1 (2.5)

Intervention and Comparator

Approximately half of the economic evaluations (23/39) were model-based [196, 207, 211-226] compared to trial-based evaluations. A range of interventions were identified, all containing individual behaviour change elements (Figure 4.2). A large proportion (25/39) (including all treatment interventions) were focused exclusively on behaviour change techniques, the rest combined individual behaviour change elements with either an environmental component (modifying the local environment e.g. active school transport) [115, 206, 215-219, 227] or a policy component (population-wide legislative or fiscal interventions such as banning unhealthy food advertising or a physical education policy) [211, 213, 214, 221, 224, 228]. Approximately half of the interventions (21/39; 12 prevention and 9 treatment) targeted a combination of physical activity and dietary behaviours [17, 115, 196, 205, 212, 218, 222-227, 229-234], the rest focused on either physical activity [206, 211, 215-217, 219, 228, 235-237] or dietary habits only [196, 207, 213, 214, 221, 238].

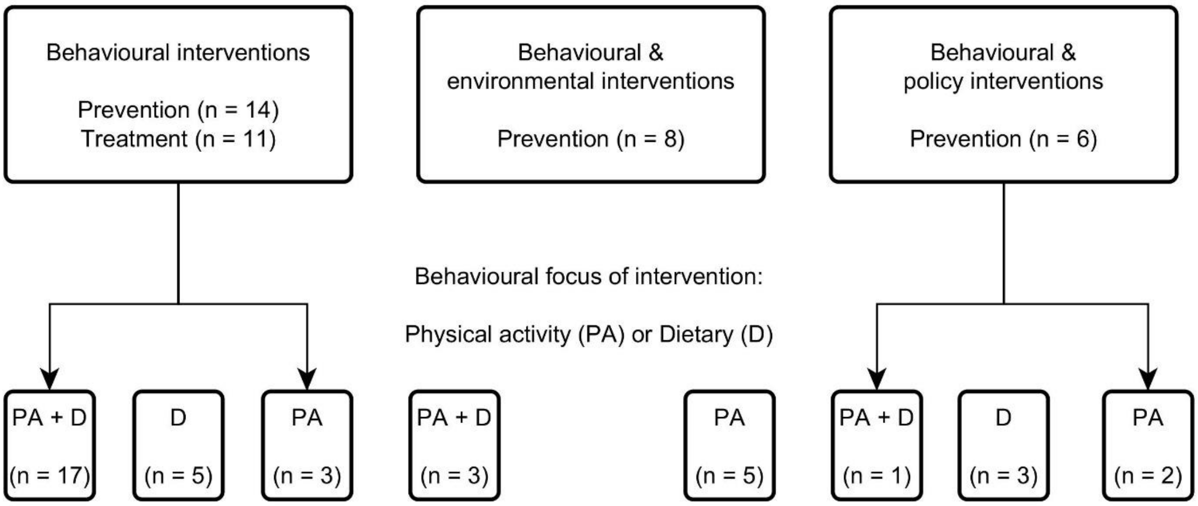

Figure 4.2 - Summary of the interventions

The intensity of the interventions differed considerably. For prevention interventions, this ranged from one session per 3 months [229] to approximately 2 sessions per month [235]; and for treatment interventions, ranged from one [205] to 12 sessions per week [230]. The duration of the interventions also differed, ranging from 8 months [237] to 4 years [236] for prevention studies; and from 3 months [205] to 1 year [230] for treatment studies. Overall therefore, the treatment interventions were generally more intensive but delivered over a shorter time period compared to prevention interventions. The comparison or control group was not always clearly specified but was assumed to be "usual care" in most of the studies (33/39) and often the studies did not justify their rationale for choosing the comparator.

Country and Setting

The evaluations were spread across a wide range of countries and study settings. The vast majority (38/39) originated from high-income countries, mainly Australasia (Australia (n=15) [196, 206, 214-218, 226, 229, 234] and New Zealand (n=2) [220, 227]), with 13 from North America (USA (n=12) [115, 207, 211-213, 221-224, 228, 230, 233], Canada (n=1) [231]) and 8 from Europe (UK (n=4) [205, 219, 225, 232], Germany (n=2) [235, 236], Finland (n=1) [238], Spain (n=1) [237]). Only one study was from a developing country context - China [17]. In terms of study setting, the majority of prevention interventions (21/28) were school-based [17, 115, 206, 207, 212, 215-218, 220, 222, 223, 227, 233, 235-237] and for treatment interventions, most (9/11) took place in clinical settings [196, 205, 225, 226, 230-232, 234, 238].

4.3.2 Review of Economic Evaluation Methods

A detailed account of the economic evaluation methods are presented in Appendix 1.4 (i) - 1.4 (iv), Appendix 1.5 (i) - 1.5 (iv), and Appendix 1.6 (i) - 1.6 (iv).

Type of Economic Evaluation and Measures of Effectiveness

Focusing on the methods of economic evaluation, most studies performed a CEA using raw or standardised BMI as a measure of clinical outcome (26/39) (18 prevention and 8 treatment), whilst a few used other outcomes (e.g. cost per case of overweight/obesity prevented, cost per unit increase in MET minutes, reduction in body fat). Within these studies, more than half (16/26) conducted a CEA only [17, 115, 206, 211, 224, 225, 227-233, 235, 237, 238], whereas the rest combined a CEA approach with a CUA [205, 213-219, 221, 226] using QALYs as the outcome measure. Eleven further studies conducted a CUA only [196, 207, 212, 220, 222], while only two used a CCA [234, 236]. The vast majority of trial-based economic evaluations (15/16) did not use QALYs/DALYs whereas the model-based evaluations (n=20) tended to report QALYs/DALYs as the main health outcome measure. When QALYs were used, the age of the participants was between 6 and 11 years in the trial-based economic evaluation [205], and between 2 and 19 years in the model-based economic evaluations. Educational attainment outcomes were not explored in any of the studies. Approximately half of the studies did not justify the choice of form of economic evaluation related to the question and the outcome measure selected.

A pattern with preferred type of economic evaluation by country context was apparent. Within Australasia (13/17) a CUA or a combination of CUA and CEA [196, 214-218, 220, 226] was most popular, whereas the majority of studies from North America (7/13) [115, 211, 224, 228, 230, 231, 233], and the only study from China [17] conducted a CEA only. Across Europe, only UK-based studies used CUA [205, 219]. In terms of study setting, a CEA was most common in clinical settings (7/9), whereas within school settings a mixed approach was applied with around half conducting a CEA (12/22). There was no clear pattern found in terms of approach taken to evaluate prevention or treatment interventions. Similarly, mixed approaches were applied to evaluate the different type of interventions. Slightly more studies used QALYs/DALYs to evaluate "behavioural interventions" (14/25), compared with "behavioural and environmental" (5/8) [215-219] and "behavioural and policy" (3/6) [213, 214, 221] interventions, but there was no clear pattern identified.

Evaluation Perspective Taken

An economic evaluation is of most importance to policy makers as it points out to whom the costs and benefits occur. Most (35/39) studies clearly reported the study perspective. The majority (n=28) were from a societal perspective. Interestingly, none of the UK studies [205, 219, 225, 232], compared to most of those conducted within Australia and the USA, applied a societal perspective. Two studies reported using a health care perspective, but from the data reported it was clear that wider societal costs were included within a secondary analysis [205, 234].

For all interventions that included either a policy or environmental component (12/14), the perspective was societal, whereas for interventions focused exclusively on individual behaviour change a combination of societal (17/25) and healthcare (6/25) perspectives was undertaken. A societal perspective was also adopted by the vast majority of interventions in school settings (19/22), compared to less than half of the interventions in clinical settings (4/9) [196, 226, 230, 234].

Time Horizon Considered and Type of Modelling Approach Taken

The time horizon within economic evaluation refers to the duration over which outcomes and costs are measured and this should be explicit and justified. The evaluated duration of intervention delivery should ideally reflect how the intervention would be applied in real life.

The time durations for the trial-based economic evaluations were predicted by the period of the trial. Of interest this ranged from 8 months [237] to 6 years [236] in the prevention studies; and from 10 months [233] to 15 months [234] for the treatment interventions. For the model-based evaluations, the time horizon was more at the analysts discretion and within this review ranged from at least 10 years (n=5) [211, 213, 221, 223, 224] to a lifetime (15/23) [196, 207, 214-218, 220, 225, 226]. The time horizon was also found to be much shorter within clinical settings (6/9) [205, 230-232, 234, 238] compared to the other study settings such as schools for example. The majority of the studies did not justify their choice of time horizon. The vast majority of model-based studies (20/23) were from Australia and the USA, with the rest from the UK and New Zealand [219, 220, 225]. The time horizon was also found to vary by intervention setting with clinical settings having the shortest time horizon (6/9) [205, 230-232, 234, 238].

With respect to modelling, the vast majority of model-based studies (18/23) applied Markov modelling [196, 211, 213-218, 220, 221, 224-226] compared to decision tree [207, 212, 219, 222, 223]. The majority of the model-based studies did not justify their model choice and the description of model details was suboptimal in most of them.

Choice of Discount Rate

For the majority of the trial-based studies (10/16) (4 prevention and 6 treatment), discounting was not appropriate as the time horizons considered were relatively short (less than one year) [17, 115, 205, 230-233, 235, 237, 238]. For all the trial-based studies of more than one year, all reported using a discount rate in accordance with the relevant country guidelines apart from one prevention trial from New Zealand [227], which used a 5% discount rate per year for costs, rather than the 3.5% discount rate per year for both costs and outcomes recommended [239]. Most model-based studies (22/23) applied a discount rate for both costs and outcomes (3% per year for Australia (n=12) [196, 214-218, 226], the USA (n=8) [207, 211-213, 221-224] and the UK (n=1) [225], and 3.5% per year for New Zealand (n=1) [220]. Interestingly, the rates used for studies from Australia and the UK were not in accordance with their respective country guidelines (which is 5% per year for Australia according to Pharmaceutical Benefits Advisory Committee (PBAC) and 3.5% per year for the UK according to NICE) [161, 240]. However, different state governments in Australia recommend different rates and the discount rate used in the included Australian and UK studies was consistent with the US panel recommendations [162]. Most of the studies did not justify their choice of discount rate.

Methods for Collecting and Estimating Resource Use/Costs

Half of the trial-based evaluations (8/16) (4 prevention and 4 treatment) reported their methods for collecting resource use [205, 206, 229, 230, 232, 234-236], while only 10 out of 23 model-based evaluations (9 prevention and 1 treatment) did so [207, 211, 213, 215-218, 221, 224, 226].

As expected, the choice of inclusion of a particular type of cost varied considerably according to the study purpose, perspective, setting and the nature of the intervention being evaluated. Costs tended to be categorised into programme delivery, direct medical (e.g. healthcare visits), direct non-medical (e.g. travel time/cost for participants) and indirect (e.g. productivity losses because of parents' absence from work). In line with recommendations for CEA [241], the development/set up costs were not considered in the vast majority of studies, apart from one trial-based prevention study from the USA [228].

Of the 9 studies (5 prevention and 4 treatment) that included indirect costs incurred by parents [115, 205, 215, 216, 218, 226, 230, 234, 236], these were mainly from Australia (n=5) and most of them were for preventive "behavioural" interventions within a school-based setting (5/9). Also, direct non-medical costs were reported by 4 prevention studies from Australia [215, 218, 226, 234] and 1 treatment study from the USA [230]. Most of these types of costs (3/5) were for "behavioural" interventions implemented within a clinical setting.

Sensitivity Analysis Undertaken

The majority of the trial-based studies (10/16) conducted a deterministic sensitivity analysis to assess the robustness of the results [115, 205, 206, 227, 229, 234-238], while the other 6 prevention/treatment studies did not conduct any type of sensitivity analysis. Three of these 6 studies were from the USA [228, 230, 233], while the others were from Canada, China and the UK [17, 231, 232].

Most of the model-based studies (22/23) apart from the study by Pringle et al (2010) from the UK [219], conducted at least one type of sensitivity analysis. The majority of these studies (n=20) conducted both DSA and PSA, while one treatment study from the UK within a hospital-community setting conducted only a DSA [225] and the other prevention study from the USA within a school setting conducted only a PSA [212]. Almost half of the studies did not justify the choice of the variables for sensitivity analysis.

4.3.3 Narrative Synthesis of Cost-Effectiveness Evidence

The most common method for presenting cost-effectiveness evidence was the ICER (30/39). The vast majority of the studies (33/37), excluding the CCA ones, reported results that were cost-effective. Some of these (13 of the model-based prevention/treatment studies including 5 by Carter et al (2009)), [196, 213, 214, 219, 221-225] illustrated cost saving results. For instance, Long et al (2015) concluded that a sugar-sweetened beverage excise tax would increase benefits in terms of DALYs averted and result in healthcare cost savings in the USA [213]. Almost half of these 13 studies that illustrated cost-savings were from Australia, followed by 5 from the USA and 2 from the UK. None of the trial-based evaluations reported cost saving results, probably due to shorter time horizons. Whilst the findings are not directly comparable between studies due to the heterogeneous nature of the methods used, all of the studies which evaluated interventions targeting only dietary habits (8/8) and the majority of the studies targeting both physical activity and dietary habits (19/21) indicated cost-effective or cost saving results. However, the studies which focused on only physical activity indicated a proportionally smaller number of cost-effective or cost saving results (7/10). Furthermore, the evidence suggests that the majority of behavioural interventions supported by a policy intervention (4/6) were cost-saving [213, 214, 221, 224].

A small number of studies (n=4) [205, 215-217] reported interventions to not be cost-effective. The UK trial-based treatment study [205], which targeted a combination of physical activity and dietary habits with the aim to reduce weight gain in children with obesity remained cost-ineffective using a CEA/CUA approach regardless of the choice of perspective. Also, the 3 model-based studies that targeted only physical activity were not cost-effective, for example, the "Walking School Bus" programme which had a high cost of delivery coupled with low participation rates [215].

4.3.4 Quality Assessment of the Included Studies

The quality of reporting the economic evaluations was assessed using the Drummond checklist - a 35-item instrument with a total of 3 domains. The expected possible responses to each question were: Y=Yes, N=No, NC=Not clear, N.A.=Not Applicable. Full details of the quality assessment are presented in Appendix 1.7 (i) – 1.7 (iv). None of the included studies fulfilled all of the quality criteria. The most positive aspect of the quality assessment is that none of the studies was ranked as "worthless". Most studies fulfilled a large number of the quality criteria and only a small number of the studies were poor. One challenge regarding the quality assessment was that quality was judged based on the published data only and there might be a difference in what has been reported and what has actually been done. So a bad scoring study might just be due to lack of transparency rather than lack of quality.

Certain criteria were simply not applicable to each respective study (e.g. items 12–15, due to different perspectives chosen), while others were not reported. As there are 16 trial-based studies, all the questions dealing with the quality of models had to be answered with "N.A." for them. The three criteria which were least well addressed were the rationale for the comparator, the justification for the choice of discount rate, and the model choice. The description of model details was suboptimal in most of the model-based studies. Whilst the time horizon for each study was generally well specified, most studies omitted to provide reasons for choice. Additionally, approximately half of the studies did not justify the choice of economic evaluation nor offered justification for what was explored within a sensitivity analysis.

4.4 Discussion

This is the first study to conduct a rigorous review of the methods for economic evaluation and to determine how these methods vary by setting, country and intervention type. Therefore this review provides an important contribution to the current knowledge gap in a relevant and fast growing field.

In this systematic review, it was found that most of the studies (38/39) were from high income countries. Only one evaluation was identified from a developing country context [17], where the obesity epidemic is rapidly increasing. It was also found that economic evaluation of obesity interventions in children and adolescents is an expanding area of research, with a third of included studies being published within the last 5 years. Whilst growing attention is being given to effective and well-conducted intervention studies to prevent or treat obesity in children [198, 199], few interventions to date have been subjected to economic evaluation and therefore the number of published economic evaluations is relatively small. Also, the interventions which have been evaluated in terms of their economic credentials are narrow in terms of their content; most related to health promotion programs, with only a couple targeting tax legislation.

The review identified some emerging patterns. The results suggest that among the published economic evaluations, there was no consistent measure of outcomes. Around half of the studies reported clinical (e.g. BMI), rather than health-related outcome measures commonly used within economic evaluation (QALYs/DALYs). The dearth of trial-based studies that included QALYs or DALYs (1/16) [205] suggests that the measurement of these types of outcomes within obesity trials is not firmly established or may relate to reported lack of sensitivity of utility-based HRQoL instruments to changes in overweight/obesity in younger children [42]. This heterogeneity of outcome measures will hinder comparability of cost-effectiveness.

The overall pattern to emerge from the results of one systematic review suggests that there is a weak negative association between obesity and educational attainment in children and young people. I.e. higher weight is weakly associated with lower educational attainment [242]. Obesity is also associated with other variables, such as socio-economic status, and when these are taken into consideration, they may better explain much of the negative association between obesity and attainment [242]. The results from longitudinal studies indicate that, in general, obesity in children was associated with lowered educational outcomes [243, 244], especially for girls. It was also found that having friends with overweight and obesity drives down the odds of educational success. Attendance at a higher SES school or a school with a lower percentage of minority students was positively associated with the odds of college attendance and obtaining an undergraduate degree [243].

Educational attainment outcomes were not explored in any of the studies. This is because the measurement of these types of outcomes within obesity trials is not firmly established globally. Also, as educational outcome data are very rarely considered in economic evaluations, guidelines for including these data are limited. It is challenging to estimate what a given society may be willing to pay for a unit gain in an educational outcome. It might be more appropriate to use a CBA or a CCA to report improvement in educational outcomes alongside costs but not to try and offset the two against each other. A more disaggregated analysis (e.g. a CCA) alongside a CEA and CUA would give a 'list' rather than offset against the costs.

No evaluation applied a CBA approach. Consideration of broader outcomes going beyond the health sector allows for inclusion of costs and effects from multiple sectors and is particularly relevant for obesity intervention. This is an emerging area of development within economic evaluation and efforts are being made to adapt methodologies to promote the use of CBA [245]. These approaches have been recommended by the UK Treasury guidance to evaluate (usually non-health) public sector projects [246].

Model-based evaluations offer the opportunity to improve the generalisability of results as they combine data from a variety of sources. However the findings from five of the model-based evaluations identified within this review were based on small samples [207, 212, 219, 222, 223] and only one of these offered data based on a lifetime horizon. Furthermore, all of the model-based evaluations were for interventions that targeted individual health behaviours and were therefore highly dependent on cultural, infrastructural and other system-related aspects. So the generalisability of results to other contexts, particularly from developed to developing country settings, would be questionable [247]. The majority of the papers did not make explicit mention of procedures for checking their models. Despite associated assumptions with modelling studies, the studies evaluated are important as model-based health economic evaluations are today widely accepted as policy-making tools that can inform resource allocation decisions. Almost half of the model-based studies chose a lifetime perspective.

Most trial-based and model-based evaluations in this review applied recommended discount rates in accordance with the relevant country guidelines. Methods for collecting resource use and the type of cost included were found to vary across the studies. In particular, the indirect costs of overweight and obesity (e.g. productivity losses) were not generally collected alongside the trials. It is considered good practice to report results both with and without indirect costs. Including indirect costs (e.g. costs incurred by families) has the potential to alter the treatment recommendations.

The narrative synthesis of the economic evidence and the quality assessment of the included studies are useful for informing health economists/modellers and the direction for future research in this area. In terms of judging cost-effectiveness of interventions, context-specific assessment is problematic as there are different thresholds for cost-effectiveness in different countries. For example, in the UK, NICE recommends a threshold willingness to pay of £20,000-£30,000 per QALY [30], by contrast in Australia the recommendation is AU$ 50,000 per QALY [215] and in many countries there are no clearly defined thresholds at all. Whilst most interventions in this review appear cost-effective using standard rules of cost-effectiveness, there is substantial variation by intervention design.

4.4.1 Review of Recent Papers Published since the Original Search

The same search strategy used in the initial search for this study was used in September 2019 to update the systematic review. Studies published between May 2017 and September 2019 were sought. Four studies were found, which were in line with the eligibility criteria. Therefore, these articles were included in a brief analysis and synthesis for updating this systematic review as below.

Consistent with the earlier results, these 4 studies (2 prevention, 2 treatment) were also from high income countries [248-251] and only one of these used QALYs as an outcome measure [249], with the rest reporting only clinical outcome measures (e.g. BMI). Two of the studies used a societal perspective [248, 250]. Three conducted a sensitivity analysis to assess the robustness of the results [248-250]. The only model-based study was likely to be cost-effective [248]. Among the trial-based evaluations two, one from the UK within a school setting [249] and one from Sweden with a child health centre setting [250], were neither effective nor cost-effective using clinical outcomes. The UK study was cost-effective using QALYs (ICER: £26,815 per QALY gained from a public sector perspective), however there was a high level of uncertainty as demonstrated by the net-benefit equation and the corresponding CEAC [249]. The other trial-based study, which was from Denmark, is one of the first studies to have assessed the cost-effectiveness of a camp-based obesity management intervention programme [251]. This study compared a standard group with a weight management camp group and, after 12 months, showed favourable effects in the intervention arm (decreased BMI). However, the camp group was more costly.

4.4.2 Comparison with Previous Systematic Reviews

The finding of this study that most interventions were cost-effective or even cost-saving, is similar to those reported by two other reviews [200, 204], with some overlap between included studies. Other reviews have focused on particular age groups (e.g. pre-schoolers [101]), specific interventions (e.g. only physical activity [203]), or particular outcomes (e.g. anthropometric measurements [101]). Two additional reviews from Australia [196] and the US [201] used the Assessing Cost-Effectiveness (ACE) obesity approach to summarise and compare the cost-effectiveness of a range of interventions. However, none of the previous studies reviewed the methods of the economic evaluations in the way it has been outlined in this review.

4.4.3 Strengths and Limitations of this Review

One of the important strengths of this review is the comprehensive search strategy applied encompassing a broad range of electronic bibliographic databases of published studies and the grey literature (six additional studies were identified). Furthermore, the results were not limited to only those published in English (two non-English publications identified) and there were no country restrictions (there was one publication from china as a developing country), resulting in a more complete review than those published previously. Also, the formal quality assessment of the economic evaluations undertaken adds strength to the conclusions. The vast majority of the studies were found to be of very good reporting quality.

The review had some limitations. As the focus was on full economic evaluations, some important data contained within partial evaluations may have been missed. Further limitations relate to the shortcomings of the included studies and underlying evidence base. There was heterogeneity in both the methods used and with the type of intervention being evaluated, which made synthesising the evidence base challenging. Not all included studies used the same definition of obesity, which may impact on the results. Most of the included studies reported an economic evaluation for an intervention that had previously been reported as clinically effective. It is possible that any trial which had ineffective results did not conduct an economic evaluation or, if they did, failed to get it published, introducing potential publication bias.

4.5 Conclusions

This systematic review in this chapter suggests that current economic evaluations are mainly set in developed countries and the majority focus on the prevention of obesity in children, compared to treatment. The findings of this study show that the majority of published economic evaluations are for interventions with an individual behaviour change component. The majority, particularly “behavioural and policy” preventive interventions, were cost-effective, even cost-saving. However, this review found that relatively few policy interventions designed to prevent obesity have been subject to a rigorous economic evaluation. The review found heterogeneity with respect to methods applied. So, to improve the evidence base further and to enhance comparability across interventions, we recommend a consistent and expanded form of economic evaluation which captures both health and non-health costs and consequences beyond health-gain.

In general, the systematic review results showed that the following main gaps exist in the current literature:

- Number of reported economic evaluations of obesity prevention interventions in developing countries is very low.
- Inclusion of societal costs and outcomes in economic evaluations that are relevant to family members are rarely included within economic evaluations.
- Number of CUAs of trial-based studies comparing obesity prevention strategies, whose results can be compared with other public health programmes is low across the world.

In the next chapter, the methodological challenges of conducting an economic evaluation within a Chinese setting and including spillover effects are explored. This is done by evaluating the CHIRPY DRAGON intervention and undertaking an economic evaluation from a public sector and societal perspective using both clinical and economic outcome measures. The aim is to provide robust evidence on the cost-effectiveness of the school- and family-based prevention programme when compared to usual practice in China, and to highlight the methodological challenges.

CHAPTER 5. COST-EFFECTIVENESS OF THE CHIRPY DRAGON OBESITY PREVENTION INTERVENTION IN CHINESE PRIMARY SCHOOL-AGED CHILDREN: A CLUSTER-RANDOMISED CONTROLLED TRIAL

5.1 Introduction

Chapter 4 reported that the majority of preventive interventions were cost-effective. However the review found heterogeneity with respect to methods applied. Current economic evaluations are mainly set in developed countries and the measurement of QALYs within obesity trials is not firmly established.

As discussed in chapters 3 and 4, QALYs are commonly used as the unit of assessment to make judgements about the relative cost-effectiveness of competing interventions [36] and to inform resource allocation decisions worldwide [37]. As obesity prevention often involves lifestyle interventions which have costs and consequences that fall outside the health care sector, a societal perspective for evaluation is usually recommended [26]. This means that all relevant resource use/costs and consequences are measured, outlining how these fit within a given sector [27].

As discussed in chapter 2, to address the evidence gap of what interventions to implement to prevent childhood obesity in China, and to address the methodological challenges of conducting economic evaluation within this setting, the CHIRPY DRAGON study was developed using guidelines from the UK MRC Framework for complex interventions [21, 22], in consultation with parents, grandparents, teachers and school staff.

This is a case study used to explore and develop methods for the economic evaluation of school-based interventions to prevent obesity in children in LMICs. This chapter will conduct a comprehensive economic evaluation (from a public sector and societal perspective using both clinical and economic outcome measures) alongside the CHIRPY DRAGON trial to provide robust evidence on the cost-effectiveness of this intervention programme compared to usual practice in China. In doing so, it will highlight, the methodological challenges of conducting an economic evaluation within a Chinese setting and explore and investigate the impact of including spillover effects. Where possible, the results are reported using CHEERS guidelines [48] (Appendix 2.1).

As the CHIRPY DRAGON trial has been described in full in chapter 2, only a brief overview of the trial is provided (Part 1). The remainder of the methods section describes the study intervention and its delivery; how the resource use and costs are measured; outcome measures; how missing data is dealt with; and the statistical analysis plan for each of the perspectives separately. The results then follow. The results are separated into two parts: Part 2 refers to the economic evaluation from a public sector perspective and Part 3 to the economic evaluation from a societal perspective. Finally, in Part 4, the discussion and methodological challenges are reported.

5.2 Part 1: Methods

5.2.1 Details about Study Context

Trial Design and Participants

The evaluation of the CHIRPY DRAGON obesity prevention intervention, targeting children aged 6-7 years at baseline, was undertaken using a cluster randomised controlled trial (cRCT) in 40 state-funded primary schools in Guangzhou, China. Schools were randomly allocated to either the usual practice (n=20) or intervention arm (n=20). In China, primary schools have an average of four (range two to eight) year-one classes per school. Each class consists of approximately 45 children. In the participating schools, all children from year-one classes (age 6-7), along with family members, were eligible for inclusion and were offered the opportunity to take part in the prevention programme. One year-one class was randomly selected from each of the participating schools to have outcome measurements taken. Written consent was sought for each study participant from parents/guardians.

CHIRPY DRAGON Study Intervention and Its Delivery

The CHIRPY DRAGON programme was a 12-month multi-component intervention which was implemented from March 2016 to March 2017. It consisted of four components targeting diet and physical activity behaviours, inside and outside of school, through nine interactive workshops, daily family home activities, and supporting school physical-activity and healthy food provision. It aimed to facilitate the development and/or maintenance of a healthy weight through improving diet and promoting physical activity in children. The intervention was delivered in 20 schools randomised to the intervention arm by five full-time trained Chinese project staff (known as CHIRPY DRAGON teachers/researchers). This means that each of the five CHIRPY DRAGON teachers employed and trained by the research team were responsible for the facilitation, coordination and delivery of the programme activities in four intervention schools. There were tailored activities for various target groups, including children; main carers (parents/ guardians or grandparents); school staff (principals, class teachers, physical education (PE) teachers, school meal director) and school lunch providers (including catering staff (managers and workers)). School staff had direct roles in supporting the delivery of the CHIRPY DRAGON intervention programme.

The four core components within the CHIRPY DRAGON intervention were:

Component 1: improving childhood obesity related knowledge and behaviour among children and their main carers

Component 2: improving the nutritional quality of school lunch provision (generally a set lunch box for each child)

Component 3: increasing children's physical activity levels outside school

Component 4: increasing children's physical activity levels inside school

Component 1 consisted of four activities. These were:

1) Two interactive educational workshops per year (in the term one and two), held in school, for main carers with a leaflet to take home.
2) Four interactive educational workshops/classroom sessions per year, held in school, for children.
3) Ongoing family-wide healthy behavioural challenges and child self-monitoring during the week, held at home, for children and their main carers.
4) One cross-generation health quiz event, held in school during the second term, for children and their main carers.

Main carers' workshops focused on correcting common misperceptions identified through the formative research in relation to child healthy weight and healthy behaviours, and introducing practical parenting tips for encouraging healthy behavioural change in children. Children's workshops focussed on key messages related to healthy eating and an active lifestyle. To encourage adoption and maintenance of the promoted healthy behaviours for this component, the CHIRPY DRAGON teachers asked children to set individual goals, which were challenging but achievable, based on their current habits within their family and rotate them every two weeks. These included restricting themselves to less than two hours of sedentary screen-based activities per day; eating at least five portions of vegetables and fruit per day and reducing consumption of drinks and snacks containing high levels of fat and/or sugar. All three behaviour challenges required written feedback from their main carers, using specially designed cards ("fun cards"), to be returned to CHIRPY DRAGON teachers for evaluation alongside each child's self-monitoring records. The cards were used to encourage main carers' involvement and support in behaviour challenges. After reviewing the cards and judging them objectively, based on children's behavioural outputs, the CHIRPY DRAGON teachers rewarded best performance and improvements. Children were given CHIRPY DRAGON stamps for meeting goals or making good behaviour progress. These stamps were collected by children and were shown on a reward board at the back of each classroom. Each term, the three children who collected the most stamps, in each class were presented with an incentive prize.

Component 2 involved introducing five school lunch improvement targets to lunch providers (catering staff from companies who were responsible for producing meals for intervention schools). These were jointly developed by Chinese nutrition experts, obesity prevention specialists within the research team and Chinese school meal providers (including both managers and catering workers e.g. chefs). They were tested for feasibility and acceptability in a feasibility trial. The aims were:

1) To restrict provision of deep-fried meals to no more than once a week.
2) To decrease portion size for rice or noodles in younger children.
3) To provide children with vegetables every day.
4) To reduce high fat and processed meats from the school menu.
5) To ensure the amount of cooking oil, salt and added sugar in each school meal was not more than one-third of the recommended daily intake for the Chinese children (6-7 years old) [252].

These aims allowed maximum flexibility for local adoption/adaptation in response to changes in costs, seasonal food availability and local resources. The implementation of these aims was achieved through inviting representatives of each catering team, two (one manager and one chef) from each company, to an introductory meeting held at the Guangzhou CDC. Here it was explained what the catering team were required to do and how the CDC would support them. Regular supportive school lunch evaluation and constructive feedback in relation to the improvement objectives was given weekly to the lunch providers by the CHIRPY DRAGON teachers. Each school was scored on a weekly basis using a range of 0 - 5 by CDC [5 - excellent; 0 - bad] according to how well they performed against each of the improvement goals.

Component 3 promoted easy, fun and safe family-friendly physical games, exercises or sports activities for both children and their parents. These, which could be undertaken with minimal equipment and space at home, were taught and tried out during one taster session in every class each school term. Each family was assigned a “healthy behavioural challenge” to play one of the games learnt or to engage in any other sports or activities they preferred, outside of school for a minimum of 30 minutes each weekend. This behaviour challenge required written parental feedback, using the same fun cards and procedures as Component 1. Again, the CHIRPY DRAGON teachers rewarded best performance and improvements.

Component 4 consisted of improving the implementation of the Chinese national requirement of 'One-Hour of Physical Activity on Campus every School Day' in intervention schools. CHIRPY DRAGON teachers facilitated a 30-minute meeting at each school involving school principals, class-level head teachers, PE teachers and student representatives. They discussed their current situation, identified barriers to implementing the national requirement, which could be modified, as well as improvement opportunities. Additionally, monthly action goals and plans which were measurable, achievable and maintained or exceeded the national standard were set by the school staff. Continuous evaluation and feedback was provided by the CHIRPY DRAGON teachers on a monthly basis.

All intervention components are summarised in Table 5.1.

Comparator

Schools allocated to the comparator/control arm continued with their ongoing standard provision (usual practice) during the full trial period with no access to any of the CHIRPY DRAGON intervention activities and resources.

Table 5.1 - Overview of the CHIRPY DRAGON programme

Target groups	Activities	Number of sessions/workshops and setting
Component 1: To improve childhood obesity related knowledge, skills and behaviours among children and their main carers		
Main carers (parents/guardians or grandparents)	(A) Interactive educational workshops for main carers	Two, school-based
Children	(B) Interactive educational activities for children	Four, school-based
Children and their main carers	(C) Family-wide healthy behavioural challenges and child self-monitoring (during the week)	Three daily challenges rotated every two weeks throughout the intervention year, home-based
	(D) Health knowledge quiz for main cares and children	One, school-based
Component 2: To improve the nutritional quality of school lunch provision (usually a set lunch box for each child)		
School lunch providers including catering staff	(A) Introduce school lunch improvement objectives which were set jointly by researchers and school lunch providers and then tested by school lunch providers (including both commercial suppliers and school funded catering units)	One introduction meeting held in the Guangzhou CDC
	(B) Supportive school lunch evaluation and feedback in relation to the improvement objectives	Continuous throughout the intervention year, school-based
Component 3: To increase children's physical activity level outside school		
Children and their parents	(A) Fun and active family games taught and tried in school	Two, school-based
	(B) Assign home work (a family-wide healthy behavioural challenge) - practice the games learnt or any other non-sedentary activities involving the parents (at the weekend)	At least 30 minutes of the challenge every weekend, home-based
Component 4: To increase children's physical activity level in school		
Children and school staff	(A) Situation analysis in relation to current implementation of the Chinese national standard of having one-hour physical activity on campus every school day	One staff meeting held at each school
	(B) Setting monthly goals (measurable and achievable) and action plans to meet, maintain or exceed the national standard and continuous evaluation and feedback	Monthly meetings held throughout the intervention year, school-based

Li et al. (2019) [129]

5.2.2 Resource Use and Costs

Rationale for Costing

Costs collected focused on those that were likely to vary between the intervention and control arm.

The costs linked to the intervention were divided into three categories:

- Development/set-up
- Implementation
- Delivery/running

Table 5.2 summarises the cost items associated with each intervention category.

Category one consisted of costs related to the development and set-up of the intervention. Category two comprised costs regarding the implementation of the intervention. Category three costs were associated with the delivery of the intervention.

According to standard practice, the base case analysis assumed that the intervention was in a 'running state' and therefore only costs associated with the delivery of the intervention were included. All of the other costs (set-up and implementation) were, however, reported separately and implementation costs were fully explored within the sensitivity analysis.

Table 5.2 - Cost items by intervention category

Category 1: Intervention development/set-up	Category 2: Intervention implementation	Category 3: Intervention delivery/running
• Research staff time for development of the school teacher handbook explaining intervention • Hiring of a designer to optimise the presentation of intervention materials (leaflets and illustration media) • Researcher preparation time for CHIRPY DRAGON teacher training • Time and travel costs related to staff meeting at each school to discuss their current situation about children's physical activity	• Time and travel costs connected with the CHIRPY DRAGON teachers training to deliver workshops/sessions • Initial printing of school teacher handbooks • Time and travel costs related to the intervention set-up meeting to explain about the intervention components to school staff • Time and travel costs related to the catering team meeting for introducing five school lunch improvement objectives	• Labour: CHIRPY DRAGON teachers' time and workshop assistants' time • Intervention materials used during workshops/sessions • Delivery fee for reward boards and loudspeakers • Office stationery • Relevant printing • Incentives: incentive prizes for meeting family healthy behaviour challenges and performance recognition certificates for catering teams • CHIRPY DRAGON teachers' transport • CHIRPY DRAGON teachers' telephone

Resource Use and Costs

Resource use was collected from both the public sector and societal perspectives. For the former, all public sector resource use was multiplied by the relevant unit cost (Yuan currency), obtained from different Chinese sources such as the Guangzhou CDC, or valued at market prices (e.g. purchase order receipts collected by the trial team), to calculate the total cluster (school)-level cost. The cluster-level costs were then averaged across the number of classes and average number of children per class (n=45), to derive average class level and individual child costs for the intervention arm.

All unit costs were reported in Chinese Yuan at a 2016-2017 price base. Gross Domestic Product Purchasing Power Parities (GDP PPPs) [253] were used to convert Yuan into Pound/Dollar (the conversion of Yuan to Dollar rate is 3.55 and to Pound rate is 5.05 using the Organisation for Economic Co-operation and Development (OECD) Data in 2017) [254]. [For details on the use of GDP PPPs - see Appendix 2.2.

Public-Sector Resource Use Collection and Unit Costs (Delivery of the Intervention)

Given the school-based multi-faceted nature of the intervention, when using a public sector perspective, only costs falling on the schools were included. There were four main components to the intervention. Study-specific resource use forms were developed for each of these and CHIRPY DRAGON teachers were asked to complete these during the intervention period (March 2016 – March 2017). A single standardized form was used to record all working time spent by each CHIRPY DRAGON teacher on the various intervention activities including their administration time. Each activity was denoted by a different letter on the form. An explanation of what the expected time requirements were for each intervention activity is presented in

Appendix 2.3, while the CHIRPY DRAGON teachers' record of minutes worked per week on each intervention activity are presented in Appendices 2.4 i – iii.

Detailed information of all resource use collected and associated unit costs applied is presented in Table 5.3.

Component 1: Improving childhood obesity related knowledge and behaviour among children and their main carers (predominantly parents and grandparents)

CHIRPY DRAGON teachers recorded how much time was spent on the delivery of the workshops, classroom sessions/quizzes and on reviewing/providing feedback about children's performances during the week-only family healthy behavioural challenges associated with this intervention component, using a standardized form. Only the CHIRPY DRAGON teacher was involved in delivering the workshops/sessions. The materials used within the workshops/classroom sessions (e.g. food used in child carers workshops) or through healthy behavioural challenges (e.g. the CHIRPY DRAGON Reward Boards, used to record children's achievements as well as the involvement of their main carers) were purchased by each CHIRPY DRAGON teacher either to be delivered or taken to a school in person. Any costs associated with delivering materials to the schools to facilitate the workshops were either directly calculated from an actual delivery fee or an estimation of the teacher transportation cost. Receipts for the materials used within the workshops/sessions (e.g. printing educational leaflets) were logged and costed.

Component 2: Improving the nutritional quality of school lunch provision

CHIRPY DRAGON teachers recorded how much time was spent providing regular school lunch evaluation and feedback on the five school-lunch improvement objectives. This evaluation was undertaken by direct observation (once a month); review of daily lunch record forms (completed by lead chef) and photographs of lunch boxes (submitted weekly by catering or school staff). These materials were reviewed and scored weekly by the CHIRPY DRAGON teachers against the five pre-specified school lunch improvement goals. Reward certificates were issued to catering teams for best performance. Receipts for printing these were logged and costed.

Component 3: Increasing children's physical activity outside of school

CHIRPY DRAGON teachers recorded how much time was spent on the delivery of the school-based family friendly games (taster sessions) and on reviewing/providing feedback about children's performances in the weekend-only family healthy behavioural challenges associated with this intervention component. On most occasions it was only the CHIRPY DRAGON teacher who was involved in delivering these workshops/sessions. However, on a few occasions when there was extra help from workshop assistants, these costs were also included. The materials used within the workshops/classroom sessions (e.g. finger board used in physically active family friendly games) or through the healthy behavioural challenge (e.g. CHIRPY DRAGON Reward Board or fun cards) were purchased by each CHIRPY DRAGON teacher either to be delivered or taken to the school in person. As before, any costs associated with delivering materials to the schools were either directly calculated from an actual delivery fee or an estimation of the teacher transportation cost. Receipts for the materials (e.g. strips of cloth), used within the workshops/sessions were logged and costed.

Component 4: Increasing children's physical activity level in school

Monthly meetings were held in each intervention school at which CHIRPY DRAGON teachers provided constructive feedback and action goals. Plans were reviewed and redefined for each subsequent month. CHIRPY DRAGON teachers reviewed record forms and scored each school's performance against the goals (scores ranged from 0 to 3). CHIRPY DRAGON teachers also recorded how much time was spent attending these meetings. The costs associated with the CHIRPY DRAGON teacher's travel to each school to attend these monthly sessions was estimated.

Details for all unit costs applied to each component/activity of the intervention is outlined in Table 5.3.

Unit Costs Associated with the Teaching Staff

The monthly salary paid to CHIRPY DRAGON teachers was based on an estimation of the standard monthly salary rate for primary school teachers in China, provided by Guangzhou CDC. This was 4,500 Yuan (salary) + 1,517 Yuan (employer contribution for social security, insurance and pension) = 6,017 Yuan. The annual salary was calculated by multiplying the monthly rate by 12 months (72,204 Yuan). The annual salary was based on a contract of 1,440 hours per year. Thus, to estimate the cost per hour, the annual salary was divided by 1,440. Annual salary scales were not available for teaching assistants. A unit cost for approximately 2 hours, was derived from an estimation of average salary levels in China provided by Guangzhou CDC.

Other Unit Costs

For components 1 and 3, the unit costs for the workshop materials, printing costs, and incentive prizes were based on the purchase price of each item and where relevant, included the delivery fee. For component 2, the cost of the incentives for the catering team (recognition certificates) was based on the purchase price. Office stationary was costed using the purchase price of each item, and then spread equally across the 4 components. The cost of telephone calls made by the CHIRPY DRAGON teachers, in relation to intervention delivery, were estimated by averaging the monthly cost of calls, and then dividing that equally across the 4 components. Finally, the unit cost for the CHIRPY DRAGON teachers' transport was based on the average price of a train ticket from the office to each school, obtained from the China National Railways per month and was charged to components 1, 3 and 4 by considering the annual number of workshops/sessions, within each component.

Table 5.3 - Resource use and unit costs (Yuan, 2016/17 Prices), public sector perspective

Type	Resource use item for intervention delivery	Unit cost (Yuan)	Quantity for 12 month	Total	Source
Office stationery	Ink pads (Components 1, 2, 3 and 4)	4.9/each	4	19.6	Online (taobao.com)
	Permanent markers (Components 1, 2, 3 and 4)	1.2/each	12	14.4	Online (taobao.com)
Printing	Colourful educational leaflets for parents and families (Component 1)	0.58/each	1950	1131	Feida Tu Wen store
	Family healthy behaviour challenges fun cards (Components 1 and 3)	0.15/each	27000	4050	Feida Tu Wen store
	Stickers (with CHIRPY DRAGON logo) (Components 1 and 3)	0.14/each	3700	500	Caiyi Bangong Haocai store
	Record cards of individual performance (Components 1 and 3)	6.5/each	111	721.5	Wanmei Chongyin store
	Illustrative photo cards for child workshops (first semester) (Component 1)	0.4/each card	2030	819	Wanmei Chongyin store
	Illustrative photo cards for child workshops (second semester) (Component 1)	0.4/each card	1190	476	Wanmei Chongyin store
	Illustrative photo cards for child carers workshops (Component 1)	0.35/each card	7300	2580	Wanmei Chongyin store
	Teaching boards for child carers workshops (Shahe School) (Component 1)	0.15/each note	40	6	Wanmei Chongyin store
	Family healthy behaviour challenges reward board (Components 1 and 3)	0.29/each	3000	850	Feida Tu Wen store
Labour	Workshops assistant (each practice = 2 hours) (Component 3)	30.5/each practice	77	2348	Standard monthly salary rate (Guangzhou CDC)
	Hour of CHIRPY DRAGON teachers time (Components 1, 2, 3 and 4)	50/hr	708.5	35425	Standard rates provided by Guangzhou CDC
Delivery fee	Family healthy behaviour challenges reward board (Components 1 and 3)	20/delivery	2	40	Guangjun Tourism
	Loudspeakers (Components 1 and 3)	12/delivery	2	24	Guangzhou transportation group taxi Co. Ltd

Table 5.3 - Resource use and unit costs (Yuan, 2016/17 Prices), public sector perspective (continued)

Type	Resource use item for intervention delivery	Unit cost (Yuan)	Quantity for 12 month	Total	Source
Workshops materials	PowerPoint remote control (Component 1)	32/each	7	224	Online (taobao.com)
	Canister (Component 1)	1/each	30	30	Online (taobao.com)
	Measuring spoon (Component 1)	6/each	3	18	Online (taobao.com)
	AA batteries (Component 1)	3.1/each	25	77	Jinli store
	Stamp (Components 1 and 3)	30/each	3	90	Jinli store
	Transparent plastic bag (Component 1)	0.01/each	200	19.4	Online (taobao.com)
	Suitcase (Components 1 and 3)	99/each	2	198	Online (taobao.com)
	Cart (Component 1 and 3)	36.8/each	3	110.4	Online (taobao.com)
	Paper plates for child workshops (Component 1)	0.19/each plate	1500	295	Online (taobao.com)
	Electronic scale (Component 1)	23/each	3	69	Online (taobao.com)
	Food used in child carer workshops (Component 1)	19.8/each workshop	40	793	Local market, Wal-Mart
	Canister (larger) (Component 1)	2.42/each	6	14.5	Online (taobao.com)
	Loudspeakers (Components 1 and 3)	160/each	5	800	Online (taobao.com)
	Stamp (with CHIRPY DRAGON logo) (Components 1 and 3)	22.5/each	10	225	Qingqing Wenxue store
	Balloon (Component 3)	0.1/each	2000	200	Online (taobao.com)
	Finger board used in physically active family	37.5/each	2	75	Online (taobao.com)
	Pencils (Components 1 and 3)	6/each	5	30	Online (taobao.com)
	Rubber band (Component 3)	0.008/each	2000	15	Online (taobao.com)
	Megaphone (Components 1 and 3)	40/each	6	240	Online (taobao.com)
	Strips of cloth (Component 3)	1/each	100	100	Qingqing Wenxue store
Incentives	Incentive prizes for meeting family healthy behaviour challenges (Components 1 and 3)	9.7/each	991	9620	Jinli store, Qingqing Wenxue store
	Recognition certificates for catering teams (Component 2)	14.8/each	33	489	Online (taobao.com)
Teachers telephone	CHIRPY DRAGON teachers average cost of mobile phone (Components 1, 2, 3 and 4)	269.8/month	10	2698	China mobile
Teachers transport	CHIRPY DRAGON teachers average cost of transport (Components 1, 3 and 4)	744.25/month	12	8931	China National Railways

Societal Resource Use Collection and Unit Costs (Delivery of the Intervention)

Components 1 and 3 of the intervention included workshop elements which parents and other live-in adult family members were invited to attend. Family-specific resource use questionnaires were developed and used to obtain information. To form part of an exploratory analysis, these questionnaires were also used to help to understand families' expenditure patterns by collecting data on household expenditure on food and other items. The data collected on household expenditure is not included in the societal economic evaluation – reasons explained later. All these self-reported questionnaires were completed by parents/main carers at home.

School Lunch Cost

The cost of providing lunch each day was recorded by the catering team in 38 schools (20 control and 18 intervention). The other two intervention schools did not have a lunch provision during the intervention year.

Generally all children eat school lunch (rather than a packed lunch) when it is provided. To estimate the mean (standard deviation (SD)) daily cost of lunch for each diner, the school mean lunch cost was divided by the average number of diners on each school day. The average daily cost of lunch per diner was then calculated for the control and intervention schools separately. As the intervention period was one year, for analysis purposes, this average cost was then multiplied by the number of days (250) which schools were open during the 12 month trial period to obtain the annual average lunch cost per child for the intervention versus control schools. To calculate the incremental cost of lunch paid for by the children's families, the annual average lunch cost was linked to the children's data. The overall average lunch cost per year for the intervention group was assumed for the two intervention schools which did not have lunch provision during the intervention year.

Parents/Other Family Members' Attendance Time/Cost

For the intervention workshops, either parents or main carers were invited to attend. On average, two parents or main carers per child attended each workshop. The travel costs for these parents/main carers to attend the workshops, were not collected. This was because most families lived close so travel costs were likely to be negligible.

A summary of parents/other family members' attendance time for these workshops is shown in Table 5.4.

Table 5.4 - Parents/main carers' workshop attendance time

	Workshop's name	**Term1 (hours)**	**Term2 (hours)**	**Total time spent at each workshop type across the whole trial duration (hours)**
Component 1 (main carers)	Education	1	1	2
	Family quiz	-	1	1
Component 3 (parents)	Family game	0.5	0.5	1

Using questionnaires, family members of children from all the 20 intervention schools were asked what they would have been doing if not attending the workshop, to measure the opportunity cost of their time. The alternative activities they could select from were defined as: 'at work' or 'not at work'. As it was not possible to reliably source salary details for all the occupations provided, the population average salary was used to estimate the value of lost time

[255]. For those who would not have been at work, the national minimum wage [256] was assumed as a valid cost of leisure time [257, 258]. The data revealed that overall, across all the workshops, 61% of the parents/grandparents would have otherwise been at paid work and therefore 39% of them were using unpaid time. To measure the average cost of time spent by each family member attending the workshop the following method was applied:

Step 1:

Average cost of workshop time for each family member in employment = average hourly wage rate * workshop hours over trial duration = A

Average cost of workshop time for each family member not in employment = average hourly leisure rate * workshop hours over trial duration = B

Step 2:

Number of paid family members attending the workshop = 61% * total attendees based on the process evaluation data = C

Number of unpaid family members attending the workshop = 39% * total attendees based on the process evaluation data = D

Step 3:

Average cost of each family member time = (A*C) + (B*D) / total attendees based on the process evaluation data = E

Steps 1 through to 3 were applied to calculate the average cost for each family member who attended the workshops. This cost was also assumed for family members with missing data.

Household Expenditure

The following information, on monthly household expenditure of families within the control and intervention groups at both baseline and at 12 months follow up was collected:

1) Household electricity and gas bills
2) Transport (e.g. car fuel, parking and bus/train tickets)
3) Recreation and culture (e.g. visiting a park, cinema or theatre; karaoke; sports centre/GYM membership)
4) Shopping for food and non-alcoholic drinks (soft drinks)
5) Alcoholic drinks, tobacco and narcotics
6) Eating in restaurants (not for business purposes) and small sit-in outlets
7) Clothing, footwear and other accessories
8) Communications (e.g. phone and internet bills)
9) Household goods (both buying and repairing) and services (e.g. house servant wages)
10) Other goods and services (e.g. hair dressing, body massage, facial treatments etc.)
11) Education (e.g. child interest lessons, after school childcare and adult learning courses)
12) Healthcare (e.g. medical treatments & medication and food supplements)

An open-ended questionnaire was used to collect information about the amount of money which each family spent per month on each of these categories. This approach revealed families' consumption behaviour and revealed WTP values per month across the categories in the intervention and control groups at both baseline and at 12 months follow up. Data was also collected on total family income in the intervention and control groups at both baseline and at 12 months follow up, using an open-ended question design.

More detailed questions on the weekly food household expenditure was also collected to measure the weekly pattern of food expenditure when compared to the monthly food expenditure pattern. This information was collected separately (for month and per week) using two different sets of questions as follows.

Monthly

1) Shopping for food and non-alcoholic drinks (soft drinks)
2) Eating in restaurants (not for business purposes) and small sit-in outlets

Weekly

1) Total food expenditure (groceries, eating out and takeaways)
2) Fruit and vegetables (fresh, frozen or canned)
3) Ready meals (microwave meals), frozen fast food and takeaways

The figures for weekly category (1) includes but is not limited to those for categories (2) and (3). The purpose of including the weekly food household expenditure was for behavioural change evaluation at the weekly level. A questionnaire with a payment scale was used to collect the weekly household expenditure and families were asked to circle a response. Then, midpoints of the ranges for the three categories were used to assign a value for each family. In the interval midpoint WTP model, it is assumed that the family's WTP is distributed within the given interval [259]. Since a family's true point valuation lies somewhere in the interval between the chosen value and the next higher one, this assumption is reasonable [259]. However, the WTP values calculated by this method are always relatively higher compared with those obtained using other methods. It is plausible that this method is an optimistic estimation [259].

The mean (SD) of monthly and weekly expenditures on each of the different categories and mean (SD) of family income were calculated at each time point (baseline and 12 months follow up) for all participants, control and intervention groups. The percentage/proportion of income spent on the different monthly household expenditure categories, was then calculated for both the control and intervention arms.

Development and Implementation Unit Costs

The justification of and source for the unit costs applied to the development and implementation categories are outlined in Table 5.5 and detailed below.

Unit Costs Associated with the Research Staff

These costs were associated with the development (i.e. school teacher handbook and researcher preparation time for CHIRPY DRAGON teachers' training) and implementation (i.e. training CHIRPY DRAGON teachers to deliver workshops/sessions) of the intervention. To calculate the unit costs for each research associate; research fellow; and senior research fellow role, the 2016-2017 Guangzhou CDC academic research staff salary scales were used (between 7,500 and 9,200 Yuan). These salaries were then converted to an hourly rate assuming that research staff works 7.5 hours per day excluding weekends, university 'closed days' and public holidays.

Unit Costs Associated with Catering Teams (Managers/Chefs) and Designer

The monthly salary rate of catering staff (36 managers/chefs from 18 catering companies) was based on an estimation of the standard monthly salary rate in China provided by Chinese colleagues. This was 10,000 Yuan. The annual salary rate was then calculated by multiplying the monthly rate by 12 months (120,000 Yuan). The annual salary rate was based on a contract of 1,350 hours per year. Thus, to estimate the cost per hour, the annual salary was divided by 1,350. For the costs associated with the development of the intervention, a designer was employed to optimise the presentation of intervention materials (leaflets and illustration media). The cost of this designer was also included at an agreed fixed price of 20,000 Yuan for 2 months.

Other Unit Costs

School principals (n=20) and class-level head teachers (n=20) were provided with a detailed programme handbook which explained all intervention activities and the support needed from them. The unit cost for the printing of handbooks was based on the purchase price of each handbook. This was included in the implementation costs.

The intervention set-up meeting, which was for 20 school principals, was part of the implementation costs and held in China Guangzhou CDC. With regard to the school principals meeting time, their attendance was part of their existing/routine duties rather than additional work or unpaid work therefore no meeting time or cover cost has been calculated for them. The meeting lasted two hours and was facilitated by one of the CHIRPY DRAGON teachers. The cost of this is included in the implementation costs based on the previously calculated hourly rate. All the school principals travelled to the meeting by underground. The unit cost for their return train tickets was based on the China National Railways price of each ticket. The printed school teacher handbooks were handed out at this meeting.

The catering team introduction meeting, for managers/chefs from the catering companies, was part of the implementation costs and held in China Guangzhou CDC. The meeting lasted one hour. The catering staff were paid to attend based on their previously calculated hourly rate. A lead CHIRPY DRAGON teacher facilitated this meeting and this was, again, costed at the appropriate rate. The catering staff all travelled by underground and the same unit cost as above was applied.

Table 5.5 - Development, implementation resource use and unit costs (Yuan, 2016/17 Price)

Type	Resource use item	Unit cost (Yuan)	Quantity for 12 months	Total	Source
Teachers time costs	CHIRPY DRAGON teachers time (Development and implementation)	50/hr	150.7	7535	Standard monthly salary rate (Guangzhou CDC)
Research team time costs	Senior research fellow time (Development)	72/hr	2	4251.6	Academic research staff salary scales (Guangzhou CDC)
	Research fellow time (Development and implementation)	67/hr	26.45		
	Research associate time (Development and implementation)	62/hr	39.2		
Catering team time costs	Catering team time (Implementation)	74/hr	36	2664	Standard monthly salary rate (CHIRPY DRAGON Chinese staff)
Hiring a designer	Designing diagrams (Development)	10000/month	2	20000	An agreed fixed price
Printing	Handbooks (Implementation)	0.4/each	40	16	Feida Tu Wen store
Teachers, research staff, catering team and school principals travel costs	Underground return train ticket (Development and implementation)	4/each	90	360	China National Railways

5.2.3 Outcome Measures

All outcomes were collected at the individual level. Assessments were undertaken in each school by independent and trained assessors (research staff) who were blind to allocation, using standardised procedures and instruments at baseline (start of intervention) and first follow-up (end of intervention). Data on participants' date of birth and gender were obtained from school records.

Public Sector Outcome Measures

The primary clinical outcome for effectiveness was the difference in BMI standard deviation scores (z scores) between arms at completion of the 12-month intervention. There is no agreed consensus on the minimal clinically important difference for BMI z-score. Clinical studies in obese children have shown improvements in clinical measures such as blood pressure and lipid levels with differences as small as 0.1 units in BMI z-score [260]. Other studies have suggested that a clinically important difference is 0.125 units [260]. However, for prevention interventions, at a population level, even smaller values could be clinically important.

The primary economic outcome measure was QALYs. No other clinical/non-clinical outcomes were considered for the economic evaluation as only the most commonly used outcomes (QALYs/BMI) were explored. Other outcomes including body fat%; waist circumference; eating behaviours; physical activity; and sedentary behaviours were explored for the intervention clinical effectiveness. Other important outcomes such as educational attainment would have been useful to consider, but these data were not collected within the trial.

Anthropometric Measurements

Height and weight measurements were undertaken without shoes and in light clothing. Standing height was measured at least twice with a TGZ-type height tester (Dalian). Weight was measured with an electronic scale (JH-1993T, weighing Apparatus Co. Ltd., Dalian, China). BMI was calculated as weight in kilograms divided by the square of height in metres (kg/m2). The WHO 2007 Growth charts, which are most widely used and known in the region, were used to calculate BMI z-scores and categorise the children into underweight, healthy weight, overweight and obese groups [261].

Measurement of HRQoL

CHU-9D and PedsQL, which are both generic instruments, were chosen for the measurement of HRQoL. Children were invited to complete both researcher-administered questionnaires (CHU-9D and PedsQL) at the same time at each time point when they were attending workshops.

Although there is no gold standard for measuring utility-based HRQoL in primary school-aged children, previous research has shown the CHU-9D, a recently developed instrument, is the most appropriate choice [38]. As a utility-based instrument, it is preference-based. It is designed for application in cost-effectiveness analyses of prevention, treatment and service programmes targeted at young people where the QALY is the desired outcome measure [183].

The PedsQL is a non-preference based instrument and therefore does not apply any explicit weighting between item domains. It therefore cannot be used to generate utility values for the construction of QALYs. However, it would be expected to produce HRQoL values which move in the same direction as the utility measures. This instrument is a widely used HRQoL instrument validated for use with young children over 5 years old. It has good reliability and validity in both sick and healthy populations [189, 190].

CHU-9D

The Chinese version of the CHU-9D (CHU9D-CHN) instrument was used to collect HRQoL information for the children [262]. Individual responses from the questionnaires were transformed into utility weights derived from a UK general population sample using an algorithm developed by Stevens et al [39, 184]. This presents a possible utility value set of between 0.33 (worst health state) and 1 (best health state). QALYs were calculated for each individual child, over the 12 month period, using the area under the curve approach which uses the trapezium rule [263]. The Chinese-specific preference weights applied in a sensitivity analysis as the Chinese version tariff was still under development at the time of analysis [264, 265].

PedsQL

For this study the validated Chinese version of PedsQL 4.0 instrument was used [266]. Provided data were available for at least half of the relevant items, the mean score for each of the four domains was then calculated by summing the values for the relevant items and dividing by the number of items answered. This was repeated including “mean of all items” for the total score, “mean of physical functioning items” for the physical health score and “mean of emotional, social and school functioning items” for the psychosocial health score.

Assessing QALY and BMI z-score Differences

To control for differences in baseline utility values between the intervention and control arms, pre-specified school- and child-level covariates were adjusted for, based on a statistical analysis plan [263]. Using regression analysis, there was a control for differences in cluster level variables used in the randomisation (provision of mid-morning snacks and having access to an indoor activity room); characteristics of children ((socio-demographic: gender and mother education level) and (health behaviour factors: consumption of fruit and vegetables; unhealthy snacks and sugar added drinks; minutes/day MVPA and sedentary time)); and baseline CHU-9D utility scores between the intervention and control arm [126].

Therefore, three models were applied:

- A linear regression model (unadjusted model)
- Mixed linear regression model (account for clustering and controlling for baseline utility)
- Mixed linear regression model (account for clustering, controlling for baseline utility and pre-specified school- and child-level covariates)

The first model, linear regression of QALYs, is an unadjusted model. The data however were clustered therefore the second model adopts a hierarchical approach to account for clustering whilst also controlling for baseline utility differences. The third and final model, and the one which was used for the main analysis, adds the pre-specified school- and child-level covariates to model two. This model therefore adjusts for clustering, baseline utility and the covariates specified within the analysis plan.

Regarding BMI z-scores, treatment effects (mean difference in BMI z-scores between the two arms at 12 months follow up) were tested using a mixed linear model adjusting for the child baseline BMI z-score and clustering by school [267]. The analysis was adjusted for pre-specified school- and child-level covariates as well.

Societal Outcome Measures

For the societal perspective, the analysis included QALY gains/losses falling on adult household members.

Measurement of HRQoL

EQ-5D-3L

The validated Chinese version of the EQ-5D-3L was chosen for the measurement of adult HRQoL [191]. Parents and other adult household members were asked to complete the EQ-5D-3L at home at each time point (baseline and 12 months follow up). Individual response permutations to the EQ-5D-3L were used to calculate health index scores (utility scores) based on the UK value set. This presents a possible utility value set of between -0.59 (worst health state) and 1 (best health state). QALYs were calculated for each individual parent/grandparent, over the 12 month period, using the standard area under the curve approach [36]. Parental/grandparental utility scores and QALYs were only estimated from data where the same parent/grandparent completed both the baseline and the follow up measures.

The Chinese tariff scores were applied for parents/grandparents in a sensitivity analysis. There were two potential published papers to facilitate this analysis [268, 269]. Liu et al (2014) [268] was used as it included respondents from urban China whereas the other paper [269] only included respondents from both urban and rural China, which was less representative of the CHIRPY DRAGON trial population.

Assessing QALY Differences

To control for differences in baseline utility scores between the intervention and control arm, pre-specified parent-level covariates were adjusted for, based on a pre-specified statistical analysis plan [263]. Using regression analysis, there was a control for differences in characteristics of parents ((socio-demographic: gender and mother/father education level); and baseline EQ-5D-3L utility scores between the intervention and control arm [126].

Therefore, three models were applied:

- A linear regression model (unadjusted model)
- Mixed linear regression model (account for clustering and controlling for baseline utility)
- Mixed linear regression model (account for clustering, controlling for baseline utility and pre-specified parents-level covariates)

5.2.4 Missing Data and Multiple Imputation

Resource use data, using a public sector perspective, was collected at the cluster level whereas the outcome data and the resource use data from a societal perspective, were collected at the individual level. As a result, the reasons for missing data differed for these two types.

Public Sector Missing Data and Multiple Imputation

Given a very high retention rate and a high level of data completeness (0% missing for the resource use data; less than 4% for the children's outcome data) for the public sector perspective analysis, there was no need to use multiple imputation methods to account for missing data. Therefore, a modified intention to treat (ITT) approach was used. This means all analyses were conducted on participants with non-missing data in their original randomisation groups.

Societal Missing Data and Multiple Imputation

For the societal perspective analysis, although almost 25% of the outcome data for parents/grandparents were missing, multiple imputation for the base-case analysis was not required. This was for two main reasons:

- The covariates in the model were fairly complete.
- The baseline characteristics of the study participants were well balanced between the groups with available and missing outcome data.

These analyses were therefore also conducted using the modified ITT approach. Missing data for the resource use regarding school lunch costs and workshop attendance time costs for parents and other family members were assumed.

The impact of not imputing the missing data for the societal perspective analysis was explored in a sensitivity analysis. This was to avoid any loss of efficiency or potential bias of the results with the exclusion of participants with missing data [270, 271]. Health index scores for children (CHU-9D) and parents/grandparents (EQ-5D-3L) were imputed at baseline and 12 months [272] using predictive mean matching multiple imputation. The baseline health and relevant co-variates (gender, mother/father education level, provision of mid-morning snacks, having access to an indoor activity room, daily consumptions of fruit and vegetables, weekly servings of unhealthy snacks and sugar added drinks, objectively measured time in MVPA and sedentary time (minutes/day)) for each imputed variable were selected separately and included within the imputation model [273]. Predictive mean matching was chosen as it is reported superior for imputing continuous variables which are not normally distributed [274]. This method has the advantage that it produces imputed values which are more like real values because the imputed values are real values which are “borrowed” from individuals with real data [274].

The level of missing data was less than 16% for the monthly and weekly household expenditure, and monthly income data. A modified ITT approach was used as this analysis was not included in the economic evaluation and was only exploratory.

5.2.5 Statistical/Economic Evaluation Methods

An initial data analysis was performed to clean and validate the data entered in the dataset. The analysis of cost-effectiveness was undertaken according to current best practice methods for conducting economic evaluation alongside cluster randomised controlled trials [275].

A mixed linear model to analyse the primary clinical outcome was used. Logistic or mixed linear model functions were used to analyse binary or continuous primary outcomes (QoL) respectively. For analysing utility scores, mixed models were used. The cost data was highly skewed therefore a gllamm model (gamma log link) was used which allowed the data distribution to be specified.

For continuous outcomes, the distribution of residuals for normality to conduct linear regression analysis was checked. For all outcomes, the residuals were normally distributed therefore no transformations were used. Means (SD) and 95% confidence intervals (CIs) were calculated for continuous variables which were normally distributed. For the outcomes that were not normally distributed, median (Inter-Quartile Range (IQR)) and 95% CIs were calculated.

The economic analyses took an incremental approach and it was assumed that there were no costs associated with the control arm. Since a time horizon of 1 year was used, costs and outcomes were not discounted [36]. Both a CUA and a CEA were conducted. The results were expressed through the ICER based on the fully adjusted costs and effects.

Within the Chinese setting, as there is no equivalent threshold value for how much decision makers are willing to pay for a unit gain in QALY, recommendations were made using the established UK and US threshold values as a reference point [161, 162]. For instance in the UK, ICER values under £20,000-£30,000 per QALY is deemed a cost-effective use of resources [161]. In addition, given the lack of a specified threshold for China, both 1xGDP and 3xGDP per capita thresholds, recommended by WHO, were used in the analyses [165, 166]. All analyses were conducted in Stata 13.

Public Sector Perspective for Economic Evaluation

The primary analyses from a public sector perspective included:

- A CUA estimating the cost per QALY gained.
- A CEA evaluating the cost per change in BMI (z-score).

The uncertainty around the ICER was assessed through the use of the NBR framework [163, 164] using non-parametric bootstrapping and decision uncertainty presented using CEACs whilst controlling for any baseline differences, clustering effects and co-variates [164]. CEACs were developed to estimate the probability of the intervention being cost-effective across a range of values of WTP for an extra QALY.

Sensitivity Analyses

Three sensitivity analyses were undertaken to assess the robustness of the results to assumptions made in the analysis:

- Sensitivity analysis 1: including costs associated with implementing the intervention

The base case analysis assumed that the intervention was in a 'running state' and therefore only costs associated with the delivery of the intervention were included. In this sensitivity analysis the implementation costs were included to assess the impact on the ICER.

- Sensitivity analysis 2: using the Chinese tariff to estimate QALYs

Individual responses from the CHU-9D questionnaires were transformed into utility scores using the Chinese tariff.

- Sensitivity analysis 3: varying the class size to only including consented children

Within the trial, parents had to provide consent for their children to have measurements taken. On average, there were 41 children with parental consent within each class. Sensitivity analysis 3 assumed an average class based on number of children who consented to measurement instead of average total class size (n = 45).

Societal Perspective for Economic Evaluation

The secondary analysis from a societal perspective included a CUA estimating the societal cost per QALY gained for children and household members combined. Children and family members' QoL data were linked and matched using a 'multiplier' approach. This involved adjusting children's QALYs for family spillover effects using the following steps [276-278]:

Step 1:

Mean incremental QALYs per child was calculated. **(CQ)**

Step 2:

Mean incremental QALYs for each family member was calculated. **(FQ)**

A judgement was made regarding the mean number of family members per child and was assumed to be two because, on average, two family members (parents/grandparents) attended the intervention workshops. **(n)**

Step 3:

The multiplier for each child was then calculated as:

$$[1 + (n*FQ / CQ)]$$

Additionally, an allowance (a figure of around 1.1) was made for the spillovers displaced by the intervention [276, 277]. For further discussion and technical explanation see Al Janabi et al (2016) [276, 277].

The multiplier approach avoids averaging all benefits out. It represents the fact that family benefits are additional from the same intervention not that children and family members are receiving the intervention separately.

The multiplier approach was not applied to the resource use data as it was not possible to link each component of the family-related costs to the related child. Instead, the costs were simply summed and averaged assuming that each child had at least two family members attend the workshops.

For the societal perspective therefore the base-case ICER was calculated by applying the following formula:

$$\frac{\text{mean incremental public sector costs}}{CQ} * \frac{1.1}{[1+(n*FQ/CQ)]} * \frac{\text{mean incremental societal costs}}{\text{mean incremental public sector costs}}$$

Sensitivity Analyses

Four sensitivity analyses were undertaken to assess the robustness of the results to assumptions made in the analysis:

- Sensitivity analysis 1: including costs associated with implementing the intervention
- Sensitivity analysis 2: using the Chinese tariff to estimate QALYs
- Sensitivity analysis 3: varying the class size to only include consented children
- Sensitivity analysis 4: using predictive mean matching multiple imputation

Exploratory Household Expenditure and Income Analyses

In addition to the economic evaluation, an exploratory analysis on household expenditure was undertaken. The change in proportion (%), CI and p-value of income spent on the different monthly household expenditure categories between the intervention and control groups were calculated. These were done with no adjustment, adjustment for baseline, and adjustment for baseline and different covariates to control for differences in families' characteristics (mother and father education level) using regression analysis.

The result of this analysis may be interesting as, in theory, the intervention could have had substitution effects on consumption behaviour as families spend proportionally more on food (e.g. purchasing fruit and vegetables), and perhaps less on alcohol or eating out. A common perception is that healthy eating is more expensive and this analysis investigated how the families adjusted their spending habits and adapted to the intervention.

Sub-group analysis was not conducted as the family income data was less accurate (not detailed enough to facilitate such an analysis).

A suitable model could be used to assess the long-term effects and cost-effectiveness. However, modelling was not applied for the research as it was beyond the scope of this thesis.

5.3 Part 2: Results: Economic Evaluation (Public Sector Perspective)

5.3.1 Participant Flow during the Trial

No schools dropped out of the trial. In total, 1641 children were recruited and randomized to 20 intervention (n= 832) and 20 control (n= 809) schools. 794 children (95.4%) from the intervention group and 768 (95.0%) from the control group were included in the primary outcome analysis. An overview of the trial participants is presented in Table 5.6.

Table 5.6 - CHIRPY DRAGON programme participants

	Intervention	**Usual Practice**
Eligible children from the consented schools (September 2015)	1799	
Children consented	1641	
Children measured at baseline (September-December 2015)	1630	
Number of children after randomisation		
Children consented	832	809
Children measured at baseline	826	804
Children measured (first follow-up) (April-July 2017)	805	781
Children included in analysis for primary outcome	794	768

5.3.2 Baseline Characteristics of the Study Participants

The baseline characteristics of the study participants were well balanced between the two groups (Table 5.7). The mean age of the children was 6.1 years (SD= 0.35) and 54.5% were male. More than a third of parents did not have a university education. Approximately 18% of the children were either overweight (10.8%) or living with obesity (7.1%); comparable to national data for overweight/obesity in the same age group (20.4%) [127].

Table 5.7 - Baseline characteristics of children participating in CHIRPY DRAGON study

	Intervention group (20 schools) n=832	Control group (20 schools) n=809
Age (years)	6·15 (0·36)	6·14 (0·35)
Gender		
Male	463 (55·6%)	431 (53·3%)
Female	369 (44·4%)	378 (46·7%)
Mother education level		
Lower education		
None	1 (0·1%)	1 (0·1%)
School education (Primary and Middle schools)	167 (20·5%)	137 (17·8%)
Occupation college	160 (19·6%)	132 (17·2%)
Higher education		
University education (Undergraduate level)	434 (53·3%)	433 (56·3%)
Postgraduate education	53 (6·5%)	66 (8·6%)
Father education level		
Lower education		
None	0 (0%)	1 (0.1%)
School education (Primary and Middle schools)	131 (15.8%)	122 (16.1%)
Occupation college	162 (19.9%)	132 (17.2%)
Higher education		
University education (Undergraduate level)	440 (54.9%)	407 (52.3%)
Postgraduate education	76 (9.4%)	107 (14.3%)
Anthropometric measures		
Weight (kg)	22.30 (4.32)	22.19 (4.28)
Height (cm)	119.77 (5.47)	119.49 (5.50)
Weight status*		
Thinness	37 (4·5%)	44 (5·5%)
Normal weight	637 (77·5%)	610 (76·6%)
Overweight	92 (11·2%)	83 (10·5%)
Obese	56 (6·8%)	59 (7·4%)

Table 5.7 - Baseline characteristics of children participating in CHIRPY DRAGON study (continued)

	Intervention group (20 schools) n=832	**Control group (20 schools) n=809**
Daily average servings of fruit and vegetables, median [IQR]	3·00 [2·00-4·00]	3·00 [2·00-4·00]
Weekly average servings of unhealthy snacks and sugar-added drinks~, median [IQR]	2·50 [0·00-4·50]	2·00 [0·00-3·50]
Objectively measured time in MVPA (minutes/24hours)	64·68 (30·79)	67·91 (29·12)
Objectively measured sedentary time (minutes/24 hours)	440·26 (90·11)	442·77 (87·01)

Data are mean SD or n (%), unless specified as median [IQR]
*based on WHO 2007 Growth Chart
Unhealthy snack consumption is estimated as the sum of average servings of salty high fat snacks (e.g. crisp, deep fried snacks), sweet high fat snacks (e.g. chocolates, cake, ice cream, and biscuits), candies and sugared beverages (e.g. carbonated drinks) in the previous week,
MVPA: moderate to vigorous physical activity

Li et al. (2019) [129]

5.3.3 Delivery Fidelity/Adherence and Attendance Rate

A high delivery fidelity/adherence was achieved and CHIRPY DRAGON teachers successfully delivered the intended number of workshops/quizzes/active game tasters, meetings and other activities [129]. Overall, 78% of the intervention events were delivered for the intended time length. Child attendance rates for the workshops ranged between 98% and 99%. Family members of 88% of the children attended school-based intervention activities aimed at child carers [129]. No reports were received on adverse events related to the intervention [129].

5.3.4 Impact of Intervention on Children's HRQoL (CHU-9D)

Table 5.8 outlines the response rate for the CHU-9D instrument. Missing/invalid data for this outcome was very low at both measurement points. There were no differences between the two study groups in completeness of this outcome measure.

Table 5.8 - Number of consented children with completed CHU9D

Time Point	CHU-9D completed all participants (n (%))	CHU-9D completed control group (n (%))	CHU-9D completed intervention group (n (%))
Baseline	1605 (97.8%)	793 (98%)	812 (97.6%)
12 months follow up	1587 (96.8%)	781 (96.6%)	806 (97%)

Table 5.9 presents the mean utility scores at each time point (using the UK tariff) for the base-case analysis. There was no baseline imbalance for utility scores, but nevertheless all incremental analyses were adjusted for baseline. At 12 months follow up, the mean utility scores for the intervention group was slightly higher than the control group. The mean utility scores at each time point (using the Chinese tariff for sensitivity analysis) were consistently lower compared to using the UK tariff (Appendix 2.5).

Table 5.9 - CHU9D utility scores at each time point (UK tariff)

	CHU-9D utility scores								
	All participants			Control group			Intervention group		
Time	Mean (SD)	(95% CI)		Mean (SD)	(95% CI)		Mean (SD)	(95% CI)	
Baseline (N= 1605)	0.937 (.069)	0.934	0.940	0.936 (.069)	0.931	0.941	0.938 (.068)	0.933	0.942
12 months follow up (N= 1587)	0.933 (.061)	0.929	0.936	0.928 (.064)	0.923	0.933	0.937 (.058)	0.933	0.941

Examining the Impact of the Intervention on QALYs

Table 5.10 describes the unadjusted mean QALYs (using the UK tariff). At 12 months follow up, the intervention group accrued 0.937 QALYs compared to 0.932 QALYs for the control group. The unadjusted mean QALYs (using the Chinese tariff) were consistently lower compared to using the UK tariff (Appendix 2.6).

Table 5.10 - Unadjusted QALYs accrued (CHU9D, UK tariff)

	Unadjusted QALYs accrued								
	All participants			Control group			Intervention group		
Time	Mean (SD)	(95% CI)		Mean (SD)	(95% CI)		Mean (SD)	(95% CI)	
12 months follow up (N= 1554)	0.935 (.051)	0.932	0.937	0.932 (.052)	0.928	0.936	0.937 (.049)	0.934	0.941

Incremental Analysis – Effectiveness

Table 5.11 describes the incremental difference in mean QALYs (using the UK tariff) between the intervention and control group for the data with no adjustment; adjustment for clustering and baseline differences; and adjustments for clustering, baseline differences and the pre-specified covariates using mixed effect linear regression models. In the unadjusted model, the mean QALY difference for 12 months follow up was in favour of the intervention but did not reach statistical significance (mean difference (MD) = 0.005, 95% CI: -0.003 to 0.014, p = 0.252) whereas when controlling for baseline utility and clustering, there was a significant difference in favour of the intervention group between the groups at 12 months follow up (MD = 0.004, 95% CI: 0.000 to 0.007, p = 0.034). Furthermore, after controlling for baseline utility, clustering, and the co-variates, there was some evidence of difference in the QALY for 12 months follow up (MD = 0.004, 95% CI: 0.000 to 0.008, p = 0.056) at borderline significance level. The QALYs attained using the Chinese tariff were higher compared to using the UK tariff and the results were statistically significant for both baseline and further adjusted models (Appendix 2.7).

Table 5.11 - Incremental difference in QALYs (CHU9D, UK tariff)

	No adjustment				**Adjusted for clustering and baseline utility [a]**				**Adjusted for clustering, baseline utility, co-variates [b]**			
Time	**Mean difference**	**(95% CI)**		**P-value**	**Mean difference**	**(95% CI)**		**P-value**	**Mean difference**	**(95% CI)**		**P-value**
12 months follow up	0.005	-0.003	0.014	0.252	0.004	0.000	0.007	**0.034**	0.004	0.000	0.008	0.056

[a]: Adjusted for baseline outcome. [b]: Adjusted for baseline outcome, pre-specified school- (i.e. whether the school provides mid-morning snack, whether the school has an indoor activity room) and child-level covariates (sex, mother education level, daily average servings of fruit and vegetables, weekly servings of unhealthy snacks and sugar added drink, objectively measured time in MVPA (minutes/24 hours) and objectively measured sedentary time (minutes/24 hours).

5.3.5 Impact of Intervention on BMI z-score

Table 5.12 outlines the response rate for the BMI z-score. Missing/invalid data for this outcome was very low at both measurement points. There were no differences between the two study groups having measurements taken.

Table 5.12 - Number of consented children with completed BMI z-score

Time Point	BMI z-score completed all participants (n (%))	BMI z-score completed control group (n (%))	BMI z-score completed intervention group (n (%))
Baseline	1618 (98.6%)	796 (98.4%)	822 (98.8%)
12 months follow up	1581 (96.3%)	777 (96%)	804 (96.6%)

Incremental Analysis - Effectiveness

There was no baseline imbalance for BMI z-score, but nevertheless all incremental analyses were adjusted for baseline. Table 5.13 describes the mean difference in BMI z-score between the intervention and control group. Overall, at 12 months (end of intervention period), the mean BMI z-score was significantly lower in the intervention compared to the control group, MD = - 0.13, 95% CI: -0.26 to 0.00, p = 0.048 in the baseline adjusted model; and MD = - 0.13, 95% CI: -0.26 to -0.01, p = 0.041 in the further adjusted model.

Table 5.13 - Adjusted differences for BMI z-score between groups at first follow up

Follow up outcome variable N = total participants	Intervention arm		Control arm		Intervention vs control (BL adjusted)[a]		Intervention vs control (Further adjusted)[b]	
	BL	**FU1**	**BL**	**FU1**	**FU1**		**FU1**	
	Mean (SD)		Mean (SD)		MD (95 % CI)	p value	MD (95% CI)	p value
BMI z-score N = 1581	-0.13 (1.30)	-0.35 (1.22)	-0.13 (1.30)	-0.23 (1.34)	-0.13 (-0.26 to 0.00)	**0.048**	-0.13 (-0.26 to -0.01)	**0.041**

SD: Standard deviation. CI: Confidence interval. BL: baseline. [a]: Adjusted for baseline outcome. [b]: Adjusted for baseline outcome, pre-specified school- (i.e. whether the school provides mid-morning snack, whether the school has an indoor activity room) and child-level covariates (sex, mother education level, daily average servings of fruit and vegetables, weekly servings of unhealthy snacks and sugar added drink, objectively measured time in MVPA (minutes/24 hours) and objectively measured sedentary time (minutes/24 hours).

Li et al. (2019) [129]

5.3.6 Resource Use and Costs

A breakdown of the resources used for the development, implementation and delivery of the intervention is presented in Tables 5.14 to 5.16. Regarding the implementation and delivery costs, the total mean cost per child, class and school is displayed in Tables 5.15 and 5.16 respectively. However, regarding the development costs, only the total costs are presented in Table 5.14.

Relative to the delivery costs, both the development and (in particular) implementation costs were low. With regard to the development costs, the largest cost component was hiring a designer to optimise the presentation of the intervention materials (leaflets and illustration media). CHIRPY DRAGON staff time made up the largest component of the implementation costs. Of the four main intervention components, the cheapest was improving the nutritional quality of school lunches. The most expensive was on improving childhood obesity related knowledge and behaviour among children and their main carers, which accounted for more than two thirds of the intervention costs.

For the delivery of the intervention, the most expensive costs were related to labour (CHIRPY DRAGON teachers' and workshop assistants' time) and printing. The rest of the costs were related to CHIRPY DRAGON teachers' transport, incentives, intervention materials used during workshops, CHIRPY DRAGON teachers' telephone costs, delivery fee and office stationary respectively.

Table 5.14 - Resource use and cost associated with development of intervention

Component	Resource type	Total resource use	Total cost, Yuan
Development of school teacher handbook (Explanation of the intervention)	**Research staff time (hours)** Research Fellow Senior Research Fellow	 6 2	 402 144
Hiring of a designer **(**To optimise the presentation of developed intervention materials (leaflets and illustration media))	**Designer time (month)**	2	20000
Researcher preparation time **(**For CHIRPY DRAGON teachers training)	**Research staff time (hours)** Research Associate Research Fellow	 12 6	 744 402
Staff meeting at each school **(**To discuss their current situation about children's physical activity)	**CHIRPY DRAGON teachers time (hours)** **CHIRPY DRAGON travel costs (transport)** Return train ticket	10 20	500 80
			22272 Yuan (6273 $, 4410 £)

Notes: Total number of intervention schools **(n = 20)**, Total number of intervention classes **(n = 85),** Total number of intervention consented classes **(20),** Assumed average class size **(45),** Total number of intervention consented children **(n = 832)**

Table 5.15 - Mean resource use and cost associated with implementation of intervention

Component	Resource type	Mean annual resource use per class	Mean annual cost per class, Yuan	Average cost per child (assuming average class size of 45), Yuan
Workshops/sessions training (CHIRPY DRAGON teachers training to deliver workshops/sessions)	**Research staff time (hours)**			
	Research Associate	0.32	18.69	0.42
	Research Fellow	0.17	11.98	0.27
	Research associate and research fellow travel costs			
	Return train ticket	0.16	0.65	0.01
	CHIRPY DRAGON teachers time (hours)	1.62	81.17	1.80
Initial printing of handbooks	**Number of handbooks**	0.47	0.18	0.004
Intervention set-up meeting (To explain about the intervention components to school staff)	**CHIRPY DRAGON teachers time (hours)**	0.02	1.18	0.03
	School principal travel costs (transport)			
	Return train ticket	0.24	0.94	0.02
Catering team introduction meeting (Managers and chefs meeting for introducing five school lunch improvement objectives)	**CHIRPY DRAGON teachers time (hours)**	0.01	0.59	0.01
	Catering team time (hours)			
	Managers and chefs time	0.42	31.34	0.69
	Managers and chefs travel costs (transport)			
	Return train ticket	0.42	1.69	0.04
Total mean intervention implementation cost per school			630.74 Yuan (177.67 $, 124.89 £)	
Total mean intervention implementation cost per class			148.41 Yuan (41.80 $, 29.39 £)	
Total mean intervention implementation cost per child assuming a class of 45			3.29 Yuan (0.92 $, 0.65 £)	
Total mean intervention implementation cost per consented child (averagely 41 per class)			3.62 Yuan (1.02 $, 0.71 £)	

Notes: Total number of intervention schools **(n = 20)**, Total number of intervention classes **(n = 85),** Total number of intervention consented classes **(20),** Assumed average class size **(45),** Total number of intervention consented children **(n = 832)**

Table 5.16 - Mean resource use and costs associated with delivery of intervention

Intervention component	Resource type	Mean annual resource use per class	Mean annual cost per class, Yuan	Average cost per child (assuming average class size of 45), Yuan
	Labour: CHIRPY DRAGON teachers time (hours)			
	Interactive educational activities for main carers *	4.85	242.5	5.38
	Interactive educational activities for children	2.32	115.88	2.58
	Quiz for main cares and children *	0.9	45	1
	Family-wide healthy behavioural challenges	2.56	128.38	2.85
	Office stationery			
	Ink pads	0.01	0.06	0.001
	Permanent markers	0.04	0.04	0.0008
	Printing			
	Colourful educational leaflets for parents and families *	97.5	56.55	1.25
	Family healthy behaviour challenges fun cards	238.12	35.73	0.79
	Stickers (with CHIRPY DRAGON logo)	32.64	4.41	0.10
Component 1: Improving childhood obesity related knowledge and behaviour among children and their main carers	Record cards of individual performance	0.98	6.37	0.14
	Illustrative photo cards for child workshops (first semester)	23.88	9.64	0.21
	Illustrative photo cards for child workshops (second semester)	14	5.6	0.12
	Illustrative photo cards for child carers workshops *	365	129	2.87
	Teaching boards for child carers workshops (Shahe School) *	2	0.3	0.007
	Family healthy behaviour challenges reward board	26.47	7.50	0.17
	Delivery fee			
	Family healthy behaviour challenges reward board	0.017	0.21	0.005
	Loudspeakers *	0.08	0.96	0.021
	Workshops materials			
	PowerPoint remote control	0.08	2.64	0.06
	Canister	0.35	0.35	0.008
	Measuring spoon	0.04	0.21	0.005
	AA batteries	0.29	0.91	0.02
	Stamp	0.03	0.79	0.017

	Transparent plastic bag	2.35	0.23	0.005
	Suitcase	0.02	1.86	0.04
	Cart	0.03	1.04	0.02
	Paper plates for child workshops	17.64	3.47	0.08
	Electronic scale	0.04	0.81	0.02
	Canister (larger)	0.07	0.17	0.004
	Loudspeakers *	0.2	32	0.71
	Stamp (with CHIRPY DRAGON logo)	0.09	1.99	0.04
	Pencils	0.04	0.26	0.006
	Megaphone	0.06	2.26	0.05
	Food presenting in child carers workshops	0.56	9.35	0.20
	Incentives: Family healthy behavioural challenges			
	Incentive prize	8.24	95.32	2.12
	Telephone			
	CHIRPY DRAGON teacher telephone	0.03	7.93	0.18
	Transport			
	CHIRPY DRAGON teacher transport	0.06	46.70	1.04
Component 1 total mean intervention running/delivery cost per school			2549.52 Yuan (718.17 $, 504.85 £)	
Component 1 total mean intervention running/delivery cost per class			996.42 Yuan (280.68 $, 197.33 £)	
Component 1 total mean intervention running/delivery cost per child assuming a class of 45			22.12 Yuan (6.23 $, 4.38 £)	
Component 1 total mean intervention running/delivery cost per consented child (on average 41 per class)			24.30 Yuan (6.84 $, 4.81 £)	
Component 2: Improving the nutritional quality of school lunch provision	**Labour: CHIRPY DRAGON teachers time (hours)**			
	Supportive regular evaluations and feedbacks to the catering teams	0.01	0.59	0.01
	Office stationery			
	Ink pads	0.01	0.06	0.001
	Permanent markers	0.04	0.04	0.0008
	Incentives: Catering teams			
	Recognition certificate	0.51	5.75	0.13
	Telephone			
	CHIRPY DRAGON teacher telephone	0.03	7.93	0.18

Component 2 total mean intervention running/delivery cost per school			61.1 Yuan (17.21 $, 12.09 £)	
Component 2 total mean intervention running/delivery cost per class			14.37 Yuan (4.05 $, 2.85 £)	
Component 2 total mean intervention running/delivery cost per child assuming a class of 45			0.32 Yuan (0.09 $, 0.06 £)	
Component 2 total mean intervention running/delivery cost per consented child (on average 41 per class)			0.35 Yuan (0.10 $, 0.07 £)	
Component 3: Increasing children's physical activity level outside school	**Labour: CHIRPY DRAGON teachers time (hours)**			
	Physically active family friendly games learnt and practiced at school for children and their parents *	4.52	226.25	5.03
	Family-wide healthy behavioural challenges	0.86	42.8	0.95
	Workshops assistant time (hours) *	3.85	117.4	2.61
	Office stationery			
	Ink pads	0.01	0.06	0.001
	Permanent markers	0.04	0.04	0.0008
	Printing			
	Family healthy behaviour challenges fun cards	79.41	11.91	0.26
	Stickers (with CHIRPY DRAGON logo)	10.88	1.47	0.03
	Record cards of individual performance	0.33	2.12	0.05
	Family healthy behaviour challenges reward board	8.82	2.49	0.06
	Delivery fee			
	Family healthy behaviour challenges reward board	0.006	0.07	0.002
	Loudspeakers *	0.02	0.24	0.005
	Workshops materials			
	Stamp *	0.04	1.12	0.02
	Suitcase *	0.02	1.98	0.04
	Cart *	0.03	1.1	0.02
	Loudspeakers *	0.05	8	0.18
	Stamp (with CHIRPY DRAGON logo) *	0.12	2.81	0.06
	Balloon *	100	10	0.22
	Finger board used in physically active family friendly games *	0.1	3.75	0.08
	Pencils *	0.06	0.37	0.008
	Rubber band *	100	0.75	0.02
	Megaphone *	0.06	2.4	0.053
	Strips of cloth *	5	5	0.11
	Incentives: Family healthy behavioural challenges			
	Incentive prize	3.41	15.71	0.35

	Telephone			
	CHIRPY DRAGON teacher telephone	0.03	7.93	0.18
	Transport			
	CHIRPY DRAGON teacher transport *	0.06	49.61	1.10
Component 3 total mean intervention running/delivery cost per school			790.41 Yuan (222.65 $, 156.51 £)	
Component 3 total mean intervention running/delivery cost per class			515.38 Yuan (145.18 $, 102.05 £)	
Component 3 total mean intervention running/delivery cost per child assuming a class of 45			11.43 Yuan (3.22 $, 2.26 £)	
Component 3 total mean intervention running/delivery cost per consented child (on average 41 per class)			12.57 Yuan (3.54 $, 2.49 £)	
Component 4: Increasing children's physical activity level in school	**Labour: CHIRPY DRAGON teachers time (hours)**			
	Monthly meeting with relevant school staff and student representatives	0.16	8.23	0.18
	Office stationery			
	Ink pads	0.01	0.06	0.001
	Permanent markers	0.04	0.04	0.0008
	Telephone			
	CHIRPY DRAGON teacher telephone	0.03	7.93	0.18
	Transport			
	CHIRPY DRAGON teacher transport	0.08	58.37	1.30
Component 4 total mean intervention running/delivery cost per school			317.23 Yuan (89.36 $, 62.81 £)	
Component 4 total mean intervention running/delivery cost per class			74.63 Yuan (21.02 $, 14.77 £)	
Component 4 total mean intervention running/delivery cost per child assuming a class of 45			1.66 Yuan (0.47 $, 0.33 £)	
Component 4 total mean intervention running/delivery cost per consented child (on average 41 per class)			1.82 Yuan (0.51 $, 0.36 £)	
Total mean intervention running/delivery cost per school			3718.26 Yuan (1047.39 $, 736.28 £)	
Total mean intervention running/delivery cost per class			1600.8 Yuan (449.73 $, 317 £)	
Total mean intervention running/delivery cost per child assuming a class of 45			35.53 Yuan (10.01 $, 7.04 £)	
Total mean intervention running/delivery cost per consented child (averagely 41 per class)			39.04 Yuan (10.97 $, 7.73 £)	

Notes: Total number of intervention schools **(n = 20)**, Total number of intervention classes **(n = 85),** Total number of intervention consented classes **(20),** Assumed average class size **(45),** Total number of intervention consented children **(n = 832).** Mean cost per class: the total cost for delivery across the whole intervention arm for the specific resource type divided by the number of classes (either 85 or 20)**.** Some parts of the intervention, which involved family members, were only delivered to intervention consented classes (1 class per school). These costs were collected at class level and were related to interactive educational activities for main carers, quiz for main carers and children; and physically active family friendly games learnt and practiced at school for children and their parents. These resource type costs are indicated with ***** in the table. The rest of the intervention was delivered to all year one children (85 classes in 20 schools).

Incremental Analysis of Cost (Yuan, US Dollar, Pound)

Given the assumed 'no costs' associated with the control, the intervention was statistically significantly more expensive than the control: 35.53 Yuan (£7.04/US$10.01) per child who received the intervention (assuming an average class size of 45).

5.3.7 Cost-Utility Analysis

The ICER associated with the base case was £1,760 (US$2,502) per QALY gained, which is far below the £20,000 and $50,000 per QALY thresholds for cost-effectiveness in the UK and US respectively [30, 279]. In addition, even using a 1xGDP per capita threshold (US$19,000), recommended by WHO, the ICER is far below the threshold for cost-effectiveness [165, 166]. Figure 5.1 shows the net-benefits associated with the intervention at different levels of WTP.

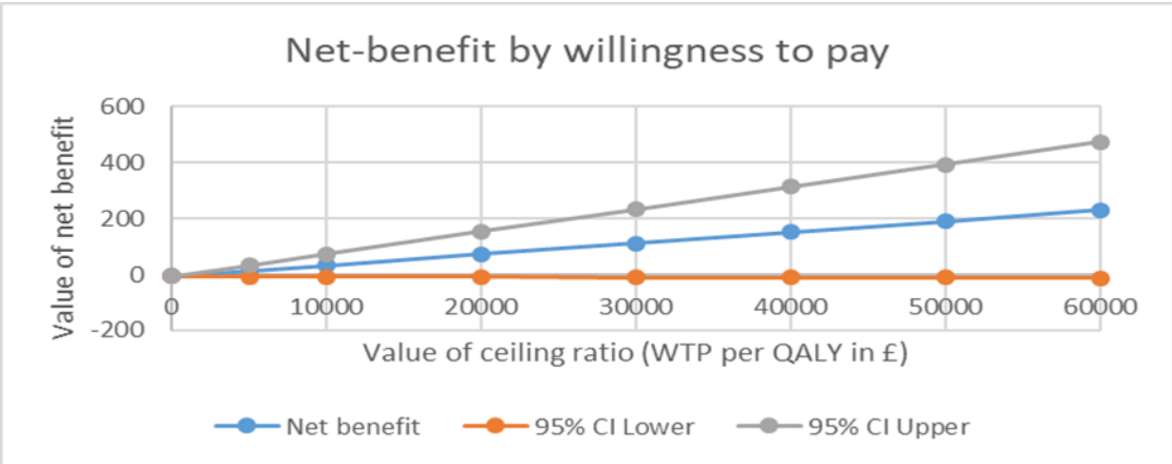

Figure 5.1 - Net-benefit of intervention at different WTP levels (base case analysis)

The decision uncertainty is presented using Cost Effectiveness Acceptability Curve for the further adjusted model. Figure 5.2 shows the probability of the intervention being cost-effective at different levels of WTP. At the UK NICE recommended threshold, there is a 96% probability of the intervention being cost-effective. Even at a WTP of nearly £5,000 per QALY, the probability of cost-effectiveness remains high at 90%.

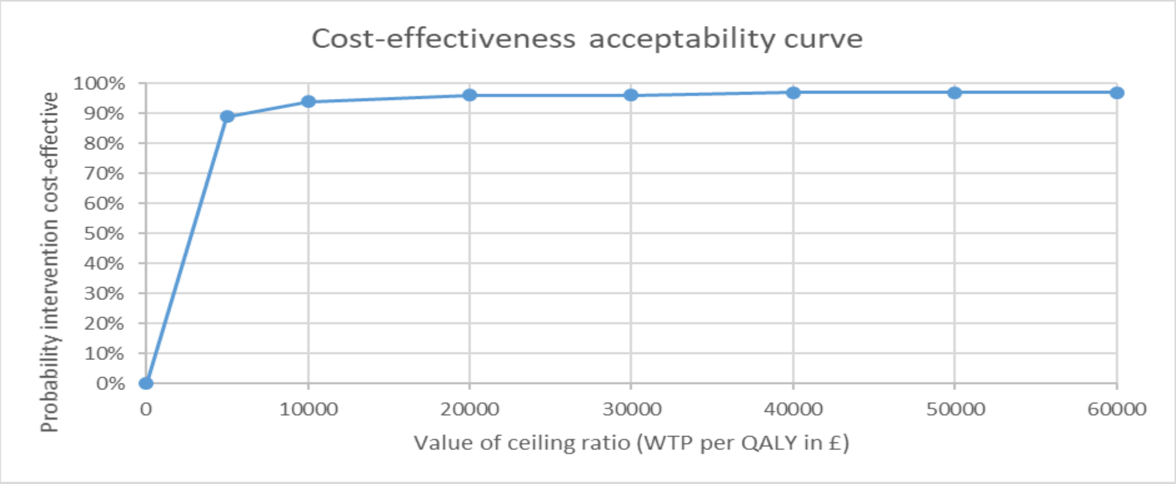

Figure 5.2 - CEAC (base case analysis)

5.3.8 Sensitivity Analyses Regarding Cost-Utility Analysis

Sensitivity Analysis 1: Implementation Costs Included

The inclusion of the implementation costs resulted in a slight increase in the ICER associated with the intervention and produced almost the same CEAC (Figure 5.3). The addition of the implementation costs increased average costs by 3.29 Yuan (£0.65/$0.92), increasing costs to 38.82 Yuan (£7.69/$10.93) per child. This increase in cost had little impact on the overall recommendation from the economic evaluation, increasing the ICER to just £1,922/$2,732 per QALY. Even at a WTP of £10,000 per QALY, the probability of the intervention being deemed cost-effective is 94%.

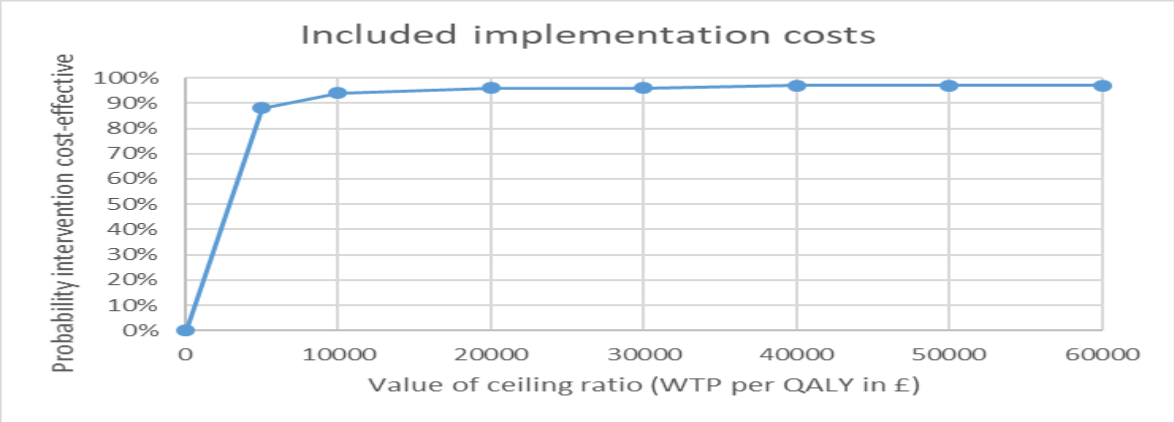

Figure 5.3 - CEAC (sensitivity analysis 1: implementation costs included)

Sensitivity Analysis 2: Chinese Tariff Applied for Utility Values

The second sensitivity analysis tested the sensitivity of the results to using the Chinese tariff to estimate QALYs. The average QALY increased to 0.006 per child. This rise had a small impact on the results, decreasing the ICER to just £1,173/$1,668 per QALY. The CEAC (Figure 5.3) shows the impact of the slightly higher levels of QALY moving the CEAC slightly upwards, particularly at the lower levels of WTP, in comparison with the base-case analysis. However, even at a WTP of £10,000 per QALY, the probability of the intervention being deemed cost-effective is 96%.

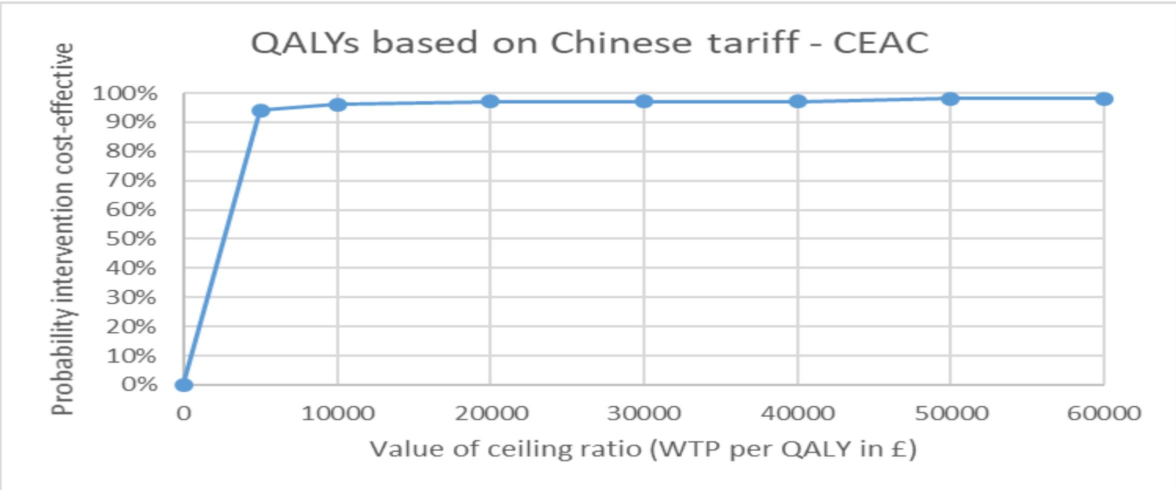

Figure 5.4 - CEAC (sensitivity analysis 2: Chinese tariff applied for utility values)

Sensitivity Analysis 3: Class Size only Included Consented Children

In this scenario, on average there were 41 children with parental consent to undertake measurements within each class. As a result, the cost increased to 39.04 Yuan (£7.73/$10.97) per child. This produced almost the same CEAC (Figure 5.5) with the ICER, rising to just £1,932/$2,742 per QALY. Again even at a WTP of £10,000 per QALY, the probability of the intervention being deemed cost-effective is 94%.

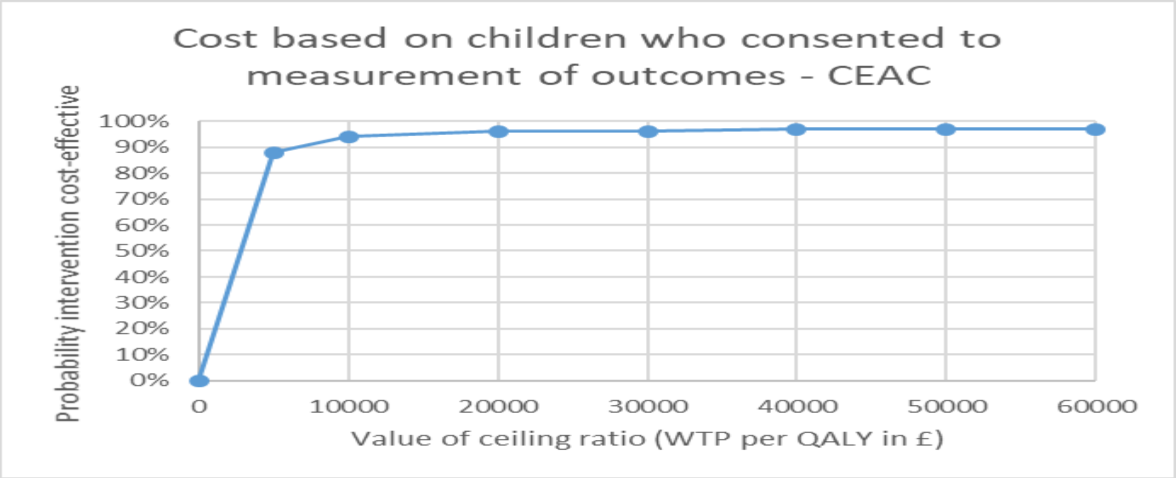

Figure 5.5 - CEAC (sensitivity analysis 3: class size only included consented children)

5.3.9 Cost-Effectiveness Analysis

The significant difference between arms made it possible to assess the cost per BMI z-score change. The ICER was £54 (US$77) per BMI z-score change.

5.4 Part 3: Results: Economic Evaluation (Societal Perspective)

5.4.1 Impact of Intervention on Parents/Household Members' HRQoL (EQ-5D-3L)

Table 5.17 outlines the response rate for the EQ-5D-3L instrument. Missing/invalid data for this outcome was almost 25% at both measurement points. As explained in part 2, section 5.2.4.2, multiple imputation for the base-case analysis was not required.

Table - 5.17 Number of consented parents/household members with completed EQ-5D-3L

Time point	EQ-5D-3L completed all participants (n (%))	EQ-5D-3L completed control group (n (%))	EQ-5D-3L completed intervention group (n (%))
Baseline	1235 (75.3%)	596 (73.7%)	639 (76.9%)
12 months follow up	1226 (74.7%)	584 (72.2%)	642 (77.2%)

Table 5.18 presents the mean utility scores at each time point (using the UK tariff) for the base-case analysis. There was no baseline imbalance for utility scores, but nevertheless all incremental analyses were adjusted for baseline. At 12 months follow up, the mean utility scores for the intervention group was slightly higher than the control group. The mean utility scores at each time point (using the Chinese tariff for sensitivity analysis) were consistently higher compared to using the UK tariff (Appendix 2.8).

Table 5.18 - EQ-5D-3L utility scores at each time point (UK tariff)

	EQ-5D-3L utility scores (parents/grandparents)								
	All participants			**Control group**			**Intervention group**		
Time	**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**	
Baseline (N= 1235)	0.961 (.083)	0.957	0.966	0.961 (.085)	0.954	0.968	0.962 (.081)	0.956	0.968
12 months follow up (N = 1226)	0.968 (.078)	0.963	0.972	0.965 (.075)	0.959	0.972	0.969 (.081)	0.963	0.976

Examining the Impact of the Intervention on QALYs

Table 5.19 describes the unadjusted mean QALYs (using the UK tariff). At 12 months follow up, the intervention group accrued 0.966 QALYs compared to 0.965 QALYs for the control group. The unadjusted mean QALYs (using the Chinese tariff) were consistently higher compared to using the UK tariff (Appendix 2.9).

Table 5.19 - Unadjusted QALYs accrued (EQ-5D-3L, UK tariff)

	Unadjusted QALYs accrued								
	All participants			**Control group**			**Intervention group**		
Time	**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**	
12 months follow up (N= 1224)	0.966 (.064)	0.962	0.969	0.965 (.061)	0.960	0.970	0.966 (.066)	0.961	0.971

Incremental Analysis – Effectiveness

Table 5.20 describes the incremental difference in mean QALYs (using the UK tariff) between the intervention and control group for the data with no adjustment; adjustment for clustering and baseline differences; and adjustments for clustering, baseline differences and the pre-specified covariates using mixed effect linear regression models. The mean QALY difference, using the three models, was in favour of the intervention for 12 months follow up but did not reach statistical significance. The QALYs attained were very similar compared to using the UK tariff (only marginally lower using the Chinese tariff with further adjusted model). These results were also not statistically significant using the three models (Appendix 2.10). After conducting multiple imputation, the results remain similar to those pre-imputation (Appendix 2.11).

Table 5.20 - Incremental difference in QALYs (EQ-5D-3L, UK tariff)

	No adjustment				Adjusted for clustering and baseline utility				Adjusted for clustering, baseline utility, co-variates			
Time	**Mean difference**	**(95% CI)**		**P-value**	**Mean difference**	**(95% CI)**		**P-value**	**Mean difference**	**(95% CI)**		**P-value**
12 months follow up	0.001	-0.006	0.008	0.784	0.002	-0.002	0.006	0.329	0.002	-0.002	0.007	0.421

5.4.2 Resource Use and Costs

A breakdown of the mean cost of lunch per diner per day and per year in each of the intervention and control schools is presented in Table 5.21. As the cost of lunch was paid for by the household, these costs assume a societal perspective. The total mean annual cost of lunch per diner was higher in the intervention schools 1765 Yuan (£349.50/$497.18) compared to the control schools 1637.5 Yuan (£324.25/$461.26). A comparison of the mean cost of lunch per diner per day in Yuan in the intervention versus the control schools is presented in Figure 5.6. As can be seen, interestingly, the variation in the mean cost is larger in the control schools compared to the intervention schools.

Table 5.21 - Mean (SD) cost of lunch per diner per day and per year in each school

	School	Mean (SD) cost of lunch per school per day over the 12 month trial period in Yuan	Mean (SD) number of diners per school per day over the 12 month trial period	Mean (SD) cost of lunch per diner per day in Yuan *	Average cost of lunch per diner per year in Yuan **	Average cost of lunch per diner per year in US£	Average cost of lunch per diner per year in £
Intervention group	1	3740.49 (218.47)	596 (0)	6.28 (0.36)	1570	442.25	310.89
	2	3419.88 (330.16)	465.95 (37.91)	7.33 (0.26)	1832.5	516.19	362.87
	3	4149.05 (142.41)	643.76 (2.92)	6.44 (0.21)	1610	453.52	318.81
	4	3667.24 (2641.25)	362.78 (202.49)	8.24 (3.38)	2060	580.28	407.92
	5	4807.07 (144.92)	892.92 (10.58)	5.38 (0.16)	1345	378.87	266.34
	6	1494.82 (97.28)	265.93 (5.58)	5.62 (0.33)	1405	395.77	278.21
	7	4149.42 (178.72)	756.75 (13.45)	5.48 (0.21)	1370	385.91	271.28
	8	7077.36 (143.39)	1040.79 (20.54)	6.8 (0.03)	1700	478.87	336.63
	9	4620.75 (1387.71)	568.78 (163.84)	8.09 (0.39)	2022.5	569.72	400.49
	10	8553.28 (44.18)	1006.26 (5.19)	8.5 (0)	2125	598.59	420.79
	11	7834.54 (5640.71)	784.94 (55.14)	9.38 (0.44)	2345	660.56	464.35
	12	4622.03 (847.66)	477.97 (23.38)	9.7 (1.8)	2425	683.09	480.19
	13	1996.13 (1446.23)	275.73 (260.63)	8.22 (1.27)	2055	578.87	406.93
	14	4296.01 (557.45)	641.96 (73.29)	6.69 (0.37)	1672.5	471.12	331.18
	15	828.07 (28.16)	148.01 (3.93)	5.59 (0.07)	1397.5	393.66	276.73
	16	4691.11 (546.03)	873.51 (71.59)	5.38 (0.5)	1345	378.87	266.34
	17	4625.72 (613.50)	597.59 (7.22)	7.88 (0.23)	1970	554.92	390.09
	18	4105.75 (277.65)	668.98 (42.80)	6.13 (0.18)	1532.5	431.69	303.46
Total Mean (SD)		**4469.12 (2433.28)**	**631.24 (257.88)**	**7.06 (1.66)**	**1765**	**497.18**	**349.50**

	School	Mean (SD) cost of lunch per school per day over the 12 month trial period in Yuan	Mean (SD) number of diners per school per day over the 12 month trial period	Mean (SD) cost of lunch per diner per day in Yuan *	Average cost of lunch per diner per year in Yuan **	Average cost of lunch per diner per year in US£	Average cost of lunch per diner per year in £
Control group	1	2763.93 (0.17)	500 (0)	5.53 (0)	1382.5	389.44	273.76
	2	757.42 (179.26)	110.159 (18.18)	6.94 (1.56)	1735	488.73	343.56
	3	331.66 (0.23)	60 (0)	5.52 (0)	1380	388.73	273.27
	4	2159.56 (25.65)	398.2 (3.88)	5.42 (0.02)	1355	381.69	268.32
	5	9355.67 (1147.67)	1325.12 (34.47)	7.05 (0.79)	1762.5	496.48	349.01
	6	7579.64 (426.62)	1363.23 (48.39)	5.56 (0.38)	1390	391.54	275.25
	7	3815.64 (680.06)	543.85 (25.84)	7.02 (1)	1755	494.36	347.52
	8	7625.85 (649.41)	1013.52 (35.94)	7.51 (0.47)	1877.5	528.87	371.78
	9	1957.52 (96.7)	351.46 (7.54)	5.57 (0.31)	1392.5	392.25	275.74
	10	2006.05 (592.53)	550 (0)	3.64 (1.07)	910	256.34	180.19
	11	3599.47 (376.53)	599.58 (19.83)	6.01 (1.38)	1502.5	423.24	297.52
	12	3531 (17.39)	379.10 (10.51)	9.32 (0.25)	2330	656.34	461.38
	13	5659.1 (234.21)	908.57 (10.48)	5.77 (0.29)	1442.5	406.34	285.64
	14	5580.32 (1028.29)	606.37 (14.68)	9.18 (1.56)	2295	646.47	454.45
	15	1107.29 (125.06)	138.52 (15.53)	7.99 (0.1)	1997.5	562.67	395.54
	16	2525.53 (340.44)	331.98 (39.48)	7.63 (0.96)	1907.5	537.32	377.72
	17	5765.19 (841.25)	1027.33 (5.77)	5.61 (0.82)	1402.5	395.07	277.72
	18	3238.51 (749.09)	384.49 (89.85)	8.43 (0.45)	2107.5	593.66	417.32
	19	1309.62 (222.08)	272.28 (9.71)	4.82 (0.87)	1205	339.43	238.61
	20	2271.04 (0.09)	395 (0)	5.74 (0)	1435	404.22	284.16
Total Mean (SD)		**3944.63 (2675.93)**	**607.43 (399.02)**	**6.55 (1.57)**	**1637.5**	**461.26**	**324.25**

Notes: * Mean cost per day was measured by taking the average of the total daily cost of lunch divided by the number of diners on each given day. ******Average cost of lunch per diner per day was multiplied by the number of days (250) when schools were open during the 12 month trial period to obtain estimated annual lunch cost per child.

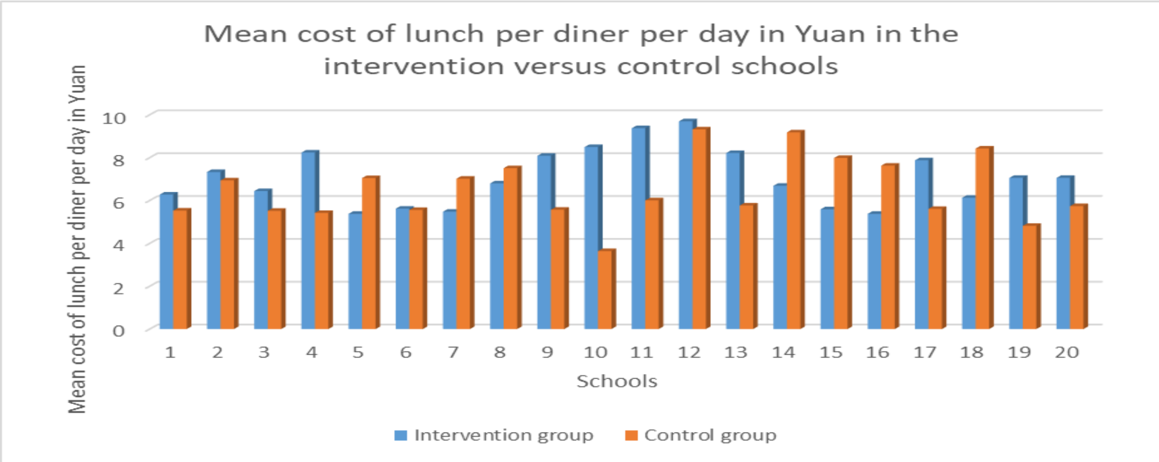

Figure 5.6 - Mean cost of lunch per diner per day in intervention versus control schools

A breakdown of the cost of the parents/main carers' workshop attendance time is presented in Table 5.22. The data revealed that overall, across all the workshops, 61% of the parents/grandparents would have otherwise been at paid work, therefore 39% of them were using unpaid time. As on average, two parents or main carers attended the workshops, the total mean cost of family members' time was therefore 373.92 Yuan (£74.4 /$105.32) per child. With regard to this cost, the largest proportion was on interactive educational workshops for main carers which accounted for half of the family members' time costs.

Incremental Analysis of Cost (Yuan, US Dollar, Pound)

The intervention was statistically significantly more expensive than the control: 536.95 Yuan (£106.33/US$151.25) per child/family who received the intervention (assuming an average class size of 45).

Table 5.22 - Parents/main carers' workshop attendance time cost

	Workshop	Average cost of workshop time for each family member in employment [A]	Average cost of workshop time for each family member not in employment [B]	Number of paid family members attending the workshop [C]	Number of unpaid family members attending the workshop [D]	Average cost of each family member time [E]
Component 1 (main carers)	Education	65 * 2	18.3 * 2	61% * 683	39% * 683	93.48 Yuan (18.6
	Family quiz	65 * 1	18.3 * 1	61% * 491	39% * 491	46.74 Yuan (9.3 £)
Component 3 (parents)	Family game	65 * 1	18.3 * 1	61% * 680	39% * 680	46.74 Yuan (9.3 £)
Total						186.96 Yuan (37.2 £/52.66 $)

Notes:

[E] **=** Average cost of each family member time = (A*C) + (B*D) / total attendees based on the process evaluation data
[A] = Average cost of workshop time for each family member in employment = average hourly wage rate * workshop hours over trial duration
[B] = Average cost of workshop time for each family member not in employment = average hourly leisure rate * workshop hours over trial duration
[C] = Number of paid family members attending the workshop = 61% * total attendees based on the process evaluation data
[D] = Number of unpaid family members attending the workshop = 39% * total attendees based on the process evaluation data

5.4.3 Cost-Utility Analysis

To account for family member effects, the multiplier for each child was calculated as:

[1 + (2 * (0.002) / 0.004)]

The base-case ICER was therefore calculated using the following formula:

$$\frac{£7.04}{0.004} * \frac{1.1}{2} * \frac{£106.33}{£7.04}$$

In the base-case model, the impact from including household member QALYs and household costs increased the ICER from £1,760 (US$2,502) to £14,620 (US$20,796) per QALY gained, which is still well below the established UK and US thresholds for cost-effectiveness [30, 279]. Using a 1xGDP per capita threshold (US$19,000), recommended by WHO, the ICER is slightly above the threshold for cost-effectiveness. However, when a 3xGDP per capita threshold (US$ 57,000) is used the ICER is well below the threshold for cost-effectiveness [165, 166].

5.4.4 Sensitivity Analyses

The four sensitivity analyses had the following impact on the ICER: implementation costs included (ICER increased to £14,709/$20,923 per QALY), Chinese tariff applied for utility values (ICER remained the same as base-case £14,620/$20,796 per QALY), using consented children only (ICER increased to £16,047/$22,823 per QALY) and using predictive mean matching multiple imputation (ICER increased to £16,709/$23,767 per QALY).

5.4.5 Exploratory Analysis of Household Expenditure

Appendices 2.12 i – iii outline the response rate for the household expenditure and income data. Missing/invalid data for these were less than 16% at both measurement points. There were no differences between the two study groups in completeness of the data.

The mean household expenditure at each time point is presented in Appendices 2.13 i – ii and Table 5.23 presents the mean income data at each time point. There was no baseline imbalance for these data. At 12 months follow up, there were no noticeable differences in mean expenditure between the intervention and control groups over a range of different expenditure categories.

Table 5.23 - Monthly mean (SD) of total income of families in Yuan

	Baseline			**12 months follow up**		
	All Participants	**Control group**	**Intervention group**	**All participants**	**Control Group**	**Intervention group**
Total income of families	19048.24 (17180.51)	18752.87 (16428.8)	19323.54 (17858.8)	19411.04 (15725.78)	19589.94 (15758.07)	19241.63 (15703.99)

Table 5.24 and Figure 5.7 present the proportion (%) of household monthly income spent on different items. There was no baseline imbalance for these data, nevertheless all incremental analyses were adjusted for baseline. As can be seen, in general, families spent the greatest proportion of their income on food and non-alcoholic drinks, and the least on alcoholic drinks, tobacco and narcotics.

Table 5.25 describes the incremental difference in the proportion of monthly income spent on the different categories between the intervention and control groups, using three models: i) no adjustment; ii) adjustment for baseline differences; and iii) adjustments for baseline differences and the pre-specified covariates using mixed linear regression models. The differences were small and none of them were statistically significant.

Table 5.24 - Proportion of income spent monthly on the different household expenditure

Categories of family expenditure	Baseline		12 months follow up	
	Control	**Intervention**	**Control**	**Intervention**
Electricity/gas	4.6%	4.5%	4.8%	4.5%
Transport	7%	6.9%	6.4%	6.7%
Recreation	3%	2.8%	2.8%	2.9%
Food/non-alcoholic drinks	13.7%	15.2%	13.7%	14.5%
Alcoholic drinks, tobacco/narcotics	1.9%	1.9%	1.9%	1.8%
Eating out	5.6%	5.9%	5.5%	5.6%
Clothing	5.8%	6.1%	6.1%	5.9%
Communications	3%	3%	2.9%	2.8%
Household goods/services	3.7%	4.3%	4.6%	4.7%
Other goods/services	1.9%	2.3%	2.2%	2.1%
Education	7.9%	7.3%	10.2%	9.8%
Healthcare	3.1%	3%	3.1%	2.9%
Total	61.2%	63.2%	64.2%	64.2%

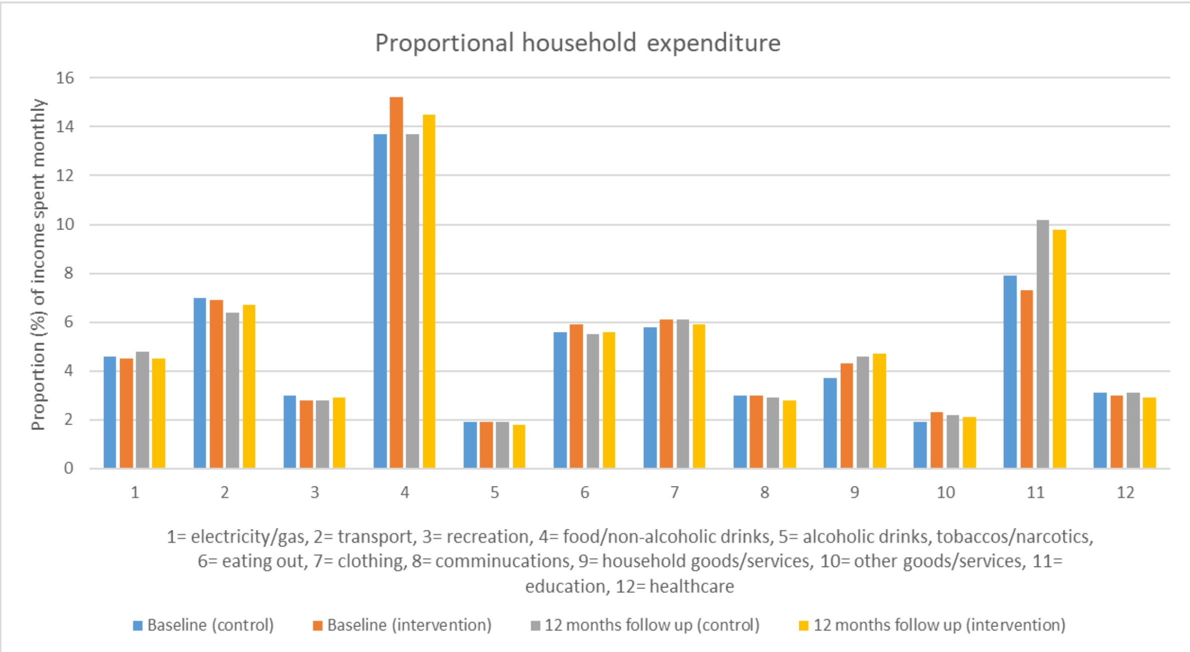

Figure 5.7 - Proportional household expenditure

Table 5.25 - Incremental difference in different categories of family monthly expenditure

Categories of monthly family expenditure	No adjustment				Adjusted for baseline				Adjusted for baseline and co-variates			
	% difference	(95% CI)		P-value	% difference	(95% CI)		P-value	% difference	(95% CI)		P-value
Electricity/gas	-0.3%	-1.3%	0.7%	0.577	-0.1%	-0.9%	0.8%	0.900	-0.3%	-1.1%	0.5%	0.435
Transport	0.3%	-0.4%	1.0%	0.457	0.3%	-0.5%	1.0%	0.507	0.3%	-0.5%	1.0%	0.482
Recreation	0.1%	-0.4%	0.5%	0.806	0.2%	-0.1%	0.6%	0.133	0.2%	-0.1%	0.6%	0.179
Food/non-alcoholic drinks	0.7%	-0.7%	2.2%	0.318	0.6%	-0.8%	2%	0.395	0.5%	-0.9%	1.9%	0.478
Alcoholic drinks, tobacco/narcotics	-0.1%	-0.5%	0.2%	0.523	-0.2%	-0.6%	0.2%	0.429	-0.3%	-0.7%	0.1%	0.396
Eating out	0.1%	-0.5%	0.6%	0.863	0.1%	-0.5%	0.5%	0.970	0.1%	-0.4%	0.5%	0.933
Clothing	-0.2%	-0.1%	0.5%	0.526	-0.1%	-0.5%	0.8%	0.683	-0.1%	-0.6%	0.8%	0.809
Communications	-0.1%	-0.3%	0.2%	0.712	-0.1%	-0.2%	0.4%	0.560	-0.1%	-0.3%	0.3%	0.944
Household goods/services	0.1%	-0.5%	0.4%	0.981	0.3%	-0.7%	0.5%	0.373	0.3%	-0.8%	0.6%	0.427
Other goods/services	-0.1%	-0.3%	0.2%	0.441	-0.2%	-0.4%	0.3%	0.176	-0.2%	-0.5%	0.2%	0.182
Education	-0.3%	-0.8%	0.7%	0.680	-0.2%	-0.6%	0.3%	0.772	-0.3%	-0.7%	0.5%	0.822
Healthcare	-0.2%	-0.7%	0.2%	0.398	-0.2%	-0.6%	0.2%	0.415	-0.2%	-0.6%	0.2%	0.364

5.5 Part 4: Discussion

5.5.1 Statement of Principal Findings

The economic evaluation from a public sector perspective showed that the CHIRPY DRAGON intervention had a relatively low cost and significant intervention benefits over the course of 12 months, suggesting it is highly cost-effective. The intervention is cost-effective using the conventional decision making rules within a CEA and CUA. Broadening the evaluative space to include household costs and QALYs had the effect of increasing the ICER however the intervention remained cost-effective. Both these results were robust in sensitivity analyses.

The impact of the intervention on proportional spend on different household categories was minimal and none of the differences were statistically significant.

5.5.2 Methodological Challenges of Conducting Economic Evaluation

A number of methodological challenges were encountered from undertaking an economic evaluation within a Chinese setting. All costs had to be converted into either UK pounds or US dollars using GDP PPPs. All unit costs had to be sourced that were relevant to a Chinese context. Using the appropriate value-set for either the CHU-9D or EQ-5D-3L as, although there is a Chinese version, the recommendation was to use the UK set since the Chinese value set was still under development. Furthermore, as economic evaluation is uncommon in China and due to the lack of an equivalent threshold value for a Chinese setting, the established UK and US threshold values for a QALY were used to judge cost-effectiveness. It is unclear whether these values reflect society's valuation of a QALY gain within a Chinese setting.

This study also highlighted some methodological challenges related to the inclusion of spillover effects. First, the spillover data was not complete due to non-response. Second, as household members' resource use and outcome data are very rarely considered in economic evaluations, guidelines for including these data were limited therefore an exploratory analysis was only possible.

5.5.3 Strengths and Limitations of this Study

Strengths include the large sample size (1641 children), standardised data collection procedures as part of the RCT, training for the CHIRPY DRAGON teachers, very good follow up rates, the low level of attrition throughout the follow up period, and the use of a pre-specified analysis plan which took account of school clustering. The intervention programme was well delivered and received. A very low level of missing data regarding resource use and outcomes, using a public sector perspective, proved to be a significant strength. This was largely because purposely employed staff collected the data, which was collected alongside the trial as recommended by best practice [22, 36, 275]. Where possible, the results of this economic evaluation was reported using CHEERS guidelines [48]. Detailed analysis of the resource items was provided. Furthermore, this study is one of the very few economic evaluations of obesity prevention studies worldwide and the first in China, which collected utility-based HRQoL information in children as young as 6 years and family members to calculate QALYs, and included societal costs. It used both the UK and Chinese tariffs for calculating the utility scores. This study reported the ICER from two alternative perspectives and included both clinical and economic outcomes. This enabled comparison with other studies. Moreover, this is the first CUA study worldwide to consider health spillover effects generated from a behavioural obesity intervention using a multiplier approach. As a result it may provide policy makers with additional useful information when making policy decisions. The intuition behind the multiplier approach is that there is a bigger health dividend for the population than is represented just by considering children's QALYs and therefore this wider health dividend should be captured within an economic evaluation [276, 277].

The study also had some limitations. One potential limitation relates to the way HRQoL information was collected from children. There may have been an influence on how children completed the questionnaire as items and possible responses within the CHU-9D were read to children, on a one-to-one basis, by the interviewers (research staff). This could have led to responder-bias [42]. However, given the young age of the participants, this collection strategy was chosen to optimise data quality and completion. Furthermore, interviewers were blind to allocation, minimising any differential bias. A further limitation was that the number of responses from household members was smaller than the number of children. This may result in a lack of statistically significant results regarding household members' outcomes. Furthermore, both children's and household members' incremental QALYs were estimated separately before aggregating the mean estimates. In future studies, where the number of children and household members are more similar, we would recommend using a dyadic approach. The advantage of dyadic analysis, compared to the multiplier approach, for including health spillovers is that it enables a probabilistic sensitivity analysis to be conducted to explore cost-effectiveness uncertainty. Additionally, there was a lack of information on the salaries of parents/grandparents. Instead, the Chinese population average salary was applied to estimate the value of lost time due to participation in the trial. Furthermore, the analysis was constrained by the time horizon of the intervention. Whilst the intervention appears to be cost-effective and able to obtain benefits for both clinical and economic outcomes in children for a relatively low cost, the sustainability of these effects remains unknown. Moreover, there is an ongoing methodological debate concerning the use of the CHU-9D to capture HRQoL in young children and whether this instrument is responsive to change in weight status [280]. Finally, as this is a behavioural intervention, therefore highly dependent on cultural, infrastructural and other system-related aspects, the generalisability of results to other contexts, particularly to developed country settings, could be questionable [247].

5.5.4 Comparison with Other Studies

Based on the most recent systematic review of economic evaluations reported in this thesis [47], there is no consistent measure of outcomes across published evaluations. Most reports of trial-based economic evaluations of school-based child obesity interventions used clinical outcome improvements (e.g. BMI or waist circumference) [17, 115, 206, 227, 235, 237] and studies based on QALYs gained were limited [249]. This heterogeneity of outcome measures hinders comparability of cost-effectiveness. Also, in terms of judging cost-effectiveness of interventions, context-specific assessment is problematic as there are different thresholds for cost effectiveness in different countries. Apart from one study [249], the rest of these school-based obesity interventions appeared cost-effective using a ‘cost per weight-specific outcome’. However, without thresholds for obesity-related outcomes, it is difficult to judge value for money. Currently, for example, there is no national or international threshold on WTP for the prevention of a BMI gain in childhood.

The cost-effectiveness analysis results showed that the ICER from a public sector perspective was £54 (US$77) per BMI z-score change. This was lower than two previous trial-based intervention studies which used BMI z-score as their measure of effectiveness: one Chinese study, targeting dietary habits and physical activity in children 6-13 years, (US$ 249.3 per BMI z-score change) [17]; the other Australian, targeting physical activity in adolescents 13-16 years, (AU$ 563 per 10 % reduction in BMI z-score) [206]. Neither of these studies included indirect costs. Contrary to this study, a similar study in the UK, targeting dietary habits and physical activity in children 6-7 years, was not cost-effective using BMI z-score [249].

Contextual factors including differences in the stage of the childhood obesity epidemic and cultural factors, as well as intervention differences (e.g. target, components and how these were delivered) may have contributed to the observed differences in findings. It has been determined that obesity prevention interventions are more effective when delivered by dedicated staff rather than classroom teachers [281]. The staff employed to deliver the intervention in this trial were well accepted by schools and their costs were incorporated in the economic evaluation.

In terms of understanding the impact of the intervention on multiple outcomes, a more disaggregated analysis (e.g. a CCA) alongside a CEA and CUA could have been undertaken. This would give a 'list' rather than offset against the costs. Thus decision makers would understand how the benefits are distributed across the different sectors (health and education) and this could then act as a tool to facilitate cross-sectoral decision making. Using a CCA could also give decision makers confidence that the data are valid to use as the basis for resource allocation decisions.

5.5.5 Implications for Practice, Policy and Research

The research in this thesis demonstrates the feasibility of collecting and including household members' cost and outcome data in cost utility analysis. However, as household members' resource use and outcome data are very rarely considered in economic evaluations, guidelines for how to include these data are limited. For preventive public health interventions (e.g. obesity prevention) the use of a societal perspective in economic evaluations is more complex compared to health perspective, and the vast majority of consequences and therefore costs prevented fall in the future. To capture all possible societal costs and effects, data from various sources (e.g. clinical, epidemiological, and economic) are needed. The development of a toolbox including good-practice guidelines for intervention developers and evaluators would be useful in order to collect relevant data alongside trials. All methods for estimating societal costs should explicitly be documented and reported to provide information that will inform future evaluations and policy making.

The economic evaluation results can inform future research and policy decisions in China. Future studies need to evaluate whether the effects and the cost-effectiveness results are sustainable in the long term, and a suitable model could be used to assess the long-term cost-effectiveness.

5.6 Conclusion

The economic evaluation from a public sector perspective suggested that this intervention, delivered through schools and families with high implementation adherence, was highly cost-effective. This study demonstrated the feasibility of collecting and including household members' cost and outcome data in cost utility analysis. In this case the intervention did not impact significantly on household members' health, but inclusion of household spillovers may make a difference in other contexts. Including societal costs and effects increased the ICER, however the intervention remained cost-effective using established cost-effectiveness thresholds. Although this chapter contributes robust applied evidence, a number of methodological issues remains unknown. The next chapter reports on the following methodological evidence gaps using data from the CHIRPY DRAGON trial:

- The association between weight status and HRQoL
- The construct validity of the CHU-9D when used in a Chinese population

CHAPTER 6. EXPLORING HOW WEIGHT STATUS RELATES TO HRQOL; AND AN ASSESSMENT OF THE CONSTRUCT VALIDITY OF THE CHU-9D IN SCHOOL-AGED CHILDREN: EVIDENCE FROM THE CHINESE TRIAL

6.1 Introduction

Chapter 5 reported that the CHIRPY DRAGON obesity prevention intervention, delivered through schools and families with high implementation adherence, was highly cost-effective. The intervention was deemed cost-effective using the conventional decision making rules within a CEA and CUA.

There is evidence to suggest that HRQoL is affected by culture [282], may differ by gender [283] and is positively associated with socio economic status [42]. As discussed in chapter 2, in some communities obesity is not recognized as a problem and is associated with good health, so HRQoL may be less influenced by obesity in these communities [82, 284, 285]. Obesity trends follow a different pattern in China compared with high-income countries (which are at a more advanced stage of the obesity epidemic) with the risk of obesity being greater in children from higher socio-economic backgrounds and much greater in boys compared to girls [10, 11]. It is important to explore further how this relates to utility-based HRQoL in this population as this directly impacts on QALY measurement for cost-utility analysis.

HRQoL is influenced by culture, gender and socioeconomic status and is, obviously, impaired in ill-health. However, obesity may not be perceived as ill-health and has a different meaning to different cultures, gender and social groups. Therefore it is important to examine how weight status relates to HRQoL. QALYs are used as the unit of assessment to make judgements about the relative cost-effectiveness of competing interventions [36] and require an understanding of the relationship between weight and health related QoL when measured in utility terms. QALYs are also used to inform resource allocation decisions in other country-settings [37]. This is why it is important to analyse how weight status relates to HRQoL in this unique country context.

To date, very few studies examining the relationship between weight status and utility-based HRQoL in children have been conducted and of the few that do exist, they are predominantly from western or high income countries [42, 286]. This chapter directly addresses this evidence gap to support the interpretation of QALY results and provides further information within economic analysis.

The **first aim (a)** of this chapter was to examine how children's weight status relates to their HRQoL in China. The objectives were to:

- Examine the relationship between HRQoL and weight status, gender and socio-economic status.
- Examine the relationship between weight status and HRQoL, adjusting for age, gender and socio-economic status.
- Examine whether any relationship between weight status and HRQoL differs by gender.

As discussed in chapter 3, the CHU-9D is a recently developed paediatric utility measure for application in the economic evaluation of prevention and treatment interventions. Although there is emerging evidence regarding the psychometric properties of the CHU-9D instrument [41, 42, 44], more evidence is required with respect to its validity for use in different age groups and country settings considering different tariffs. This is important because the measure may have different construct validity in different populations.

The **second aim (b)** within this chapter was to assess the construct validity of the CHU-9D instrument in primary school-aged children in a Chinese setting, with the objectives being:

- To determine the discriminant validity of the instrument (discriminant validity refers to the degree with which the instrument discriminates between groups with known differences [287] (e.g. socio-economic status) [41, 42]).
- To determine the convergent validity of the instrument (convergent validity refers to the degree to which two theoretically related measures of construct are correlated [287]).

To facilitate this assessment, the CHU-9D was compared to the PedsQL instrument which is a widely used, validated generic HRQoL measure in children.

As the CHIRPY DRAGON trial and its measurements have been described in full in chapters 2 and 5, only a brief overview of the trial and its measurements are provided below as part of the methods. The remainder of the methods section describes the statistical analyses plan for each of the two aims separately.

6.2 Methods

6.2.1 Trial Design and Participants

The analysis presented uses data from the CHIRPY DRAGON cluster-randomised controlled trial assessing effectiveness and cost-effectiveness of a childhood obesity prevention intervention in Guangzhou, China [126]. Children were aged 6-7 years old at baseline, and initially followed up 12 months later. At baseline, a range of measurements were undertaken, including the PedsQL; CHU-9D; height; weight; gender; age; and socio-economic factors (described in detail in chapter 5). This chapter only uses data collected at baseline. Utility-based HRQoL was measured using the Chinese-translated version of the CHU-9D instrument. General HRQoL was measured using the validated Chinese version of the PedsQL instrument. Mother/father's education level was collected through a parent completed questionnaire and used as a proxy for socio-economic status. Within a sensitivity analysis, the mother/father's employment status was used as an alternative proxy for socio-economic status.

6.2.2 Statistical Analyses

Descriptive Statistics of the Sample Characteristics

Descriptive statistics for the variables of interest which were reported extensively in chapter 5 are summarised in this chapter. All statistical analyses were undertaken in 2019, using Stata version 13.

Factors Associated with HRQoL

The relationship between HRQoL (measured using the CHU-9D combined with the UK and Chinese tariffs, and PedsQL) and weight status category (defined as overweight/obese vs. healthy/underweight or underweight vs. normal weight, overweight, obese); and with gender were examined using descriptive analyses. HRQoL was also assessed in relation to socio-economic status using the mother/father's education level coded as a binary variable ((did or did not attend university) and a categorical variable (school education, occupation college, university undergraduate education, university postgraduate education)). For the mother/father's employment status, used in a sensitivity analysis, this was coded as a binary variable ((did or did not work) and a categorical variable (working full time, working part time, unemployed or looking for work, looking after the family/house, other)).

Differences in HRQoL scores between groups were estimated using either the Kruskal-Wallis test (across all levels of categorical variables), or the non-parametric test for trend (across ordered categories of a variable). Non-parametric tests were used because the HRQoL variables did not follow a normal distribution (based on Kolmogorov–Smirnov test).

A linear mixed regression model (with random effect for school), adjusted for potential confounders (age, gender and mother/father's education) was used to compare the CHU-9D utility values (using the UK and Chinese tariffs) between the two weight status groups (overweight/obese as compared with healthy/underweight).

Relationship between Weight Status and HRQoL by Gender

This analysis was used to assess if any relationship between HRQoL and weight status differed in boys versus girls.

Construct Validity

Discriminant Validity

Three tests were used to explore the discriminant validity: firstly, statistical tests of difference were used to determine if the CHU-9D instrument was able to discriminate between groups with known differences. Studies from both a UK and Australian setting report a lower HRQoL for children from worse socio-economic backgrounds [41, 42, 280], therefore socio-economic status was used for this analysis.

Secondly, the sample was split according to the median PedsQL total score. The mean (SD) and median CHU-9D utility values (using the UK and Chinese tariffs) were compared for children who had a score either on/above, or below, this median PedsQL score, using the t-test.

Thirdly, it was examined how well the mean PedsQL scores corresponded with the options for each of the CHU-9D dimensions, and for this, the mean PedsQL total score was estimated for each level of CHU-9D response on every dimension. This analysis was done with the expectation that the mean PedsQL total score would decrease linearly with increasing severity on each of the CHU-9D dimensions.

Convergent Validity

Convergent validity was explored, using statistical tests of association, to determine how the CHU-9D correlated with the PedsQL measure.

Graphical means (scatter plots), along with fitted regression line and 95 % CIs, for the CHU-9D utility values (using the UK and Chinese tariffs) and the PedsQL total scores were used to show the relationship between the instruments.

Then, using the Spearman's rho statistic, the correlation coefficient between the CHU-9D utility values (using the UK and Chinese tariffs) and the PedsQL total scores was calculated. Spearman's Rank correlation coefficient Rs is a technique which can be used to summarise the strength and direction (negative or positive) of a relationship between two instruments. The result is always between 1 and -1. The meaning of the strength of the correlation using the guide for the value of Rs [288] is shown in Table 6.1.

Table 6.1 - Meaning of the strength of the correlation for the value of coefficient Rs

Value of coefficient Rs (positive or negative)	Meaning
0.00 to 0.19	A very weak correlation
0.20 to 0.39	A weak correlation
0.40 to 0.69	A moderate correlation
0.70 to 0.89	A strong correlation
0.90 to 1.00	A very strong correlation

The content and coverage of the two instruments were further estimated by examining the correlation between individual CHU-9D dimensions and the theoretically similar PedsQL domains (Table 6.2).

Table 6.2-Mapping of CHU9D dimensions against theoretically similar PedsQL domains

PedsQL instrument	CHU9D instrument
Physical functioning	Pain, Tired, Sleep, Daily routine
Emotional functioning	Worried, Sad, Annoyed
Social functioning	Ability to join in activities
School functioning	School work/Home work

6.3 Results

6.3.1 Participant Characteristics

Full data (including PedsQL total score and its sub-scales; CHU-9D dimensions and utility value; height and weight (converted to BMI z-score and weight status); gender; age; and parents' education level) were available for 1539 children (93.8% of those who consented and participated in study measurements) and are described in Table 6.3. Data on parental employment status was available for 1539 children and is presented in Appendix 3.1.

The mean age of the children was 6.6 years (SD= 0.42) and 54% were male. Around a third of parents did not have a university education. The mean BMI z-score was -0.12 (SD=1.29), whilst more than 17% of the children were either overweight (10.7%) or living with obesity (7.2%); comparable to national data for overweight/obesity in the same age group (20.4%) [127]. The mean utility scores of the total sample was, on average, slightly higher for CHU-9D using the UK tariff (mean = 0.937 (SD= 0.068) compared to using the Chinese tariff (mean = 0.920 (SD= 0.094). The mean total PedsQL score was 82.92 (SD= 11.21).

Table 6.3 - Characteristics of the study population

Characteristics	
Gender: n (%)	
Male	831 (54.0)
Female	708 (46.0)
Age (years): mean (SD)	6.6 (0.42)
Measures of socio-economic status	
Maternal university education: n (%)	
Yes	963 (62.6)
No	576 (37.4)
Maternal education level: n (%)	
1 School education	296 (19.2)
2 Occupation college	280 (18.2)
3 University undergraduate education	847 (55.1)
4 University postgraduate education	116 (7.5)
Paternal university education: n (%)	
Yes	1005 (65.3)
No	534 (34.7)
Paternal education level: n (%)	
1 School education	247 (16.2)
2 Occupation college	287 (18.6)
3 University undergraduate education	824 (53.5)
4 University postgraduate education	181 (11.7)
Weight status: n (%)	
Underweight	75 (4.9)
Healthy weight	1189 (77.2)
Overweight	165 (10.7)
Obese	110 (7.2)
Underweight/Healthy weight compared to Overweight/Obese: n (%)	
Underweight/Healthy weight	1264 (82.1)
Overweight/Obese	275 (17.9)
BMI: mean (SD)	15.45 (2.13)
BMI z-score: mean (SD)	-0.12 (1.29)
CHU-9D mean score (SD)	
CHU-9D: using UK tariff	0.937 (0.068)
CHU-9D: using Chinese tariff	0.920 (0.094)
PedsQL mean score (SD)	
PedsQL Total scale score	82.92 (11.21)
PedsQL Physical functioning	83.67 (13.15)
PedsQL Psychosocial functioning	82.52 (12.36)
PedsQL Emotional functioning	81.69 (17.54)
PedsQL Social functioning	84.09 (15.30)
PedsQL School functioning	81.77 (15.36)

6.3.2 Relationship between HRQoL and Weight Status, Gender and SES

Table 6.4 summarises the CHU-9D utility values and PedsQL total scores according to the weight status, socio-economic status and gender of the children. The direction of the relationships were similar between instruments.

Of interest, the mean and median utility scores using both UK and Chinese tariffs and mean and median PedsQL total scores were all marginally higher for children who were overweight/obese compared to those who were not.

The CHU-9D using both UK and Chinese tariffs discriminated by gender ($p = 0.003$ and $p = 0.004$ respectively), with girls having slightly higher mean and median utility scores. The mean and median PedsQL total score in girls was also slightly higher than that in boys, although these differences were not statistically significant.

All HRQoL measures were also marginally higher in children whose parents did not have a university education (lower socio-economic status) compared to those who did. However, these differences were not statistically significant. The analyses were re-run using parental employment status as an alternative proxy for socio-economic status and the results were similar (Appendix 3.2).

Table 6.4 - Mean (SD), median (IQR) for CHU9D, PedsQL scores based on characteristics

	Number (%)	CHU-9D Utility, UK tariff	CHU-9D Utility, Chinese tariff	PedsQL total score
		Mean (SD), Median (IQR)	**Mean (SD), Median (IQR)**	**Mean (SD), Median (IQR)**
Gender				
Male	831 (54.0)	0.932 (0.072), 0.952 (0.897-1.000)	0.914 (0.098), 0.939 (0.873-1.000)	82.29 (11.72), 83.69 (75.00-91.30)
Female	708 (46.0)	0.943 (0.063), 0.963 (0.909-1.000)	0.927 (0.089), 0.955 (0.881-1.000)	83.66 (10.54), 85.86 (77.17-91.30)
p-value*		**0.003***	**0.004***	0.06
Mother's university education				
Yes	963 (62.6)	0.936 (0.068), 0.956 (0.903-1.000)	0.920 (0.091), 0.943 (0.876-1.000)	82.58 (11.29), 83.69 (76.08-91.30)
No	576 (37.4)	0.938 (0.068), 0.963 (0.903-1.000)	0.921 (0.099), 0.952 (0.874-1.000)	83.49 (11.07), 85.86 (77.17-91.30)
p-value*		0.27	0.42	0.08
Mother education level				
1 School education	296 (19.2)	0.937 (0.070), 0.963 (0.895-1.000)	0.921 (0.096), 0.953 (0.870-1.000)	83.06 (11.18), 85.86 (76.08-91.30)
2 Occupation college	280 (18.2)	0.940 (0.067), 0.963 (0.907-1.000)	0.919 (0.102), 0.945 (0.879-1.000)	83.95 (10.96), 85.86 (78.26-91.30)
3 University undergraduate education	847 (55.1)	0.937 (0.068), 0.958 (0.903-1.000)	0.920 (0.091), 0.943 (0.876-1.000)	82.58 (11.37), 83.69 (76.08-91.30)
4 University postgraduate education	116 (7.5)	0.932 (0.070), 0.952 (0.901-1.000)	0.919 (0.092), 0.942 (0.885-1.000)	82.59 (10.71), 84.23 (75.00-89.13)
p-value**		0.27	0.36	0.19
Father's university education				

Yes	1005 (65.3)	0.936 (0.068), 0.955 (0.902-1.000)	0.920 (0.091), 0.943 (0.876-1.000)	82.90 (11.06), 84.78 (76.08-91.30)
No	534 (34.7)	0.939 (0.069), 0.963 (0.904-1.000)	0.921 (0.100), 0.955 (0.876-1.000)	82.97 (11.51), 85.86 (76.08-91.30)
p-value*		0.17	0.38	0.61
Father education level				
1 School education	247 (16.2)	0.931 (0.075), 0.963 (0.892-1.000)	0.911 (0.110), 0.943 (0.864-1.000)	82.27 (11.65), 83.69 (75.00-91.30)
2 Occupation college	287 (18.6)	0.946 (0.062), 0.963 (0.915-1.000)	0.928 (0.090), 0.955 (0.882-1.000)	83.57 (11.36), 85.86 (76.08-92.39)
3 University undergraduate education	824 (53.5)	0.937 (0.067), 0.960 (0.903-1.000)	0.921 (0.090), 0.943 (0.877-1.000)	83.11 (11.14), 84.78 (76.08-91.30)
4 University postgraduate education	181 (11.7)	0.932 (0.072), 0.952 (0.897-1.000)	0.916 (0.096), 0.943 (0.870-1.000)	81.91 (10.65), 83.69 (76.08-89.13)
p-value**		0.42	0.63	0.53
Weight status groups				
Underweight	75 (4.9)	0.942 (0.067), 0.963 (0.908-1.000)	0.923 (0.092), 0.938 (0.873-1.000)	82.47 (12.06), 85.86 (72.82-92.39)
Healthy weight	1189 (77.2)	0.936 (0.069), 0.962 (0.900-1.000)	0.919 (0.095), 0.943 (0.876-1.000)	82.84 (11.13), 83.69 (76.08-91.30)
Overweight	165 (10.7)	0.941 (0.064), 0.963 (0.909-1.000)	0.925 (0.086), 0.955 (0.874-1.000)	83.18 (11.65), 85.86 (76.08-91.30)
Obese	110 (7.2)	0.939 (0.071), 0.962 (0.914-1.000)	0.921 (0.096), 0.943 (0.890-1.000)	83.69 (10.94), 86.95 (77.17-91.30)
p-value**		0.73	0.89	0.29
Weight status groups				
Underweight/Healthy weight	1264 (82.1)	0.936 (0.069), 0.963 (0.901-1.000)	0.919 (0.095), 0.943 (0.875-1.000)	82.82 (11.18), 83.69 (76.08-91.30)
Overweight/Obese	275 (17.9)	0.940 (0.067), 0.964 (0.909-1.000)	0.923 (0.090), 0.944 (0.876-1.000)	83.38 (11.35), 85.86 (76.08-91.30)
p-value**		0.38	0.66	0.27

*Kruskal-Wallis test; **non-parametric test for trend

Table 6.5 shows the results of the linear mixed regression model which compared the CHU-9D utility score between the two weight status groups, adjusted for potential confounders (age, gender and mother/father's education). The results were similar to the unadjusted analyses as children who were overweight or obese had a marginally higher CHU-9D utility value, using UK and Chinese tariffs, compared to underweight/healthy weight but this association was not statistically significant. Girls had slightly higher mean CHU-9D utility values compared to boys ($p = 0.001$ and $p = 0.003$ for UK and Chinese tariffs respectively), whilst children whose parents had a university education reported a lower HRQoL (not statistically significant).

Relationship between Weight Status and HRQoL by Gender

Table 6.6 summarises the CHU-9D utilities and PedsQL total scores by gender further classified into weight status and socio-economic groups. As before, although the direction of the relationships between each instrument and; weight status and socio-economic status were very similar, the mean CHU-9D utilities using the UK tariff were, on average, slightly higher than those using the Chinese tariff.

When children were categorised by gender, there was no evidence of differentiating HRQoL, using either instrument, in either of the gender groups according to their weight status and parent's university education.

Table 6.5 - Results of linear mixed model to assess variation in CHU9D

Variables	CHU-9D utility score UK tariff			CHU-9D utility score Chinese tariff		
	Mean difference	**95 % CI**	**p-value**	**Mean difference**	**95 % CI**	**p-value**
Age (months)	0.001	(0.000, 0.001)	**0.01****	0.001	(0.000, 0.002)	**0.01****
Weight						
Underweight/Healthy weight	**-**					
Overweight/Obese	0.005	(-0.003, 0.014)	0.25	0.004	(-0.007, 0.016)	0.45
Gender						
Male	**-**					
Female	0.011	(0.005, 0.018)	**0.001****	0.014	(0.004, 0.023)	**0.003****
Mother's university Education						
No	**-**					
Yes	-0.003	(-0.012, 0.005)	0.46	-0.005	(-0.017, 0.007)	0.41
Father's university Education						
No	**-**					
Yes	-0.001	(-0.010, 0.007)	0.80	-0.002	(-0.009, 0.014)	0.66

**Significant at $p = 0.05$

Table 6.6 - Mean (SD), median (IQR) for CHU9D, PedsQL scores by sex characteristics

	n	Boys			n	Girls		
		Mean (SD), Median (IQR) CHU-9D Utility, UK tariff	**Mean (SD), Median (IQR) CHU-9D Utility, Chinese tariff**	**Mean (SD), Median (IQR) PedsQL total score**		**Mean (SD), Median (IQR) CHU-9D Utility, UK tariff**	**Mean (SD), Median (IQR) CHU-9D Utility, Chinese tariff**	**Mean (SD), Median (IQR) PedsQL total score**
Weight status groups								
Underweight Healthy weight	641	0.931 (0.072) 0.951 (0.897-1.000)	0.913 (0.099) 0.938 (0.874-1.000)	82.10 (11.74) 83.69 (75.00-91.30)	623	0.942 (0.065) 0.963 (0.904-1.000)	0.926 (0.091) 0.955 (0.880-1.000)	83.56 (10.54) 84.78 (77.17-91.30)
Overweight /Obese	190	0.936 (0.072) 0.963 (0.903-1.000)	0.918 (0.096) 0.943 (0.872-1.000)	82.93 (11.68) 85.86 (76.08-91.30)	85	0.951 (0.051) 0.963 (0.914-1.000)	0.935 (0.072) 0.955 (0.891-1.000)	84.41 (10.58) 86.95 (79.34-91.30)
p-value**		0.29	0.41	0.28		0.45	0.79	0.38
Weight status groups								
Underweight	35	0.923 (0.070) 0.929 (0.877-1.000)	0.899 (0.096) 0.922 (0.815-1.000)	79.93 (14.55) 83.69 (69.56-92.39)	40	0.959 (0.060) 0.989 (0.924-1.000)	0.944 (0.084) 0.998 (0.913-1.000)	84.70 (8.97) 86.95 (76.08-92.39)
Healthy weight	606	0.932 (0.072) 0.951 (0.900-1.000)	0.914 (0.099) 0.939 (0.876-1.000)	82.23 (11.56) 83.69 (75.00-91.30)	583	0.940 (0.065) 0.963 (0.902-1.000)	0.925 (0.091) 0.953 (0.875-1.000)	83.48 (10.64) 84.78 (77.17-91.30)
Overweight	108	0.937 (0.069) 0.963 (0.893-1.000)	0.918 (0.094) 0.953 (0.860-1.000)	82.56 (12.26) 85.86 (76.08-91.30)	57	0.950 (0.052) 0.963 (0.914-1.000)	0.939 (0.065) 0.955 (0.896-1.000)	84.35 (10.38) 85.86 (79.34-91.30)
Obese	82	0.934 (0.077) 0.951 (0.914-1.000)	0.918 (0.100) 0.943 (0.890-1.000)	83.41 (10.92) 86.41 (76.08-91.30)	28	0.953 (0.051) 0.963 (0.916-1.000)	0.928 (0.086) 0.940 (0.883-1.000)	84.53 (11.15) 87.77 (80.97-91.30)
p-value**		0.27	0.29	0.22		0.84	0.58	0.57
Mother's university education								
Yes	508	0.931 (0.072) 0.951 (0.901-1.000)	0.914 (0.094) 0.938 (0.876-1.000)	81.95 (11.72) 83.69 (75.00-91.30)	455	0.942 (0.063) 0.963 (0.907-1.000)	0.926 (0.087) 0.955 (0.878-1.000)	83.28 (10.75) 84.78 (77.17-91.30)
No	323	0.934 (0.072) 0.963 (0.893-1.000)	0.915 (0.104) 0.943 (0.865-1.000)	82.82 (11.73) 85.86 (76.08-91.30)	253	0.944 (0.064) 0.965 (0.914-1.000)	0.928 (0.092) 0.956 (0.889-1.000)	84.34 (10.14) 85.86 (78.26-91.30)
p-value*		0.25	0.43	0.21		0.66	0.70	0.20
Father's university education								
Yes	532	0.931 (0.072) 0.951 (0.896-1.000)	0.913 (0.095) 0.938 (0.871-1.000)	82.20 (11.45) 83.69 (75.00-91.30)	473	0.942 (0.063) 0.963 (0.909-1.000)	0.926 (0.086) 0.953 (0.882-1.000)	83.61 (10.56) 84.78 (78.26-91.30)
No	299	0.935 (0.072) 0.963 (0.902-1.000)	0.915 (0.104) 0.951 (0.875-1.000)	82.46 (12.21) 84.78 (76.08-91.30)	235	0.945 (0.063) 0.963 (0.908-1.000)	0.927 (0.094) 0.955 (0.881-1.000)	83.68 (10.53) 85.86 (76.08- 92.39)
p-value*		0.15	0.29	0.44		0.54	0.82	0.94

6.3.3 Construct Validity

Discriminant Validity

As reported in section 6.3.2, HRQoL using both the CHU-9D and the PedsQL, was marginally higher in children from lower versus higher socio-economic background.

The discriminant validity of the CHU-9D instrument was further explored. Table 6.7 shows that the mean (SD) utility values (using UK and Chinese tariffs) were significantly lower for children who had a PedsQL total score less than the median value, compared to those with scores greater than or equal to the median.

Table 6.7 - CHU9D utility by PedsQL total score

Group	**n**	**Mean (SD) CHU-9D Utility, UK tariff**	**Mean (SD) CHU-9D Utility, Chinese tariff**
PedsQL total score <median	793	0.909 (0.075)	0.881 (0.106)
PedsQL total score >=median	746	0.967 (0.043)	0.961 (0.056)
p-value		**<0.001**	**<0.001**

Table 6.8 summarises the mean PedsQL total scores across the dimension levels of the CHU-9D. The majority of children reported themselves in good health, with the largest proportion reporting themselves at the highest level for all dimensions of the CHU-9D. In general, the mean PedsQL total scores corresponded well, decreasing mostly linearly with increasing levels of severity on each dimension of the CHU-9D. This result was statistically significant (p-value = <0.001) for each of the dimensions.

Table 6.8 - Mean PedsQL total scores by each level of CHU9D dimension

CHU-9D Dimensions	Level	n (%)	Mean (SD) PedsQL total score	p-value*
Worried	No	1375 (89.4)	83.93 (10.56)	**<0.001**
	A little bit	88 (5.8)	78.11 (12.01)	
	A bit	34 (2.2)	71.59 (12.61)	
	Quite	23 (1.4)	67.81 (13.93)	
	Very	19 (1.2)	70.65 (11.23)	
Sad	No	1439 (93.5)	83.59 (10.74)	**<0.001**
	A little bit	55 (3.7)	76.17 (11.42)	
	A bit	22 (1.4)	69.81 (15.32)	
	Quite	7 (0.4)	79.65 (13.17)	
	Very	16 (1.0)	65.72 (13.98)	
Pain	No	1353 (87.9)	83.91 (10.60)	**<0.001**
	A little bit	97 (6.4)	77.71 (12.16)	
	A bit	49 (3.2)	74.46 (11.25)	
	Quite	28 (1.8)	73.72 (14.71)	
	Very	12 (0.7)	70.10 (17.89)	
Tired	No	1235 (80.3)	84.71 (10.25)	**<0.001**
	A little bit	171 (11.1)	78.01 (10.80)	
	A bit	69 (4.5)	74.85 (11.19)	
	Quite	31 (2.0)	70.53 (14.75)	
	Very	33 (2.1)	70.06 (13.80)	
Annoyed	No	1400 (91.0)	83.84 (10.68)	**<0.001**
	A little bit	63 (4.1)	75.70 (11.02)	
	A bit	33 (2.1)	70.69 (11.97)	
	Quite	18 (1.2)	74.03 (13.66)	
	Very	25 (1.6)	72.30 (14.24)	
School work/Home work	No problems	1213 (78.8)	84.72 (10.34)	**<0.001**
	A few problems	156 (10.1)	79.86 (10.15)	
	Some problems	129 (8.5)	74.23 (11.09)	
	Many problems	33 (2.1)	70.75 (14.53)	
	Can't do	8 (0.5)	59.64 (12.10)	
Sleep	No problems	1291 (83.9)	84.05 (10.71)	**<0.001**
	A few problems	145 (9.4)	78.83 (11.38)	
	Some problems	43 (2.8)	75.15 (11.03)	
	Many problems	21 (1.4)	74.94 (9.07)	
	Can't do	39 (2.5)	73.49 (14.90)	
Daily routine	No problems	1410 (91.6)	83.67 (10.78)	**<0.001**
	A few problems	100 (6.6)	76.55 (10.77)	
	Some problems	19 (1.2)	70.50 (16.21)	
	Many problems	9 (0.5)	62.92 (15.84)	
	Can't do	1 (0.1)	81.52 (.)	
Ability to join in activities	Any	906 (58.8)	85.37 (10.20)	**<0.001**
	Most	238 (15.4)	81.81 (11.81)	
	Some	186 (12.1)	78.35 (11.59)	
	A few	147 (9.5)	77.42 (11.56)	
	No	62 (4.2)	78.14 (10.21)	

*Non-parametric test for trend

Convergent Validity

Figure 6.1 shows a scatter plot comparison of the relationship between the CHU-9D utility values (using UK tariff) and the PedsQL total scores. Some anomalies are apparent. For instance, one child reported a high CHU-9D utility score of 0.963, yet had a low PedsQL total score of 34.78. However, in general, there is a moderate association between the instruments with higher CHU-9D utility values corresponding with higher PedsQL total scores (the CHU-9D utility values and PedsQL total scores converging towards the highest end of the scale).

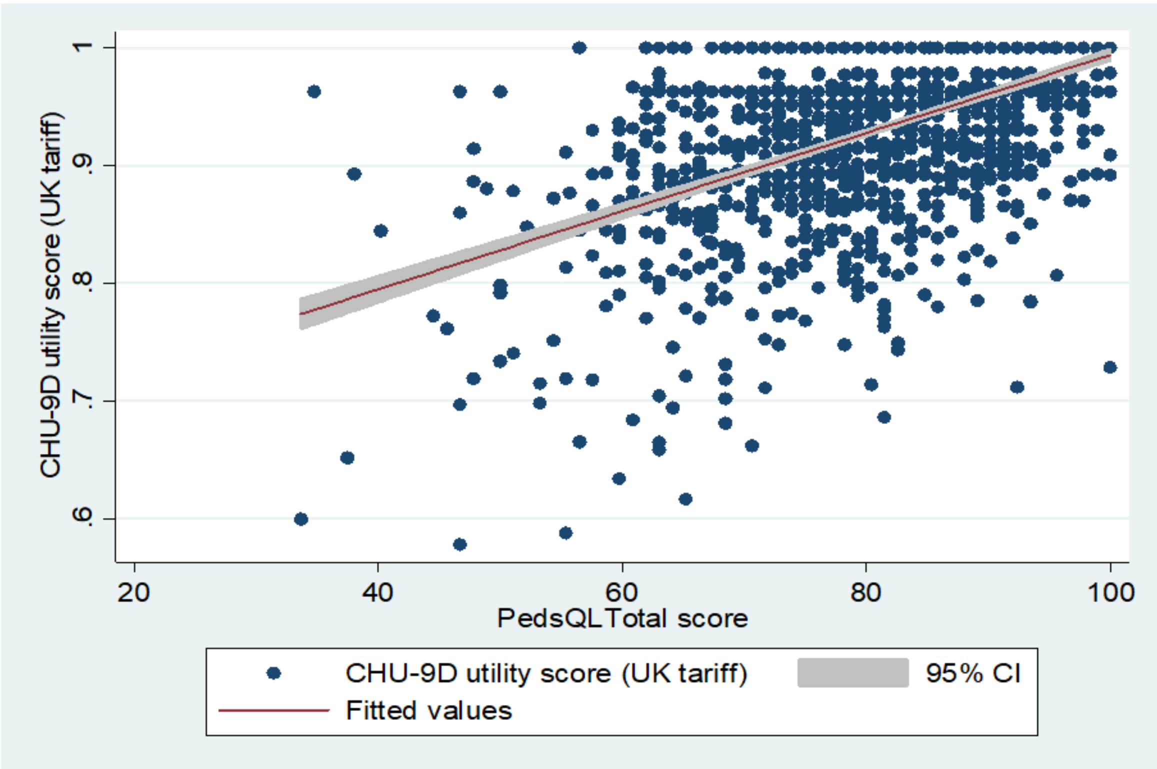

Figure 6.1 - Relationship between CHU9D utility scores, UK tariff, and PedsQL scores

Figure 6.2 shows a scatter plot comparison of the relationship between the CHU-9D utility values (using Chinese tariff) and the PedsQL total scores. Figure 2 is similar to Figure 1 but some wider anomalies are apparent. For instance, one child reported a high CHU-9D utility score of 0.996, yet had a low PedsQL total score of 34.78, and another child reported a low CHU-9D utility score of 0.535, yet had a high PedsQL total score of 82.60. However, in general, again there is a moderate association between the instruments with higher CHU-9D utility values corresponding with higher PedsQL total scores (the CHU-9D utility values and PedsQL total scores converging towards the highest end of the scale).

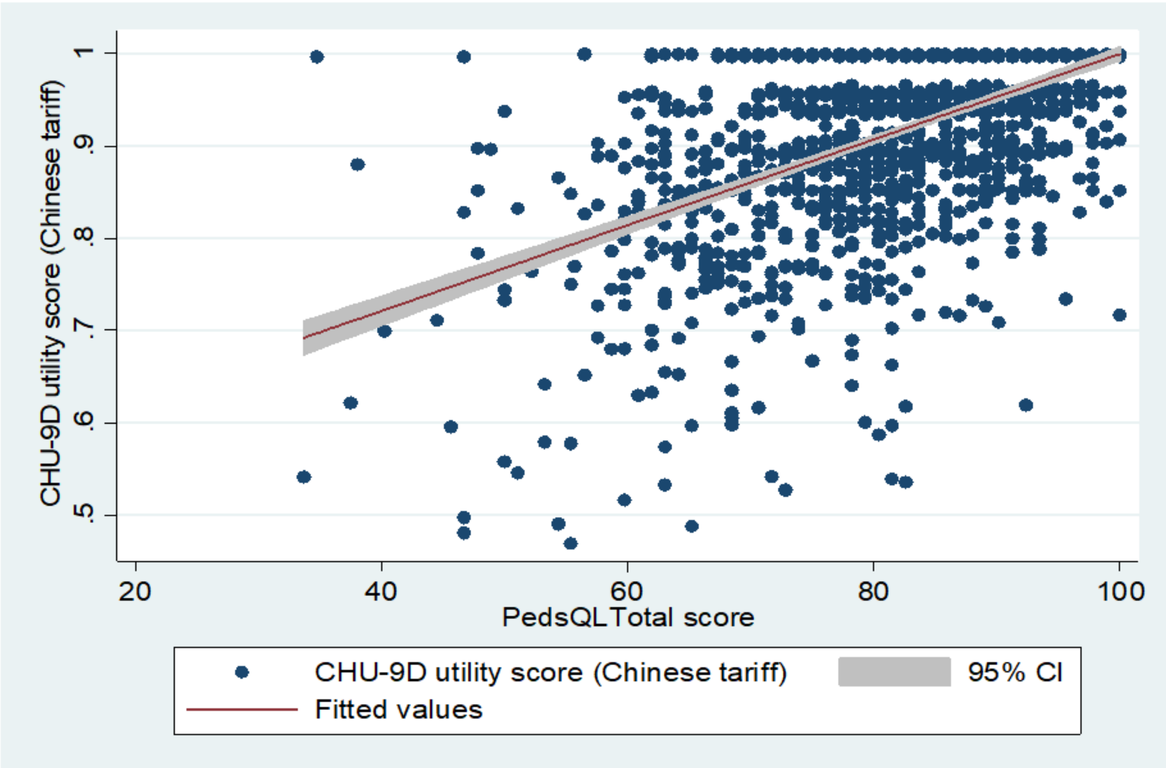

Figure 6.2 - Relationship between CHU9D utility scores, Chinese tariff, and PedsQL scores

Table 6.9 summarises the relationship between the distribution of CHU-9D utility scores and the PedsQL total scores in terms of Spearman's correlation coefficients. Overall, the correlation between the CHU-9D utility values (using both UK and Chinese tariffs) and PedsQL total scores showed a statistically significant moderate positive correlation (Rs = 0.5221, p = <0.001) and (Rs = 0.5316, p = <0.001) respectively.

Table 6.9 - Correlation between CHU9D utility score and PedsQL total score

CHU-9D utility score	Correlation with PedsQL total score	Spearman's rho	p-value
Using UK tariff	PedsQL total score	0.52	**<0.001**
Using Chinese tariff	PedsQL total score	0.53	**<0.001**

*Both were significant at 0.01 level

The content and coverage of the two instruments were further compared by examining the correlation between each of the CHU-9D dimensions and the theoretically similar PedsQL domain functioning scores (Table 6.10).

Using conventional cut-off values for Spearman's rho, each CHU-9D dimension was either weakly, or very weakly correlated with each of the predetermined PedsQL domain functioning scores (there was very weak evidence for convergent validity in relation to dimensions and domains: between pain and physical functioning, between sleep and physical functioning, between daily routine and physical functioning). Since the CHU-9D dimensions are labelled with 1 as highest level and 5 as lowest level, the signs on the coefficients were consistently negative. All correlations were significant at the 0.01 level.

Table 6.10 - Correlation between CHU9D dimensions and PedsQL domain scores

CHU-9D dimension	**Correlation with PedsQL domain functioning score**	**Spearman's rho**	**p-value***
Worried	Emotional functioning	-0.24	**<0.001**
Sad	Emotional functioning	-0.22	**<0.001**
Pain	Physical functioning	-0.18	**<0.001**
Tired	Physical functioning	-0.25	**<0.001**
Annoyed	Emotional functioning	-0.22	**<0.001**
School work/Home work	School functioning	-0.25	**<0.001**
Sleep	Physical functioning	-0.14	**<0.001**
Daily routine	Physical functioning	-0.18	**<0.001**
Ability to join in activities	Social functioning	-0.21	**<0.001**

*All were significant at 0.01 level

The paired comparison of the CHU-9D utility scores, using UK and Chinese tariffs, is presented in Table 6.11. The difference between the means was found to be statistically significant ($p < 0.001$). In addition, exactly the same proportion of the total sample reported themselves in full health (utility value = 1.0) for the CHU-9D using both UK and Chinese tariffs (32.1%, n = 494). The Spearman's rho indicated a very strong level of agreement overall between the CHU-9D, using the UK and Chinese tariffs.

Table 6.11 - Paired comparison of CHU9D utility scores using UK and Chinese tariff

Mean (SD), Median (IQR) CHU-9D Utility, UK tariff	**Mean (SD), Median (IQR) CHU-9D Utility, Chinese tariff**	**Difference between means**	**Spearman's Rank correlation coefficient Rs**
0.937 (0.068) 0.963 (0.903-1.000)	0.920 (0.094) 0.943 (0.876-1.000)	0.017*	0.95

*p-value < 0.001

6.4 Discussion

Mapping algorithms provide an empirical tool for estimating CHU-9D index scores and for undertaking CUA within clinical studies that have only collected PedsQL data. They are valid for children aged 5 years or older (preferably up to a maximum of 13 years old) [289]. It has been shown that mapping algorithms also provide an empirical tool for estimating health utilities in childhood when EQ-5D data are not available [290]. These can be used to inform future economic evaluations of childhood interventions. They are likely to be robust for a population of children aged 11-15 years whose characteristics are comparable to the study by khan et al. 2014 [290]. However, the performance of these algorithms in younger children with different clinical characteristics remains to be evaluated and they are, therefore, not useful for the CHIRPY DRAGON study. Mapping is always regarded as a 'second best option' to empirically collecting the data and as the CHU-9D data was empirically collected within the CHIRPY DRAGON study, mapping was not applied.

6.4.1 Statement of Principal Findings

Aim (a)

The findings suggest that HRQoL in this study population was marginally higher in children who were overweight/living with obesity compared to children in healthy weight, although these associations were not statistically significant. Girls reported significantly higher HRQoL, compared to boys, using both the CHU-9D and the PedsQL.

When children were categorised by gender, there was no evidence of differentiating HRQoL, using either instrument.

Aim (b)

Regarding the discriminant validity, it was found that the CHU-9D discriminated according to the median PedsQL total score. Furthermore, the mean PedsQL total scores decreased mostly linearly with increasing levels of severity on each dimension of the CHU-9D. However, contrary to studies conducted in Western countries, and although not statistically significant, it was found that HRQoL, using both the CHU-9D and the PedsQL, was higher in children whose parents had lower levels of education, compared to those whose parents were university educated.

With respect to convergent validity, although there was a moderate significant positive correlation between CHU-9D utility values (using both UK and Chinese tariffs) and PedsQL total scores, the correlation between individual CHU-9D dimensions and the theoretically similar PedsQL domains were weak or very weak.

6.4.2 Strengths and Limitations of this Study

Strengths include the large sample size (1539 children), diverse population (selected to include a range of socio-economic backgrounds) and standardised data collection procedures as part of the randomised controlled trial.

Furthermore, this study is one of the very few studies worldwide and the first study in China, which collected utility-based HRQoL information in children as young as 6 years. It used both UK and Chinese tariffs for calculating the utility scores and reports on the psychometric properties of the CHU-9D in direct comparison to the widely used PedsQL instrument.

The study had some limitations to note however: in this chapter, the “underweight” and “healthy weight” children were pooled into one weight category as a very small number of children in the sample (5%) were measured as “underweight”. Although some population studies have reported that the HRQoL of underweight children was generally no different from those with healthy weight [291], this could not be explored in this study as the underweight sample size was so small.

Another concern about the findings of this study was that data analysis was limited to data collected as part of the trial. For example, discriminant validity of the CHU-9D could not be assessed compared to an obesity-specific quality of life measure. However, the PedsQL as a ‘gold standard’ is a widely used HRQoL instrument validated for use with young children in diverse populations [189, 190].

6.4.3 Comparison with Other Studies

Aim (a)

There is no robust evidence on the direction of the relationship between weight status and utility-based HRQoL in this population. This is compounded by the challenging nature of measuring utility in a paediatric population more generally [292]. In four previous studies which have explored this relationship in children (three UK-based studies using the CHU-9D [42, 280, 286] and one US-based study using the HUI instrument [188]), the opposite direction of effect was found (lower HRQoL in participants with overweight/obesity compared with their underweight/healthy weight counterparts). However, like this study, the results were not statistically significant (no evidence of a negative relationship between health utility and weight status in children aged 5-6 years [42], aged 6-7 years [286], aged 5-10 years [280] or in children and adolescents aged 5-18 years [188] was found). In contrast, however, the findings of one recent study from Australia using the CHU-9D in children aged 9-12 years [293] and one study from the UK using the EQ-5D-Y in children aged 11-15 years [294], found a significant negative relationship between weight status and health utility. Among all these studies, a higher utility-based HRQoL in children who were overweight/obese compared to underweight/healthy weight was only found in this Chinese study. Therefore, despite these reports of a negative relationship between utility-based HRQoL and being overweight in children, the direction of the relationship is not robust, the evidence for understanding this relationship is inconsistent and is mixed in terms of whether this effect reaches statistical significance.

The weak relationship between weight status and utility-based HRQoL may be attributed to the CHU-9D not being sensitive enough to detect a difference in very young children as it was originally developed for use with children aged 7–11 years [41]. Although the findings of a UK-based study suggested the instrument to be acceptable and feasible to administer for children aged 6–7 years [38], there are still concerns with regard to the instrument's reliability in young children [38, 295]. A wide range of previous studies demonstrate that childhood obesity is associated with lower HRQoL when non-utility instruments are used [102, 294, 296-299]. However, the findings are not consistent and, for example, in addition to this study, another study from China [127] and one from the UK [42] found no significant relationship between weight status and HRQoL measured using the PedsQL. Cultural differences may play a role. Most Chinese parents and, moreover, grandparents, aspire for children to be overweight, as this is taken to be a sign of health, growth and prosperity [82, 300]. Obesity trends follow a different pattern in China compared with high-income countries with the risk of obesity being greater in children from higher socio-economic backgrounds. The lack of association may also be related to the fact that co-morbidities attached to obesity do not substantially affect utility in this age group and, possibly, it is only once these children approach adolescence that the effects of obesity starts to impact negatively on utility-based HRQoL.

Within a UK-based study, the opposite direction of effect was found compared to this study regarding the relationship between utility-based HRQoL and gender, but there was no statistical difference between utility values and gender in children aged 5–6 years [42]. Within an Australian-based study, the same direction of effect was found compared to this study, but, again, there was no statistical difference between utility values and gender in children [41].

Aim (b)

Regarding the discriminant validity, two of the findings were in line with a previous study reported from a UK setting [42]. However, unlike studies from the West (a UK study using the CHU-9D and PedsQL in children aged 5-6 years, and an Australian study using the CHU-9D in children aged 11-17 years) [41, 42], there was no evidence of lower HRQoL in those from a lower socio-economic background - and the direction of effect suggested that any association was the reverse of that observed in other studies. A Chinese study setting reported a statistically significant trend for higher HRQoL scores (using PedsQL) in children with increased years of parental education [127].

With respect to the convergent validity, the findings were similar to the previous study in the UK [42]. The weak, or very weak correlation between the individual dimensions of each instrument might be because these individual dimensions actually describe something that is quite specific and different while appearing quite similar.

Sensitivity to change was not part of the psychometric analysis. As only the relationship between weight status and HRQoL, along with the construct validity of the CHU-9D, was checked, only baseline data was used. Every other study whose results were compared with these findings, also used baseline data.

6.4.4 Implications for Practice, Policy and Research

Aim (a)

The results of this chapter has methodological and policy implications in terms of how the cost-effectiveness of childhood obesity interventions is measured in children aged 6-7 years. Obesity prevention and treatment interventions tend to target young populations, therefore information about how weight status is associated with HRQoL in utility terms in this age group is useful for the design of economic evaluations. Within health economic studies conducted globally, utility values are often used to derive QALYs to inform resource allocation decisions. To help inform the methods of economic evaluations alongside clinical trials of childhood obesity prevention and treatment interventions, future studies need to determine the relationship between weight status and utility-based HRQoL in different age groups, and across different country settings. In addition, it is recommended that future studies aiming to prevent obesity in young children (age 6-12) do not rely solely on utility-based HRQoL measures for economic evaluation, and capture clinical or wellbeing outcomes as well. This is because: CHU-9D might not be sensitive enough to detect a difference in very young children [41]. Also, there are still concerns with regard to the instrument's reliability when used with young children [38, 295]. Furthermore, it remains unknown how the co-morbidities associated with obesity affect utility in this age group and, possibly, it is only once these children approach adolescence that the effects of obesity start to impact negatively on utility-based HRQoL.

Aim (b)

Overall, the findings provide support for the discriminant and convergent validity of the CHU-9D when used as a utility score within a Chinese population aged 6-7 years. This is because the CHU-9D discriminated according to the PedsQL median score, and the mean PedsQL total scores decreased linearly with increasing levels of severity on each dimension of the CHU-9D. However, there still remains uncertainty, as the CHU-9D dimensions were only weakly correlated with theoretically similar PedsQL dimensions. So we recommend future studies continue to test the validity of the CHU-9D in different age groups and country settings.

6.5 Conclusion

Aim (a)

The results of this chapter suggest that HRQoL using both CHU-9D and PedsQL instruments is slightly higher among children who are overweight/obese compared to underweight/healthy weight. However, this difference is not statistically significant. Findings of sub-group analyses are consistent with the analysis on whole groups. Some studies from high-income countries suggest that overweight/obesity in children is negatively associated with utility-based HRQoL. However, the extent of the relationship, how it varies across age groups, and how this translates to utility-based HRQoL across different settings is as yet under researched.

Aim (b)

This chapter contributes utility data generated from a large Chinese sample of children. It reports on the psychometric properties of the CHU-9D instrument. The findings support the discriminant and convergent validity of the CHU-9D, as a measure of utility-based HRQoL for application in economic evaluation of prevention interventions within Chinese children aged 6-7 years.

In the following chapter, a discussion of the findings of the whole thesis is reported. This includes a summary of the findings, discussion on applied findings to inform policy development, reflections on the methods used for conducting the economic evaluation, implications for policy making and recommendations for future research.

CHAPTER 7. DISCUSSION

7.1 Introduction

To address the evidence gap of what interventions to implement to prevent childhood obesity in China, and to address the methodological challenges of conducting economic evaluation within this setting, the CHIRPY DRAGON study was developed.

The primary aim of this thesis was to contribute to the methodology of conducting economic evaluation in LMICs and to estimate the cost-effectiveness of a school- and family-based childhood obesity intervention in a Chinese setting. For this purpose, trial data was used and the economic evaluation methods were derived from a combination of published literature and guidelines for conducting economic evaluation. A comprehensive economic evaluation was conducted (from a public sector and societal perspective using both clinical and economic outcome measures). The economic evaluation results reflect the costs and benefits of preventing childhood obesity through schools in China. The methodological challenges of conducting an economic evaluation within a Chinese setting and including spillover effects were explored. The uncertainty surrounding the results was fully explored and reported to help inform policy recommendations and to plan future research.

The motivations for undertaking the research in this thesis came from the increasing concerns regarding the high and fast growing prevalence of childhood obesity in China [12, 58]; and CDC recommendations, which called for research on the costs and benefits of strategies to prevent childhood obesity.

As was shown in the systematic review presented in chapter four of this thesis, there are relatively few published economic evaluations of obesity prevention intervention studies (the only study in China was not conducted from a societal perspective and only included clinical outcome measures). Furthermore, the systematic review findings showed that the majority of published economic evaluations are for interventions with an individual behaviour change component. The review found heterogeneity with respect to methods applied making a synthesis of findings challenging.

Obesity is a particularly interesting condition to consider for two main reasons: first, the condition is chronic and some symptoms occur later in life; and second, the condition is known to greatly impact on both health and non-health aspects of individuals' lives. With the condition of obesity, an assessment of QoL is one of the primary indicators of prevention success, so the secondary aim of this thesis was to consider the suitability of economic outcome measures for interventions targeting school-aged children in China. The extra-welfarist outcome measure, the QALY, is currently recommended by UK decision-makers such as NICE. This is why it was chosen as the main economic outcome measure in the trial. For the secondary aim, first, how weight status relates to HRQoL was explored; then the construct validity of the CHU-9D was assessed. For both of these aims, the baseline data from the CHIRPY DRAGON trial was used.

This discussion starts by revisiting the aims of this thesis; and provides a summary of the key findings from the entire thesis perspective and discusses them within the context of the wider literature. This is followed by a discussion of the applied findings to inform policy development and reflections on the methods used for conducting the economic evaluation, highlighting the main strengths and limitations of the approach. The final sections discuss the implications for current policy making, and suggest future research recommendations.

7.2 Summary of the Findings

This section reports the main findings and discusses them with regard to current available evidence. Both the systematic review and the CHIRPY DRAGON obesity prevention programme followed pre-specified, peer-reviewed protocols which were published in scientific journals [126, 210]. The findings were reported in accordance with PRISMA guidelines [209] for a systematic review, and CHEERs guidelines [48] for economic evaluation. Findings have been disseminated through two publications; two conference papers and two oral presentations at an international conference and four poster presentations including one pitch poster presentation at national conferences.

The research has made several original and pertinent contributions to the literature on obesity research, both in terms of the methodological approach and also for the information made available to inform policy decision making with respect to childhood obesity prevention. These relate to the:

- Undertaking of a rigorous systematic review of methods, study quality and results of trial-based and model-based economic evaluations for childhood and adolescent obesity interventions which found heterogeneity with respect to methods applied.
- Describing the methodological challenges both of conducting this first economic evaluation of an obesity prevention intervention (CHIRPY DRAGON) in a Chinese setting and of including spillover costs/effects in the evaluation. The methodological challenges relate to methods for converting costs; sourcing unit costs; dealing with lack of an appropriate value set; dealing with clustering; managing lack of equivalent threshold values for outcome gains; and measuring household costs and outcome data.

- Undertaking an economic evaluation in a Chinese setting from both a public sector and societal perspective.
- Exploring the suitability of economic outcome measures within a Chinese childhood population.

Systematic Review

A systematic review of methods, study quality and results of trial-based and model-based economic evaluations for childhood and adolescent obesity interventions was presented in chapter 4. The results showed that current economic evaluations are mainly set in developed countries and the majority focus on the prevention of obesity in children, compared to treatment. Moreover, the findings of this review showed that the majority of published economic evaluations are for interventions with an individual behaviour change component. The majority, particularly "behavioural and policy" preventive interventions, were cost-effective, even cost-saving. However, this review found that relatively few policy interventions designed to prevent obesity have been rigorously evaluated from an economic perspective. The review found heterogeneity with respect to methods applied. So, to improve the evidence base further and to enhance comparability across interventions, a consistent and expanded form of economic evaluation which captures both health and non-health costs and consequences beyond health-gain was recommended. The systematic review also raised concerns regarding the generalisability of results to other contexts as interventions targeted individual health behaviours which are highly dependent on cultural, infrastructural and other system-related aspects.

In general, the systematic review results showed that the following main gaps exist in the current literature:

- Economic evaluations of obesity prevention intervention in developing countries.
- Inclusion of societal costs and outcomes in economic evaluations that are relevant to family members of the children affected by obesity.
- CUAs of trial-based studies comparing obesity prevention strategies, whose results can be used by decision makers and compared with other public health programmes.

Economic Evaluation of CHIRPY DRAGON Programme

As mentioned, some gaps were found in the findings of the systematic review presented in chapter 4. In order to fill these, an economic evaluation of the CHIRPY DRAGON programme was undertaken, which is presented in chapter 5. The aim was to explore the methodological challenges of conducting a comprehensive economic evaluation, including spillover effects, within a Chinese setting. This was achieved using the CHIRPY DRAGON trial as a case study.

In total, 1641 children were recruited and randomized to 20 intervention (n= 832) and 20 control (n= 809) schools. The intervention consisted of four components targeting diet and physical activity behaviours of children and their families in primary schools in Guangzhou, China. The 12-month programme, delivered by trained project staff known as CHIRPY DRAGON teachers, included (i) educational and skills-based workshops aimed at children aged 6-7 years and their parents or grandparents to promote physical activity and healthy eating; (ii) a school food improvement component involving school caterers; and physical activity initiatives (iii) within and (iv) outside school. Control schools continued with usual activities.

The costs linked to the intervention were divided into three categories: development, implementation and delivery. Unit costs were identified from Chinese sources. GDP PPPs were used to convert Yuan into Pounds and Dollars. The clinical measure of effectiveness was BMI z-score. The economic outcome measure was QALYs. Utility-data was collected using the CHU-9D for children and EQ-5D-3L for carers, applying the UK and the Chinese value set for both. The reasons for missing data differed for the resource use and outcome data. As the missing data was low, modified intention to treat approach was used.

Since a time horizon of 1 year was used, costs and outcomes were not discounted, as recommended by NICE. For the public sector perspective, the cost-effectiveness was estimated based on cost per QALY and BMI z-score change. For this, the ICER was calculated based on the fully adjusted costs and effects. In the absence of an agreed Chinese threshold for the value of a QALY, decision uncertainty was assessed using established UK and US thresholds, and presented using CEAC. Three sensitivity analyses were undertaken to assess the robustness of the results to assumptions made in the analysis. For the societal perspective, children's QALYs were adjusted for family spillover effects using a 'multiplier' approach, developed by Al Janabi et al (2016). The family costs were simply averaged for each child by assuming they had at least two family members attend the workshops. Then, the ICER was calculated. Four sensitivity analyses were undertaken.

The economic evaluation from a public sector perspective showed that the CHIRPY DRAGON intervention had a relatively low cost and significant intervention benefits over the course of 12 months, suggesting it was highly cost-effective. The intervention was cost-effective using the conventional decision making rules within a CEA and CUA. Broadening the evaluative space to include household costs and QALYs had the effect of increasing the ICER however the intervention remained cost-effective. The ICER did not change substantially in sensitivity analyses: a maximum of £16,709/$23,767 per QALY from a societal perspective when predictive mean matching multiple imputation was applied.

According to the systematic review reported in chapter 4 [47], with the exception of one study [249], all reported evaluations of school-based obesity interventions appear cost-effective using a 'cost per weight-specific outcome'. However, without thresholds for obesity-related outcomes, it is difficult to judge value for money. The ICER for the CHIRPY DRAGON programme was lower than two previous trial-based intervention studies which used BMI z-score as their measure of effectiveness: one Chinese study, targeting dietary habits and physical activity in children 6-13 years [17]; the other, an Australian study, targeting physical activity in adolescents 13-16 years [206]. However, a similar study in the UK, targeting dietary habits and physical activity in children 6-7 years, was not cost-effective using BMI z-score as the outcome measure. The UK study was cost-effective using QALY outcome (ICER: £26,815 per QALY gained from a public sector perspective), however there was a high level of uncertainty as demonstrated by the net-benefit equation and the corresponding CEAC [249].

Association between Weight and HRQoL; and Construct Validity of CHU-9D

The aims of chapter 6 were (a) to examine how children's weight status relates to their HRQoL and (b) to assess the construct validity of the CHU-9D instrument.

The results suggested that HRQoL using both CHU-9D and PedsQL instruments was slightly higher among children who were overweight/obese compared to underweight/healthy weight. However, this difference was not statistically significant. The findings of the sub-group analyses were consistent with the analysis on whole groups. Some studies from high-income countries suggest that overweight/obesity in children is negatively associated with utility-based HRQoL. However, the extent of the relationship, how it varies across age groups and how this translates to utility-based HRQoL across different settings is as yet under researched.

Chapter 6 reported on the psychometric properties of the CHU-9D instrument and showed support for the discriminant and convergent validity of the CHU-9D, as a measure of utility-based HRQoL for application in economic evaluation of prevention interventions within Chinese children aged 6-7 years. Regarding the discriminant validity, two of the findings were in line with a previous study reported from a UK setting [42]. However, unlike studies from the West [41, 42], there was no evidence of lower HRQoL in children from a lower socio-economic background – in fact the direction of effect was the reverse. This was in contrast to another Chinese study setting that reported a statistically significant trend for higher HRQoL scores (using PedsQL) in children with increased years of parental education [127].

With respect to the convergent validity of the CHU-9D, the findings were similar to the previous study in the UK [42].

7.3 Applied Findings to Inform Policy Development

The following section discusses the applied findings from the whole thesis to inform policy development.

The Novelty of CHIRPY DRAGON

Today, childhood obesity prevention is well-established and a growing research field. In 2009, when the CHIRPY DRAGON intervention study began development, it was among one of the very few trials, with a rigorous and theoretically informed intervention development process [301], to evaluate the effectiveness and cost-effectiveness of an obesity prevention intervention programme in children as young as 6 years and their family members, outside of high income countries.

CHIRPY DRAGON was a study promoting healthy eating and physical activity behaviour, closely embedded in a school setting with a cluster RCT design, feasible to implement in such a setting. There was adherence to international guidelines for trial design and implementation. The trial included a wide range of outcomes which were objectively assessed, where possible, and there were high follow up rates. In addition, a comprehensive process evaluation which demonstrated high fidelity and engagement was provided. Process evaluation helps to further identify facilitators or hindering factors on the pathway between intervention and effects [302].

CHIRPY DRAGON Effectiveness in Relation to other Studies

Findings of this trial added to the current knowledge base in relation to the effectiveness of childhood obesity prevention interventions within a global context. International research (including the updated published Cochrane review which included trials, approximately 90% of which were conducted in high income countries) has shown that well-designed and well-implemented school obesity prevention interventions were effective in the reduction of BMI in children [116, 117]. A systematic review of the effectiveness of preventive school-based obesity interventions in LMICs has demonstrated that interventions which focused on combining dietary and physical activity initiatives were effective in the reduction of BMI in children [124].

A few Chinese studies have demonstrated that comprehensive school-based interventions, which combined diet and physical activity, were effective [15-17]. However, the findings from the CHIRPY DRAGON trial are at variance with recently completed, similar and well conducted trials in a UK context, one of which used similar development and evaluation methodology [118-120]. These UK trials found no evidence for the effectiveness of school based prevention interventions. This highlights the importance of taking into account country 'context' in determining intervention effectiveness.

Generalisability of the Findings

The intervention would probably be generalisable to other urban areas of China given the centrally managed education system and similar cultures across the country, although some degree of local adaptions might be beneficial. However, the generalisability of the findings to rural and migrant children would be challenging.

Given the differences in contextual factors (including differences in the stage of the childhood obesity epidemic, dissimilar national health care systems), cultural factors and intervention differences (e.g. target, components and how these were delivered), a translation of economic findings from China to another country seems challenging and of limited use. For example, there might be implications for how to replicate any specific components of the intervention in another context. The CHIRPY DRAGON intervention targeted individual health behaviours which are highly dependent on cultural, infrastructural and other system-related aspects. For example, there is a more hierarchical structure and respect for schools and teachers in China compared to high income countries. So the generalisability of the results to other contexts, particularly from China to developed country settings, could be questionable [247].

There have been rapid socioeconomic and nutritional transitions in many urban Chinese populations which have contributed to the rising prevalence of overweight and obesity among Chinese school-aged children over a short period of time. Although the prevalence of obesity in rural areas of China has also increased over the past decades, it is not as high as in urban areas and there are growing economic and health inequities between urban and rural residents [57]. The CHIRPY DRAGON population may not be entirely representative of the general Chinese population in terms of different factors such as socio-economic status.

According to a recent qualitative study, lack of influence from grandparents; fewer opportunities for unhealthy snacking; and less pressure for academic attainment, which leads to more active play, were found as potential "protective" factors for obesity among migrant children [61]. However, lack of parental monitoring after school and unsafe neighbourhoods reduced physical activity in migrant communities. Two further barriers which restricted child physical activity in the migrant community were limited home space and cultural differences, which inhibited interactive play with local children [61]. Understanding the perceived contributors of obesity can provide valuable insights to plan or modify preventive interventions in different populations of China.

Short Term versus Long Term Cost-Effectiveness Evidence

This thesis undertook a trial-based economic evaluation which demonstrates short-term evidence and the analysis is constrained by the one year time horizon of the intervention. A longer time horizon is particularly relevant to the evaluation of obesity prevention interventions, where health benefits and cost savings may be visible into adulthood. Having a short time horizon can potentially underestimate differences between the intervention and control groups. Trial-based economic evaluations may give either an under- or overestimation and might therefore be of limited use. Whilst the intervention appears to be cost-effective and able to obtain benefits for both clinical and economic outcomes in children for a relatively low cost, the sustainability of these behaviours remains unknown.

However, this is one of the very few economic evaluations of obesity prevention studies worldwide and the first in China, which collected utility-based HRQoL information in children as young as 6 years and family members to calculate QALYs, and included societal costs. The intent was to open the debate for new ways of tackling the obesity issue in children.

In the context of obesity, it seems important to model prevention over a long time horizon (preferably life time). This is because interventions to prevent obesity have a wider effect and impact on costs over time and the outcomes may only be realised long after the trial has finished. To capture the long term effects of an intervention, and to evaluate whether the effects and the cost-effectiveness are sustainable in the long term a model should be used. Model-based economic evaluation can capture the uncertainty linked to any assumptions made and could also improve the generalization of results obtained in one setting to other settings [174]. Model-based health economic evaluations are today widely accepted as policy-making tools that can inform resource allocation decisions. However, there are some severe flaws with regard to current model-based economic evaluations in the field of childhood obesity. Firstly, the risk of obesity changes with individuals' attitudes and lifestyles and the sustainability of these behaviours are unknown. Secondly, findings from modelling studies are only as good as the data input, and high-quality input data for the costs and effects in the field of childhood obesity prevention intervention are sparse. Thirdly, to be able to extrapolate the effects over a life-time horizon, data on benefit maintenance are needed from early childhood to adulthood.

In general, the limitations of models are that they are a simplification of reality. Therefore, when a model developer translates reality into a series of definitive pathways there will always be some loss of information. One of the biggest challenges with modelling complex interventions is finding the balance between incorporating all relevant interactions and pathways, without overcomplicating the model structure. According to Squires and Boyd (2019), there are five key challenges with regard to modelling public health interventions: (1) incorporating equity; (2) extrapolating multi-component intervention effectiveness beyond study data; (3) modelling behaviour of individuals; (4) capturing relevant complex relationships of a complex system; and (5) capturing relevant non-health costs and benefits and the relationship between human and social determinants [303]. However, there are two approaches which could help to address these issues: (1) adopting an iterative approach to the evaluation, using early-stage decision modelling, to guide primary data collection; (2) using a conceptual modelling framework to guide the model development process [303].

Modelling was not applied as it was beyond the scope of this thesis.

7.4 Reflection on the Economic Evaluation Methods Taken

Economic evaluation should be conducted with respect to guidelines on methodology. This following section summarises reflections on the methodological approaches taken within the thesis for conducting economic evaluation.

Perspective

Where possible, recent literature advocates for a societal approach [47], given the nature of obesity and the public health strategies used to prevent/treat it. One of the most important strengths of this thesis is that it has attempted a societal perspective.

By using a societal perspective, a number of methodological challenges were encountered. As household members' resource use and outcome data are very rarely considered in economic evaluations, guidelines for including these data were limited. A further limitation was that the number of responses from household members was smaller than the number of children. Additionally, there was a lack of information on the salaries of parents/grandparents. Instead, the Chinese population average salary was assumed in order to estimate the value of lost time and therefore it is possible parental productivity losses might have been either under- or overestimated. Furthermore, the multiplier approach was not applied to the resource use data as it was not possible to link each component of the family-related costs to the respective child.

The research in this thesis demonstrates the feasibility of collecting and including household members' cost and outcome data in cost utility analysis. In this case the intervention did not impact significantly on household members' health, but inclusion of household spillovers may make a difference in other contexts.

Population

Doing research in children has some extra challenges. One potential challenge relates to the way HRQoL information was collected from children. There may have been an influence on how children completed the questionnaire as items and possible responses within the CHU-9D and PedsQL were read to children, on a one-to-one basis, by the interviewers (research staff). This could have led to responder-bias [42]. Moreover, because of the lack of validated dietary assessment instruments for Chinese children, tools developed and validated for English children were adapted [304]. However, similar adapted tools to assess dietary behaviours of Chinese children were previously used in the same city [305].

Measures of Effectiveness and Type of Economic Evaluation

Economic evaluations of obesity prevention interventions should be conducted using both clinical (e.g. BMI z-score) and economic (QALYs) outcomes.

Clinical outcomes are easier to measure, however harder to compare to other interventions across health and non-health settings. CEA is useful to compare interventions, which target the same health condition, and is particularly useful in a clinical setting. Although a CEA is an extra-welfarist evaluation, a major limitation of it for decision makers is its inability to directly compare cost-effectiveness of interventions across various areas of health conditions, or sectors of the economy, due to the disease-specific nature of the outcome measure used [46].

The QALY measure is universal. Therefore, various programmes across different health conditions which are evaluated using QALYs can be compared. Although, the ‘reference case’ approach, applied to a traditional HTA, takes an extra-welfarist perspective using outcomes expressed in QALYs to maximise health subject to a budget constraint [30], this cost utility approach offers limited support for public health decision makers. This is partly for considering health as the only relevant outcome, and ignoring the production of wider benefits which are not captured in the QALY. Also, there might be a lack of sensitivity of utility-based HRQoL instruments to changes in overweight/obesity in younger children [42]. Although there is no gold standard for measuring utility-based HRQoL in primary school-aged children, previous research has shown the CHU-9D, a recently developed instrument, is the most appropriate choice [38]. As a utility-based instrument, it is preference-based. It is designed for application in cost-effectiveness analyses of prevention, treatment and service programmes targeted at young people where the QALY is the desired outcome measure [183].

Researchers in the obesity field should not rely on a single measure and should use both clinical and economic measures in order to strengthen the evidence [47].

Resource Use and Costs

Public health priorities vary from country to country, and also from region to region. Like many other countries, China suffers from a scarcity of public health resources and decision makers need to prioritise spending towards policies that offer the greatest value for money [23]. Prevalence of overweight and obesity is high therefore large amounts of resources need to be invested in treatment and prevention interventions. Given the scarce public resources, economic evaluations are needed to aid decision-makers in prioritising and determining how and where to get the best value for money. Economic evaluations are well-established and fairly straightforward for therapeutic interventions. However, for preventive public health interventions (e.g. obesity prevention) the use of economic evaluations is more complex, where the vast majority of costs and consequences fall in the future and outside the health care sector.

To capture all possible costs and effects, data from various sources (e.g. clinical, epidemiological, and economic) are needed. The development of a toolbox including good-practice guidelines for intervention developers and evaluators would be useful in order to collect relevant data alongside the trial. While benefits of interventions are often captured in a standardized way, resource use data is usually collected using non-standardised resource use questionnaires which are difficult to compare with each other because of the heterogeneity of cost categories.

All methods for estimating costs should explicitly be documented and reported, like this study, to provide information that will inform future evaluations and policy making. In the CHIRPY DRAGON trial, the approach taken regarding the costing of the programme was thorough. The costs linked to the intervention were divided into three categories: development, implementation and delivery. According to standard practice, the base case analysis assumed that the intervention was in a 'running state' and therefore only costs associated with the delivery of the intervention were included. All of the other costs (set-up and implementation) were, however, reported separately and implementation costs were fully explored within the sensitivity analysis. A single standardized form was used to record all working time spent by each CHIRPY DRAGON teacher on the various intervention activities including their administration time. It has been determined that obesity prevention interventions are more effective when delivered by dedicated staff rather than classroom teachers [281]. The staff employed to deliver the intervention in this trial were well accepted by schools and their costs were incorporated in the economic evaluation. All costs were converted into either UK pounds or US dollars using GDP PPPs. All unit costs were sourced relevant to a Chinese context. However, more data are needed regarding indirect costs that already occur in childhood (e.g. educational attainment) [26].

7.5 Implications for Policy Making

The decision of what interventions to fund falls to the policy-maker. Evidence-based decision making has been advocated and evidence is available in terms of the implications of each choice in order to make the right decision. However, there is uncertainty regarding these decisions. Also, the choice requires value judgements, which may be weak or strong, depending on the available evidence.

The literature suggests school-based interventions delivered in high income country settings are cost-effective [47]. The cost-effectiveness results reported in this thesis, using both QALYs and clinical outcome (BMI z-score) as a measure of effectiveness, showed that this intervention was highly cost-effective in preventing obesity in China. Including societal costs and effects, using the decision-maker recommended EQ-5D-3L, increased the ICER, however the intervention remained cost-effective using established cost-effectiveness thresholds. Moreover, the sensitivity analyses presented in this thesis showed that the intervention was cost-effective.

Such cost-effective results should prompt decision-makers in China to take action towards developing and implementing childhood obesity prevention interventions in schools. Important considerations should focus on what should be offered.

The results of chapter 6 of this thesis have methodological and policy implications in terms of how the cost-effectiveness of childhood obesity interventions is measured in children aged 6-7 years. Obesity prevention and treatment interventions tend to target young populations, therefore information about how weight status is associated with HRQoL in utility terms in this age group is useful for the design of economic evaluations. Within health economic studies conducted globally, utility values are often used to derive QALYs to inform resource allocation decisions. To help inform the methods of economic evaluations alongside clinical trials of childhood obesity prevention and treatment interventions, future studies need to determine the relationship between weight status and utility-based HRQoL in different age groups, and across different country settings. The findings of chapter 6 provides support for the discriminant and convergent validity of the CHU-9D within Chinese children aged 6-7 years.

7.6 Future Research Recommendations

The research developed within this thesis makes a valuable and novel contribution to the existing literature of obesity research. However, it also serves to outline the following recommendations for future research.

As has been highlighted in the literature, obesity can be prevented/treated through "behavioural", "behavioural and environmental" and "behavioural and policy" interventions. Most countries have implemented "behavioural" interventions. However, "behavioural and policy" interventions are encouraged.

The research in this thesis demonstrates the feasibility of collecting and including household members' cost and outcome data in cost utility analysis in LMICs. However, as household members' resource use and outcome data are very rarely considered in economic evaluations, guidelines for how to include these data are limited. For preventive public health interventions (e.g. obesity prevention) the use of a societal perspective in economic evaluations is more complex compared to health perspective, and the vast majority of consequences and therefore costs prevented fall in the future. To capture all possible societal costs and effects, data from various sources (e.g. clinical, epidemiological, and economic) are needed. The development of a toolbox including good-practice guidelines for intervention developers and evaluators would be useful in order to collect relevant data alongside trials. All methods for estimating societal costs should explicitly be documented and reported to provide information that will inform future evaluations and policy making. In the CHIRPY DRAGON trial, the approach taken regarding the costing of the programme was thorough. As economic evaluation is uncommon in LMICs and due to the lack of an equivalent threshold value for most of these settings, established threshold values for a QALY alongside GDP per capita threshold should be used to judge cost-effectiveness.

Currently, there are no thresholds for obesity-related outcomes. This may be an area for future research. Also, a threshold value for how much decision makers are willing to pay for a unit gain in a QALY in a Chinese setting may be an area for future research. In addition, more effort should be placed on the inclusion of QALY measures for children in trial-based evaluations worldwide and spill-over effects (e.g. towards siblings) need to be further investigated worldwide. Furthermore, more research is needed on the indirect (e.g. educational attainment) costs of obesity/overweight in childhood in China and worldwide.

In general, the methodological challenges of conducting an economic evaluation of an obesity prevention intervention in LMICs (e.g. converting costs, sourcing unit costs, lack of an appropriate value set, dealing with clustering, lack of equivalent threshold value for a QALY) and including spillover costs/effects (e.g. measuring household costs and outcome data) need to be further investigated.

Future research on obesity intervention could benefit from taking a CBA approach. Obesity is a complex issue that involves socio-demographic determinants such as age, ethnicity, geography, lifestyle and religious/cultural traditions. To consider broader outcomes going beyond health and to account for inequalities, efforts are being made to adapt methodologies within the health economics community [245]. Consideration of broader outcomes going beyond the health sector allows for inclusion of costs and effects from multiple sectors and is particularly relevant for obesity intervention. This is an emerging area of development within economic evaluation and efforts are being made to adapt methodologies to promote the use of CBA [245] and to account for non-health opportunity costs in CUA [28]. These approaches have been recommended by the UK Treasury guidance to evaluate (usually non-health) public sector projects [246].

Future economic evaluations of childhood obesity prevention interventions should adopt a longer time horizon, because health benefits and cost savings may manifest themselves even into adulthood.

Moreover, considering the previous point regarding extrapolating over the long term, it would be valuable to undertake a suitable model to make full use of all routinely collected data as well as data generated from experimental settings.

The association between weight status and HRQoL in different age groups, and across different country settings needs to be further investigated. In addition, it is recommended that future studies aiming to prevent obesity in young children (age 6-12) do not rely solely on utility-based HRQoL measures for economic evaluation, and capture clinical or wellbeing outcomes as well. Future studies also need to further test the validity of the CHU-9D or other similar utility-based paediatric instruments in different age groups and country settings using different tariff value sets.

The CHIRPY DRAGON study used dedicated (rather than school) staff. This helped to maximise the consistency and quality of implementation as Chinese school teachers are often overloaded and, in this context, the extra workload would not be welcomed or sustainable. The results of the economic evaluation included staff training and employment costs and still showed evidence of cost effectiveness. The sensitivity analyses and other discussion points suggest the evidence of effectiveness and cost-effectiveness is likely to be generalizable to a wider set of school settings in China. These findings can help inform Chinese obesity policy and the education sector. However, in terms of understanding the impact of the intervention on multiple outcomes, a more disaggregated analysis (e.g. a CCA) alongside a CEA and CUA could have been undertaken. This would give a 'list' rather than offset against the costs. Thus decision makers would understand how the benefits are distributed across the different sectors (health and education) and this could then act as a tool to facilitate cross-sectoral decision making. Using a CCA could also give decision makers confidence that the data are valid to use as the basis for resource allocation decisions.

Multi-sectoral interventions which focus on the different root causes of obesity (e.g. neighbourhood influence) are necessary. These need a shift in responsibilities and a stronger political commitment in the fight against obesity. Although schools are an ideal setting to deliver population-based interventions, these types of interventions might not be sufficiently intense to impact the school, the family environment, and the weight status of children - as was experienced in a few recent studies from Europe (e.g. UK). Although the CHIRPY DRAGON prevention intervention was cost-effective, this could change as China becomes more developed and the obesity epidemic more established. Over time, school based interventions may no longer be sufficient, as is seen increasingly in higher income settings. Therefore, focus on more upstream determinants of obesity and using whole systems approaches to complex public health issues like obesity, as well as realistic reviews of the literature, may be a good way of conducting future research in China and worldwide. It needs to be acknowledged that complex public health research faces various challenges and therefore requires consideration of different research methods.

7.7 Conclusion

This thesis estimated the cost-effectiveness of the ‘CHIRPY DRAGON’ obesity prevention intervention in Chinese primary school-aged children from both a public sector and societal perspective, and raised relevant questions to be addressed in future research. The results of the economic evaluation from a public sector perspective showed that this intervention was highly cost-effective. Including societal costs and effects increased the ICER, however the intervention remained cost-effective using established cost-effectiveness thresholds.

HRQoL using both CHU-9D and PedsQL instruments was slightly higher among children who were overweight/obese compared to underweight/healthy weight which highlights the need for future research that takes account of cultural context. The findings of this thesis support the discriminant and convergent validity of the CHU-9D, as a measure of utility-based HRQoL for application in economic evaluation of prevention interventions within Chinese children aged 6-7 years. However, future studies need to further test the validity of the CHU-9D in different age groups and country settings using different tariff value sets.

The research has made several original and pertinent contributions to the literature on obesity research, both in terms of the methodological approach and also for the information made available to inform policy makers. Given that the obesity crisis persists, new approaches may be needed, and further research required to prevent obesity in children.

APPENDICES

APPENDIX 1. Systematic Review of Economic Evaluations: Search Strategy, Drummond Checklist, Data Extraction, and Quality Assessment of Studies

Appendix 1.1 - Search Strategy

MEDLINE (Ovid)

1. exp Obesity/
2. Obese.mp.
3. exp Overweight/
4. (BMI or body mass index).af.
5. Weight gain/
6. (Overweight or over weight or obesity or adipose).af.
7. exp Child/
8. exp Infant/
9. (Child* or adolescen* or infant*).af.
10. Schoolchild*.mp.
11. exp Adolescent/
12. (Boys or girls or youth or youths).af.
13. (Teenage* or young person).af.
14. (Nutrition adj2 intervent*).af.
15. (Obesity adj2 prevent* or treat*).af.
16. Counsel?ing.mp.

17. exp support groups/
18. exp Health Behaviour.mp.
19. exp Life Style/
20. exp Delivery of Health Care/
21. exp Social Support/
22. exp Family Practice/
23. exp Parent-Child Relations/
24. Food Habits .mp.
25. exp Diet therapy/
26. exp Food Preferences/
27. exp Exercise therapy/
28. Physical activit*.mp.
29. Economic Evaluat*.mp.
30. Cost*.ti.
31. Cost?Benefit*.mp.
32. Cost?Utilit*.mp.
33. Cost?Effective*.mp.
34. exp "costs and cost analysis"/
35. 1 or 2 or 3 or 4 or 5 or 6
36. 7 or 8 or 9 or 10 or 11 or 12 or 13
37. 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28
38. 29 or 30 or 31 or 32 or 33 or 34
39. 35 and 36 and 37 and 38
40. Limit 39 to (yr=”2001-Current”)

EMBASE (Ovid)

1. exp Obesity/
2. Obese.mp.
3. exp Overweight/
4. (BMI or body mass index).af.
5. Weight gain/
6. (Overweight or over weight or obesity or adipose).af.
7. exp Child/
8. exp Infant/
9. (Child* or adolescen* or infant*).af.
10. Schoolchild*.mp.
11. exp Adolescent/
12. (Boys or girls or youth or youths).af.
13. (Teenage* or young person).af.
14. (Nutrition adj2 intervent*).af.
15. (Obesity adj2 prevent* or treat*).af.
16. Counsel?ing.mp.
17. exp support groups/
18. exp Health Behaviour.mp.
19. exp Life Style/
20. exp Delivery of Health Care/
21. exp Social Support/
22. exp Family Practice/
23. exp Parent-Child Relations/

24. Food Habits .mp.

25. exp Diet therapy/

26. exp Food Preferences/

27. exp Exercise therapy/

28. Physical activit*.mp.

29. Economic Evaluat*.mp.

30. Cost*.ti.

31. Cost?Benefit*.mp.

32. Cost?Utilit*.mp.

33. Cost?Effective*.mp.

34. exp "costs and cost analysis"/

35. 1 or 2 or 3 or 4 or 5 or 6

36.7 or 8 or 9 or 10 or 11 or 12 or 13

37. 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28

38. 29 or 30 or 31 or 32 or 33 or 34

39. 35 and 36 and 37 and 38

40. Limit 39 to (yr=”2001-Current”)

PsycINFO

1. exp Obesity/

2. Obese.mp.

3. exp Overweight/

4. (BMI or body mass index).af.

5. Weight gain/

6. (Overweight or over weight or obesity or adipose).af.
7. exp Child/
8. exp Infant/
9. (Child* or adolescen* or infant*).af.
10. Schoolchild*.mp.
11. exp Adolescent/
12. (Boys or girls or youth or youths).af.
13. (Teenage* or young person).af.
14. (Nutrition adj2 intervent*).af.
15. (Obesity adj2 prevent* or treat*).af.
16. Counsel?ing.mp.
17. exp support groups/
18. exp Health Behaviour.mp.
19. exp Life Style/
20. exp Delivery of Health Care/
21. exp Social Support/
22. exp Family Practice/
23. exp Parent-Child Relations/
24. Food Habits .mp.
25. exp Diet therapy/
26. exp Food Preferences/
27. exp Exercise therapy/
28. Physical activit*.mp.
29. Economic Evaluat*.mp.

30. Cost*.ti.

31. Cost?Benefit*.mp.

32. Cost?Utilit*.mp.

33. Cost?Effective*.mp.

34. exp "costs and cost analysis"/

35. 1 or 2 or 3 or 4 or 5 or 6

36. 7 or 8 or 9 or 10 or 11 or 12 or 13

37. 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28

38. 29 or 30 or 31 or 32 or 33 or 34

39. 35 and 36 and 37 and 38

40. Limit 39 to (yr=”2001-Current”)

Web of Science

1. TS= (Obesity OR obese OR overweight)
2. TS= (Child OR infant OR schoolchild* OR adolescent)
3. TS= (Interven* OR prevent* OR therapeutics OR counseling OR “primary health care” OR “preventive health services” OR “health behaviour” OR “life style” OR “health knowledge, practice, attitudes” OR “delivery of health care” OR “social support” OR “family practice” OR “parent-child relations” OR “food habits” OR “food preferences” OR exercise OR sports)
4. TS= (“Economic evaluat*” OR costs* OR “cost?benefit*” OR “cost?utilit*” OR “cost?effective*”)
5. #1 AND #2 AND #3 AND #4 Timespan 2001-2017

CINAHL Plus

S1. (MH “Obesity+”)

S2. “obese”

S3. “overweight”

S4. (MH "Child+")

S5. (MH "Infant+")

S6. "schoolchild*"

S7. "adolescent"

S8. "Interven*"

S9. "prevent*"

S10. (MH "Therapeutics+")

S11. "counseling"

S12. (MH "Primary Health Care")

S13. "preventive health services"

S14. (MH "Health Behavior")

S15. (MH "Life Style+")

S16. "health knowledge, practice, attitudes"

S17. "delivery of health care"

S18. "social support"

S19. (MH "Family Practice")

S20. (MH "Parent-Child Relations")

S21. (MH "Food Habits")

S22. (MH "Food Preferences")

S23. (MH "Exercise+")

S24. (MH "Sports+")

S25. "Economic evaluat*"

S26. "costs*"

S27. "cost?benefit*"

S28. “cost?utilit*”

S29. “cost?effective*”

S30. S1 OR S2 OR S3

S31. S4 OR S5 OR S6 OR S7

S32. S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24

S33. S25 OR S26 OR S27 OR S28 OR S29

S34. S30 AND S31 AND S32 AND S33

S35. Limit S34 to Publication Year: 2001-2017

EconLit

S1. (MH “Obesity+”)

S2. “obese”

S3. “overweight”

S4. (MH "Child+")

S5. (MH "Infant+")

S6. "schoolchild*"

S7. "adolescent"

S8. "Interven*"

S9. "prevent*"

S10. (MH "Therapeutics+")

S11. "counseling"

S12. (MH "Primary Health Care")

S13. "preventive health services"

S14. (MH "Health Behavior")

S15. (MH "Life Style+")

S16. "health knowledge, practice, attitudes"

S17. "delivery of health care"

S18. "social support"

S19. (MH "Family Practice")

S20. (MH "Parent-Child Relations")

S21. (MH "Food Habits")

S22. (MH "Food Preferences")

S23. (MH "Exercise+")

S24. (MH "Sports+")

S25. "Economic evaluat*"

S26. "costs*"

S27. "cost?benefit*"

S28. “cost?utilit*”

S29. “cost?effective*”

S30. S1 OR S2 OR S3

S31. S4 OR S5 OR S6 OR S7

S32. S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24

S33. S25 OR S26 OR S27 OR S28 OR S29

S34. S30 AND S31 AND S32 AND S33

S35. Limit S34 to Publication Year: 2001-2017

CRD (DARE, NHS EED, HTA)

1. MeSH DESCRIPTOR Obesity EXPLODE ALL TREES
2. (Obese) OR (Overweight): any field
3. MeSH DESCRIPTOR Child EXPLODE ALL TREES
4. MeSH DESCRIPTOR Infant EXPLODE ALL TREES
5. (Schoolchild*): any field
6. MeSH DESCRIPTOR Adolescent EXPLODE ALL TREES
7. (Interven*) OR (prevent*): any field
8. MeSH DESCRIPTOR Therapeutics EXPLODE ALL TREES
9. MeSH DESCRIPTOR Counseling EXPLODE ALL TREES
10. MeSH DESCRIPTOR Primary Health Care EXPLODE ALL TREES
11. MeSH DESCRIPTOR Preventive Health Services EXPLODE ALL TREES
12. MeSH DESCRIPTOR Health Behavior EXPLODE ALL TREES
13. MeSH DESCRIPTOR Life Style EXPLODE ALL TREES
14. (Health knowledge, practice, attitudes): any field
15. MeSH DESCRIPTOR Delivery of Health Care EXPLODE ALL TREES
16. MeSH DESCRIPTOR Social Support EXPLODE ALL TREES
17. MeSH DESCRIPTOR Family Practice EXPLODE ALL TREES
18. MeSH DESCRIPTOR Parent-Child Relations EXPLODE ALL TREES
19. (Food Habits): any field

20. MeSH DESCRIPTOR Food Preferences EXPLODE ALL TREES

21. MeSH DESCRIPTOR Exercise EXPLODE ALL TREES

22. MeSH DESCRIPTOR Sports EXPLODE ALL TREES

23. MeSH DESCRIPTOR Costs and Cost Analysis EXPLODE ALL TREES

24. MeSH DESCRIPTOR Economics EXPLODE ALL TREES

25. (Cost) OR (Economic): any field

26. (#1 or #2) and (#3 or #4 or #5 or #6) and (#7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22) and (#23 or #24 or #25) From 2001-2017

CENTRAL and CDSR

1. "MeSH descriptor: [Obesity] explode all trees
2. Obese or overweight: ti, ab.kw (Word variations have been searched)
3. "MeSH descriptor: [Child] explode all trees
4. "MeSH descriptor: [Infant] explode all trees
5. Schoolchild*: ti, ab.kw (Word variations have been searched)
6. "MeSH descriptor: [Adolescent] explode all trees
7. Interven* or prevent*: ti, ab.kw (Word variations have been searched)
8. "MeSH descriptor: [Therapeutics] explode all trees
9. "MeSH descriptor: [Counseling] explode all trees
10. "MeSH descriptor: [Primary Health Care] explode all trees
11. "MeSH descriptor: [Preventive Health Services] explode all trees
12. "MeSH descriptor: [Health Behavior] explode all trees
13. "MeSH descriptor: [Life Style] explode all trees

14. Health knowledge, practice, attitudes: ti, ab.kw (Word variations have been searched)
15. "MeSH descriptor: [Delivery of Health Care] explode all trees
16. "MeSH descriptor: [Social Support] explode all trees
17. "MeSH descriptor: [Family Practice] explode all trees
18. "MeSH descriptor: [Parent-Child Relations] explode all trees
19. "MeSH descriptor: [Food Habits] explode all trees
20. "MeSH descriptor: [Food Preferences] explode all trees
21. "MeSH descriptor: [Exercise] explode all trees
22. "MeSH descriptor: [Sports] explode all trees
23. "MeSH descriptor: [Costs and Cost Analysis] explode all trees
24. "MeSH descriptor: [Economics] explode all trees
25. Cost or economic: ti, ab.kw (Word variations have been searched)
26. (#1 or #2) and (#3 or #4 or #5 or #6) and (#7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22) and (#23 or #24 or #25) From 2001-2017, in other reviews or economic evaluations

Appendix 1.2 - Drummond Checklist for Critically Appraising Relevant Studies

Drummond checklist for assessing primary economic evaluations				
Study design				
1 The research question is stated				
2 The economic importance of the research question is stated				
3 The viewpoint (s) of the analysis are clearly stated and justified				
4 The rationale for choosing alternative programmes or interventions compared is stated				
5 The alternatives being compared are clearly described				
6 The form of economic evaluation used is stated				
7 The choice of form of economic evaluation is justified in relation to the questions addressed				
Data collection				
8 The source (s) of effectiveness estimates used are stated				
9 Details of the design and results of effectiveness study are given (if based on a single study)				
10 Details of the methods of synthesis or meta-analysis of estimates are given (if based on a synthesis of a number of effectiveness studies)				
11 The primary outcome measure (s) for the economic evaluation are clearly stated				
12 Methods to value benefits are stated				
13 Details of the subjects from whom valuations were obtained were given				
14 Productivity changes (if included) are reported separately				
15 The relevance of productivity changes to the study question is discussed				
16 Quantities of resource use are reported separately from their unit costs				
17 Methods for the estimation of quantities and unit costs are described				
18 Currency and price data are recorded				
19 Details of currency of price adjustments for inflation or currency conversion are given				
20 Details of any model used are given				
21 The choice of model used and the key parameters on which it is based are justified				

Analysis and interpretation of results				
22 Time horizon of costs and benefits is stated				
23 The discount rate (s) is stated				
24 The choice of discount rate (s) is justified				
25 An explanation is given if costs and benefits are not discounted				
26 Details of statistical tests and confidence intervals are given for stochastic data				
27 The approach to sensitivity analysis is given				
28 The choice of variables for sensitivity analysis is justified				
29 The ranges over which the variables are varied are justified				
30 Relevant alternatives are compared				
31 Incremental analysis is reported				
32 Major outcomes are presented in a disaggregated as well as aggregated form				
33 The answer to the study question is given				
34 Conclusions follow from the data reported				
35 Conclusions are accompanied by the appropriate caveats				

Notes: Y= Yes; N=No; NC= Not clear; N.A. = Not applicable

Appendix 1.3 (i) - 1.3 (iv) - Data Extraction (Details about Study Context)

Appendix 1.3 (i) - Details about study context (trial-based prevention studies) (alphabetically sorted)

Authors	Year	Country	Study design	Setting	Target population/ age group	N (analytical sample)	Parents/ guardians included	Intervention overview /target	Intervention aim /mode of delivery	Comparator
Hayes et al. [229]	2014	Australia	RCT	Home	Up to age 2 years, boys/girls from a mixed-weight group	324 parents with infants	Yes: Not for indirect and direct non-medical costs	8 one-to-one consultations with education and advice on feeding, nutrition and physical activity	Prevention /nurse	Usual care, plus home safety information sent by mail
Kesztyus et al. [235]	2011	Germany	RCT	School	7-8 years, boys/girls from a mixed-weight group	945 children	Yes: parents involved but not costed	28 units, health education, physical activity breaks	Prevention /teacher	Usual care
Krauth et al. [236]	2013	Germany	Cohort	School	6-10 years, boys/girls from a mixed-weight group	660 children	Yes	3 additional lessons per week regarding physical activity	Prevention /teacher	Usual care
Martinez et al. [237]	2011	Spain	RCT	School	9-10 years, boys/girls from a mixed-weight group	1,409 children	Yes: parents involved but not costed	3 sessions, school-based physical activity program	Prevention /teacher	Usual care
McAuley et al. [227]	2010	New Zealand	RCT	School-community	5–12 years, boys/girls from a mixed-weight group	279 children	Yes: parents involved but not costed	A pilot program for nutrition and physical activity	Prevention /activity coordinator	Usual care

Appendix 1.3 (i) - Details about study context (trial-based prevention studies) (alphabetically sorted) continued

Authors	Year	Country	Study design	Setting	Target population/ age group	N (analytical sample)	Parents/ guardians included	Intervention overview /target	Intervention aim /mode of delivery	Comparator
Meng et al. [17]	2013	China	RCT	School	6-13 years, boys/girls from a mixed-weight group	8, 301 children	Yes: parents involved but not costed	6 times, nutrition education for children, parents and teachers, physical activity intervention and comprehensive	Prevention /teacher	Usual care
Peterson et al. [228]	2008	USA	Cross-sectional	State	12–18 years, boys/girls from a mixed-weight group	3,782 adolescents	No	The get up and do something media campaign (Television and/or billboards) for physical activity	Prevention (policy) /media	Usual care
Sutherland et al. [206]	2016	Australia	RCT	School-community	13-16 years, boys/girls from a mixed-weight group	1,150 adolescents	Yes: Not for indirect and direct non-medical costs	Seven physical activity promotion strategies and six additional strategies	Prevention /trained teacher	Usual care
Wang et al. [115]	2008	USA	RCT	School (after hours)	6-10 years, boys/girls from a mixed-weight group	182 children	Yes	After school environment program: physical activity, healthy snacks	Prevention /coordinator	Usual care

Notes: RCT = randomised controlled trial

Appendix 1.3 (ii) - Details about study context (trial-based treatment studies) (alphabetically sorted)

Authors	Year	Country	Study design	Setting	Target population/ age group	N (analytical sample)	Parents/ guardians included	Intervention overview /target	Intervention aim /mode of delivery	Comparator
Epstein et al. [230]	2014	USA	RCT	Primary care	8-12 years, boys/girls with obesity/ overweight	50 children with parents	Yes	Family-based behavioural treatment, 15 sessions (12 weekly, 2 biweekly and 1 monthly): diet, physical activity and behaviour change for both treatment groups	Treatment /staff	Separate group treatment (parent and child)
Goldfield et al. [231]	2001	Canada	RCT	Primary care	8-12 years, boys/girls with obesity	24 children with parents	Yes: Not for indirect and direct non-medical costs	Group treatment, 13 sessions (8 weekly, 4 bi-weekly, and 1 monthly): diet, physical activity and behaviour change for both treatment groups	Treatment /counselling degree	Mixed family-based behavioural treatment
Hollinghurst et al. [232]	2013	UK	RCT	Primary care/ home/ hospital	5-16 years and 9-17 years, boys/girls with obesity	143 children and adolescents	Yes: Parents involved but not costed	Every 3 months, nurse-led, input from dietitian and exercise specialist and an intensive intervention Mandometer	Treatment /doctor, nurse, exercise specialist, dietitian	Hospital, Consultant-led care with discretionary input from dietitian and exercise specialist

Appendix 1.3 (ii) - Details about study context (trial-based treatment studies) (alphabetically sorted) continued

Authors	Year	Country	Study design	Setting	Target population/ age group	N (analytical sample)	Parents/ guardians included	Intervention overview /target	Intervention aim /mode of delivery	Comparator
Janicke et al. [233]	2009	USA	RCT	Community	8-14 years, boys/girls with obesity/ overweight	76 children	Yes: Not for indirect and direct non-medical costs	Parent only behavioural intervention, group sessions: weekly: 8, bi-weekly: 4/ diet and physical activity for both treatment groups	Treatment /post-doctoral psychologist and graduate students in clinical psychology	Family-based behavioural intervention,
Kalavainen et al. [238]	2009	Finland	RCT	Primary care	7–9 years, boys/girls with obesity	70 children with parents	Yes: Not for indirect and direct non-medical costs	Routine counselling treatment, 2 appointments for children: diet	Treatment /nurses, nutritionists	Group treatment, 15 separate sessions for parents and children: diet
Robertson et al. [205]	2017	UK	RCT	NHS primary care	6-11 years, boys/girls with obesity/ overweight	128 children with 137 parents/ carers	Yes	1 per week, parenting skills, social and emotional development/ physical activity and diet	Treatment /intervention team	Usual care
Wake et al. [234]	2008	Australia	RCT	Primary care	5–9 years, boys/girls with obesity/ overweight	163 children with parents	Yes	Training of GP (3 times 2.5 h), 4 consultations over a 12-week period/physical activity, diet	Treatment /GP	Usual care

Appendix 1.3 (iii) - Details about study context (model-based prevention studies) (alphabetically sorted)

Authors	Year	Country	Study design	Setting	Target population/ age group	N (analytical sample)	Parents/ guardians included	Intervention overview /target	Intervention aim /mode of delivery	Comparator
Barrett et al. [211]	2015	USA	Cohort	State's school	6–11 years, boys/girls from a mixed-weight group	17.6 million children	No	Active physical education policy (Active PE)	Prevention (policy) /teachers	Usual care
Brown et al. [212]	2007	USA	Cohort	School	8-11 years, Boys/girls from a mixed-weight group	423 children	No	Physical education, school food service modification, family- and home-based program	Prevention /teacher, trainer	Usual care
Carter et al. [196]	2009	Australia	RCT	School	7–11 years, boys/girls from a mixed-weight group	595, 000 children over 5 years (119,000 each year)	Yes: Parents involved but not costed	Education programme to reduce sugar sweetened drink consumption	Prevention /trained project staff	Usual care
Carter et al. [196]	2009	Australia	Cohort	School	6 years, boys/girls from a mixed-weight group	114, 630 children	Yes: Parents involved but not costed	Education to improve nutrition and physical activity, with an active physical education	Prevention /teacher	Usual care

Appendix 1.3 (iii) - Details about study context (model-based prevention studies) (alphabetically sorted) continued

Authors	Year	Country	Study design	Setting	Target population/ age group	N (analytical sample)	Parents/ guardians included	Intervention overview /target	Intervention aim /mode of delivery	Comparator
Carter et al. [196]	2009	Australia	RCT	School	8–10 years, boys/girls from a mixed-weight group	268, 600 children	Yes: parents involved but not costed	Education programme to reduce television viewing of snacks	Prevention /teacher	Usual care
Carter et al. [196]	2009	Australia	Cohort	School	6 years, boys/girls from a mixed-weight group	114, 630 children	Yes: parents involved but not costed	Education to improve nutrition and physical activity, without an active physical education	Prevention /teacher	Usual care
Graziose et al. [207]	2016	USA	RCT	School	10-11 years, boys/girls from a mixed-weight group	769 children	No	24 lessons, obesity prevention nutrition education	Prevention /trained teacher	Usual care
Long et al. [213]	2015	USA	Cohort	State	2-19 years, boys/girls from a mixed-weight group	74 million children	No	Sugar-sweetened beverage excise tax/diet	Prevention (policy) /government, industry	Usual care
Magnus et al. [214]	2009	Australia	RCT	State	5-14 years, boys/girls from a mixed-weight group	2.4 million children	Yes: parents involved but not costed	Removing TV advertising of energy-dense nutrition-poor (EDNP) food and beverages	Prevention (policy) /media	Usual care

Appendix 1.3 (iii) - Details about study context (model-based prevention studies) (alphabetically sorted) continued

Authors	Year	Country	Study design	Setting	Target population/ age group	N (analytical sample)	Parents/ guardians included	Intervention overview /target	Intervention aim /mode of delivery	Comparator
Moodie et al. [215]	2009	Australia	Cohort	School-community	5-7 years, boys/girls from a mixed-weight	7, 840 children	Yes	Walking school bus program, encouraging physical activity	Prevention /volunteer conductors	Usual care
Moodie et al. [216]	2010	Australia	Cohort	School (after hours)	5-11 years, boys/girls from a mixed-weight	99, 000 children	Yes	Active after-school communities, physical activity	Prevention /organizations	Usual care
Moodie et al. [217]	2011	Australia	Cohort	School-community	10–11 years, boys/girls from a mixed-weight	267, 700 children	Yes: parents involved but not costed	Travel smart school, promotion of physical activity	Prevention /teachers	Usual care
Moodie et al. [218]	2013	Australia	Quasi-experimental,	School-community	4-12 years, boys/girls from a mixed-weight	2, 184 children	Yes	The be active eat well program, diet and physical activity	Prevention /community service	Usual care
Pringle et al. [219]	2010	UK	_	Community	10-17 years, boys/girls from a mixed-weight	343 children and adolescents	Yes: Not for indirect and direct non-medical costs	Free swimming activities, campaigns, exercise classes, motivational interviews	Prevention /trainer	Other interventions
Rush et al. [220]	2014	New Zealand	RCT	School	6-8 years and 9-11 years, boys/girls from a mixed-weight	2, 474 younger and 2, 330 older children	Yes: parents involved but not costed	Project Energize: Multicomponent physical activity and nutrition	Prevention /organizations	Usual care

Appendix 1.3 (iii) - Details about study context (model-based prevention studies) (alphabetically sorted) continued

Authors	Year	Country	Study design	Setting	Target population/ age group	N (analytical sample)	Parents/ guardians included	Intervention overview /target	Intervention aim /mode of delivery	Comparator
Sonneville et al. [221]	2015	USA	Cohort	State	2-19 years, boys/girls from a mixed-weight group	74 million Children and adolescents	No	Elimination of the tax subsidy of TV advertising costs for nutritionally poor foods and beverages advertised	Prevention (policy) /industry	Usual care
Wang et al. [222]	2003	USA	RCT	School	10-14 years, girls from a mixed-weight group	620 children	No	Lessons, sport materials, wellness, teacher training, targeting diet and physical activity, reduction of TV viewing time	Prevention/teacher, trainer	Usual care
Wang et al. [223]	2011	USA	RCT	School	10-14 years, girls from a mixed-weight group	480 children	No	Lessons, sport materials, wellness, teacher training, targeting diet and physical activity	Prevention/teacher, trainer	Usual care
Wright et al. [224]	2015	USA	Cohort	State	2.5-5 years, boys/girls from a mixed-weight group	3.7 million children	No	Early care and education policy change, physical activity, diet and reduction of TV viewing time	Prevention (policy) /Child care trainers	Usual care

Notes: RCT = randomised controlled trial

Appendix 1.3 (iv) - Details about study context (model-based treatment studies) (alphabetically sorted)

Authors	Year	Country	Study design	Setting	Target population/ age group	N (analytical sample)	Parents/ guardians included	Intervention overview /target	Intervention aim /mode of delivery	Comparator
Carter et al. [196]	2009	Australia	Cohort	School	7–10 years, boys/girls with obesity/ overweight	17, 000 children over 4 years, (4 200 each year)	Yes: Parents involved but not costed	Multifaceted targeted programme, diet and physical activity	Treatment/ teacher	Usual care
Carter et al. [196]	2009	Australia	RCT	Primary Care	10–11 years, boys/girls with obesity	5, 800 children	Yes: Parents involved but not costed	Primary care-based program, children with obesity and their parents, diet	Treatment/ GPs, paediatricians, psychologists, dietitians	Usual care
Hollingworth et al. [225]	2012	UK	RCT	Hospital-community	4-5 years and 10-11 years, boys/girls with obesity/ overweight	9, 956 younger and 9, 698 older children	Yes: Parents involved but not costed	Interventions aimed at modifying behaviour, diet and/or physical activity	Treatment/GP, paediatricians, nurse, exercise specialist, dietitian	Usual care or minimal intervention
Moodie et al. [226]	2008	Australia	RCT	Primary Care	5–9 years, boys/girls with obesity/ overweight	9, 685 children	Yes	Training of GP (3 times 2.5 h), 4 consultations over a 12-week period/physical activity, diet	Treatment/GP	Usual care

Notes: GP = general practitioner; RCT = randomised controlled trial

Appendix 1.4 (i) - 1.4 (iv) - Data Extraction (Detailed Account of the Economic Evaluation Methods A)

Appendix 1.4 (i) - Detailed account of the economic evaluation methods A (trial-based prevention studies) (alphabetically sorted)

Authors and year	Measures of effectiveness/ study type	Type of modelling approach	Study perspective	Duration of intervention/ follow-up	Time horizon	Price year	Currency unit	Discount rate
Hayes et al. 2014 [229]	Reduction in BMI (z-score), unit BMI avoided/CEA	N.A.	Health care funder	2 years/ 2 years after randomisation	2 years	2012	AU$	Costs: 5% Effects: 5%
Kesztyus et al. 2011 [235]	Reduction in BMI, cm WC and unit WHtR prevented/CEA	N.A.	Societal	1 year/ 1 year after randomisation	1 year	2008	€	N.A.
Krauth et al. 2013 [236]	Reduction in BMI, (increase in physical activity: the measure was not specified)/CCA	N.A.	Societal	4 years/ 4, 5 and 6 years after intervention	6 years	No price year	€	Not stated
Martinez et al. 2011 [237]	Percent point decrease in triceps skinfold thickness and body fat/CEA	N.A.	Societal and institutional	8 months/ 8 months after randomisation	8 months	2005	€	N.A.
McAuley et al. 2010 [227]	Reduction in BMI (z-score), cm WC prevented, weight gain prevented, HRQoL/CEA	N.A.	Societal	2 years/ 2 and 4 years after intervention	4 years	2006	NZ$	Costs at 5% Effects: Not reported
Meng et al. 2013 [17]	Reduction in BMI, BMI (z-score), overweight and obesity case avoided/CEA	N.A.	Social	1 year/ 1 year after randomisation	1 year	2010	RMB/ US$	N.A.
Peterson et al. 2008 [228]	(Increase in physical activity: the measure was not specified)/CEA	N.A.	Not specified	Not reported	Not reported	No price year	US$	N.A.
Sutherland et al. 2016 [206]	MVPA (min/day) gained, BMI unit avoided and reduction in BMI (z-score)/CEA	N.A.	Societal	2 years/ 1 year (mid-intervention), 2 years after	2 years	2014	AU$	Not stated
Wang et al. 2008 [115]	Reduction in body fat/CEA	N.A.	Societal	1 year/ 1 year after randomisation	1 year	2003	US$	N.A.

Appendix 1.4 (ii) - Detailed account of the economic evaluation methods A (trial-based treatment studies) (alphabetically sorted)

Authors and year	Measures of effectiveness/ study type	Type of modelling approach	Study perspective	Duration of intervention/ follow-up	Time horizon	Price year	Currency unit	Discount rate
Epstein et al. 2014 [230]	BMI change and weight for children and parents/CEA	N.A.	Societal (payer plus participant costs)	1 year/ 1 year after randomisation	1 year	No price year	US$	N.A.
Goldfield et al. 2001 [231]	Reduction in BMI (z-score) and percentage overweight/CEA	N.A.	Not specified	6 months/ 6 and 12 months after randomisation	1 year	No price year	$	N.A.
Hollinghurst et al. 2013 [232]	Reduction in BMI sd/CEA	N.A.	Healthcare	1 year/ 1 year after randomisation	1 year	No price year	£	N.A.
Janicke et al. 2009 [233]	Reduction in BMI/CEA	N.A.	Not specified	4 months/ 4 and 10 months after randomisation	10 months	No price year	US$	N.A.
Kalavainen et al. 2009 [238]	Reduction in weight for height and BMI/CEA	N.A.	Service provider (healthcare)	6 months/ 6 and 12 months after randomisation	1 year	2004	€	N.A.
Robertson et al. 2017 [205]	Reduction in waist z-score, body fat, WC, MVPA (min/day) gained, change in BMI (z-score) and QALYs gained/CEA, CUA	N.A.	NHS and PSS (healthcare)	3 months/ 3 and 12 months after randomisation	1 year	2013, 2014	£	N.A.
Wake et al. 2008 [234]	Reduction in BMI, parent-reported physical activity and dietary habits/CCA	N.A.	Societal	9 months/ 9 and 15 months after randomisation	15 months	2003	AU$	Not stated

Notes: BMI = body mass index; CCA = cost-consequence analysis; CEA = Cost-effectiveness analysis; CUA = cost-utility analysis; QALYs = quality-adjusted life years; MVPA = moderate-to-vigorous physical activity; NHS = National Health Service; PSS = personal social services; WC = waist circumference; N.A. = not applicable

Appendix 1.4 (iii) -Detailed account of the economic evaluation methods A (model-based prevention studies) (alphabetically sorted)

Authors and year	Measures of effectiveness/ study type	Type of modelling approach	Study perspective	Duration of intervention/ follow-up	Time horizon	Price year	Currency unit	Discount rate
Barrett et al. 2015 [211]	Reduction in BMI and obesity-related healthcare expenditure, increase in minutes of MVPA and MET-hours/CEA	Markov model	Societal	2 years/ 2 years	10 years	2014	US$	Costs: 3% Effects: 3%
Brown et al. 2007 [212]	Cases of adult overweight prevented, QALYs saved/CUA	Decision tree	Societal	3 years/	25 years From age 40 to 64	2004	US$	Costs: 3% Effects: 3%
Carter et al. 2009 (4 the same) [196]	BMI unit saved, DALYs saved/CUA	Markov model	Societal	1 year/	Lifetime	2001	AU$	Costs: 3% Effects: 3%
Graziose et al. 2016 [207]	Reduction in adult obesity, QALYs saved/CUA	Decision tree	Societal	1 year/	Lifetime	2012	US$	Costs: 3% Effects: 3%
Long et al. 2015 [213]	Changes in BMI, reductions in disease burden and healthcare expenditures, DALYs averted and QALYs gained/CEA, CUA	Markov model	Societal	2 years/ 2 years	10 years	2014	US$	Costs: 3% Effects: 3%
Magnus et al. 2009 [214]	BMI unit saved, DALYs saved/CEA, CUA	Markov model	Societal	1 year/	Lifetime	2001	AU$	Costs: 3% Effects: 3%
Moodie et al. 2009 [215]	BMI unit saved, DALYs saved, increase in physical activity (MET) and energy expenditure/CEA, CUA	Markov model	Societal	1 year/	Lifetime	2001	AU$	Costs: 3% Effects: 3%
Moodie et al. 2010 [216]	BMI unit saved, DALYs saved, increase in physical activity (MET) and energy expenditure/CEA, CUA	Markov model	Societal	1 year/	Lifetime	2001	AU$	Costs: 3% Effects: 3%

Appendix 1.4 (iii) - Detailed account of the economic evaluation methods A (model-based prevention studies) (alphabetically sorted) continued

Authors and year	**Measures of effectiveness/ study type**	**Type of modelling approach**	**Study perspective**	**Duration of intervention/ follow-up**	**Time horizon**	**Price year**	**Currency unit**	**Discount rate**
Moodie et al. 2011 [217]	BMI unit saved, DALYs saved, increase in physical activity (MET) and energy expenditure/CEA, CUA	Markov model	Societal	1 year/	Lifetime	2001	AU$	Costs: 3% Effects: 3%
Moodie et al. 2013 [218]	Reduction in BMI, DALYs saved/CEA, CUA	Markov model	Societal	3 years/	Lifetime	2006	AU$	Costs: 3% Effects: 3%
Pringle et al. 2010 [219]	Change in MPA, QALYs saved/CEA, CUA	Decision tree	Not specified	Not reported	Not reported	2003	£	N.A.
Rush et al. 2014 [220]	Reduction in BMI, QALYs saved, increased life expectancy/CUA	Markov model	Health treatment payer (Health care)	2 years/ 5 years	Lifetime	2011	NZ$	Costs: 3.5% Effects: 3.5%
Sonneville et al. 2015 [221]	Reduction in BMI, reductions in disease burden, healthcare expenditures and QALYs gained/CEA, CUA	Markov model	Societal	2 years/ 2 years	10 years	2014	US$	Costs: 3% Effects: 3%
Wang et al. 2003 [222]	Cases of adult overweight prevented, QALYs saved/CUA	Decision tree	Societal	2 years/	25 years From age 40 to 65	1996	US$	Costs: 3% Effects: 3%
Wang et al. 2011 [223]	DWCB avoided, QALYs saved/CUA	Decision tree	Societal	2 years/	10 years	2010	US$	Costs: 3% Effects: 3%
Wright et al. 2015 [224]	Unit BMI avoided, reduction in obesity-related healthcare expenditure/CEA	Markov model	Societal	2 years/ 2 years	10 years	2014	US$	Costs: 3% Effects: 3%

Notes: BMI = body mass index; CEA = Cost-effectiveness analysis; DALYs = disability-adjusted life years; DWCB = disordered weight control behaviours; QALYs = quality-adjusted life years; MVPA = moderate-to-vigorous physical activity; MPA = moderate physical activity; N.A. = not applicable

Appendix 1.4 (iv) - Detailed account of the economic evaluation methods A (model-based treatment studies) (alphabetically sorted)

Authors and year	Measures of effectiveness/ study type	Type of modelling approach	Study perspective	Duration of intervention/ follow-up	Time horizon	Price year	Currency unit	Discount rate
Carter et al. 2009 (2 the same) [196]	BMI unit saved, DALYs saved/CUA	Markov model	Societal	1 year/	Lifetime	2001	AU$	Costs: 3% Effects: 3%
Hollingworth et al. 2012 [225]	Reduction in BMI sd, life year gained/CEA	Markov model	NHS (healthcare)	1 year/	Lifetime	2009	£	Costs: 3% Effects: 3%
Moodie et al. 2008 [226]	BMI unit saved, DALYs saved/CEA, CUA	Markov model	Societal	1 year/	Lifetime	2001	AU$	Costs: 3% Effects: 3%

Notes: BMI = body mass index; CEA = Cost-effectiveness analysis; CUA = cost-utility analysis; DALYs = disability-adjusted life years; NHS = National Health Service; NHF = National Health Forum

Appendix 1.5 (i) - 1.5 (iv) - Data Extraction (Detailed Account of the Economic Evaluation Methods B)

Appendix 1.5 (i) - Detailed account of the economic evaluation methods B (trial-based prevention studies) (alphabetically sorted)

Authors and year	Methods for estimating/collecting resource use	Cost categories	Largest cost drivers	Excluded costs	Average costs per participant	Funding source
Hayes et al. 2014 [229]	Local health district records, patient-level data linkage	Programme delivery, direct medical	Hospitalisation and doctor visits	Research and development, birth, evaluation or administration of the clinical trial	AU$ 1, 309	Academic
Kesztyus et al. 2011 [235]	Official statistics of the state of Bavaria	Programme delivery	Scientific coordinator	Development, scientific evaluation, classroom time	€ 24.09	Academic
Krauth et al. 2013 [236]	Questionnaire, school admin	Programme delivery, indirect	Training	Not stated	€ 619	Academic
Martinez et al. 2011 [237]	Not stated	Programme delivery, labour	Personnel (coordinator)	Parents' care costs	€ 269.83	Academic
McAuley et al. 2010 [227]	Not stated	Programme delivery	coordinator	Research and development, planning phase, time costs of the children and their parents	NZ$ 1, 281	Academic
Meng et al. 2013 [17]	Not stated	Programme delivery, labour, money, evaluation	Materials	Not stated	Combined: RMB 182.4 (US$ 26.8), nutrition: RMB 52.8 (US$ 7.8), PA: RMB 52.3 (US$ 7.7)	Academic

Appendix 1.5 (i) - Detailed account of the economic evaluation methods B (trial-based prevention studies) (alphabetically sorted) continued

Authors and year	Methods for estimating/collecting resource use	Cost categories	Largest cost drivers	Excluded costs	Average costs per participant	Funding source
Peterson et al. 2008 [228]	Not stated	Development, media production and placement	Not stated	Not stated	Per person to become more active: Individual sections: US$ 5.11- 153.19 Whole: US$ 8.87	Not stated
Sutherland et al. 2016 [206]	Using market rates, Australian Bureau of Statistics, Industrial Relations Commission of NSW/project records	Programme delivery	Consultant	Research and development, potential effects on healthcare costs	AU$ 394	Academic
Wang et al. 2008 [115]	Not stated	Programme delivery, usual after-school care costs without intervention, indirect	Personnel	Not stated	US$ 956	Academic

Appendix 1.5 (ii) - Detailed account of the economic evaluation methods B (trial-based treatment studies) (alphabetically sorted)

Authors and year	Methods for estimating/collecting resource use	Cost categories	Largest cost drivers	Excluded costs	Average costs per participant	Funding source
Epstein et al. 2014 [230]	Tracking and recording by staff members, Google maps calculations	Programme delivery, direct medical, direct non-medical, indirect	Treatment time	Recruitment	Cost per family: FBT: US$ 1, 448, PC-1: US$ 2, 260, PC-2: US$ 2, 124	Academic
Goldfield et al. 2001 [231]	Not stated	Programme delivery, direct medical	Salary	Reduced cost of medical care, purchasing new clothes, time costs for being physically active	Cost per family: group treatment: US$ 491 Mixed treatment: US$ 1, 390	Academic
Hollinghurst et al. 2013 [232]	Patient-level data linkage	Programme delivery, direct medical	Mandometer device	Development of the Mandometer and staff training	Mandometer group: £ 1, 749 (SD £ 243), primary care group: £ 301 (SD £76), hospital groups: £ 263 (SD £ 88) and £ 209 (SD £ 81)	Academic
Janicke et al. 2009 [233]	Not stated	Programme delivery, direct medical	Group leaders	Research (assessment, recruitment), participants (travel, purchasing healthier foods)	Family-based group: US$ 872, Parent-only: US$ 521	Academic

Appendix 1.5 (ii) - Detailed account of the economic evaluation methods B (trial-based treatment studies) (alphabetically sorted) continued

Authors and year	Methods for estimating/collecting resource use	Cost categories	Largest cost drivers	Excluded costs	Average costs per participant	Funding source
Kalavainen et al. 2009 [238]	Not stated	Programme delivery, labour, direct medical	Labour	Research component, participating families	Cost per family: group treatment: € 327, routine counselling: € 61	Academic
Robertson et al. 2017 [205]	Questionnaire and secondary national tariff sets	Programme delivery, direct medical, indirect	Hospital visits, salary (GP)	Not stated	£ 998	Academic
Wake et al. 2008 [234]	3 main sources: the Live, Eat and Play (LEAP) team records, practice audit, and parent written questionnaires at 9 months	Programme delivery, direct medical, direct non-medical, indirect	Practice	Set-up, research and development, training	AU$ 705	Academic

Appendix 1.5 (iii) - Detailed account of the economic evaluation methods B (model-based prevention studies) (alphabetically sorted)

Authors and year	Methods for estimating/collecting resource use	Cost categories	Largest cost drivers	Excluded costs	Average costs per participant	Funding source
Barrett et al. 2015 [211]	Beta, normal or uniform distribution form, different databases (school administrators, interventions, survey)	Programme delivery, avoided direct medical	Sets (nationally) of active PE curricula and equipment	Start-up	US$ 4.03	Academic
Brown et al. 2007 [212]	Not stated	Programme delivery, avoided direct medical, avoided indirect (productivity loss)	Promotional	Not stated	US$ 104	Academic
Carter et al. 2009 [196]	Not stated	Programme delivery	Not stated	Set-up	AU$ 28	Academic
Carter et al. 2009 [196]	Not stated	Programme delivery	Not stated	Set-up, Teacher classroom time	AU$ 473	Academic
Carter et al. 2009 [196]	Not stated	Programme delivery	Not stated	Set-up, Teacher classroom time	AU$ 103	Academic
Carter et al. 2009 [196]	Not stated	Programme delivery	Not stated	Set-up, Teacher classroom time	AU$ 211	Academic
Graziose et al. 2016 [207]	New York City Department of Education (NYCDOE) and author estimate	Programme delivery, future obesity-related medical, avoided direct medical	Teacher preparation time	Development and evaluation	US$ 111	Academic

Appendix 1.5 (iii) - Detailed account of the economic evaluation methods B (model-based prevention studies) (alphabetically sorted) continued

Authors and year	Methods for estimating/collecting resource use	Cost categories	Largest cost drivers	Excluded costs	Average costs per participant	Funding source
Long et al. 2015 [213]	Beta, normal or uniform distribution form, different databases (interventions, revenue department, bureau of labour statistics 2013)	Programme delivery, labour, avoided direct medical	Industry auditor salary	Not stated	US$ 0.68	Academic
Magnus et al. 2009 [214]	Not stated	Programme delivery, other sectors	Government regulators	Set-up	AU$ 0.54	Academic
Moodie et al. 2009 [215]	Middle of Australian public service Level 6, Australian bureau of statistics and Victorian department of education and training	Programme delivery, direct non-medical, indirect, other sectors,	Education e.g. programme coordinator	Set-up, research and development implementation	AU$ 2, 908	Academic
Moodie et al. 2010 [216]	Middle of Australian public service level 6 and Victorian department of education and training	Programme delivery, indirect, other sectors	Sport and recreation	Set-up, research and development, implementation, external evaluation and maintenance	AU$ 488.5	Academic
Moodie et al. 2011 [217]	Middle of Australian public service level 6, Victorian department of education and training	Programme delivery, other sectors	Education e.g. Central coordinator	Set-up, research and development	AU$ 49.68	Academic
Moodie et al. 2013 [218]	Australian bureau of statistics, Victorian department of education and training, etc.	Programme delivery, direct non-medical, indirect, other sectors	Personnel time	Student time, spin-off activities, changes in the physical activity and eating patterns of participating families	AU$ 344	Academic
Pringle et al. 2010 [219]	Not stated	Programme delivery,	Primary care referral	Not stated	_	Academic

Appendix 1.5 (iii) - Detailed account of the economic evaluation methods B (model-based prevention studies) (alphabetically sorted) continued

Authors and year	Methods for estimating/collecting resource use	Cost categories	Largest cost drivers	Excluded costs	Average costs per participant	Funding source
Rush et al. 2014 [220]	Not stated	Programme delivery, avoided direct medical	Not stated	Set-up and development, indirect, out-of-pocket,	NZ$ 44.96	Academic
Sonneville et al. 2015 [221]	Normal or beta distribution form, different databases (bureau of labour statistics 2013, etc)	Programme delivery, labour, avoided direct medical	Industry auditor salary	Not stated	US$ 0.015	Academic
Wang et al. 2003 [222]	Not stated	Programme delivery, avoided direct medical, avoided indirect (productivity loss)	Subject teachers	Classroom time	US$ 28	Academic
Wang et al. 2011 [223]	Not stated	Programme delivery, avoided direct medical	Subject teachers	Not stated	US$ 184.27	Academic
Wright et al. 2015 [224]	Different databases (bureau of labour statistics 2013, etc)	Programme delivery, avoided direct medical	Supervising and training	Not stated	US$ 1.29	Academic

Appendix 1.5 (iv) - Detailed account of the economic evaluation methods B (model-based treatment studies) (alphabetically sorted)

Authors and year	Methods for estimating/collecting resource use	Cost categories	Largest cost drivers	Excluded costs	Average costs per participant	Funding source
Carter et al. 2009 [196]	Not stated	Programme delivery	Not stated	Set-up	AU$ 129	Academic
Carter et al. 2009 [196]	Not stated	Programme delivery	Not stated	Set-up,	AU$ 1,896	Academic
Hollingworth et al. 2012 [225]	Not stated	Programme delivery, lifetime treatment, obesity-related diseases	Salary (GP)	Not stated	£108 - 662	Academic
Moodie et al. 2008 [226]	Middle of Australian public service Level 6, LEAP trial, etc.	Programme delivery, direct medical, direct non-medical, indirect	Project coordinator	Set-up, research and development, resultant changes in patient behaviour	AU$ 650.5	Academic

Appendix 1.6 (i) - 1.6 (iv) - Data Extraction (Detailed Account of the Economic Evaluation Methods C)

Appendix 1.6 (i) - Detailed account of the economic evaluation methods C (trial-based prevention studies) (alphabetically sorted)

Authors and year	ICER/average cost per benefit	Uncertainty analysis	Sensitivity analysis type	Sensitivity analysis	Cost-effective
Hayes et al. 2014 [229]	AU$ 4, 230 per unit BMI avoided, AU$ 631 per 0.1 reduction in BMI (z-score)	_	DSA	Adjustments in nurse travel time	Likely to be
Kesztyus et al. 2011 [235]	€ 11.11 per WC cm prevented € 18.55 per WHtR unit prevented	_	DSA	Teachers individual working time to prepare the lessons, difference in effects tested at a 10, 20 and 30% lower value	Likely to be
Krauth et al. 2013 [236]	N.A.	_	DSA	_	N.A.
Martinez et al. 2011 [237]	ICER: No € 500 per 1% decrease in triceps skinfold thickness	_	DSA	Differences in costs (modification of the venue cost)	Likely to be
McAuley et al. 2010 [227]	ICER: No NZ$ 664–1708 per kg of weight gain prevented (depending on age),	_	DSA	Differences in weight z-score (ranged from 0.5 to 1.0 in the youngest children and 1.4 to 2.4 in the oldest children)	Likely to be
Meng et al. 2013 [17]	Combined intervention: US$ 120.3 per 1 kg/m2 BMI reduction, US$ 249.3 per BMI z-score change (BAZ), US$ 1308.9 per one overweight and obesity case avoided	_	Not stated	N.A.	Likely to be
Peterson et al. 2008 [228]	ICER: No Entire campaign: US$ 4.01: to see the ad, US$ 7.35: to consider being more active, US$ 8.87: actually become more active, with bill-boards the most cost-effectiveness	_	Not stated	N.A.	Likely to be

Appendix 1.6 (i) - Detailed account of the economic evaluation methods C (trial-based prevention studies) (alphabetically sorted) continued

Authors and year	ICER/average cost per benefit	Uncertainty analysis	Sensitivity analysis type	Sensitivity analysis	Cost-effective
Sutherland et al. 2016 [206]	AU$ 56 per additional minute of MVPA, AU$ 1 per MET hour gained per person per day, AU$ 1, 408 per BMI unit avoided, AU$ 563 per 10 % reduction in BMI z-score	_	DSA	Higher and lower estimate of the assumed opportunity cost, varying the magnitude of the effect size, extending the benefit of physical activity recess and lunchtime activities to students beyond the target year, extending the benefit of multiple strategies to all students Scenario: State wide rollout (current model), state wide roll out – Alternative (real world) model	Likely to be
Wang et al. 2008 [115]	US$ 317 per 0.76% body fat reduction	_	DSA	Changing the per capita usual after-school care costs (ranging from US$ 5.00 to US$ 10.00)	Likely to be

Notes: BAZ = BMI (z-score); BMI = body mass index; DSA = deterministic sensitivity analysis; MVPA = moderate-to-vigorous physical activity; WC = waist circumference; WHtR = waist-to-height ratio; N.A. = not applicabl

Appendix 1.6 (ii) - Detailed account of the economic evaluation methods C (trial-based treatment studies) (alphabetically sorted)

Authors and year	ICER/average cost per benefit	Uncertainty analysis	Sensitivity analysis type	Sensitivity analysis	Cost-effective
Epstein et al. 2014 [230]	ICER: No Children: FBT US$ 209.17 per % over BMI, PC1 US$ 1, 036.50 per % over BMI, PC2 US$ 973.98 per % over BMI, Parents: FBT US$ 132.97 per pound (lb), PC1 US$ 373.53 per pound (lb), PC2 US$ 351.00 per pound (lb)	_	Not stated	N.A.	Likely to be
Goldfield et al. 2001 [231]	ICER: No US$ 1, 000 per 10% overweight reduction US$ 1, 000 per 0.6 decrease in BMI z-score	_	Not stated	N.A.	Likely to be
Hollinghurst et al. 2013 [232]	£ 432 per 0.1 reduction in BMI sd	_	Not stated	N.A.	Likely to be
Janicke et al. 2009 [233]	ICER: No Family-based group: US$ 758 per 0.10 decrease in BMI z-score, Parent-only: US$ 579 per 0.10 decrease in BMI z-score	_	Not stated	N.A.	Likely to be
Kalavainen et al. 2009 [238]	€ 53 per 1% decrease in weight for height € 266 per 0.1 decrease in BMI	_	DSA	Group treatment costs: salaries of two group leaders included in costs	Likely to be

Appendix 1.6 (ii) - Detailed account of the economic evaluation methods C (trial-based treatment studies) (alphabetically sorted) continued

Authors and year	ICER/average cost per benefit	Uncertainty analysis	Sensitivity analysis type	Sensitivity analysis	Cost-effective
Robertson et al. 2017 [205]	£ 552, 175 per QALY saved, £ - 3, 935 per unit change in BMI (z-score)	_	DSA	'programme completers': families that participated in 5 or more sessions, multiple imputation of all missing cost and outcomes data, alternative sources and inputs for EQ-5D utility values	Unlikely to be
Wake et al. 2008 [234]	N.A.	_	DSA	Baseline: Value of parents' time, equal parent's time, unit cost of GP visit, economies of scale Combinations:	N.A.

Notes: BMI = body mass index; DSA = deterministic sensitivity analysis; FBT = family-based behavioural treatment; QALYs = quality-adjusted life years; PC = parent and child; N.A. = not applicabl

Appendix 1.6 (iii) - Detailed account of the economic evaluation methods C (model-based prevention studies) (alphabetically sorted)

Authors and year	**ICER/average cost per benefit**	**Uncertainty analysis**	**Sensitivity analysis type**	**Sensitivity analysis**	**Cost-effective**
Barrett et al. 2015 [211]	US$ 401 per unit BMI avoided: 2 years US$ 1, 720 per BMI unit reduced: 10 years Reduction of healthcare costs by $ 60.5 million: 10 years	PSA	DSA, PSA	Physical activity and BMI changes, more PE time, cost of intervention	Likely to be
Brown et al. 2007 [212]	US$ 900 per QALY saved	_	PSA	Both overall and Hispanics (Cases of adult overweight prevented, QALYs saved, medical costs averted, costs of lost labour productivity averted	Likely to be
Carter et al. 2009 [196]	AU$ 5, 000 per DALY saved	PSA	DSA, PSA	_	Dominant
Carter et al. 2009 [196]	AU$ 1, 800 per DALY saved	PSA	DSA, PSA	_	Dominant
Carter et al. 2009 [196]	AU$ 5, 100 per DALY saved	PSA	DSA, PSA	_	Dominant
Carter et al. 2009 [196]	AU$ 5, 600 per DALY saved	PSA	DSA, PSA	_	Likely to be
Graziose et al. 2016 [207]	US$ 275 per QALY saved	PSA	DSA, PSA	Relapse into adulthood, intervention is effective only for Hispanic and black students, intervention is effective only for male students	Likely to be

Appendix 1.6 (iii) - Detailed account of the economic evaluation methods C (model-based prevention studies) (alphabetically sorted) continued

Authors and year	ICER/average cost per benefit	Uncertainty analysis	Sensitivity analysis type	Sensitivity analysis	Cost-effective
Long et al. 2015 [213]	US$ 8.54 per unit BMI avoided: 2 years,	PSA	DSA, PSA	Change in SSB consumption and BMI, cost of implementing SSB excise tax	Dominant
Magnus et al. 2009 [214]	AU$ 5.00 per BMI unit saved, AU$ 3.70 per DALY saved	PSA	DSA, PSA	BMI and cost changes	Dominant
Moodie et al. 2009 [215]	AU$ 87, 000 per BMI unit saved, AU$ 760, 000 per DALY saved	PSA	DSA, PSA	Reduce costs, improve capacity utilisation and recruitment, increase participants receiving benefit, combine scenarios	Unlikely to be
Moodie et al. 2010 [216]	AU$ 8, 200 per BMI unit saved, AU$ 82, 000 per DALY saved	PSA	DSA, PSA	Reduction in the number of sites and co-ordinators, application of the same wage rate to all site co-ordinators (school, OSHC), combination scenarios, all participants receive full intervention benefit	Unlikely to be
Moodie et al. 2011 [217]	AU$ 13, 000 per BMI unit saved, AU$ 117, 000 per DALY saved	PSA	DSA, PSA	joint cost attribution across multiple objectives, broadening of the benefit to include other children in the school	Unlikely to be
Moodie et al. 2013 [218]	AU$ 576 per BMI unit saved, AU$ 29, 798 per DALY saved	PSA	DSA, PSA	Alternative decay of effect, if only 50% of children received the benefit	Likely to be
Pringle et al. 2010 [219]	ICER: No £ 47 - 509 per QALY gained, £ 260 - 2, 786 per completer improving at least one MPA	_	Not stated	N.A.	Dominant

Appendix 1.6 (iii) - Detailed account of the economic evaluation methods C (model-based prevention studies) (alphabetically sorted) continued

Authors and year	ICER/average cost per benefit	Uncertainty analysis	Sensitivity analysis type	Sensitivity analysis	Cost-effective
Rush et al. 2014 [220]	NZ$ 24, 690 per QALY saved: older children NZ$ 30, 438 per QALY saved: younger children	_	DSA, PSA	Varied conditions for younger and older children (varying of the cost of intervention, BMI, the annual discount rate and the horizon of the model)	Likely to be
Sonneville et al. 2015 [221]	US$ 1.16 per unit BMI avoided: 2 years,	PSA	DSA, PSA	Differences in BMI associated with the number of fast food advertising messages seen, cost of intervention	Dominant
Wang et al. 2003 [222]	US$ 4, 035 per QALY saved	_	DSA, PSA	Cases of adult overweight prevented, years of healthy life, annual discount rate, medical care costs averted, annual workdays lost averted	Dominant
Wang et al. 2011 [223]	US$ 2, 966 per QALY saved	PSA	DSA, PSA	Percentage of girls with DWCB who had SED, progression probability, long-term medical costs per BN patients, HRQoL of BN patients, time to recovery	Dominant
Wright et al. 2015 [224]	US$ 57.80 per BMI unit avoided: 2 years Net healthcare cost savings of $ 51.6 million: 10 years	PSA	DSA, PSA	Time spent in care, alternative policy adherence estimates and outcomes	Dominant

Notes: BMI = body mass index; BN = Bulimia Nervosa; DALYs = disability-adjusted life years; DSA = deterministic sensitivity analysis; DWCB = disordered weight control behaviours; HRQoL = health-related quality of life; QALYs = quality-adjusted life years; MPA = moderate physical activity; PE = physical education; PSA = probabilistic sensitivity analysis; SED = sub-diagnostic eating disorders; SSB = sugar sweetened beverage; N.A. = not applicable

Appendix 1.6 (iv) - Detailed account of the economic evaluation methods C (model-based treatment studies) (alphabetically sorted)

Authors and year	**ICER/average cost per benefit**	**Uncertainty analysis**	**Sensitivity analysis type**	**Sensitivity analysis**	**Cost-effective**
Carter et al. 2009 [196]	AU$ 3, 300 per DALY saved	PSA	DSA, PSA	_	Dominant
Carter et al. 2009 [196]	AU$ 1, 500 per DALY saved	PSA	DSA, PSA	_	Dominant
Hollingworth et al. 2012 [225]	£ 400 per 0.13 reduction in BMI sd, £ 13, 589 per life year gained	_	DSA	BMI sd (minimal, median, or maximal effect size) and intervention cost (low, moderate and high)	Dominant
Moodie et al. 2008 [226]	AU$ 4, 670 per DALY saved	PSA	DSA, PSA	Full maintenance of the BMI benefit into adulthood/vs. half maintenance, outlier removal, delivery of intervention, (family attendance, etc), recruitment rates	Likely to be

Notes: BMI = body mass index; DALYs = disability-adjusted life years; DSA = deterministic sensitivity analysis; PSA = probabilistic sensitivity analysis

Appendix 1.7 (i) - Critically appraising trial-based prevention studies (alphabetically sorted)

	Hayes et al. 2014 [229]	Kesztyus et al. 2011 [235]	Krauth et al. 2013 [236]	Martinez et al. 2011 [237]	McAuley et al. 2010 [227]	Meng et al. 2013 [17]	Peterson et al. 2008 [228]	Sutherland et al. 2016 [206]	Wang et al. 2008 [115]
1	Y	Y	Y	Y	Y	Y	Y	Y	Y
2	Y	Y	Y	Y	NC	Y	NC	Y	NC
3	Y	Y	Y	Y	Y	Y	Y	Y	Y
4	Y	NC	N	NC	N	N	N	NC	N
5	Y	Y	Y	Y	Y	Y	NC	Y	Y
6	Y	Y	Y	Y	Y	Y	Y	Y	Y
7	Y	N	N	N	Y	NC	Y	Y	NC
8	Y	Y	Y	Y	Y	Y	Y	Y	Y
9	Y	Y	Y	Y	Y	Y	NC	Y	Y
10	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
11	Y	Y	Y	Y	Y	Y	NC	Y	Y
12	N.A.	Y	Y	Y	Y	Y	N.A.	Y	Y
13	N.A.	Y	Y	Y	N	Y	N.A.	Y	N
14	N.A.	NC	N	NC	N	N	N.A.	NC	N
15	N.A.	N	N	N	N	N	N.A.	N	N
16	Y	Y	NC	Y	Y	N	N	Y	Y
17	Y	Y	NC	Y	NC	N	N	Y	NC
18	Y	Y	NC	Y	Y	Y	NC	Y	Y
19	Y	N	N.A.	N	Y	N	N	N	N
20	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
21	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
22	Y	Y	Y	Y	Y	Y	N	Y	Y
23	Y	N.A.	N.A.	N.A.	NC	N.A.	N.A.	N	N.A.
24	Y	N.A.	N.A.	N.A.	NC	N.A.	N.A.	N.A.	N.A.
25	N.A.	N.A.	N.A.	N.A.	N	N.A.	N.A.	N	N.A.
26	Y	Y	N	NC	Y	Y	N	Y	NC
27	Y	Y	Y	Y	Y	N	N	Y	Y
28	Y	N	N	N	NC	N.A.	N.A.	Y	NC
29	Y	N	N	N	NC	N.A.	N.A.	Y	NC
30	Y	Y	N	Y	Y	N.A.	N.A.	Y	Y
31	Y	Y	N.A.	N	N	Y	N	Y	Y
32	Y	Y	N.A.	N	N	Y	N	Y	Y
33	Y	Y	Y	Y	Y	Y	Y	Y	Y
34	Y	Y	Y	Y	Y	Y	Y	Y	Y
35	Y	Y	Y	Y	Y	Y	NC	Y	Y

Notes: Y= Yes; N=No; NC= Not clear; N.A. = Not applicable

Appendix 1.7 (ii) - Critically appraising trial-based treatment studies (alphabetically sorted)

	Epstein et al. 2014 [230]	Goldfield et al. 2001 [231]	Hollinghurst et al. 2013 [232]	Janicke et al. 2009 [233]	Kalavainen et al. 2009 [238]	Robertson et al. 2017 [205]	Wake et al. 2008 [234]
1	Y	Y	Y	Y	Y	Y	Y
2	Y	NC	Y	Y	Y	Y	Y
3	Y	Y	Y	Y	Y	Y	Y
4	Y	N	N	N	N	NC	N
5	Y	Y	Y	Y	Y	Y	NC
6	Y	Y	Y	Y	Y	Y	Y
7	NC	N	NC	N	Y	NC	Y
8	Y	Y	Y	Y	Y	Y	Y
9	Y	Y	Y	Y	Y	Y	Y
10	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
11	Y	Y	Y	Y	Y	Y	Y
12	Y	N.A.	N.A.	N.A.	N.A.	N.A.	Y
13	Y	N.A.	N.A.	N.A.	N.A.	N.A.	NC
14	NC	N.A.	N.A.	N.A.	N.A.	N.A.	NC
15	N	N.A.	N.A.	N.A.	N.A.	N.A.	N
16	N	NC	Y	Y	Y	NC	NC
17	N	N	Y	N	N	Y	NC
18	NC	NC	NC	NC	Y	Y	Y
19	N	N	N	N	N	Y	N.A.
20	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
21	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
22	Y	Y	Y	Y	Y	Y	Y
23	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
24	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
25	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
26	Y	Y	NC	NC	Y	Y	NC
27	N	N	N	N	Y	Y	Y
28	N.A.	N.A.	N.A.	N.A.	NC	Y	Y
29	N.A.	N.A.	N.A.	N.A.	NC	Y	Y
30	N.A.	N.A.	N.A.	N.A.	Y	Y	Y
31	N	N	Y	N	Y	Y	N.A.
32	N	N	Y	N	Y	Y	N.A.
33	Y	Y	Y	Y	Y	Y	Y
34	Y	Y	Y	Y	Y	Y	Y
35	Y	Y	Y	Y	Y	Y	Y

Notes: Y= Yes; N=No; NC= Not clear; N.A. = Not applicable

Appendix 1.7 (iii) - Critically appraising model-based prevention studies (alphabetically sorted)

	Barrett et al. 2015 [211]	Brown et al. 2007 [212]	Carter et al. 2009 [196]	Graziose et al. 2016 [207]	Long et al. 2015 [213]	Magnus et al. 2009 [214]	Moodie et al. 2009 [215]	Moodie et al. 2010 [216]
1	Y	Y	Y	Y	Y	Y	Y	Y
2	Y	NC	NC	Y	Y	Y	Y	N
3	Y	Y	Y	Y	Y	Y	Y	Y
4	N	N	N	Y	N	N	N	Y
5	Y	Y	Y	Y	Y	Y	Y	Y
6	Y	Y	Y	Y	Y	Y	Y	Y
7	NC	N	Y	Y	NC	N	N	N
8	Y	Y	Y	Y	Y	Y	Y	Y
9	Y	Y	Y	Y	Y	Y	Y	Y
10	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
11	Y	Y	Y	Y	Y	Y	Y	Y
12	Y	Y	Y	Y	Y	Y	Y	Y
13	NC	NC	NC	NC	NC	N	NC	NC
14	N	NC	N	NC	N	N	NC	NC
15	N	N	N	N	N	N	N	N
16	Y	NC	N	Y	Y	N	Y	NC
17	Y	NC	N	Y	Y	N	Y	Y
18	Y	Y	Y	Y	Y	Y	Y	Y
19	N	N	Y	N	N	N	Y	Y
20	Y	NC	NC	Y	Y	NC	NC	NC
21	NC	N	N	N	NC	N	N	N
22	Y	Y	Y	Y	Y	Y	Y	Y
23	Y	Y	Y	Y	Y	Y	Y	Y
24	N	N	Y	Y	N	N	Y	Y
25	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
26	Y	NC	Y	Y	Y	NC	Y	Y
27	Y	Y	Y	Y	Y	Y	Y	Y
28	Y	Y	N	N	N	N	Y	Y
29	Y	Y	N	N	N	N	Y	Y
30	Y	Y	N	Y	Y	Y	Y	Y
31	Y	Y	Y	Y	Y	Y	Y	Y
32	Y	Y	Y	Y	Y	Y	Y	Y
33	Y	NC	Y	Y	Y	Y	Y	Y
34	Y	Y	Y	Y	Y	Y	Y	Y
35	Y	NC	NC	Y	Y	Y	Y	Y

Notes: Y= Yes; N=No; NC= Not clear; N.A. = Not applicable

Appendix 1.7 (iii) - Critically appraising model-based prevention studies (alphabetically sorted) continued

	Moodie et al. 2011 [217]	Moodie et al. 2013 [218]	Pringle et al. 2010 [219]	Rush et al. 2014 [220]	Sonneville et al. 2015 [221]	Wang et al. 2003 [222]	Wang et al. 2011 [223]	Wright et al. 2015 [224]
1	Y	Y	Y	Y	Y	Y	Y	Y
2	N	Y	NC	Y	Y	NC	Y	Y
3	Y	Y	Y	Y	Y	Y	Y	Y
4	Y	Y	N	NC	N	N	N	N
5	Y	Y	NC	Y	Y	Y	Y	Y
6	Y	Y	Y	Y	Y	Y	Y	Y
7	Y	Y	NC	Y	NC	NC	NC	NC
8	Y	Y	Y	Y	Y	Y	Y	Y
9	Y	Y	NC	Y	Y	Y	NC	Y
10	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
11	Y	Y	NC	Y	Y	Y	Y	Y
12	Y	Y	N.A.	N.A.	Y	Y	Y	Y
13	NC	Y	N.A.	N.A.	NC	NC	NC	NC
14	NC	Y	N.A.	N.A.	NC	N	NC	N
15	N	Y	N.A.	N.A.	N	N	N	N
16	Y	Y	NC	N	Y	Y	Y	Y
17	Y	Y	NC	NC	Y	NC	NC	Y
18	Y	Y	Y	Y	Y	Y	Y	Y
19	Y	Y	N	N	N	N	N	N
20	NC	NC	NC	NC	Y	Y	NC	Y
21	N	N	N	N	NC	N	N	NC
22	Y	Y	N	Y	Y	Y	Y	Y
23	Y	Y	N	Y	Y	Y	Y	Y
24	Y	Y	N.A.	N	N	N	N	N
25	N.A.	N.A.	N	N.A.	N.A.	N.A.	N.A.	N.A.
26	Y	Y	N	Y	Y	Y	Y	Y
27	Y	Y	Y	Y	Y	Y	Y	Y
28	Y	NC	N	NC	N	Y	NC	Y
29	Y	NC	N	NC	N	Y	NC	Y
30	Y	Y	N.A.	Y	Y	Y	Y	Y
31	Y	Y	N	Y	Y	Y	Y	Y
32	Y	Y	N	Y	Y	Y	Y	Y
33	Y	Y	NC	Y	Y	NC	Y	Y
34	Y	Y	Y	Y	Y	Y	Y	Y
35	Y	Y	NC	Y	Y	Y	Y	Y

Notes: Y= Yes; N=No; NC= Not clear; N.A. = Not applicable

Appendix 1.7 (iv) - Critically appraising model-based treatment studies (alphabetically sorted)

	Carter et al. 2009 [196]	Hollingworth et al. 2012 [225]	Moodie et al. 2008 [226]
1	Y	Y	Y
2	NC	Y	Y
3	Y	Y	Y
4	N	N	Y
5	Y	Y	Y
6	Y	Y	Y
7	Y	NC	Y
8	Y	Y	Y
9	Y	Y	Y
10	N.A.	N.A.	N.A.
11	Y	Y	Y
12	Y	N.A.	Y
13	NC	N.A.	Y
14	N	N.A.	Y
15	N	N.A.	N
16	N	N	Y
17	N	N	Y
18	Y	Y	Y
19	Y	N	Y
20	NC	NC	NC
21	N	NC	N
22	Y	Y	Y
23	Y	Y	Y
24	Y	N	Y
25	N.A.	N.A.	N.A.
26	Y	NC	Y
27	Y	Y	Y
28	N	N	Y
29	N	N	Y
30	N	Y	Y
31	Y	Y	Y
32	Y	Y	Y
33	Y	Y	Y
34	Y	Y	Y
35	NC	Y	Y

Notes: Y= Yes; N=No; NC= Not clear; N.A. = Not applicable

APPENDIX 2. Economic Evaluation alongside CHIRPY DRAGON Trial: CHEERS Guidelines, Methods and Results Tables

Appendix 2.1 - CHEERS Guidelines for Reporting Economic Evaluations

Section/item	Item No	Recommendation
Title and abstract		
Title	1	Identify the study as an economic evaluation, or use more specific terms such as “cost-effectiveness analysis” and describe the interventions compared.
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base-case and uncertainty analyses), and conclusions.
Introduction		
Background and objectives	3	Provide an explicit statement of the broader context for the study.
		Present the study question and its relevance for health policy or practice decisions.
Methods		
Target population and subgroups	4	Describe characteristics of the base-case population and subgroups analysed including why they were chosen.
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.

Section/item	Item No	Recommendation
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.
Measurement of effectiveness	11a	*Single study–based estimates*: Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.
	11b	*Synthesis-based estimates*: Describe fully the methods used for the identification of included studies and synthesis of clinical effectiveness data.
Measurement and valuation of preference-based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.
Estimating resources and costs	13a	*Single study–based economic evaluation*: Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.
	13b	*Model-based economic evaluation***:** Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.
Choice of model	15	Describe and give reasons for the specific type of decision-analytic model used. Providing a figure to show model structure is strongly recommended.

Section/item	Item No	Recommendation
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytic model.
Analytic methods	17	Describe all analytic methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (e.g., half-cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.
Results		
Study parameters	18	Report the values, ranges, references, and if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.
Characterizing uncertainty	20a	*Single study–based economic evaluation*: Describe the effects of sampling uncertainty for estimated incremental cost, incremental effectiveness, and incremental cost-effectiveness, together with the impact of methodological assumptions (such as discount rate, study perspective).
	20b	*Model-based economic evaluation*: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.
Characterizing heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.

Section/item	Item No	Recommendation
Discussion		
Study findings, limitations, generalizability, and current knowledge	22	Summarize key study findings and describe how they support the conclusions reached. Discuss limitations and the generalizability of the findings and how the findings fit with current knowledge.
Other		
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other nonmonetary sources of support.
Conflicts of interest	24	Describe any potential for conflict of interest among study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors' recommendations.

Note. For consistency, the CHEERS statement checklist format is based on the format of the CONSORT statement checklist.

Appendix 2.2 - Details for the Use of GDP PPPs

Exchange rates are an unsatisfactory means of converting between currencies because they can vary considerably within a short timeframe. Instead, PPPs were used to convert the costs of goods and services priced in different currencies to UK costs [253]. PPPs are rates of currency conversion that equalise the purchasing power of different currencies by eliminating the differences in price levels between countries. In their simplest form, PPPs show the ratio of prices, in national currencies, of the same good or service in different countries and can eliminate some of the drawbacks of using exchange rates [254]. However, there is uncertainty as to which type of PPP, health service specific or related to GDP, is the more appropriate conversion method. Health PPPs are calculated using only the prices of a basket of health related goods and services whereas GDP PPPs are based on the prices of a basket of all goods in the economy. Previous attempts to establish the stability of either health PPP or GDP PPP conversion factors have come to differing conclusions. The Department of Health register of cost effectiveness studies recommends the use of GDP PPPs, though others argue that the choice makes no difference [253].

Appendix 2.3 - An Explanation of What the Expected Time Requirements were for Each Intervention Activity

Activity identification letter	Intervention activity	Standard duration, in minutes, of the activity (according to protocol)	Notes	
E	Children workshops **(Component 1)**	40	If it takes more than 45 minutes or less than 35 minutes, please record the actual time spent	One child workshop = E1 = 40 minutes
J	Main carers workshops **(Component 1)**	60	If it takes more than 65 minutes or less than 55 minutes, please record the actual time spent	One childcare workshop = J1 = 60 minutes
Q	Physically active family friendly games learnt and practiced at school **(Component 3)**	30	If it takes more than 35 minutes or less than 25 minutes, please record the actual time spent	One event of Q = Q1 = 30 minutes
K	Cross-generation family quiz **(Component 1)**	60	If it takes more than 65 minutes or less than 55 minutes, please record the actual time spent	One event of K = K1 = 60 minutes
S	Reviewing (and/or giving feedback on) the performance of family healthy behavioural challenges **(Component 1 and 3)**	Record the actual time spent		
Z	Monthly meeting with relevant school staff and student representatives **(Component 4)**	Record the actual time spent		
X	Supportive regular evaluations and feedbacks to the catering teams **(Component 2)**	Record the actual time spent		

Appendix 2.4 i - CHIRPY DRAGON Teachers' Record of Minutes Worked per Week

Time	Minutes worked per week for delivering the intervention				
2016 (28/03-01/04)		Z:47	Z:25	Z:22	Z:25
2016 (04/04-08/04)	E:40, J:120 Q1:115, Z:35, X:15	E:120, J:180, Q:90, Z:60	E:80, J1:92, Q1:55, Z:55, S:20	E:80, J:120, Q:120, Z:40, S:10	E:120, J1:171, Q1:105, Z:55
2016 (11/04-15/04)	E:80, J:300, Q1:165 S:20	E:160, J:180, Q1:85 S:70	E:80, J:120, Q1:53 S:20	E:120, J1:151, Q:120 S:10	E1:211, J1:289, Q:210 S:25
2016 (18/04-22/04)	E:160, Q1:75, S:20 X:20	E:200, J:180, Q:90 S:153	E:160, J:120, Q1:85 S:20	E:200, J1:69, Q1:100 S:10	E1:82, J1:110, Q1:80, S:10, X:30
2016 (25/04-29/04)	J:120, E:160, Q1:95, S:50	E:240, J:120, Q:60 S:146	E:240, J1:45, Q1:40 S:20, Z:80	E:80, S:10	E:240, S:275
2016 (02/05-06/05)	E:120, J:60, Q1:40 K:60, Z:25, S:50	E:120, J:120, Q:60, Z:100, S:60	k:60, S:40	E:40, J1:41, Q:30, Z:70 S:10	E:120, J1:150, Q1:126 S:45, Z:40
2016 (09/05-13/05)	E:160, J1:105, Q1:130 S:20	E:160, J:180, Q1:80 S:160	E:280, J:120, Q1:80 S:90	E:120, J:60, Q:60 Z:30, S:10	E:120, J1:215, Q1:126 Z:90
2016 (16/05-20/05)	E1:282, J:120, Q1:100 Z:20, S:60	E:80, S:225	E:40, J1:45, Q1:40 S:50	E:80, J:60, Q:60 S:10	E1:271, J:120, Q1:84 S:155
2016 (23/05-27/05)	E:280, Q:60, S:55	E:80, S:250	E:240, J1:125, Q1:40 S:30	E:240, J1:100, Q:60 S:10	E:120, J1:50, Q1:42 S:45
2016 (30/05-03/06)	Z:20, S:65	Z:25, S:87	E:80, J1:45, Q1:40 S:30	E:240, S:10	E:160, S:20
2016 (06/06-10/06)	S:70	Z:60, S:205	S:20		S:360
2016 (13/06-17/06)	S:30	S:85	E:80, S:10		S:20
2016 (20/06-24/06)		S:60			S:145

Appendix 2.4 ii - CHIRPY DRAGON Teachers' Record of Minutes Worked per Week (Continued)

Time	Minutes worked per week for delivering the intervention				
2016 (19/09-23/09)	E:240	E:320	E:320	E:320	E:320
2016 (26/09-30/09)	E:280	E:200, S:182	E:200, S:480	E:160, S:120	E:240, S:120
2016 (10/10-14/10)	E:200, J1:170, S:120 Q1:195	E:360, J1:195, Q1:130 S:106	E:200, J1:50, S: 340	E:40, S:750	E:240, J:240, Q1:160 S:240
2016 (17/10-21/10)	E:240, J1:100, S:50 Q1:130	E:80, J1:115, Q1:125 S:30	E1:135, J1: 145, Q1:50 S:40	E:160, S:430	E:120, J:120, Q1:80 S:70
2016 (24/10-28/10)	E:120, J1:105, S:55 Q1:152	E1:251, J:60, Q:150 S:95	E:40, J1:195, Q:120 S:230	E:80, J:120, S:830	E:160, J:60, S:230
2016 (31/10-04/11)	K:150, S:165	K:130, S:212	S:100	K:60, S:480	
2016 (07/11-11/11)	K:150, S:95	K:120, S:118	K:110, S:120	K:120, S:540	S:160
2016 (14/11-18/11)	K:50	K:50, S:51	K:50, S:100	S:420	S:40
2016 (21/11-25/11)	K:60, S:50	S:30	S:110	S:270, K:60	
2016 (28/11-02/12)				S:160	
2016 (05/12-09/12)				S:100	S:60
2016 (12/12-16/12)	S:75			S:300	S:170
2016 (19/12-23/12)	S:130	S:148	S:150	S:320	S:340
2016 (26/12-30/12)	S:140	S:130	S:160	S:320	S:430
2017 (03/01-06/01)				S:50	S:110
2017 (09/01-13/01)				S:30	K:50
2017 (27/02-03/03)		S:20	S:120		S:80

Appendix 2.4 iii - CHIRPY DRAGON Teachers' Record of Minutes Worked per Week (Continued)

Time	Minutes worked per week for delivering the intervention					
2017 (06/03-10/03)	S:40	S:130	S:87	S:60	S:60	
2017 (13/03-17/03)	S:170	S:68	S:120	S:140	S:70	
2017 (20/03-24/03)	S:140	S:190	S:70	S:180	S:80	S:25
2017 (27/03-31/03)	S:140	S:170	S:65	S:110	S:90	S:20
2017 (04/04-07/04)	S:140		S:135	S:110	S:80	S:95
2017 (10/04-14/04)	S:100		S:50	S:100	S:100	S:30
2017 (17/04-21/04)	S:100			S:80	S:80	S:25
2017 (24/04-28/04)			S:60			

Notes: Working hours for delivering the intervention: 42,510 minutes (708.5 hours), **Working cost:** 708.5*50 = 35,425 Yuan
E: 11,830 m (197 h) *50 = 9,850 C; **J:** 5,835 m (97 h) *50 = 4,850 C; **Q:** 5,432 m (90.5 h) *50 = 4,525 C; **K:** 1,095 m (18 h) *50 = 900 C; **S:** 17,482 m (291 h) *50 = 14,550 C; **Z:** 805 m (14 h) *50 = 700 C; **X:** 65 m (1 h) * 50 = 50 C

Appendix 2.5 - CHU9D Utility Scores at Each Time Point (Chinese Tariff)

	CHU-9D utility scores								
	All participants			**Control group**			**Intervention group**		
Time	**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**	
Baseline (N= 1605)	0.920 (.094)	0.916	0.925	0.919 (.094)	0.913	0.926	0.921 (.094)	0.914	0.927
12 months follow up (N= 1587)	0.919 (.085)	0.915	0.923	0.913 (.089)	0.906	0.919	0.925 (.082)	0.919	0.931

Appendix 2.6 - Unadjusted QALYs Accrued (CHU9D, Chinese Tariff)

	Unadjusted QALYs accrued								
	All participants			**Control group**			**Intervention group**		
Time	**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**	
12 months follow up (N= 1554)	0.919 (.071)	0.916	0.923	0.916 (.072)	0.911	0.921	0.923 (.069)	0.918	0.928

Appendix 2.7 - Incremental Difference in QALYs (CHU9D, Chinese Tariff)

	No adjustment				**Adjusted for clustering and baseline utility [a]**				**Adjusted for clustering, baseline utility, co-variates [b]**			
Measurement time	**Mean difference**	**(95% CI)**		**P-value**	**Mean difference**	**(95% CI)**		**P-value**	**Mean difference**	**(95% CI)**		**P-value**
12 months follow up	0.007	-0.005	0.019	0.296	0.006	0.000	0.011	**0.044**	0.006	0.000	0.012	**0.046**

[a]: Adjusted for baseline outcome. [b]: Adjusted for baseline outcome, pre-specified school- (i.e. whether the school provides mid-morning snack, whether the school has an indoor activity room) and child-level covariates (sex, mother education level, daily average servings of fruit and vegetables, weekly servings of unhealthy snacks and sugar added drink, objectively measured time in MVPA (minutes/24 hours) and objectively measured sedentary time (minutes/24 hours).

Appendix 2.8 - EQ-5D-3L Utility Scores at Each Time Point (Chinese Tariff)

	EQ-5D-3L utility scores (parents/grandparents)								
	All participants			**Control group**			**Intervention group**		
Time	**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**	
Baseline (N= 1235)	0.969 (.066)	0.965	0.973	0.969 (.065)	0.964	0.974	0.970 (.067)	0.964	0.974
12 months follow up (N = 1226)	0.975 (.059)	0.972	0.978	0.974 (.057)	0.969	0.978	0.977 (.060)	0.972	0.981

Appendix 2.9 - Unadjusted QALYs Accrued (EQ-5D-3L, Chinese Tariff)

	Unadjusted QALYs accrued								
	All participants			**Control group**			**Intervention group**		
Time	**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**		**Mean (SD)**	**(95% CI)**	
12 months follow up (N= 1224)	0.973 (.050)	0.969	0.975	0.972 (.047)	0.968	0.976	0.973 (.052)	0.968	0.977

Appendix 2.10 - Incremental Difference in QALYs (EQ-5D-3L, Chinese Tariff)

	No adjustment				**Adjusted for clustering and baseline utility**				**Adjusted for clustering, baseline utility, co-variates**			
Time	**Mean difference**	**(95% CI)**		**P-value**	**Mean difference**	**(95% CI)**		**P-value**	**Mean difference**	**(95% CI)**		**P-value**
12 months follow up	0.001	-0.005	0.006	0.907	0.002	-0.002	0.005	0.363	0.001	-0.001	0.004	0.450

Appendix 2.11 - Outcomes at Baseline and 12 months (Post-Imputation)

	Raw Mean (SD)		**Adjusted Mean (95% CI)**			
Outcomes	**Control group**	**Intervention group**	**Difference [a] (Intervention vs control)**	**P-value**	**Difference [b] (Intervention vs control)**	**P-value**
Baseline						
CHU-9D utility	0.935 (0.066)	0.940 (0.069)				
EQ-5D-3L utility	0.962 (0.087)	0.964 (0.082)				
At the end of the trial						
CHU-9D QALYs	0.933 (0.052)	0.938 (0.046)	0.004 (0.000 to 0.006)	**0.041**	0.003 (0.000, 0.007)	0.059
EQ-5D-3L QALYs	0.966 (0.063)	0.968 (0.066)	0.001 (-0.002 to 0.004)	0.337	0.002 (-0.002, 0.008)	0.442

SD, Standard Deviation; CI, Confidence Interval; QALYs, Quality-Adjusted Life Years;

a = adjusted for clustering and baseline utility; b = adjusted for clustering, baseline utility and co-variates

Appendix 2.12 i - Number (%) of Completed Monthly Household Expenditure for Consented Families

Categories of family expenditure	Baseline			12 months follow up		
	All participants	Control group	Intervention group	All participants	Control group	Intervention group
Electricity/gas	1470 (89.5%)	700 (86.5%)	770 (92.5%)	1492 (90.9%)	729 (90.1%)	763 (91.7%)
Transport	1504 (91.7%)	722 (89.3%)	782 (94%)	1516 (92.4%)	736 (90.9%)	780 (93.8%)
Recreation	1426 (86.9%)	686 (84.8%)	740 (88.9%)	1482 (90.3%)	715 (88.4%)	767 (92.2%)
Food/non-alcoholic drinks	1479 (90.1%)	713 (88.1%)	766 (92.1%)	1495 (91.1%)	724 (89.5%)	771 (92.7%)
Alcoholic drinks, tobacco/narcotics	1392 (84.8%)	675 (83.4%)	717 (86.2%)	1415 (86.2%)	684 (84.5%)	731 (87.9%)
Eating out	1474 (89.8%)	712 (88%)	762 (91.6%)	1505 (91.7%)	733 (90.6%)	772 (92.8%)
Clothing	1421 (86.6%)	676 (83.6%)	745 (89.6%)	1462 (89.1%)	711 (87.9%)	751 (90.3%)
Communications	1446 (88.1%)	698 (86.3%)	748 (89.9%)	1481 (90.2%)	721 (89.1%)	760 (91.3%)
Household goods/services	1381 (84.1%)	666 (82.3%)	715 (85.9%)	1399 (85.3%)	687 (84.9%)	712 (85.6%)
Other goods/services	1396 (85.1%)	678 (83.8%)	718 (86.3%)	1441 (87.8%)	700 (86.5%)	741 (89.1%)
Education	1423 (86.7%)	687 (84.9%)	736 (88.5%)	1478 (90.1%)	720 (90%)	758 (91.2%)
Healthcare	1372 (83.6%)	663 (82%)	709 (85.2%)	1420 (86.5%)	689 (85.2%)	731 (87.9%)

Appendix 2.12 ii - Number (%) of Completed Weekly Household Expenditure for Consented Families

Categories of family expenditure	Baseline			12 months follow up		
	All participants	Control group	Intervention group	All participants	Control group	Intervention group
Total food expenditure	1543 (94%)	746 (92.2%)	797 (95.8%)	1543 (94%)	751 (92.8%)	792 (95.2%)
Fruit and vegetables	1556 (94.8%)	752 (93%)	804 (96.6%)	1535 (93.5%)	750 (92.7%)	785 (94.3%)
Ready meals, fast food and takeaways	1551 (94.5%)	751 (92.8%)	800 (96.2%)	1531 (93.3%)	748 (92.5%)	783 (94.1%)

Appendix 2.12 iii - Number (%) of Completed Monthly Income for Consented Families

Time Point	All participants	Control group	Intervention Group
Baseline	1507 (91.8%)	727 (89.9%)	780 (93.8%)
12 months follow up	1431 (87.2%)	696 (86%)	735 (88.4%)

Appendix 2.13 i - Monthly mean (SD) of expense in Yuan on different households

Categories of family expenditure	Baseline			12 months follow up		
	All	Control group	Intervention group	All	Control group	Intervention group
Electricity/gas	530.44 (550.62)	543.23 (584.39)	518.82 (518.14)	539.02 (679.09)	534.78 (672.83)	543.07 (685.44)
Transport	1078.12 (1033.84)	1109.69 (1056.65)	1048.97 (1012.12)	1055.73 (1004.72)	1053.36 (1020.68)	1057.97 (990.08)
Recreation	418.92 (416.18)	429.12 (422.95)	409.46 (409.85)	423.39 (418.68)	416.64 (411.64)	429.67 (425.32)
Food/non-alcoholic drinks	1862.57 (1089.82)	1770.70 (1074.59)	1948.09 (1097.61)	1903.05 (1172.43)	1878.38 (1167.80)	1926.23 (1177.04)
Alcoholic drinks, tobacco/narcotics	243.57 (377.67)	234.98 (357.72)	251.41 (395.1)	240.19 (384.14)	240.72 (399.01)	239.67 (369.22)
Eating out	812.35 (640.03)	805.5 (644.79)	818.74 (635.91)	836.19 (663.63)	825.92 (656.98)	845.94 (670.16)
Clothing	844.47 (683.05)	832.43 (689.2)	855.4 (677.7)	890.14 (750.94)	883.53 (756.17)	896.39 (746.39)
Communications	389.53 (259.2)	387.18 (268.52)	391.72 (250.36)	393.61 (279.66)	393.02 (294.11)	394.17 (265.42)
Household goods/services	679.99 (1042.46)	646.14 (984.02)	711.43 (1093.74)	810.38 (1257.83)	827.95 (1264.99)	793.43 (1251.54)
Other goods/services	327.31 (367.45)	315.43 (346.79)	338.53 (385.84)	358.98 (409.23)	367.62 (430.26)	350.83 (388.43)
Education	1122.07 (990.59)	1167.5 (1035.66)	1079.66 (945.33)	1514.7 (1252.79)	1541.66 (1284.73)	1489.1 (1231.78)
Healthcare	380.43 (428.85)	383.18 (429.83)	377.86 (428.22)	410.58 (448.09)	426.68 (457.86)	395.4 (438.46)

Appendix 2.13 ii - Weekly mean (SD) of expense in Yuan on different households

Categories of family expenditure	Baseline			12 months follow up		
	All Participants	Control group	Intervention group	All participants	Control group	Intervention group
Total food expenditure	1064.67 (603.19)	1066.08 (604.07)	1063.36 (602.74)	1086.77 (595.86)	1079.66 (608.69)	1093.52 (583.73)
Fruit and vegetables	223.92 (118.96)	226.23 (117.09)	221.76 (120.71)	232.19 (118.72)	232.2 (119.25)	232.29 (118.29)
Ready meals, fast food and	88.16 (96.47)	87.05 (97.06)	89.21 (95.97)	98.72 (103.1)	98.02 (102.59)	99.39 (103.64)

APPENDIX 3. Assessment of Construct Validity of CHU9D: Results Tables

Appendix 3.1 - Characteristics of the Study Population

Characteristics	
Measures of socio-economic status	
Maternal work: n (%)	
Yes	1190 (77.3)
No	349 (22.7)
Maternal employment status: n (%)	
5 Working full-time	1043 (67.8)
4 Working part time	147 (9.5)
3 Unemployed or looking for work	15 (1.0)
2 Looking after the family/house	280 (18.2)
1 Other	54 (3.5)
Paternal work: n (%)	
Yes	1421 (92.3)
No	118 (7.7)
Paternal employment status: n (%)	
5 Working full-time	1378 (89.5)
4 Working part time	43 (2.8)
3 Unemployed or looking for work	10 (0.7)
2 Looking after the family/house	17 (1.1)
1 Other	91 (5.9)

Appendix 3.2 - Mean (SD), median (IQR) for CHU9D, PedsQL scores based on characteristics

	Number (%)	CHU-9D Utility, UK tariff Mean (SD), Median (IQR)	CHU-9D Utility, Chinese tariff Mean (SD), Median (IQR)	PedsQL total score Mean (SD), Median (IQR)
Mother's employment				
Yes	1190 (77.3)	0.936 (0.069), 0.963 (0.900-1.000)	0.919 (0.095), 0.943 (0.876-1.000)	82.51 (11.49), 84.09 (75.00-91.30)
No	349 (22.7)	0.940 (0.065), 0.965 (0.909-1.000)	0.923 (0.091), 0.946 (0.878-1.000)	83.04 (11.13), 84.78 (76.08-91.30)
p-value*		0.60	0.73	0.50
Mother employment status				
5 Working full-time	1043 (67.8)	0.936 (0.069), 0.962 (0.900-1.000)	0.919 (0.094), 0.943 (0.875-1.000)	82.99 (10.94), 84.78 (76.08-91.30)
4 Working part time	147 (9.5)	0.936 (0.073), 0.963 (0.893-1.000)	0.919 (0.107), 0.953 (0.881-1.000)	81.74 (11.93), 82.60 (75.00-91.30)
3 Unemployed or looking for work	15 (1.0)	0.892 (0.104), 0.915 (0.812-0.963)	0.872 (0.122), 0.892 (0.760-0.996)	81.15 (12.16), 84.78 (71.73-89.13)
2 Looking after the family/house	280 (18.2)	0.942 (0.063), 0.963 (0.914-1.000)	0.925 (0.090), 0.949 (0.881-1.000)	82.73 (11.40), 84.78 (76.08-91.30)
1 Other	54 (3.5)	0.940 (0.058), 0.952 (0.904-1.000)	0.925 (0.083), 0.939 (0.880-1.000)	83.46 (12.43), 86.95 (76.08-92.39)
p-value**		0.57	0.65	0.60
Father's employment				
Yes	1421 (92.3)	0.937 (0.068), 0.963 (0.899-1.000)	0.920 (0.094), 0.943 (0.876-1.000)	82.79 (11.23), 84.78 (76.08- 91.30)
No	118 (7.7)	0.938 (0.065), 0.964 (0.903-1.000)	0.922 (0.095), 0.950 (0.880-1.000)	84.48 (10.92), 86.95 (76.08-92.39)
p-value*		0.93	0.91	0.07
Father employment status				
5 Working full-time	1378 (89.5)	0.937 (0.068), 0.963 (0.903-1.000)	0.921 (0.093), 0.943 (0.876-1.000)	82.81 (11.25), 84.78 (76.08-91.30)
4 Working part time	43 (2.8)	0.923 (0.087), 0.952 (0.893-0.978)	0.895 (0.119), 0.937 (0.848-0.958)	82.30 (10.65), 82.60 (76.08-91.30)
3 Unemployed or looking for work	10 (0.7)	0.948 (0.040), 0.957 (0.928-0.963)	0.936 (0.054), 0.946 (0.896-0.996)	87.06 (8.20), 90.21 (83.69-91.30)
2 Looking after the family/house	17 (1.1)	0.931 (0.063), 0.940 (0.897-0.978)	0.898 (0.112), 0.916 (0.851-0.996)	85.80 (10.91), 90.21 (75.00-94.56)
1 Other	91 (5.9)	0.937 (0.068), 0 .963 (0.897-1.000)	0.924 (0.095), 0.955 (0.880-1.000)	83.95 (11.21), 85.86 (77.17-92.39)
p-value**		0.97	0.84	0.17

*Kruskal-Wallis test; **non-parametric test for trend

LIST OF REFERENCES

1. Au N. The health care cost implications of overweight and obesity during childhood. *Health Serv Res*. 2012;47(2):655-676.

2. Neovius K, Johansson K, Kark M, Neovius M. Obesity status and sick leave: a systematic review. *Obesity reviews*. 2009;10(1):17-27.

3. Basterfield L, Jones AR, Parkinson KN, Reilly J, Pearce MS, Reilly JJ, et al. Physical activity, diet and BMI in children aged 6–8 years: a cross-sectional analysis. *BMJ Open*. 2014;4(6):e005001.

4. Lobstein T, Baur L, Uauy R. Obesity in children and young people: a crisis in public health. *Obesity reviews*. 2004;5(s1):4-85.

5. Li S-m, Zhu Y. Residential mobility within Guangzhou city, China, 1990–2010: local residents versus migrants. *Eurasian Geography and Economics*. 2014;55(4):313-332.

6. Tudor-Locke C, Ainsworth BE, Adair L, Du S, Popkin B. Physical activity and inactivity in Chinese school-aged youth: the China Health and Nutrition Survey. *International Journal of Obesity*. 2003;27(9):1093-1099.

7. Zhai F, He Y, Ma G, Li Y, Wang Z, Hu Y, et al. Study on the current status and trend of food consumption among Chinese population. *Zhonghua Liuxingbingxue Zazhi*. 2005;26(7):485-488.

8. Ji CY, Chen TJ, China WGoOi. Empirical changes in the prevalence of overweight and obesity among Chinese students from 1985 to 2010 and corresponding preventive strategies. *Biomedical and Environmental Sciences*. 2013;26(1):1-12.

9. Ji CY. Report on childhood obesity in China (4) prevalence and trends of overweight and obesity in Chinese urban school-age children and adolescents, 1985-2000. *Biomedical and Environmental Sciences*. 2007;20(1):1-10.

10. Ji CY, Cheng TO. Prevalence and geographic distribution of childhood obesity in China in 2005. 2008;131(1):1-8.

11. Liu W, Liu W, Lin R, Li B, Pallan M, Cheng K, et al. Socioeconomic determinants of childhood obesity among primary school children in Guangzhou, China. *BMC Public Health*. 2016;16(1):482-489.

12. Wang Y, Yan AF, Shi X, Wang H, Wang Z, Gittelsohn J, et al. Child and parental perspectives on diet and physical activity decisions: implications for childhood obesity prevention in China. *Asia Pac J Clin Nutr*. 2017;26(5):888-898.

13. Popkin BM, Kim S, Rusev ER, Du S, Zizza C. Measuring the full economic costs of diet, physical activity and obesity-related chronic diseases. *Obesity reviews*. 2006;7(3):271-293.

14. The State Council of China. The Healthy China 2030 Planning Outline. Beijing: The People's Republic of China; 2016.

15. Jiang J, Xia X, Greiner T, Wu G, Lian G, Rosenqvist U. The effects of a 3-year obesity intervention in schoolchildren in Beijing. *Child: Care, Health and Development*. 2007;33(5):641-646.

16. Wu X, Wang J. Childhood obesity intervention study in Xuzhou. *Modern Preventive Medicine*. 2010;37(12):2225-2226.

17. Meng LP, Xu HQ, Liu AL, van Raaij J, Bemelmans W, Hu XQ, et al. The costs and cost-effectiveness of a school-based comprehensive intervention study on childhood obesity in China. *Plos One*. 2013;8(10):e77971.

18. Li M, Li S, Baur LA, Huxley RR. A systematic review of school-based intervention studies for the prevention or reduction of excess weight among Chinese children and adolescents. *Obesity Reviews*. 2008;9(6):548-559.

19. Gao Y, Griffiths S, Chan EY. Community-based interventions to reduce overweight and obesity in China: a systematic review of the Chinese and English literature. *J Public Health (Oxf)*. 2010;30(4):436-448.

20. Feng L, Wei D-M, Lin S-T, Maddison R, Mhurchu CN, Jiang Y, et al. Systematic review and meta-analysis of school-based obesity interventions in mainland China. *PloS One*. 2017;12(9):e0184704.

21. Medical Research Council. A framework for development and evaluation of RCTs for complex interventions to improve health. London: Medical Research Council; 2000.

22. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. *BMJ: British Medical Journal*. 2007;334(7591):455-459.

23. Palmer S, Raftery J. Economics notes: Opportunity cost. *BMJ: British Medical Journal*. 1999;318(7197):1551-1552.

24. Edwards RT, McIntosh E. Applied health economics for public health practice and research. Oxford: Oxford University Press; 2019.

25. Ananthapavan J, Sacks G, Moodie M, Carter R. Economics of obesity—Learning from the past to contribute to a better future. *International Journal of Environmental Research and Public Health*. 2014;11(4):4007-4025.

26. Frew E. Economic evaluation of childhood obesity interventions: Reflections and suggestions. *PharmacoEconomics*. 2016;34(8):733-740.

27. Johannesson M, Jonsson B, Jonsson L, Kobelt G, Zethraeus N. Why should economic evaluations of medical innovations have a societal perspective? 2009 [cited 2017 June 15]. Available from: https://www.ohe.org/publications/why-should-economic-evaluations-medical-innovations-have-societal-perspective.

28. Walker S, Griffin S, Asaria M, Tsuchiya A, Sculpher M. Striving for a societal perspective: A framework for economic evaluations when costs and effects fall on multiple sectors and decision makers. *Applied health economics and health policy*. 2019;17(5):1-14.

29. Edwards RT, Charles JM, Lloyd-Williams H. Public health economics: a systematic review of guidance for the economic evaluation of public health interventions and discussion of key methodological issues. *BMC public health*. 2013;13(1):1001.

30. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: National Institute for Health and Care Excellence; 2013.

31. Chen Y, He Y, Chi X, Wei Y, Shi L. Development of health technology assessment in China: New challenges. *Bioscience trends*. 2018;12(2):102-108.

32. Zhao K, Li X, Guo W, Xiao Y. Coming rapidly of age: Health technology assessment in China [Cited 2020 January 23]. Available from: https://globalforum.diaglobal.org/issue/may-2018/coming-rapidly-of-age-health-technology-assessment-in-china/.

33. Sharma K, Gautam K. Evolution of health technology assessment (HTA) in China. 2019 [Cited 2020 January 23]. Available from: https://decisionresourcesgroup.com/blog/evolution-health-technology-assessment-hta-china/.

34. Brouwer WB, Culyer AJ, van Exel NJA, Rutten FF. Welfarism vs. extra-welfarism. *J Health Econ*. 2008;27(2):325-338.

35. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. London: National Institute for Health and Care Excellence; 2014.

36. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2015.

37. Greenberg NP, Neumann P. Is the United States ready for QALYs. *Health Aff*. 2009;28:1366-1371.

38. Canaway A, Frew E. Measuring preference-based quality of life in children aged 6–7 years: a comparison of the performance of the CHU-9D and EQ-5D-Y—the WAVES Pilot Study. *Quality of Life Research*. 2013;22(1):173-183.

39. Stevens K. Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. *Qualitative Health Research*. 2010;20(3):340-351.

40. Stevens K. Assessing the performance of a new generic measure of health-related quality of life for children and refining it for use in health state valuation. *Applied Health Economics and Health Policy*. 2011;9(3):157-169.

41. Ratcliffe J, Stevens K, Flynn T, Brazier J, Sawyer M. An assessment of the construct validity of the CHU9D in the Australian adolescent general population. *Quality of Life Research*. 2012;21(4):717-725.

42. Frew E, Pallan M, Lancashire E, Hemming K, Adab P. Is utility-based quality of life associated with overweight in children? Evidence from the UK WAVES randomised controlled study. *BMC Pediatr*. 2015;15(1):211-221.

43. Chen G, Flynn T, Stevens K, Brazier J, Huynh E, Sawyer M, et al. Assessing the health-related quality of life of Australian adolescents: an empirical comparison of the child health utility 9D and EQ-5D-Y instruments. *Value in Health*. 2015;18(4):432-438.

44. Stevens K, Ratcliffe J. Measuring and valuing health benefits for economic evaluation in adolescence: an assessment of the practicality and validity of the Child Health Utility 9D in the Australian adolescent population. *Value in Health*. 2012;15(8):1092-1099.

45. Al-Janabi H, Flynn TN, Coast J. QALYs and carers. *Pharmacoeconomics*. 2011;29(12):1015-1023.

46. Palmer S, Byford S, Raftery J. Types of economic evaluation. *BMJ: British Medical Journal*. 1999;318(7194):1349.

47. Zanganeh M, Adab P, Li B, Frew E. A systematic review of methods, study quality, and results of economic evaluation for childhood and adolescent obesity intervention. *International Journal of Environmental Research and Public Health*. 2019;16(3):485-499.

48. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated health economic evaluation reporting standards (CHEERS) statement. *Cost Effectiveness and Resource Allocation*. 2013;11(1):6-11.

49. Word Health Organization. Obesity and overweight. 2018 [cited 2018 February 27]. Available from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

50. Adab P, Pallan M, Whincup PH. Is BMI the best measure of obesity? : British Medical Journal Publishing Group; 2018.

51. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *Bmj*. 2000;320(7244):1240.

52. Apovian CM. Obesity: definition, comorbidities, causes, and burden. *Am J Manag Care*. 2016;22(7):176-185.

53. Dinsdale H, Ridler C, Ells L. A simple guide to classifying body mass index in children. *National Obesity Observatory: Oxford*. 2011.

54. Onis Md, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bulletin of the World health Organization*. 2007;85:660-667.

55. Abarca-Gomez L AZ, Hamid Z, Abu- Rmeileh NME, Acosta-Cazares B, Acuin C, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128· 9 million children, adolescents, and adults. *The Lancet*. 2017;390:2627–2642. .

56. Wang Y, Lobstein T. Worldwide trends in childhood overweight and obesity. *Int J Pediatr Obes*. 2006;1(1):11-25.

57. Dearth-Wesley T, Wang H, Popkin B. Under-and overnutrition dynamics in Chinese children and adults (1991–2004). *European Journal of Clinical Nutrition*. 2008;62(11):1302-1307.

58. Zhang, Y. Report: Childhood obesity accelerating [Internet]. China Daily. 2017 [cited 09/08/2018]. Available from: http://www.chinadaily.com.cn/china/2017-05/12/content_29313556.htm.

59. Chen Y, Ma L, Ma Y, Wang H, Luo J, Zhang X, et al. A national school-based health lifestyles interventions among Chinese children and adolescents against obesity: rationale, design and methodology of a randomized controlled trial in China. *BMC Public Health*. 2015;15(1):210-220.

60. Ji CY, Cheng TO. Epidemic increase in overweight and obesity in Chinese children from 1985 to 2005. *International Journal of Cardiology*. 2009;132(1):1-10.

61. Li B, Lin R, Liu W, Chen J, Liu W, Cheng KK, et al. Differences in perceived causes of childhood obesity between migrant and local communities in China: a qualitative study. *The Lancet*. 2016;388:S3.

62. The State Council of China. The Healthy China 2030 Planning Outline Beijing: The People's Republic of China; 2016 [Available from: https://www.who.int/healthpromotion/conferences/9gchp/healthy-china/en/.

63. Li B. Diet-related NCDs in China: more needs to be done. *THE LANCET Public Health*. 2019;4(12).

64. Hoey H. Management of obesity in children differs from that of adults. *Proc Nutr Soc*. 2014;73(4):519-525.

65. World Health Organization. Global strategy on diet, physical Activity and health, recommended levels of physical activity for children aged 5 - 17 years. 2018 [cited 12 March 2018]. Available from: https://www.who.int/dietphysicalactivity/factsheet_young_people/en/.

66. Sedentary, Behaviour Research Network. Letter to the editor: standardized use of the terms" sedentary" and" sedentary behaviours". *Applied Physiology, Nutrition, and Metabolism*. 2012;37(3):540.

67. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, et al. Sedentary behavior research network (SBRN)–terminology consensus project process and outcome. *International Journal of Behavioral Nutrition and Physical Activity*. 2017;14(1):75-92.

68. World Health Organisation. Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age. 2019 [cited 2019 July 10]. Available from: https://apps.who.int/iris/handle/10665/311664.

69. Owen N, Sparling PB, Healy GN, Dunstan DW, Matthews CE. Sedentary behavior: emerging evidence for a new health risk. *Mayo Clinic Proceedings*. 2010;85(12):1138-1141.

70. Cliff DP, Hesketh KD, Vella SA, Hinkley T, Tsiros MD, Ridgers ND, et al. Objectively measured sedentary behaviour and health and development in children and adolescents: systematic review and meta-analysis. *Obesity Reviews*. 2016;17(4):330-344.

71. Cliff DP, Okely AD, Burrows TL, Jones RA, Morgan PJ, Collins CE, et al. Objectively measured sedentary behavior, physical activity, and plasma lipids in overweight and obese children. *Obesity*. 2013;21(2):382-385.

72. Saunders TJ, Chaput J-P, Tremblay MS. Sedentary behaviour as an emerging risk factor for cardiometabolic diseases in children and youth. *Canadian Journal of Diabetes*. 2014;38(1):53-61.

73. National Institute for Health and Care Excellence. Maintaining a healthy weight and preventing excess weight gain in children and adults – partial update of CG43. London: National Institute for Health and Care Excellence; 2014.

74. Pearson N, Braithwaite R, Biddle SJ, van Sluijs EM, Atkin AJ. Associations between sedentary behaviour and physical activity in children and adolescents: a meta-analysis. *Obesity Reviews*. 2014;15(8):666-675.

75. Malik VS, Schulze MB, Hu FB. Intake of sugar-sweetened beverages and weight gain: a systematic review–. *The American Journal of Clinical Nutrition*. 2006;84(2):274-288.

76. Wang X, Ouyang Y, Liu J, Zhu M, Zhao G, Bao W, et al. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. *Bmj*. 2014;349:g4490.

77. Ledoux T, Hingle MD, Baranowski T. Relationship of fruit and vegetable intake with adiposity: a systematic review. *Obesity Reviews*. 2011;12(5):e143-e150.

78. Shloim N, Edelson LR, Martin N, Hetherington MM. Parenting styles, feeding styles, feeding practices, and weight status in 4–12 year-old children: A systematic review of the literature. *Frontiers in Psychology*. 2015;6:1849.

79. Kröller K, Warschburger P. Associations between maternal feeding style and food intake of children with a higher risk for overweight. *Appetite*. 2008;51(1):166-172.

80. Faith MS, Scanlon KS, Birch LL, Francis LA, Sherry B. Parent-child feeding strategies and their relationships to child eating and weight status. *Obesity Research*. 2004;12(11):1711-1722.

81. Anzman SL, Rollins BY, Birch LL. Parental influence on children's early eating environments and obesity risk: implications for prevention. *International Journal of Obesity*. 2010;34(7):1116-1124.

82. Jingxiong J, Rosenqvist U, Huishan W, Greiner T, Guangli L, Sarkadi A. Influence of grandparents on eating behaviors of young children in Chinese three-generation families. *Appetite*. 2007;48(3):377-383.

83. Eklund L. Rethinking son preference–Gender, population dynamics and social change in the People's Republic of China: Lund University; 2011.

84. Li B, Adab P, Cheng KK. The role of grandparents in childhood obesity in China-evidence from a mixed methods study. *International Journal of Behavioral Nutrition and Physical Activity*. 2015;12(1):91-99.

85. Dietz WH. Critical periods in childhood for the development of obesity. *The American Journal of Clinical Nutrition*. 1994;59(5):955-959.

86. Yu Z, Han S, Zhu J, Sun X, Ji C, Guo X. Pre-pregnancy body mass index in relation to infant birth weight and offspring overweight/obesity: a systematic review and meta-analysis. *PloS One*. 2013;8(4):e61627.

87. Ino T. Maternal smoking during pregnancy and offspring obesity: Meta-analysis. *Pediatrics International*. 2010;52(1):94-99.

88. Schack-Nielsen L, Michaelsen KF, Gamborg M, Mortensen EL, Sørensen TI. Gestational weight gain in relation to offspring body mass index and obesity from infancy through adulthood. *International Journal of Obesity*. 2010;34(1):67-74.

89. Yu Z, Han S, Zhu G, Zhu C, Wang X, Cao X, et al. Birth weight and subsequent risk of obesity: a systematic review and meta-analysis. *Obesity Reviews*. 2011;12(7):525-542.

90. Harder T, Bergmann R, Kallischnigg G, Plagemann A. Duration of breastfeeding and risk of overweight: a meta-analysis. *Am J Epidemiol*. 2005;162(5):397-403.

91. Baird J, Fisher D, Lucas P, Kleijnen J, Roberts H, Law C. Being big or growing fast: systematic review of size and growth in infancy and later obesity. *Bmj*. 2005;331(7522):929.

92. Chen X, Beydoun MA, Wang Y. Is sleep duration associated with childhood obesity? A systematic review and meta-analysis. *Obesity*. 2008;16(2):265-274.

93. Moorcroft KE, Marshall JL, McCormick FM. Association between timing of introducing solid foods and obesity in infancy and childhood: a systematic review. *Matern Child Nutr*. 2011;7(1):3-26.

94. Serdula MK, Ivery D, Coates RJ, Freedman DS, Williamson DF, Byers T. Do obese children become obese adults? A review of the literature. *Prev Med*. 1993;22(2):167-177.

95. Wardle J, Brodersen NH, Cole TJ, Jarvis MJ, Boniface DR. Development of adiposity in adolescence: five year longitudinal study of an ethnically and socioeconomically diverse sample of young people in Britain. *Bmj*. 2006;332(7550):1130-1135.

96. Singh AS, Mulder C, Twisk JW, Van Mechelen W, Chinapaw MJ. Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obesity Reviews*. 2008;9(5):474-488.

97. Seidell JC. Obesity, insulin resistance and diabetes—a worldwide epidemic. *Br J Nutr*. 2000;83(S1):S5-S8.

98. Hubert HB, Feinleib M, McNamara PM, Castelli WP. Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. *Circulation*. 1983;67(5):968-977.

99. Grotle M, Hagen KB, Natvig B, Dahl FA, Kvien TK. Obesity and osteoarthritis in knee, hip and/or hand: an epidemiological study in the general population with 10 years follow-up. *BMC Musculoskeletal Disorders*. 2008;9(1):132-136.

100. Bianchini F, Kaaks R, Vainio H. Overweight, obesity, and cancer risk. *The Lancet Oncology*. 2002;3(9):565-574.

101. Doring N, Mayer S, Rasmussen F, Sonntag D. Economic evaluation of obesity prevention in early childhood: Methods, limitations and recommendations. *International Journal of Environmental Research and Public Health*. 2016;13 (9):911-921.

102. Schwimmer JB, Burwinkle TM, Varni JW. Health-related quality of life of severely obese children and adolescents. *Jama*. 2003;289(14):1813-1819.

103. Eisenberg ME, Neumark-Sztainer D, Story M. Associations of weight-based teasing and emotional well-being among adolescents. *Archives of Pediatrics & Adolescent Medicine*. 2003;157(8):733-738.

104. Strauss RS, Pollack HA. Social marginalization of overweight children. *Archives of Pediatrics & Adolescent Medicine*. 2003;157(8):746-752.

105. Klarenbach S, Padwal R, Chuck A, Jacobs P. Population-based analysis of obesity and workforce participation. *Obesity*. 2006;14(5):920-927.

106. Paraponaris A, Saliba B, Ventelou B. Obesity, weight status and employability: empirical evidence from a French national survey. *Econ Hum Biol*. 2005;3(2):241-258.

107. Neovius K, Rehnberg C, Rasmussen F, Neovius M. Lifetime productivity losses associated with obesity status in early adulthood. *Applied Health Economics and Health Policy*. 2012;10(5):309-317.

108. ISPOR - Improving Healthcare Decisions. China Midland - Pharmaceuticals [Cited 2020 February 03]. Available from: https://tools.ispor.org/htaroadmaps/China.asp.

109. Hamilton D, Dee A, Perry I. The lifetime costs of overweight and obesity in childhood and adolescence: a systematic review. *Obesity Reviews*. 2018;19(4):452-463.

110. Kaptchuk TJ. The double-blind, randomized, placebo-controlled trial: gold standard or golden calf? *Journal of Clinical Epidemiology*. 2001;54(6):541-549.

111. Sørensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. *Hepatology*. 2006;44(5):1075-1082.

112. Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials important? *BMJ: British Medical Journal*. 1998;316(7126):201.

113. Victora CG, Habicht J-P, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health*. 2004;94(3):400-405.

114. Puffer S, Torgerson DJ, Watson J. Cluster randomized controlled trials. *Journal of Evaluation in Clinical Practice*. 2005;11(5):479-483.

115. Wang LY, Gutin B, Barbeau P, Moore JB, Hanes Jr J, Johnson MH, et al. Cost-effectiveness of a school-based obesity prevention program. *Journal of School Health*. 2008;78(12):619-624.

116. Doak C, Visscher T, Renders C, Seidell J. The prevention of overweight and obesity in children and adolescents: a review of interventions and programmes. *Obesity Reviews*. 2006;7(1):111-136.

117. Brown T, Moore THM, Hooper L, Gao Y, Zayegh A, Ijaz S, et al. Interventions for preventing obesity in children: Do diet and physical activity strategies help prevent obesity in children (aged 0 to 18 years)? *Cochrane Database Syst Rev*. 2019(7).

118. Adab P, Pallan MJ, Lancashire ER, Hemming K, Frew E, Barrett T, et al. Effectiveness of a childhood obesity prevention programme delivered through schools, targeting 6 and 7 year olds: cluster randomised controlled trial (WAVES study). *Bmj*. 2018;360:k211.

119. Lloyd J, Creanor S, Logan S, Green C, Dean SG, Hillsdon M, et al. Effectiveness of the Healthy Lifestyles Programme (HeLP) to prevent obesity in UK primary-school children: a cluster randomised controlled trial. *The Lancet Child & Adolescent Health*. 2018;2(1):35-45.

120. Kipping RR, Howe LD, Jago R, Campbell R, Wells S, Chittleborough CR, et al. Effect of intervention aimed at increasing physical activity, reducing sedentary behaviour, and increasing fruit and vegetable consumption in children: active for Life Year 5 (AFLY5) school based cluster randomised controlled trial. *Bmj*. 2014;348:g3256.

121. Waters E, de Silva-Sanigorski A, Burford BJ, Brown T, Campbell KJ, Gao Y, et al. Interventions for preventing obesity in children. *Cochrane Database Syst Rev*. 2011(12).

122. Campbell K, Waters E, O'meara S, Summerbell C. Interventions for preventing obesity in childhood. A systematic review. *Obesity Reviews*. 2001;2(3):149-157.

123. Uijtdewilligen L, Waters C, Müller-Riemenschneider F, Lim Y. Preventing childhood obesity in Asia: an overview of intervention programmes. *Obesity Reviews*. 2016;17(11):1103-1115.

124. Verstraeten R, Roberfroid D, Lachat C, Leroy JL, Holdsworth M, Maes L, et al. Effectiveness of preventive school-based obesity interventions in low-and middle-income countries: a systematic review. *The American journal of clinical nutrition*. 2012;96(2):415-438.

125. Li B, Liu WJ, Cheng KK, Pallan M, Hemming K, Frew E, et al. Development of the theory-based Chinese primary school children physical activity and dietary behaviour changes intervention (CHIRPY DRAGON): development of a cluster-randomised controlled trial. *The Lancet*. 2016;388:S51.

126. Li B, Liu WJ, Adab P, Pallan M, Hemming K, Frew E, et al. Cluster-randomised controlled trial to assess the effectiveness and cost-effectiveness of an obesity prevention programme for Chinese primary school-aged children: the CHIRPY DRAGON study protocol. *BMJ Open*. 2017;7(11):e018415.

127. Liu W, Lin R, Liu W, Guo Z, Xiong L, Li B, et al. Relationship between weight status and health-related quality of life in Chinese primary school children in Guangzhou: a cross-sectional study. *Health and Quality of Life Outcomes*. 2016;14(1):166-123.

128. Free Learning from The Open University. A brief introduction to the Chinese education system 2019 [Cited 2020 February 26]. Available from: https://www.open.edu/openlearn/education/brief-introduction-the-chinese-education-system.

129. Li B, Pallan M, Liu WJ, Hemming K, Frew E, Lin R, et al. The CHIRPY DRAGON intervention in preventing obesity in Chinese primary school-aged children: A cluster-randomised controlled trial. *Plos Medicine*. 2019;16(11):1-20.

130. World Health Organization. Health topics: Primary health care. 2017 [cited 2017 December 03]. Available from: http://www.who.int/topics/primary_health_care/en/

131. Morris S, Devlin N, Parkin D. Economic analysis in health care: John Wiley & Sons; 2007.

132. Edwards RT, Charles JM, Lloyd-Williams H. Public health economics: a systematic review of guidance for the economic evaluation of public health interventions and discussion of key methodological issues. *BMC Public Health*. 2013;13(1):1001-1014.

133. Ngorsuraches S. Defining types of economic evaluation. *Journal of the Medical Association of Thailand= Chotmaihet thangphaet*. 2008;91:S21-27.

134. Shiell A, Donaldson C, Mitton C, Currie G. Health economic evaluation. *Journal of Epidemiology and Community Health*. 2002;56(2):85-88.

135. Black J, Hashimzade N, Myles G. A dictionary of economics: OUP Oxford; 2012.

136. Drummond MF, McGuire A. Economic evaluation in health care: merging theory with practice: OUP Oxford; 2001.

137. Coast J, Smith RD, Lorgelly P. Welfarism, extra-welfarism and capability: the spread of ideas in health economics. *Soc Sci Med*. 2008;67(7):1190-1198.

138. Johansson P-O. An introduction to modern welfare economics. Cambridge: Cambridge University Press; 1991.

139. Coast J. Maximisation in extra-welfarism: A critique of the current position in health economics. *Soc Sci Med*. 2009;69(5):786-792.

140. Mooney G. Is it not time for health economists to rethink equity and access? *Health Economics, Policy and Law*. 2009;4(02):209-221.

141. Culyer AJ. The normative economics of health care finance and provision. *Oxford Review of Economic Policy*. 1989;5(1):34-58.

142. McPake B, Normand C, Smith S. Health economics: an international perspective: Routledge; 2013.

143. McGuire A. Theoretical concepts in the economic evaluation of health care. 2001.

144. Tsuchiya A, Williams A, Drummond M, McGuire A. Welfare economics and economic evaluation. *Economic evaluation in Health Care: Merging Theory with Practice*. 2001:27-28.

145. Goranitis I, Coast J, Day E, Copello A, Freemantle N, Frew E. Maximizing health or sufficient capability in economic evaluation? A methodological experiment of treatment for drug addiction. *Medical Decision Making*. 2017;37(5):498-511.

146. Goranitis I, Coast J, Day E, Copello A, Freemantle N, Seddon J, et al. Measuring health and broader well-being benefits in the context of opiate dependence: the psychometric performance of the ICECAP-A and the EQ-5D-5L. *Value in health*. 2016;19(6):820-828.

147. Wonderling D. Introduction to health economics: McGraw-Hill Education (UK); 2011.

148. Gray A, Clarke P, Wolstenholme J, Wordsworth S. Applied Methods of Cost-Effectiveness in Health Care. Handbooks in Health Economic Evaluation. New York, USA: Oxford University Press; 2011.

149. McIntosh E. Applied methods of cost-benefit analysis in health care. Oxford: Oxford University Press; 2010.

150. Olsen JA, Smith RD. Theory versus practice: a review of 'willingness-to-pay'in health and health care. *Health Econ*. 2001;10(1):39-52.

151. Olsen JA. Aiding priority setting in health care: is there a role for the contingent valuation method? *Health Econ*. 1997;6(6):603-612.

152. Hoch JS, Dewa CS. An introduction to economic evaluation: what's in a name? *The Canadian Journal of Psychiatry*. 2005;50(3):159-166.

153. Layard P. Cost-benefit analysis: Cambridge University Press; 1994.

154. Pearce D, Atkinson G, Mourato S. Cost-benefit analysis and the environment: recent developments: Organisation for Economic Co-Operation and Development; 2006.

155. Donaldson C. The (Near) Equivalence of Cost—Effectiveness and Cost-Benefit Analyses. *Pharmacoeconomics*. 1998;13(4):389-396.

156. Garber AM, Weinstein MC, Torrance G, Kamlet M. Theoretical foundations of cost-effectiveness analysis. *Cost-effectiveness in Health and Medicine*. 1996;1:25-53.

157. York Health Economics Consortium. Cost-effectiveness plane 2016 [Cited 2020 January 28]. Available from: https://yhec.co.uk/glossary/cost-effectiveness-plane/.

158. Boyd K, Briggs A, Ducournau P, Gyldmark M, de Reydet F, Cassidy J. Statistical issues with trial data and economic modeling for cost-effectiveness evaluationIn.In: Tang, W. and Tu, X. (eds.). Modern Clinical Trial Analysis. Series: Applied Bioinformatics and Biostatistics in Cancer Research. : Springer-Verlag: New York, NY, USA,; 2013. p. 149-166.

159. Gold MR, Stevenson D, Fryback DG. HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. *Annu Rev Public Health*. 2002;23(1):115-134.

160. Shumway M. Preference weights for cost-outcome analyses of schizophrenia treatments: comparison of four stakeholder groups. *Schizophrenia Bulletin*. 2003;29(2):257-266.

161. National Institute for Health and Care Excellence. Methods for the development of NICE public health guidance. London: National Institute for Health and Care Excellence; 2012.

162. AMCP Format Executive Committee members. The AMCP Format for Formulary Submissions (version 4.0). The USA.; 2016.

163. Hoch JS, Dewa CS. Advantages of the net benefit regression framework for economic evaluations of interventions in the workplace: a case study of the cost-effectiveness of a collaborative mental health care program for people receiving short-term disability benefits for psychiatric disorders. *Journal of Occupational and Environmental Medicine*. 2014;56(4):441-445.

164. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ*. 2002;11(5):415-430.

165. Woods B, Revill P, Sculpher M, Claxton K. Country-level cost-effectiveness thresholds: initial estimates and the need for further research. *Value in Health*. 2016;19(8):929-935.

166. World Health Organization. Cost-effectiveness thresholds: pros and cons 2016 [Cited 2020 February 17]. Available from: https://www.who.int/bulletin/volumes/94/12/15-164418/en/.

167. Investopedia. Top 25 developed and developing countries [Cited 2020 February 16]. Available from: https://www.investopedia.com/updates/top-developing-countries/.

168. The World Bank. GDP per capita, PPP (current international $) [Cited 2020 February 16]. Available from: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD.

169. DAC List of ODA Recipients. Effective for reporting on 2018, 2019 and 2020 flows [Cited 2020 March 09]. Available from: http://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/DAC_List_ODA_Recipients2018to2020_flows_En.pdf.

170. Andronis L, Barton P, Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. *Health Technology Assessment*. 2009.

171. Briggs AH, O'Brien BJ. The death of cost-minimization analysis? *Health Econ*. 2001;10(2):179-184.

172. Mauskopf JA, Paul JE, Grant DM, Stergachis A. The role of cost—consequence analysis in healthcare decision—making. *Pharmacoeconomics*. 1998;13(3):277-288.

173. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Econ*. 2006;15(7):677-687.

174. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, et al. Modelling in ecomomic evaluation: an unavoidable fact of life. *Health Econ*. 1997;6(3):217-227.

175. Lipscomb J, Torrance G, Gold M, Siegel J, Russell L, Weinstein M. Time preference. Cost-effectiveness in health and medicine. 214–264. New York: Oxford University Press; 1996.

176. Smith DH, Gravelle H. The practice of discounting in economic evaluations of healthcare interventions. *International Journal of Technology Assessment in Health Care*. 2001;17(02):236-243.

177. Guide to Economic Analysis and Research (GEAR). What are the difficulties in conducting health economic evaluations in low- and middle- income countries? [Cited 2020 February 03]. Available from: http://www.gear4health.com/gear/difficulties-in-conducting-health-economic-evaluations.

178. Ungar W, Gerber A. The uniqueness of child health and challenges to measuring costs and consequences. *Economic Evaluation in Child Health*. 2010:3-32.

179. Huang I-C, Revicki DA, Schwartz CE. Measuring pediatric patient-reported outcomes: good progress but a long way to go. *Quality of Life Research*. 2014;23(3):747-750.

180. Riley AW. Evidence that school-age children can self-report on their health. *Ambulatory Pediatrics*. 2004;4(4):371-376.

181. Cremeens J, Eiser C, Blades M. Characteristics of health-related self-report measures for children aged three to eight years: a review of the literature. *Quality of Life Research*. 2006;15(4):739-754.

182. Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. *Pediatrics*. 2005;115(5):e600-e614.

183. Stevens K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Quality of Life Research*. 2009;18(8):1105-1113.

184. Stevens K. Valuation of the child health utility 9D index. *Pharmacoeconomics*. 2012;30(8):729-747.

185. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Quality of Life Research*. 2010;19(6):875-886.

186. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burström K, Cavrini G, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Quality of Life Research*. 2010;19(6):887-897.

187. Tarride J-E, Burke N, Bischof M, Hopkins RB, Goeree L, Campbell K, et al. A review of health utilities across conditions common in paediatric and adult populations. *Health and Quality of Life Outcomes*. 2010;8(1):12-23.

188. Belfort MB, Zupancic JA, Riera KM, Turner JH, Prosser LA. Health state preferences associated with weight status in children and adolescents. *BMC Pediatr*. 2011;11(1):12-19.

189. Varni JW, Seid M, Rode CA. The PedsQL™: measurement model for the pediatric quality of life inventory. *Med Care*. 1999:126-139.

190. Varni JW, Seid M, Smith Knight T, Burwinkle T, Brown J, Szer IS. The PedsQL™ in pediatric rheumatology: reliability, validity, and responsiveness of the Pediatric Quality of Life Inventory™ Generic Core Scales and Rheumatology Module. *Arthritis & Rheumatism*. 2002;46(3):714-725.

191. Wu C, Gong Y, Wu J, Zhang S, Yin X, Dong X, et al. Chinese Version of the EQ-5D Preference Weights: Applicability in a Chinese General Population. *PloS One*. 2016;11(10):e0164334.

192. Dolan P. The measurement of health-related quality of life for use in resource allocation decisions in health care. *Handbook of Health Economics*. 2000;1:1723-1760.

193. Brazier J, Rowen D. Alternatives to EQ-5D for generating health state utility values: The Decision Support Unit, University of Sheffield; 2011 [cited 2018 June 10]. Available from: https://pdfs.semanticscholar.org/d7f2/540da47b90b2510fb522f6b775e9bba49fd6.pdf.

194. Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology*. 1998;51(11):1115-1128.

195. Frew E, Breheny K. Methods for public health economic evaluation: A Delphi survey of decision makers in English and Welsh local government. *Health Econ*. 2019;28(8):1052-1063.

196. Carter R, Moodie M, Markwick A, Magnus A, Vos T, Swinburn B, et al. Assessing cost-effectiveness in obesity (ACE-Obesity): an overview of the ACE approach, economic methods and cost results. *Bmc Public Health*. 2009;9:419-430.

197. Blake-Lamb TL, Locks LM, Perkins ME, Baidal JAW, Cheng ER, Taveras EM. Interventions for childhood obesity in the first 1,000 days a systematic review. *American Journal of Preventive Medicine*. 2016;50(6):780-789.

198. Seburg EM, Olson-Bullis BA, Bredeson DM, Hayes MG, Sherwood NE. A review of primary care-based childhood obesity prevention and treatment interventions. *Curr*. 2015;4(2):157-173.

199. Wang Y, Cai L, Wu Y, Wilson R, Weston C, Fawole O, et al. What childhood obesity prevention programmes work? A systematic review and meta-analysis. *Obesity reviews*. 2015;16(7):547-565.

200. Erdol S, Mazzucco W, Boccia S. Cost effectiveness analysis of childhood obesity primary prevention programmes: a systematic review. *Epidemiology, Biostatistics and Public Health*. 2014;11(3).

201. Gortmaker SL, Long MW, Resch SC, Ward ZJ, Cradock AL, Barrett JL, et al. Cost effectiveness of childhood obesity interventions evidence and methods for CHOICES. *American Journal of Preventive Medicine*. 2015;49(1):102-111.

202. John J, Wolfenstetter SB, Wenig CM. An economic perspective on childhood obesity: recent findings on cost of illness and cost effectiveness of interventions. *Nutrition*. 2012;28(9):829-839.

203. Korber K. Quality assessment of economic evaluations of health promotion programs for children and adolescents-a systematic review using the example of physical activity. *Health Econ Rev*. 2015;5(1):35.

204. Lobstein T, Jackson-Leach R, Moodie ML, Hall KD, Gortmaker SL, Swinburn BA, et al. Child and adolescent obesity: part of a bigger picture. *Lancet*. 2015;385(9986):2510-2520.

205. Robertson W, Fleming J, Kamal A, Hamborg T, Khan KA, Griffiths F, et al. Randomised controlled trial evaluating the effectiveness and cost-effectiveness of 'families for health', a family-based childhood obesity treatment intervention delivered in a community setting for ages 6 to 11 years. *Health Technology Assessment*. 2017;21(1):1-18.

206. Sutherland R, Reeves P, Campbell E, Lubans DR, Morgan PJ, Nathan N, et al. Cost effectiveness of a multi-component school-based physical activity intervention targeting adolescents: the 'Physical Activity 4 Everyone' cluster randomized trial. *International Journal of Behavioral Nutrition and Physical Activity*. 2016;13:94-107.

207. Graziose MM, Koch PA, Wang YC, Lee Gray H, Contento IR. Cost-effectiveness of a nutrition education curriculum intervention in elementary schools. *J Nutr Educ Behav*. 2016;49(8):684-691.

208. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. England: John Wiley & Sons; 2011.

209. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine*. 2009;6(7):e1000100.

210. Zanganeh M, Adab P, Li B, Frew E. Protocol for a systematic review of methods and cost-effectiveness findings of economic evaluations of obesity prevention and/or treatment interventions in children and adolescents. *BMC Systematic Reviews*. 2018;54(7):1-7.

211. Barrett JL, Gortmaker SL, Long MW, Ward ZJ, Resch SC, Moodie ML, et al. Cost effectiveness of an elementary school active physical education policy. *American Journal of Preventive Medicine*. 2015;49(1):148-159.

212. Brown IHS, Perez A, Li YP, Hoelscher DM, Kelder SH, Rivera R. The cost-effectiveness of a school-based overweight program. *International Journal of Behavioral Nutrition and Physical Activity*. 2007;4(1):47-58.

213. Long MW, Gortmaker SL, Ward ZJ, Resch SC, Moodie ML, Sacks G, et al. Cost effectiveness of a sugar-sweetened beverage excise tax in the US. *American Journal of Preventive Medicine*. 2015;49(1):112-123.

214. Magnus A, Haby MM, Carter R, Swinburn B. The cost-effectiveness of removing television advertising of high-fat and/or high-sugar food and beverages to Australian children. *International Journal of Obesity*. 2009;33(10):1094-1102.

215. Moodie M, Haby M, Galvin L, Swinburn B, Carter R. Cost-effectiveness of active transport for primary school children - Walking School Bus program. *International Journal of Behavioral Nutrition and Physical Activity*. 2009;6(1):63-73.

216. Moodie ML, Carter RC, Swinburn BA, Haby MM. The Cost-effectiveness of Australia's active after-school communities program. *Obesity*. 2010;18(8):1585-1592.

217. Moodie M, Haby MM, Swinburn B, Carter R. Assessing cost-effectiveness in obesity: Active transport program for primary school children- TravelSMART Schools Curriculum program. *Journal of Physical Activity and Health*. 2011;8(4):503-515.

218. Moodie ML, Herbert JK, de Silva-Sanigorski AM, Mavoa HM, Keating CL, Carter RC, et al. The Cost-effectiveness of a successful community-based obesity prevention program: The Be Active Eat Well Program. *Obesity*. 2013;21(10):2072-2080.

219. Pringle A, Cooke C, Gilson N, Marsh K, McKenna J. Cost-effectiveness of interventions to improve moderate physical activity: A study in nine UK sites. *Health Education Journal*. 2010;69(2):211-224.

220. Rush E, Obolonkin V, McLennan S, Graham D, Harris JD, Mernagh P, et al. Lifetime cost effectiveness of a through-school nutrition and physical programme: Project Energize. *Obesity Research & Clinical Practice*. 2014;8(2):E115-E122.

221. Sonneville KR, Long MW, Ward ZJ, Resch SC, Wang YC, Pomeranz JL, et al. BMI and healthcare cost impact of eliminating tax subsidy for advertising unhealthy food to youth. *American Journal of Preventive Medicine*. 2015;49(1):124-134.

222. Wang LY, Yang QH, Lowry R, Wechsler H. Economic analysis of a school-based obesity prevention program. *Obesity Research*. 2003;11(11):1313-1324.

223. Wang LY, Nichols LP, Austin SB. The economic effect of planet health on preventing bulimia nervosa. *Archives of Pediatrics & Adolescent Medicine*. 2011;165(8):756-762.

224. Wright DR, Kenney EL, Giles CM, Long MW, Ward ZJ, Resch SC, et al. Modeling the cost effectiveness of child care policy changes in the US. *American Journal of Preventive Medicine*. 2015;49(1):135-147.

225. Hollingworth W, Hawkins J, Lawlor DA, Brown M, Marsh T, Kipping RR. Economic evaluation of lifestyle interventions to treat overweight or obesity in children. *International Journal of Obesity*. 2012;36(4):559-566.

226. Moodie M, Haby M, Wake M, Gold L, Carter R. Cost-effectiveness of a family-based GP-mediated intervention targeting overweight and moderately obese children. *Econ Hum Biol*. 2008;6(3):363-376.

227. McAuley KA, Taylor RW, Farmer VL, Hansen P, Williams SM, Booker CS, et al. Economic evaluation of a community-based obesity prevention program in children: The APPLE Project. *Obesity*. 2010;18(1):131-136.

228. Peterson M, Chandlee M, Abraham A. Cost-effectiveness analysis of a statewide media campaign to promote adolescent physical activity. *Health Promotion Practice*. 2008;9(4):426-433.

229. Hayes A, Lung T, Wen LM, Baur L, Rissel C, Howard K. Economic evaluation of "Healthy Beginnings" an early childhood intervention to prevent obesity. *Obesity*. 2014;22(7):1709-1715.

230. Epstein LH, Paluch RA, Wrotniak BH, Daniel TO, Kilanowski C, Wilfley D, et al. Cost-effectiveness of family-based group treatment for child and parental obesity. *Child*. 2014;10(2):114-121.

231. Goldfield GS, Epstein LH, Kilanowski CK, Paluch RA, Kogut-Bossler B. Cost-effectiveness of group and mixed family-based treatment for childhood obesity. *International Journal of Obesity*. 2001;25(12):1843-1849.

232. Hollinghurst S, Hunt LP, Banks J, Sharp DJ, Shield JP. Cost and effectiveness of treatment options for childhood obesity. *Pediatric Obesity*. 2014;9(1):E26-E34.

233. Janicke DM, Sallinen BJ, Perri MG, Lutes LD, Silverstein JH, Brumback B. Comparison of program costs for parent-only and family-based interventions for pediatric obesity in medically underserved rural settings. *J Rural Health*. 2009;25(3):326-330.

234. Wake M, Gold L, McCallum Z, Gerner B, Waters E. Economic evaluation of a primary care trial to reduce weight gain in overweight/obese children: The LEAP trial. *Ambulatory Pediatrics*. 2008;8(5):336-341.

235. Kesztyus D, Schreiber A, Wirt T, Wiedom M, Dreyhaupt J, Brandstetter S, et al. Economic evaluation of URMEL-ICE, a school-based overweight prevention programme comprising metabolism, exercise and lifestyle intervention in children. *European Journal of Health Economics*. 2013;14(2):185-195.

236. Krauth C, Liersch S, Sterdt E, Henze V, Robl M, Walter U. [Health economic evaluation of health promotion - the example "fit for pisa"]. *Gesundheitswesen*. 2013;75(11):742-746.

237. Martinez PM, Lopez MS, Bastida JL, Sotos FE, Pacheco BN, Aguilar FS, et al. Cost-effectiveness of an intervention to reduce overweight and obesity in 9-10-year-olds. The Cuenca study. *Gaceta Sanitaria*. 2011;25(3):198-204.

238. Kalavainen M, Karjalainen S, Martikainen J, Korppi M, Linnosmaa I, Nuutinen O. Cost-effectiveness of routine and group programs for treatment of obese children. *Pediatrics International*. 2009;51(5):606-611.

239. PHARMAC, the Pharmaceutical Management Agency. Prescription for Pharmacoeconomic Analysis: Methods for Cost-utility Analysis (version 2.2). New Zealand;2015.

240. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (version 4.3). Canberra: Australian Government Department of Health and Ageing; 2008.

241. Anderson R. Systematic reviews of economic evaluations: utility or futility? *Health Econ*. 2010;19(3):350-364.

242. Caird J, Kavanagh J, O'Mara-Eves A, Oliver K, Oliver S, Stansfield C, et al. Does being overweight impede academic attainment? A systematic review. *Health Education Journal*. 2014;73(5):497-521.

243. Ryabov I. Childhood Obesity and Academic Outcomes in Young Adulthood. *Children*. 2018;5(11):150.

244. Segal AB, Huerta MC, Sassi F. Understanding the effect of childhood obesity and overweight on educational outcomes: an interdisciplinary secondary analysis of two UK cohorts. *The Lancet*. 2019;394:S84.

245. Frew E. Aligning Health Economics Methods to Fit with the Changing World of Public Health. *Applied Health Economics and Health Policy*. 2017;15(3):287-289.

246. HM Treasury Guidance. The green book, appraisal and evaluation in central government treasury guidance. London: TSO.2013 [cited 2017 April 11]. Available from: http://www.fao.org/ag/humannutrition/33236-040551a7cfbc0e73909932192db580c4.pdf.

247. Musgrove P, Fox-Rushby J. Cost-effectiveness analysis for priority setting. 2nd edition. Washington (DC): World Bank;2006.

248. Sharifi M, Franz C, Horan CM, Giles CM, Long MW, Ward ZJ, et al. Cost-effectiveness of a clinical childhood obesity intervention. *Pediatrics*. 2017;140(5):e20162998.

249. Canaway A, Frew E, Lancashire E, Pallan M, Hemming K, Adab P, et al. Economic evaluation of a childhood obesity prevention programme for children: Results from the WAVES cluster randomised controlled trial conducted in schools. *PloS One*. 2019;14(7):e0219500.

250. Döring N, Zethraeus N, Tynelius P, de Munter J, Sonntag D, Rasmussen F. Economic evaluation of PriMrOse—a Trial-Based analysis of an early childhood intervention to Prevent Obesity. *Frontiers in Endocrinology*. 2018;9:104-111.

251. Larsen KT. Cost-effectiveness of a day-camp weight-loss intervention programme for children: Results based. *Children*. 2017;116:73-77.

252. The Chinese nutrition association. Chinese dietary reference intakes handbook: Beijing: Chinese Standard Publishing; 2013.

253. Gosden TB, Torgerson DJ. Economics notes: converting international cost effectiveness data to UK prices. *BMJ: British Medical Journal*. 2002;325(7358):275.

254. Organisation for Economic Co-operation and Development (OECD) Data: Purchasing power parities (PPP) [Internet]. 2018 [cited 11/05/2018]. Available from: https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm.

255. ChinaDaily. Top 10 Chinese cities with highest average monthly salary China: ChinaDaily; 2017 [cited 2018 December 17]. Available from: http://www.chinadaily.com.cn/bizchina/2017top10/2017-06/23/content_29853826_7.htm.

256. China Briefing. Guangdong’s Minimum Wages to Increase July 1 China: China Briefing; 2018 [cited 2018 December 19]. Available from: http://www.china-briefing.com/news/guangdongs-minimum-wages-rise-july-1/.

257. Amoako-Tuffour J, Martınez-Espineira R. Leisure and the opportunity cost of travel time in recreation demand analysis: a re-examination. 2015 [cited 2018 December 09]. Available from: https://mpra.ub.uni-muenchen.de/8573/1/MPRA_paper_8573.pdf.

258. Topic 1: Wage rates and the supply and demand for labour [cited 2018 December 10]. Available from: https://www.economics.utoronto.ca/jfloyd/modules/sadl.html.

259. Tian X, Yu X, Holst R. Applying the payment card approach to estimate the WTP for green food in China. *LeibnizInstitut für Agrarentwicklung in Mittel- und Osteuropa (IAMO)*. 2011.

260. Kolsgaard MLP, Joner G, Brunborg C, Anderssen SA, Tonstad S, Andersen LF. Reduction in BMI z-score and improvement in cardiometabolic risk factors in obese children and adolescents. The Oslo Adiposity Intervention Study-a hospital/public health nurse combined treatment. *BMC Pediatr*. 2011;11(1):47.

261. World Health Organization. The WHO child growth standards 2007 [

262. Xu F, Chen G, Stevens K, Zhou H, Qi S, Wang Z, et al. Measuring and valuing health-related quality of life among children and adolescents in mainland China–a pilot study. *PloS One*. 2014;9(2):e89222.

263. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ*. 2005;14(5):487-496.

264. Chen G, Xu F, Huynh E, Wang Z, Li C, Stevens K, et al., editors. Scoring the Child Health Utility 9D instrument. Estimation of a Chinese adolescent-specific tariff. Quality of Life Research; 2016: SPRINGER VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS.

265. Chen G, Xu F, Huynh E, Wang Z, Stevens K, Ratcliffe J. Scoring the Child Health Utility 9D instrument: estimation of a Chinese child and adolescent-specific tariff. *Quality of Life Research*. 2019;28(1):163-176.

266. Chen Y, He L, Mai J, Hao Y, Xiong L, Chen W, et al. Validity and reliability of pediatric quality of life inventory version 4.0 generic core scales in Chinese children and adolescents. *Zhonghua Liuxingbingxue Zazhi*. 2008;29(6):560-563.

267. Hemming K, Taljaard M, Forbes A. Modeling clustering and treatment effect heterogeneity in parallel and stepped-wedge cluster randomized trials. *Stat Med*. 2018;37(6):883-898.

268. Liu GG, Wu H, Li M, Gao C, Luo N. Chinese time trade-off values for EQ-5D health states. *Value in Health*. 2014;17(5):597-604.

269. Zhuo L, Xu L, Ye J, Sun S, Zhang Y, Burstrom K, et al. Time Trade-Off Value Set for EQ-5D-3L Based on a Nationally Representative Chinese Population Survey. *Value in Health*. 2018;21(11):1330-1337.

270. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing.... presumed at random: cost-analysis of incomplete data. *Health Econ*. 2003;12(5):377-392.

271. Marshall A, Billingham LJ, Bryan S. Can we afford to ignore missing data in cost-effectiveness analyses? *The European Journal of Health Economics*. 2009;10(1):1-3.

272. Simons CL, Rivero-Arias O, Yu L-M, Simon J. Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index? *Quality of Life Research*. 2015;24(4):805-815.

273. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Med Res Methodol*. 2014;14(1):75.

274. Allison P. Imputation by Predictive Mean Matching: Promise & Peril: Statistical Horizons; 2015 [cited 2019 March 04]. Available from: https://statisticalhorizons.com/predictive-mean-matching.

275. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *Bmj*. 2012;345:e5661.

276. Al-Janabi H, Van Exel J, Brouwer W, Coast J. A framework for including family health spillovers in economic evaluation. *Medical Decision Making*. 2016;36(2):176-186.

277. Al-Janabi H, Van Exel J, Brouwer W, Trotter C, Glennie L, Hannigan L, et al. Measuring health spillovers for economic evaluation: a case study in meningitis. *Health Econ*. 2016;25(12):1529-1544.

278. Bhadhuri A, Al-Janabi H, Jowett S, Jolly K. Incorporating household spillovers in cost utility analysis: A case study using behavior change in COPD. *International Journal of Technology Assessment in Health Care*. 2019;35(3):212-220.

279. Shepard DS. Cost-effectiveness in health and medicine. By MR Gold, JE Siegel, LB Russell, and MC Weinstein (eds). New York: Oxford University Press, 1996. *The journal of mental health policy and economics*. 1999;2(2):91-92.

280. Eminson K, Canaway A, Adab P, Lancashire E, Pallan M, Frew E. How does age affect the relationship between weight and health utility during the middle years of childhood? *Quality of Life Research*. 2018;27(6):1455-1462.

281. Baranowski T, Cullen KW, Nicklas T, Thompson D, Baranowski J. School-based obesity prevention: a blueprint for taming the epidemic. *American Journal of Health Behavior*. 2002;26(6):486-493.

282. Hamzaid H, Talib RA, Azizi NH, Maamor N, Reilly JJ, Wafa SW. Quality of life of obese children in Malaysia. *Int J Pediatr Obes*. 2011;6(5-6):450-454.

283. Bolton K, Kremer P, Rossthorn N, Moodie M, Gibbs L, Waters E, et al. The effect of gender and age on the association between weight status and health-related quality of life in Australian adolescents. *BMC Public Health*. 2014;14(1):898-906.

284. Miranda Fredriks A, Van Buuren S, Hira Sing RA, Maarten Wit J, Pauline Verloove-vanhorick S. Alarming prevalences of overweight and obesity for children of Turkish, Moroccan and Dutch origin in The Netherlands according to international standards. *Acta Paediatrica*. 2005;94(4):496-498.

285. Pallan M, Parry J, Adab P. Contextual influences on the development of obesity in children: a case study of UK South Asian communities. *Prev Med*. 2012;54(3-4):205-211.

286. Canaway A, Frew E. Is utility-based quality of life in children aged 6–7 years affected by Body Mass Index (BMI)? *International Journal of Obesity*. 2014;38(8):1146.

287. Cronbach L, Meehl P. Construct validity in psychological tests. *Psychol Bull*. 1955;52(4):281-302.

288. Barcelona Field Studies Centre. Spearman's rank correlation coefficient Rs and probability (p) value calculator. [cited 2018 November 29]. Available from: https://geographyfieldwork.com/SpearmansRankCalculator.html.

289. Lambe T, Frew E, Ives NJ, Woolley RL, Cummins C, Brettell EA, et al. Mapping the Paediatric Quality of Life Inventory (PedsQL™) Generic Core Scales onto the Child Health Utility Index–9 Dimension (CHU-9D) Score for Economic Evaluation in Children. *PharmacoEconomics*. 2018;36(4):451-465.

290. Khan KA, Petrou S, Rivero-Arias O, Walters SJ, Boyle SE. Mapping EQ-5D utility scores from the PedsQL™ generic core scales. *Pharmacoeconomics*. 2014;32(7):693-706.

291. Wallander J, Kerbawy S, Toomey S, Lowry R, Elliott MN, Escobar-Chaves SL, et al. Is obesity associated with reduced health-related quality of life in Latino, black and white children in the community? *International Journal of Obesity*. 2013;37(7):920.

292. Petrou S. Methodological issues raised by preference-based approaches to measuring the health status of children. *Health Econ*. 2003;12(8):697-702.

293. Chen G, Ratcliffe J, Olds T, Magarey A, Jones M, Leslie E. BMI, health behaviors, and quality of life in children and adolescents: a school-based study. *Pediatrics*. 2014;133(4):e868-e874.

294. Boyle SE, Jones GL, Walters SJ. Physical activity, quality of life, weight status and diet in adolescents. *Quality of Life Research*. 2010;19(7):943-954.

295. Furber G, Segal L. The validity of the Child Health Utility instrument (CHU9D) as a routine outcome measure for use in child and adolescent mental health services. *Health and Quality of Life Outcomes*. 2015;13(1):22-35.

296. Ottova V, Erhart M, Rajmil L, Dettenborn-Betz L, Ravens-Sieberer U. Overweight and its impact on the health-related quality of life in children and adolescents: results from the European KIDSCREEN survey. *Quality of Life Research*. 2012;21(1):59-69.

297. Williams J, Wake M, Hesketh K, Maher E, Waters E. Health-related quality of life of overweight and obese children. *Jama*. 2005;293(1):70-76.

298. Wille N, Bullinger M, Holl R, Hoffmeister U, Mann R, Goldapp C, et al. Health-related quality of life in overweight and obese youths: results of a multicenter study. *Health and Quality of Life Outcomes*. 2010;8(1):36-43.

299. Griffiths LJ, Parsons TJ, Hill AJ. Self-esteem and quality of life in obese children and adolescents: a systematic review. *Int J Pediatr Obes*. 2010;5(4):282-304.

300. Bell A. Designing and testing questionnaires for children. *Journal of Research in Nursing*. 2007;12(5):461-469.

301. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. *Bmj*. 2007;334(7591):455-459.

302. Moore G, Audrey S, Barker M, Bond L, Bonell C, Cooper C, et al. Process evaluation in complex public health intervention studies: the need for guidance. *BMJ Publishing Group Ltd*. 2014;68(2):101-102.

303. Squires H, Boyd K. Use of modelling in the economic evaluation of public health interventions. In: Edwards, R. T. and McIntosh, E. (eds.) Applied Health Economics for Public Health Practice and Research. Series: Handbooks in health economic evaluation. Oxford University Press. 2019.

304. Roberts K, Flaherty S. Review of dietary assessment methods in public health. *National Obesity Observatory: Oxford, UK*. 2010.

305. Li B, Adab P, Cheng KK. Family and neighborhood correlates of overweight and obesogenic behaviors among Chinese children. *International Journal of Behavioral Medicine*. 2014;21(4):700-709.