Economic Evaluation of Population Health Interventions Aimed at Children and Delivered at School

Nicole Renée Soldner Boyer

MSc Public Health, BSc Community Health

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Health Economics and Health Technology Assessment (HEHTA)

Institute of Health and Wellbeing

College of Medical, Veterinary and Life Science

University of Glasgow

October 2017

Abstract

Background: Population health interventions by their nature affect an entire population and are typically delivered outwith of health services and within the community, such as in schools. An example of such interventions are those that aim to improve children's social and emotional wellbeing, which have demonstrated effectiveness in the short-term and potentially the long-term. However, challenges arise when conducting economic evaluations of population health interventions, most notably the difficulties of identifying, measuring, and valuing broader intersectoral costs, health, and non-health outcomes. Economic evaluation in an education context is relatively novel, but could provide decision-makers with information to help them make transparent and consistent decisions about how to allocate limited funds. This thesis examined the role for economic evaluation in school-based interventions and sought to determine appropriate methods for its implementation in addition to examining appropriate child-focused outcome measures. Thus, the overarching research question asked, 'How should the cost-effectiveness of school-based, population health interventions aimed at children be determined?'

Methods: A mixed methods approach to this thesis was used:

(i) a systematic literature review and narrative synthesis to determine which evaluation methods (economic and non-economic) are currently being used in school-based population health interventions;

(ii) a case study to illustrate an economic evaluation (including cost-utility and cost-effectiveness analysis) of a school-based intervention to reflect on the advantages and disadvantages for decision making in this context; and

(iii) an exploration of outcome measures (through mapping validation) for valuing child health and social and emotional wellbeing in school-based programmes to support future evaluation work in this context.

Data for the economic evaluation and mapping validation study were available from a cluster randomised controlled trial of the Roots of Empathy programme in Northern Ireland (Ref: 10/3006/02).

Results: The systematic review found that the methods currently being utilised to evaluate school programmes are varied (including economic evaluation, cost only, and effectiveness only studies), with poor quality reporting for the economic evaluations. Of the few cost-utility analyses in school-based settings identified, none had directly measured health-related quality of life using child measures or values. The case study cost-utility analysis using Child Health Utility 9D of a school-based intervention was found to be cost-effective from the National Health Service perspective with an incremental cost-effectiveness ratio of £11,000 per quality-adjusted life year (confidence interval: -£95,500 to £147,000), however the wide confidence interval demonstrates considerable uncertainty. This uncertainty is likely due to a lack of statistically significant effect that remained at the 36-month follow-up. Cost-effectiveness analysis using child behavioural descriptive measure, the Strengths and Difficulties Questionnaire, resulted in an incremental cost-effectiveness ratio of £197 per unit decrease in total difficulties score (confidence interval: £77 to £471). The Strengths and Difficulties Questionnaire is suitable for measuring social and emotional wellbeing, but is less advantageous for cost-effectiveness decision-making as no consensus has been reached as to what a clinically meaningful change in score represents, nor has a cost-effectiveness threshold been defined. It remains uncertain how these cost-effectiveness results will be interpreted in an education decision-making context where cost-effectiveness thresholds have not been set up. The mapping validation study validated a mapping algorithm to convert the Strengths and Difficulties Questionnaire into child health utility. Using this algorithm provides an option for valuing incremental changes in health-related quality of life against a generally accepted cost-effectiveness threshold from a health service perspective.

Conclusions: Given the findings from the various aspects of work undertaken for this thesis to address population health issues, this thesis identified cost-benefit analysis as currently the most comprehensive method for determining the value for money of school-based public health interventions. Cost-benefit analysis incorporates monetary valuation of multisector outcomes in a final net benefit/loss result allowing clear, consistent, decision-making criteria to be set. Other methods such as cost-consequence

analysis, cost-utility analysis, and multi-criteria decision analysis may also be suitable depending on the decision-making context and problem. This thesis demonstrates a lack of clear decision-making criteria in place for funding allocation decisions in education (e.g. education specific cost-effectiveness thresholds). Furthermore, there is no equitable method currently in place for apportioning the cost of funding public health interventions that generate benefits for multiple sectors. From a health service perspective, directly measuring child health utility using the Child Health Utility 9D is preferred as it is the only preference-based measure developed specifically for children and valued by young people. Mean child health utility can be predicted by mapping from the Strengths and Difficulties Questionnaire. This affords the opportunity to estimate longer-term utility by utilising long-term cohort data that routinely collects the Strengths and Difficulties Questionnaire, as long-term cost-effectiveness of school-based preventive programmes is an area in need of further research. The school setting plays an important role in shaping our young people's futures. Economic evaluation of school-based population health interventions is justified, as schools need to maximise their existing resources in order to give children the best start in life.

Table of Contents

List of Tables

List of Figures

Acknowledgement

I would like to thank my supervisors, Professor Emma McIntosh and Dr Kathleen Boyd for their unwavering support and valuable guidance throughout this entire process. From the beginning, Emma encouraged me to undertake this thesis and has been a steady and accessible mentor throughout my time here at the University of Glasgow. I am lucky to have benefited from Kathleen's supervision twice now; throughout she has played a major role in guiding my development as a researcher. To them both, I am greatly indebted.

I am very grateful to Professor Paul Connolly and Dr Sarah Miller at Queen's University Belfast for their collaboration on the Roots of Empathy trial. They both were such a pleasure to work with and much of this thesis would not be possible without their corporation and support. Dr Seaneen Sloan provided detailed cost data for the RoE programme and her assistance is much appreciated.

I would also like to acknowledge Professor Helen Minnis who provided her expert opinion in matters regarding child and adolescent mental wellbeing. There were also those who provided valuable advice, feedback, and pastoral support throughout such as: Dr Marion Henderson, Professor Olivia Wu, Dr Craig Melville, and Dr Breda Cullen.

To all my wonderful colleagues and friends at HEHTA, thank you for providing a supportive environment, cups of coffee, cakes, and chat. Doing a PhD can be a lonely undertaking, thanks for being there for me.

Finally, I must thank my husband Marc for his love, support, and encouragement. You bring out the best in me and I am so grateful to have you as my partner in life. And to our future addition whom we are expecting in October, thank you for giving me that extra kick (or two) enabling me to finish in a timely manner.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signed:

Printed name: Nicole Renée Soldner Boyer

Publications, Working Papers and Presentations

The following publications, working papers and presentations were developed as part of this thesis:

Publications

Boyer NR, Miller S, Connolly P, McIntosh E. Paving the way for the use of the SDQ in economic evaluations of school-based population health interventions: an empirical analysis of the external validity of SDQ mapping algorithms to the CHU9D in an educational setting. *Quality of Life Research* 2016;**25**:913-23. http://dx.doi.org/10.1007/s11136-015-1218-x

Working Papers

Connolly P, Miller S, Kee F, Sloan S, Gildea A, McIntosh E, Boyer N, Bland M. A cluster randomised controlled trial evaluation and cost-effectiveness analysis of the Roots of Empathy schools-based programme for improving social and emotional wellbeing outcomes among 8-9 year olds in Northern Ireland. Submitted August 2016. *Under review in NIHR Journals Library PHR. Report No: 10/3006/02*

Boyer N, Miller S, Connolly P, McIntosh E. Population health economic evaluation of the school-based Roots of Empathy programme in Northern Ireland. Health Economists' Study Group, Gran Canaria, June 2016.

Boyer N, Miller S, Connolly P, McIntosh E. Children's model planes, trains and *health outcomes?* Developing a model to evaluate social and emotional outcomes on the longer-term effects of 'Roots of Empathy.' Health Economists' Study Group, Glasgow, June 2014.

Conference Presentations

Boyer N, Miller S, Connolly P, McIntosh E. Population health economic evaluation of the school-based Roots of Empathy programme in Northern Ireland. European Conference on Health Economics (EUHEA), July 2016.

Boyer N, Miller S, Connolly P, McIntosh E. Use of the Child Health Utility (CHU9D) and Strengths and Difficulties Questionnaire (SDQ) outcome measures in economic evaluations of school-based interventions: data from a cluster-randomised controlled trial in Northern Ireland. The Lancet Public Health Science Conference, Glasgow, November 2014.

Abbreviations

16D	16 Dimensional
A&E	accident and emergency
AQoL-6D	Assessment of Quality of Life Mark 2 6D adolescents
ASCOT	Adult Social Care Outcome Toolkit
AUC	area under the curve
AV	adult values
BMJ	British Medical Journal
BNF	British National Formulary
CAMHS	Child and Adolescent Mental Health Services
CBA	cost-benefit analysis
CCA	cost-consequence analysis
CE Plane	cost-effectiveness plane
CEA	cost-effectiveness analysis
CEAC	cost-effectiveness acceptability curve
CER	cost-effectiveness ratios
CHEC	Consensus on Health Economics Criteria List
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CHSCSPS	Comprehensive Health Status Classification System - Preschool
CHU9D	Child Health Utility 9 Dimensions
CI	confidence interval
CMA	cost-minimisation analysis
CPRD	Clinical Practice Research Datalink
CRD	Centre for Reviews and Dissemination
CSO	Chief Scientist Office
CUA	cost-utility analysis
D.A.R.E	Drug Abuse Resistance Education
DALY	disability-adjusted life year
DCE	discrete choice experiments
EEF	Education Endowment Foundation
EQ-5D-Y	EuroQol 5D Youth
ESRC	Economic and Social Research Council
GBP	British Pound
GDP	gross domestic product
GLM	generalised linear models
GP	general practitioner
HCHS	Hospital and Community Health Services
HIA	health impact assessment
HRQoL	health-related quality of life
HTA	health technology assessment
HUI	Health Utilities Index
ICC	intracluster/intraclass correlation coefficient
ICECAP	Investigating Choice Experiments for the Preferences of Older people – CAPability
ICER	incremental cost-effectiveness ratio
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MAPS	Mapping onto Preference-based measures reporting Standards
MAR	missing at random

MCAR	missing completely at random
MCDA	multi-criteria decisions analysis
MeSH	medical subject headings
MI	multiple imputation
MLM	multilevel model
MSG	monosodium glutamate
NHB	net health benefit
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NIMDM	Northern Ireland Multiple Deprivation Measure
NMAR	not missing at random
OECD	Organisation for Economic Cooperation and Development
OLS	ordinary least squares
ONS	Office for National Statistics
PedsQL	Pediatric Quality of Life Inventory
PHI	population health intervention
PICOS	Participants, Interventions, Comparators, Outcomes and Study design
PMM	predictive mean matching
PPP	purchasing power parities
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	patient-reported outcome measure
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life year
QHES	Quality of Health Economic Studies List
RCT	randomised controlled trial
RMSE	root mean squared error
RoE	Roots of Empathy
ROI	return on investment
SA	sensitivity analysis
SD	standard deviation
SDQ	Strengths and Difficulties Questionnaire
SEL	social and emotional learning
SEW	social and emotional wellbeing
SG	standard gamble
SIA	social impact assessment
SIGN	Scottish Intercollegiate Guidelines Network
SMC	Scottish Medicines Consortium
SMD	standardised mean difference
SROI	social return on investment
TTO	time trade-off
UK	United Kingdom
USA	United States of America
USD	US dollars
WHO	World Health Organization
WTP	willingness-to-pay

1 Introduction

In a current socio-political climate where pressures on healthcare costs ensure that physical wellbeing and reactive treatment take precedent, mental health is a key, yet often ignored component of health and wellbeing, and the younger generation is no exception.[1] Social and emotional wellbeing (SEW) allows children to build and maintain positive relationships and handle interpersonal situations constructively. It also lays the foundations for healthy behaviours and educational attainment by preventing behavioural and mental health problems from developing.[2] The importance of children's SEW is gaining increased attention in educational and policy circles with growing evidence linking early SEW to later academic performance and various health outcomes including mental health.[3-5] Research suggests social-emotional competency at a young age is associated with increased wellbeing and school performance, while problems with these competencies can lead to personal, social, and academic difficulties.[6, 7] Children with emotional and behavioural problems are more likely to develop mental health disorders,[8] be involved in crime or violence,[9] practice unsafe sex, and misuse drugs and alcohol.[10] Children with low levels of SEW may also display antisocial behaviours which have been linked to poorer overall health and increased odds of developing cardiovascular problems, wheezing, cancer, and serious injury as an adult.[11] Children with a clinical diagnosis of a mental health disorder are also more costly to society, with significantly higher public sector costs and lower overall quality of life.[12]

Given this plethora of negative outcomes arising from low SEW, the role of school-based social and emotional learning (SEL) programmes to improve SEW as a means to promote children's success in school and life are of increasing interest. SEL programmes help children recognise and manage their emotions, understand the perspective of others, and make responsible decisions.[13] Various SEL programmes have demonstrated positive impacts on social emotional competencies and academic performance, as well as reductions in problem behaviours such as antisocial conduct and hyperactivity.[14-16] The *You Can Do It! Early Childhood Education Program* (YCDI)[14] found that the programme increased social and emotional competence, wellbeing, and reading achievement, while it decreased problem behaviours such as externalising, internalising, and hyperactivity. The *Fast Track PATHS (Promoting Alternative Thinking Strategies)*[16] programme found modest positive effects on increases prosocial behaviour and decreased aggressive behaviour.

Students who participated in *Project Attitude*[15] self-reported positive results in social awareness, self-control, self-esteem, social isolation, and social anxiety; these results were replicated by teacher-report. None of the previous examples examined cost-effectiveness, and as will be seen in Chapter 3, very few SEL/SEW programmes have been evaluated with economic evaluation methods. Schools have long been recognised as an ideal setting for health education and promotion as they are efficient in reaching the majority of young people and play an important role in developing and maintaining children's social lives and interactions.[17] A recent meta-analysis (Durlak et al.) of school-based SEL programmes found participants to have significantly improved social and emotional skills, attitudes, behaviours, and academic performance.[18] In this analysis, the effects diminished at follow-up, but remained statistically significant for six months after intervention.[18] Few studies report follow-up longer than six months[18] and there is little evidence of cost-effectiveness and long-term effectiveness. The Department for Education's overall school budget in the United Kingdom (UK) is relatively protected, but does not increase in line with inflation, and projected increased student numbers results in schools needing to make up an estimated £3 billion in savings to alleviate these cost pressures.[19] These real-term reductions to publically funded education have resulted in scarce resources needing to be maximised to their full potential.

Economic evaluation (further detail in section 2.1) can help education decision-makers make more informed decisions about how to allocate limited funds. The Durlak et al. meta-analysis[18] highlighted a gap between the research on effective school-based SEL and actual practice and implementation of these programmes. Additionally, the study highlighted the need to document costs and benefits of SEL programmes as well as the fact that future studies must include cost analyses in their evaluation designs. This demonstrated need for cost-effectiveness evidence in the area of SEW identifies a gap in the current knowledge, leaving decision-makers less informed about the cost-effectiveness of new SEL programmes they might choose to implement. The majority of practical economic evaluations have been conducted in healthcare and related settings, as well as transport sectors. In healthcare settings, cost-effectiveness decisions are based on health outcomes, whereas the transport sector typically values outcomes in monetary units.[20] Decision-making across sectors (e.g. involving both health and education sectors), and how to appropriately value different sectoral outcomes are key concepts to be investigated throughout this thesis. The long-term broader impacts of school-based SEW

programmes on educational outcomes, health behaviours, adult unemployment, crime, and health related outcomes are important to identify as these potential impacts inform any comprehensive economic evaluation of SEW programmes.

SEL/SEW programmes are considered population health interventions (PHIs) because they have the potential to impact on an entire school population. Population and public health are often used interchangeably and for the purpose of this thesis, PHI is defined as '…policy and program interventions that operate within or outside of the health sector and have the potential to impact health at the population level.'[21] A formal definition of population health is given later in section 2.2. A recent systematic review of the return on investment (ROI)[a] of PHIs found cuts to public spending in high-income countries representing a false economy.[22] Short-sighted policy decisions may contribute to this phenomenon as PHIs often have broader, long-term effects to society that may never be fully realised if funding is cut due to a lack of or small effectiveness gains in short-term outcomes.[23]

Pre-school and early years interventions aimed at low-income and socially deprived children have demonstrated long-term effectiveness[24-26] and cost-effectiveness.[27] Comprehensive education, family, and health services delivered in the early pre-school years have demonstrated higher rates of high school and education completion, lower rates of juvenile and violent arrests, and fewer school dropouts at age 18 and later.[24, 28]

If school-based SEL programmes have the potential to impact on immediate and longer-term adult outcomes, investment in such programmes would appear to be warranted. However, simply allocating more money to education does not necessarily result in increased education attainment,[29] and it is important that these new SEL programmes are vigorously evaluated for cost-effectiveness, particularly in times of constrained education budgets, as is the case today. The overall aim of economic evaluation is to aid decision makers to maximise benefits, given the resources available, and make sure no resources are wasted in the process.[23] Decision-making across and between multi-sectoral budgets is a challenge (as will be described in section 2.2.3) and this thesis will examine

[a] Formal definition of return on investment give in section 2.1.1

appropriate methods to deal with these challenges as compared to decision making in the healthcare sector.

Economic evaluation within the healthcare sector has long been established, however this is rarely the case in education sectors where new education initiatives may involve significant expenditure.[29] As will be seen in Chapter 3, current examples of economic evaluation in school settings are limited and of varying quality. There is much scope for broadening decision-making and reporting of school-based economic evaluation and this thesis will discuss this through an example case study of a comprehensive economic evaluation of a school-based SEW programme. These points are significant as the novelty of school-based economic evaluation brings opportunities for more informed decision-making and resource allocation in school settings. The school is a novel setting for economic evaluation, and with this novelty comes new challenges, for example establishing the most appropriate multi-sector funding strategy when SEL programmes give rise to education and health benefits. UK guidance states, 'no standard method has yet been devised to apportion costs - and who should bear them - when more than one government department (or, indeed, local authority) is involved. This may prove particularly difficult when one national or local authority department secures the benefits of a public health intervention, but another is required to fund it' (p. 5[30]).

In education economics literature, fundamental work on human capital theory by Schultz,[31] Becker,[32] and Mincer[33] has long made the economic case for education. Human capital theory is based on the assumption that education serves as an investment into individual knowledge and skills,[34] which then contribute to individual successes in the labour market and productivity. Investment in SEL can therefore be viewed as investment in human capital as well as health and wellbeing. This investment gives rise to multiple benefits in various sectors of society as detailed above. It is therefore difficult to expect the onus of investment in SEL programmes to rely solely on the education sector. To implement a new SEL programme might require additional time and resources to be diverted away from traditional school subjects, negatively affecting students' learning in those other areas.[35] Therefore, an important question arising is, 'Who should pay for implementing PHIs when multiple sectors stand to benefit from the intervention?' The school could potentially be compensated by the health sector if the resulting health benefit is greater than the loss to other education subjects.[35] This approach, referred to

as 'cofinancing,' has been suggested by Remme, et al.[35] as a means of redistributing parts of the healthcare budget to other sectors that achieve health gains more efficiently than then would be accomplished within the health sector. This would work the other way as well with other sectors (such as education) transferring some of their budget to PHIs which generate benefits that they are interested in. This approach will be explored further in section 5.7.

Economic evaluation of PHIs provide their own separate set of challenges as they often produce health and non-health benefits which are difficult to identify and value appropriately. The longer-term outcomes produced by a preventative PHI could span multiple sectors such as health, education, justice, housing, transport, and the broader economy; identifying appropriate outcomes to measure in the short-term can be challenging in addition to valuing these health and non-health outcomes.[36] This is a clear distinction from a traditional economic evaluation in a healthcare setting where all outcomes are often more narrowly focused on health. Generally, however, the primary outcomes from a SEL programme will be health related.

Exploring the use of appropriate paediatric outcomes measures is also important for establishing the cost-effectiveness of SEL programmes because the development of outcomes specifically aimed at children has lagged behind the development of adult measures.[37] It is important to establish which child health outcomes are appropriate for measuring generic preference-based health-related quality of life (HRQoL) and intervention specific outcomes of SEL programmes as each will have implications on the type of economic evaluation that can be performed (to be covered more thoroughly in sections 2.1.3 and 2.1.4). Preference-based child HRQoL measures are useful because they can be compared within and across health related areas.[38] Research into these types of outcomes has typically been limited due to challenges of in obtaining preference-based valuation from children.[37] In establishing effectiveness of SEL programmes, decision-makers may also be interested in descriptive measures of SEW, or often times a common descriptive measure will have been used in a trial which lacked a preference-based measure. Mapping from a descriptive measure to a preference-based measure has been suggested as a way to derive utilities (and therefore make comparisons across health related areas) in situations where preference-based measures have not been collected.[39] Child outcomes research is an important area within the PHI school-based context as they

can generate a broader range of benefits; therefore, to avoid underestimating the benefits of a PHI,[35] suitable outcomes need to be identified and appropriately measured. These and other challenges (including those related to economic evaluation within a school setting) will be addressed throughout this thesis. The school plays an important role in shaping our young people's futures. With public funding consistently under stress, now more than ever, schools need to maximise their existing resources.

The remainder of this chapter will introduce Roots of Empathy (RoE), which is a school-based SEL programme that is the focus of this thesis. There are many existing SEL programmes as evidenced by the meta-analysis mentioned above (n=213), [18] and more are currently being developed. However, RoE benefits from having extensive effectiveness evidence, [40-46] as it has been an established programme for over 20 years. Its effectiveness has also been established internationally, but it has never been evaluated for cost-effectiveness so this is an important area of research that will be covered in later chapters. SEL and SEW programmes are numerous[18] making the task of choosing and implementing the right programme for individual school needs difficult for decision-makers and funding bodies. Additionally, there is the cost of implementing and running the programme that needs to be considered, so having that information combined with effectiveness evidence is a key component in the decision-making process. As will be seen in Chapter 3, very few SEW programmes have been evaluated for their cost-effectiveness (n=8 identified from systematic review), so providing one of the first economic evaluations of a SEW programme (in Chapter 4 and 5) will be key to assisting decision-makers and funders in education. Following on from the RoE introduction, section 1.2 reports the aim and research question for this thesis. The final section concludes with an outline of what will follow in each remaining chapter.

1.1 Roots of Empathy

A substantial body of evidence now exists to suggest that well designed school-based prevention programmes can be effective in improving a variety of social, health, and academic outcomes.[47, 48] Several reviews have been conducted on SEL programmes and the consensus is that they positively impact on child outcomes such as improved self-esteem, positive social behaviour, social skills, academic performance; and reduced aggressive or disruptive behaviour, conduct problems, suicide, and emotional distress.[49-52]

Roots of Empathy (RoE) is a universal school-based SEL programme that was originally developed and implemented in Canada over 20 years ago and was only recently introduced into the UK. It aims to increase empathy, prosocial behaviour, and decrease aggressive behaviour in children.[53] At the heart of the programme is the development of empathy. Empathy is the ability to identify and to some extent experience the feelings and thoughts of others. It forms the basis of helping and prosocial behaviours and is essential to building successful social relationships during all stages of life. In contrast, the absence of empathy leads a person to consider their own needs without consideration of the feelings of others resulting in asocial or antisocial behaviour.[54]

RoE is amongst a small number of named universal school-based SEL programmes that has an existing evidence base regarding its effectiveness as referenced above. A number of evaluations of RoE have been conducted to date and details are provided elsewhere.[55] RoE is delivered on a whole-class basis for a single academic year and consists of a monthly classroom visit by an infant and parent, typically recruited from the local community, whom the class 'adopts' at the start of the school year. Children learn about the infant's growth and development via interactions and observations with the infant at these monthly visits. A characteristic of RoE is that it is a mentalisation-based programme. Mentalisation is the ability to focus on mental states in oneself and others to understand behaviour.[56] The labelling of feelings and exploration of the relationship between feelings and behaviour is achieved through observation of the mother-infant interaction in the classroom. Clearly, the infant cannot communicate in words and can only express his/her feelings through their behaviour. For this reason, the infant in RoE provides an ideal opportunity for children to learn mentalisation skills through interpreting and labelling the infant's emotions. This then helps them identify and label their own emotions and those of others. They learn affective and cognitive components of empathy, enabling them to empathise with others.

In total, the programme consists of 27 lessons delivered throughout the academic year. Each month a trained RoE instructor, who is not the class teacher, visits the classroom three times for a pre-family visit; the visit of the parent and infant; and a post-family visit. In the cluster randomised controlled trial of RoE in Northern Ireland,[55] instructors undergo a total of four days intensive training that is delivered directly by a specialist RoE trainer from Canada. The specialist trainer also provides on-going mentoring support via

regular telephone calls to all instructors. In addition, on-going support is also available to each instructor through each Health and Social Care Trust's lead RoE coordinator. Each RoE lesson takes place in the classroom with the teacher present but not actively involved in delivery. The programme provides opportunities to discuss and learn about the different dimensions of empathy such as emotion identification and explanation; perspective-taking; and emotional sensitivity. The parent-infant visit serves as a springboard for discussions about understanding feelings and infant development and effective parenting practices. The intervention is highly manualised and any adaptation or tailoring of either the content or method of delivery is discouraged by the RoE organisation.

RoE is considered a PHI because if implemented year after year, the whole school population would be impacted by the intervention. PHIs can often be complex, with multiple interacting components. This can make identifying the 'active ingredients' which are responsible for the success of an intervention difficult.[57] Complexity can refer to two different constructs, the complexity of the intervention and/or the complexity of the system in which the intervention is given; distinguishing between the two can have important consequences for economic evaluation.[58] When the intervention is complex, as long as health economists can quantify the inputs and outputs appropriately it does not matter how the intervention works. However, if the system is complex, evaluating efficiencies from changing components of the system is much more complicated.[58] Implementation issues within complex systems continues to be a substantial challenge for PHIs.[59] Complex health system interventions are characterised by the presence of several characteristics such as: 1.) having several interacting components; 2.) targeting groups or organisations versus individuals; 3.) having numerous and variable outcomes; 4.) the use of feedback and a degree of flexibility or tailoring of the intervention permitted; and 5.) the effectiveness may be impacted by the behaviours of those delivering or receiving the intervention.[60] RoE is susceptible to all of these characteristics, even number 4 as it has been implemented worldwide and a certain degree of flexibility is necessary to adapt the programme to specific cultural and social contexts. This complexity means that RoE may not fit neatly within the current methods of economic evaluation which focus on maximising health gains[61] as there are other non-health outcomes which may be impacted such as those relating to education attainment. A content analysis of published evaluations of complex interventions found the interaction between the intervention and

its context to be a main source of complexity as the people involved may be key to the intervention's success.[62] The analysis also found an emphasis on moving away from the use of primary outcome measures to a multi-criteria framework that can acknowledge multiple objectives of a complex intervention. Complex interventions require complex evaluations, therefore ample time and resource needs to be allocated to the economic evaluation of such programmes.[63] The effectiveness of RoE has been established in different contexts globally, however to date the cost-effectiveness of the programme has never been evaluated. There are additional costs relating to ensuring programme fidelity when RoE is implemented outside of Canada, therefore determining cost-effectiveness is a key concern in a UK context.

As was mentioned in the previous section, economic evaluation of school-based programmes is relatively novel, yet it would provide decision-makers with important information regarding cost-effectiveness of school programmes under consideration for implementation. If establishing longer-term benefits of a programme is a key concern, appropriate outcome measures must be identified to evaluate the programme. Ideally, these measures would be established in longitudinal evidence available in the literature.

1.2 Research question and aims

This thesis will examine the role of economic evaluation in school-based interventions and determine appropriate methods and outcomes for its implementation. Specifically the overarching research question asks,

> 'How should the cost-effectiveness of school-based, population health interventions aimed at children be determined?'

To understand how cost-effectiveness should be determined, this thesis is split into three main empirical works which together aim to answer this research question. Each empirical work has an associated overall aim; these aims are to:

(i) determine what evaluation methods (economic and non-economic) are currently being used to evaluate school-based population health interventions;

(ii) illustrate a good practice example of a thorough cost-utility and cost-effectiveness analysis of a school-based intervention (the RoE programme) to reflect on the advantages of such practice and disadvantages that remain, such as decision-making in multisectoral settings; and

(iii) explore which outcomes are appropriate for children in the SEW and economic evaluation context to support future evaluation work in this context.

The first overall aim will be addressed through systematic review and narrative synthesis of evaluation methods that are currently being implemented in school-based PHIs available in the literature. The second aim will be address through a case study of a comprehensive economic evaluation of RoE. This case study will demonstrate the advantages of conducting economic evaluation in a school-based SEL setting as well as identify the issues that remain when applying the traditional methods of health economic evaluation to an intervention in an education setting. The final aim will explore appropriate outcome measurement in relation to the cost-utility and cost-effectiveness analyses of RoE to support future evaluation work including modelling long-term cost-effectiveness of SEL programmes. This work is facilitated through the validation of mapping from a SEW specific outcome measure to a generic child HRQoL measure. In addition to the overarching research question and aims, each of the three empirical works individually have their own specific aims and research questions, which are addressed separately within each section.

1.3 Thesis outline

As outlined earlier, there is a clear need for the economic evaluation of PHIs in school-based settings due to the lack of cost-effectiveness evidence in the SEW context. Additionally, because school-based PHIs are aimed at children, appropriate paediatric measures are needed for evaluation. Chapter 2 introduces these and the main concepts to be covered in this thesis. It consists of two parts; the first is an introduction to the various methods of economic evaluation. The chapter starts with a brief introduction to economics as a disciple and gives definitions for key terms used throughout this thesis. This is followed by a brief history of the early development of economic evaluation methods for healthcare programmes populated with examples throughout history prior

to the 1970s. Next, each method for economic evaluation is detailed in turn covering: cost-effectiveness analysis (CEA), cost-utility analysis (CUA), cost-benefit analysis (CBA), cost-minimization analysis (CMA), and cost-consequence analysis (CCA). Finally, the role of economic evaluation within a UK healthcare context and internationally is described. The second part of Chapter 2 details economic evaluation of PHIs, a main concept of this thesis. Determining the cost-effectiveness of PHIs is covered by detailing appropriate and emergent methodologies for evaluation of such programmes. PHIs are often preventative by nature, and thus the economic case for prevention is detailed as well as the challenges of conducting economic evaluation of these types of programmes. The final section in Chapter 2 details economic evaluation in a school setting while drawing upon educational economics literature.

Chapter 3 explores the current state of evaluation of school-based PHIs within the published and grey literature. A systematic review and narrative synthesis is presented to determine what economic and non-economic evaluation methodologies are currently being used for school-based programmes. As economic evaluation in this setting is novel, a broad approach was taken to identify all evaluation methodologies. This would help inform the practical application of an economic evaluation of the RoE programme.

Using the results from Chapter 3 (and identified gaps in the literature), a comprehensive economic evaluation of the RoE programme was designed to provide a case study of an example of one of the first comprehensive economic evaluations of a school-based SEL programme. The purpose of this case study within this thesis is to demonstrate the advantages economic evaluation can bring to school settings while identifying potential challenges to consider for future evaluations in this context. Chapters 4 detailed the methods and Chapter 5 described the results of the economic evaluation of the RoE programme. Chapter 4 starts by describing economic evaluation in child health and key considerations that differ from the evaluation of adult interventions, particularly the need for paediatric outcome measures, such as those specific to SEW as well as generic health outcome measures. The RoE trial is detailed along with the methods for the economic evaluation. Chapter 5 details the results starting with a descriptive analysis followed by the costs, outcomes, missing data analysis, and cost-effectiveness results of all sensitivity analyses performed. A thorough discussion follows highlighting the advantages and

challenges of implementing economic evaluation in a school setting, as well as the limitations and conclusion of this case study.

Determining appropriate outcomes for child-focussed economic evaluation in the SEW context is one of the overall aims of this thesis. Paediatric, child-focussed economic evaluation outside of a healthcare context (school setting) is novel and as highlighted previously, appropriate child outcomes are needed to measure these benefits. Specifically, for cost-utility analysis, UK guidance advises use of a standardised and validated preference-based HRQoL measure that has been designed specifically for use in children.[64] This is because there are risks of compromising validity and psychometric properties when modifying adult measures for use with children.[37]

Chapter 6 details the final empirical work, which examines the appropriate use of paediatric outcomes in cost-utility analysis. A commonly used non-preference based outcome measure, the Strengths and Difficulties Questionnaire (SDQ), was mapped to the generic, preference-based, Child Health Utility 9D (CHU9D) using previously developed mapping algorithms to validate and explore their generalisability in an external dataset. This has implications for the economic evaluation of future child-focussed and school-based PHIs as the SDQ is routinely collected in many large datasets. The work in this chapter, which validates the use of these mapping algorithms, allows analysts the opportunity to conduct CUA using a non-preference based outcome measure which is commonly used in SEW research. As cost-effectiveness evidence in SEW is lacking, this final empirical work provides a potential solution to allow both retrospective and prospective cost-utility analysis of SEL/SEW programmes which used the SDQ.

Chapter 7 summarises each previous chapter and discusses the strengths and limitations of each of the three methodological works in turn. A critique of the methods critically appraises the work of the author, this thesis, and that of other authors in related fields. Chapter 8 provides the overall conclusions for this body of work including implications for policy and practice, recommendations, and areas for further research.

2 Economic Evaluation and Population Health Economic Evaluation

Modern advances in health care have led to marked increases in life expectancy and quality of life.[65] However, as a result, healthcare costs are rising worldwide and many fear the rising costs will be unsustainable.[66, 67] In order to combat rising healthcare costs, government and national bodies need to make tough choices about how to organise scarce health resources and which treatments and services should be offered to the public. Economic evaluation in the healthcare context exists to aid decision makers in making these tough choices.

Economic evaluation is a relatively new discipline, with many of the methods being developed in the last 50 years, and a majority of the practical application of the methods appearing in the published literature within the last 20 to 30 years.[68] Traditional methods of economic evaluation have focused on health benefits and have often used a narrower health provider perspective, focusing on ways to value health benefits by eliciting preferences from the general public.[68] Economic evaluation of PHI's represents a marked transition from the traditional more 'clinical' evaluation.[36]

In the most general sense, population health refers to the health outcomes of a defined group of people and how those outcomes are distributed among that group. Therefore, a PHI is an initiative that affects a whole population. In the first chapter, SEL/SEW programmes were introduced as PHIs for children. Another example might be a national policy change to encourage healthier behaviour in the population such as an indoor smoking ban. An example of a PHI aimed at children might be a school-based programme to encourage healthy diet and physical activity of schoolchildren. There are distinct differences and challenges to consider when conducting economic evaluation of PHIs because there will be wider health and non-health benefits arising from these types of initiatives. This chapter introduces two main themes of this PhD thesis, traditional methods of economic evaluation in a healthcare context and considerations for how those methods should be adapted for PHIs.

The chapter starts by introducing basic fundamental concepts of economics and defines key terms used throughout this thesis. A brief history is given of the beginnings of health

economics before the discipline was formally recognised, followed by a description of the formal methods of economic evaluation and its role in national and international decision making contexts. The second part of this chapter (2.2) covers PHIs; detailing appropriate methodologies, the economic case for promoting preventive population health initiatives, and the challenges associated with the conduct of economic evaluation of PHI's.

2.1 Economic evaluation methods

2.1.1 Introduction and definitions

What is economics? A good starting point is the well-established definition by Lord Robbins in 1932; economics is 'the science which studies human behaviour as a relationship between ends and scarce means which have alternatives uses.' A more modern definition from an introductory economics textbook is simply 'the study of how society manages its scarce resources.'[69] Economics derives from the Greek word 'oikonomia' meaning 'household management.'[70]

There are many definitions of economics available; in all, the fundamental concepts are the same. Fundamental concepts of economics are scarcity and opportunity costs. Scarcity is the concept that resources are limited in such a way that there are not enough resources available to satisfy every person's wants or demands. Scarcity of societal resources is unavoidable and universal. Opportunity cost is the value of the alternative foregone. Because resources are scarce, choices have to be made between one or more options, and the value of the option foregone is an opportunity cost. For example, if a school only had space for one hour of health education in its timetable, the scarce resource is time and the value of the outcome relates to health gains. The school may be considering using the hour to provide healthy lifestyle and nutrition education, or to provide physical education. The opportunity cost in this example, is the value of the choice that is forgone (i.e. the outcomes from nutrition education or physical activity). If resources are scarce, individuals and society need to decide the most efficient way of allocating those scarce resources; understanding that there will be an opportunity cost associated with every decision. In a healthcare setting, fixed budgets mean that limited resources i.e. doctors, nurses, health technologies, need to be allocated in such a way

that society deems most cost-effective; understanding that each decision that will potentially benefit a patient group, will result in benefits forgone by other patient groups.

The World Bank defines health economics as, 'the study of how scarce resources are allocated among alternative uses for the care of sickness and the promotion, maintenance and improvement of health, including the study of how healthcare and health-related services, their costs and benefits, and health itself are distributed among individuals and groups in society.' One of the earliest definitions given by Selma Mushkin in 1958 was, 'a field of inquiry whose subject matter is the optimum use of resources for the care of the sick and the promotion of health. Its task is to appraise the efficiency of the organization of health services, and to suggest ways of improving this organization.'[71]

In free markets, laissez-faire economics (or freedom from interference) allows markets to achieve equilibrium naturally. However, the provision of health care is different, as it experiences market failure.[72] Disruptions in the supply and demand for the provision of health care contribute to this market failure. Supply for example, is restricted because entry into the healthcare market requires licensing and training. This is a barrier to healthcare supply because only medical professionals, who are trained and have specialist knowledge and information, are able to provide health care. Another market failure is referred to as asymmetric information whereby the medical professional has knowledge and information that the patient does not. The patient puts their trust in the medical professional, trusting that the treatment they receive is going to improve their health. The power in the patient-provider relationship is unbalanced with the provider holding more power due to their increased specialist knowledge. This creates inefficiencies in the market, as markets are most efficient when knowledge is perfect and shared equally by everyone. This is common in any profession, and contributes to market failure. On the demand side, demand for health care is said to be a derived demand for health, or 'good health.'[73] Derived demand is the demand for a good or service (in this case health care) which is actually a consequence of a demand for something else, i.e. good health. As such, the demand for health care is irregular, sporadic, and unpredictable. Consequently, there is a need to correct this market failure, and the study of the allocation and consumption of health care is a branch of economics termed health economics.

2.1.1.1 Priority setting, HTA, and economic evaluation in health care

Corrections to healthcare market failure might come in the form of government intervention to allocate scarce healthcare resources to maximise health. This need for government intervention is now giving rise to the need for priority setting to efficiently allocate scarce resources to meet the rising demand for health care. Rationed health care, a more politically charged way to describe priority setting, is necessary to make decisions about how to fairly allocate scarce healthcare resources, and it is a global issue. Worldwide there are differences in healthcare systems and how they are financed, but the issue of scarcity is always the same. Priority setting in some countries comes in the form of developing principles that guide prioritisation; examples include Norway, Sweden, Denmark, and the Netherlands.[74] Other countries such as the UK, New Zealand, and Israel establish bodies that make recommendations for which treatments and services should be offered in the healthcare system.[74]

In the UK, the National Institute for Health and Care Excellence (NICE) is an independent organisation that provides evidence-based guidance and advice to improve health and social care for England (and generally the rest of the UK). NICE was set up in 1999 as the National Institute for Clinical Excellence, as a special health authority to reduce variation in availability and quality of National Health Service (NHS) treatments and care.[75] NICE issues evidence-based guidance on safety, effectiveness, and cost-effectiveness of health technologies through their technology appraisals guidance. The recommendations that NICE make regarding cost-effectiveness, inform government decision-making and priority setting in health care in the UK. Scotland has its own Scottish Medicines Consortium (SMC), which provides advice to local NHS boards about the status of newly licenced drugs. The remit of the SMC's advice is confined to prescription medications only.

A health technology is any device, medication or service that aims to improve health. Examples are drugs, diagnostic procedures, medical devices such as scanning or monitoring equipment, surgical procedures, medical interventions, services, and health promoting activities. Health technology assessment (HTA), is therefore the assessment of new health technologies for safety, effectiveness, and cost-effectiveness. An early definition of health technology assessment is given below,

> 'We shall use the term assessment of a medical technology to denote any process of examining and reporting properties of a medical technology used in health care, such as safety, efficacy, feasibility, and indications for use, cost, and cost-effectiveness, as well as social, economic, and ethical consequences, whether intended or unintended.' (Institute of Medicine 1985)

A more recent definition from the HTA glossary is given as,

> 'The systematic evaluation of the properties and effects of a health technology, addressing the direct and intended effects of this technology, as well as its indirect and unintended consequences, and aimed mainly at informing decision making regarding health technologies.
>
> Note: HTA is conducted by interdisciplinary groups that use explicit analytical frameworks drawing on a variety of methods.'[76]

Economic evaluation is one way to address the cost-effectiveness component of HTA. It is a mechanism that can be used to inform resource allocation decisions. Economic evaluation is concerned with two key components: inputs and outputs, or costs and consequences.[77] When making decisions about whether or not to adopt a new healthcare technology, device, treatment, or service it is important to not only consider the cost of the new technology, but health benefits including prevention, compared to what is already currently available, additionally considering the benefits forgone from any potential displacement resulting in adoption of the new technology. Economic evaluation is defined by Dummond et al,[68] 'as the comparative analysis of alternative courses of action in terms of both their costs and consequences.' In any economic evaluation, the basic tasks include identifying, measuring, and valuing the costs and consequences of the alternatives considered. Full economic evaluations explicitly consider relative costs of the alternatives and compare them to the relative consequences.[68]

Economic evaluation is used as an input for reimbursement and decision-making.[78] The overall aim of economic evaluation is to aid decisions about efficient and equitable resource allocation by comparing cost and benefits of health intervention.[68] Resource allocation decisions in the hospital setting might include diagnostic, treatment, and patient management. For example, the use of resources for treatment of one particular condition, means that those resources cannot be used for treatment of other conditions (opportunity cost), and economic evaluation aims to help decision makers identify the most efficient and equitable allocation of limited healthcare resources. Typically in the

UK, economic evaluation will be carried out from the health (NHS) and personal social services (PSS) perspective,[64] however NICE has recognised the importance of broader societal perspectives when considering economic evaluation of public health programmes.[79] This is because outside of the hospital setting, population health programmes might include a variety of resource allocation decisions from multisectoral funding streams, e.g. which programme should be run to improve children's SEW? The funding might come partly from the local education authority and partly from the local health board, further complicating the evaluation of such programmes, and thus requiring a broader public sector or societal perspective.

This section introduced the fundamental concepts of economics such as scarcity and opportunity costs. Important definitions of key concepts of this thesis were defined including economics, health economics, HTA, and economic evaluation. Background information was given as to why the field of health economics developed (due to market failures) and the use of economic evaluation (to aid decision makers). The next section provides a brief history of economic evaluation before each type of economic evaluation is outlined in turn.

2.1.2 A brief history of economic evaluation

The field of health economics is a relatively new one with many of the economic evaluation methodologies used today being developed in the last 50 years. However, scarcity and opportunity cost in healthcare is not a new phenomenon; attempts have been made to value human life in monetary terms beginning in the Victorian era.[80] One of the earliest forms of cost-benefit analysis comes from Gary N. Calkins, writing in the *American Statistical Association* in 1891.[81] Calkins quantified the costs and effects of England's Public Health Act 1875 that included sanitary improvements to water drainage and clean water supply. He quantified the cost of the improvements in US dollars (USD) which were given as $583,500,000. He then assumed the difference in annual mortality in the 10-year period before the works and the 10-year period after the works would be directly contributed to the Act, resulting in 856,804 lives saved. The value he placed on each life saved came from an estimate from William Farr's work in *Vital Statistics*, (p.61) which was estimated at £159 per head or $770 USD (at the time). Thus, the total value of

lives saved over a 10-year period after the passage of the Act was over $650,000,000 and the benefits outweighed the costs.

Charles Value Chapin was one of the first to consider the need for what is thought of as the modern day economic evaluation.[80] He read before the Institute of Medicine, Chicago, in 1917,

> "Money is the measure of most effort, and appropriations are limited. In what way shall the appropriation for the health department be expended so as to save the most lives and prevent the most sickness? Are our municipal health departments making the best apportionment of their funds? Are health officials devoting the most effort to that which will best conserve the health of the people?"[82]

Chapin considers how institutions are slow to break away from traditions of the past, and if you started over with a new health care budget, you would probably end up with a different allocation of resources based on current knowledge of costs and effectiveness. This is still true today as old inefficiencies in the health system are difficult to break away from; it is very difficult to convince stakeholders to disinvest in traditional methods of care that are no longer cost-effective. He concludes,

> "Until there are unlimited money and unlimited talent available, let us earnestly study to do that which pays best."[82]

Selma Mushkin as mentioned in section 2.1.1, was one of the first authors to define health economics in 1958.[71] Her work stemmed from the advancing medical techniques at the time, and the challenges of financing these new advancements. The official recognition of health economics as a discipline is often credited to Kenneth Arrow[83] in his 1963 paper 'Uncertainty and the welfare economics of medical care.'[84] In his seminal paper, Arrow discusses the economics of the medical care industry (not health) and how it satisfies the needs of society in a way that differs from the 'normal' economic model; these differences stemming mainly from risks and uncertainty.[84]

Herbert Klarman was a Polish immigrant in the United States of America (USA).[80] He was a professor of public health administration from 1962 to 1969 at John Hopkins during which time he published the first health economics textbook, Economics of Health.[85] He also published an early cohort decision model for the treatment of chronic renal disease.[86]

This was the first study to apply quality adjustment to life years gained, or the first study to use a QALY as it is referred to in modern day terms. In his chronic renal disease study, a quality adjustment was applied to account for the quality differences between life after transplantation and life on dialysis.[86]

This brief history introduced some of the earliest works in economic evaluation, before the disciple had been formally recognised. One of the earliest forms of economic evaluation was a cost-benefit type analysis of England's Public Health Act 1875. In 1917, appropriately named Charles 'Value' Chaplin, recognised that institutions have a hard time breaking away from traditions of the past, creating inefficiencies in the health care system. Selma Mushkin is one of the first to define health economics in 1958 before Kenneth Arrow, who is often credited with the recognition of health economics as a discipline in 1963. Herbert Klarman wrote the first health economics textbook, and was the first make a quality adjustment of life years gained. From the 1970s onward, methods for modern economic evaluation were developed and the rest of this section details methods and definitions for modern use of the different types of economic evaluation. A number of recommendation guidelines, documents, and texts have emerged since the 1990s on the design and conduct of health economic evaluation. These recommendations have helped to standardise the basic elements of economic evaluation and analytic techniques.[37] Many countries already have their own country specific HTA guidance in place for conducting economic evaluation.[87]

The types of economic evaluation are mainly differentiated by the outcomes used to measure benefits. There are three types of full economic evaluation as classified by Drummond et al.[68] which are cost-effectiveness analysis, cost-utility analysis, and cost-benefit analysis. Cost-consequence analysis and cost-minimization analysis are not always considered full economic evaluations. Sections 2.1.3 through 2.1.7 go over each type of economic evaluation individually.

2.1.3 Cost-effectiveness analysis (CEA)

Cost-effectiveness analysis (CEA) is a type of economic evaluation where effects are measured in natural units. CEA is used in situations where a decision maker with a limited budget, is considering a limited range of options within a given field.[77] Examples of

natural units that may be used as a measure of health outcome could be 'cases detected,' 'improved mental health status,' or 'life years gained.' It is important that the health outcome chosen is a reliable measure for the desired objective.[77] Take for example the evaluation of two cancer drugs, drug A the standard drug and drug B the newly developed drug. Treatment with drug A and B share the same outcome of interest, life years gained, but they may have differential success in achieving this outcome as well as differential costs. Evaluators would be interested in the incremental cost per unit of effect, i.e. life years gained. CEAs are often expressed in terms of an incremental cost-effectiveness ratio (ICER) which is a ratio of the incremental difference in costs between two alternatives and the incremental difference in effectiveness between the same two alternatives.[88] The ICER formula is given below where Δ represents the difference in mean costs and effects between groups.

Equation 1: ICER formula

$$\boldsymbol{ICER = \frac{\Delta \mathbf{Cost}}{\Delta \mathbf{Effects}}}$$

In CEA, the result of interest is typically expressed as cost per unit of effect gained (e.g. cost per life year gained), but it could also be expressed as effect per unit of cost (e.g. life years gained per pound spent).[68] An example of a CEA might look at the number of quitters in a smoking cessation programme. The ICER would be expressed as the incremental cost per successful quitter. There are many examples of CEA published in the literature, one such example examined home visiting to improve parenting and health and social outcomes for children.[89] The outcome of interest was unit increases in maternal sensitivity and infant cooperation components of the CARE Index, an outcome that measures mother and child interaction. The results were expressed as £2,723 and £2,033 per increased unit of maternal sensitivity and infant cooperativeness respectively.[89]

The cost-effectiveness (CE) plane is used to plot the difference in effects (ΔE) along the horizontal axis against the difference in costs (ΔC) per participant along the vertical axis.[90] The plane is split into four quadrants labelled using the points of a compass NE, SE, SW, and NW (see Figure 1). If an ICER falls in the SW quadrant, the new treatment dominates in that it is more effective and less costly. If it falls in the NW quadrant, the new

treatment is said to be dominated because it is more costly and less effective. If it falls within the NE or SW quadrant a trade-off has to be made between costs and effects.[38] Some decision makers might specify a willingness-to-pay threshold (WTP) which is a value judgement usually denoted by λ. NICE for example, uses a WTP threshold of £20,000 to £30,000 per quality-adjusted life year (QALY) to determine cost-effectiveness of health technologies.[64] The QALY as a concept will be discussed further in the next section, 2.1.4. A WTP threshold can be specified for any unit of effect the decision maker deems relevant, and the amount may be context specific and based on a value judgement. In Figure 1, the line that passes through the origin of the CE plane, denoted by λ, represents a hypothetical WTP threshold; i.e. the maximum WTP per unit of effect.[77] When an ICER falls in the NE or SW quadrant, decision makers must decide if the additional health benefits of the more effective treatment are worth the additional cost. If a WTP threshold has been specified, the additional cost is capped by this ceiling value. This decision rule is expressed in Equation 2 below.

Equation 2: Cost-effectiveness decision rule

$$\boldsymbol{Decide\ to\ implement\ new\ programme\ if}\text{: } \frac{\Delta\mathbf{Cost}}{\Delta\mathbf{Effects}} < \boldsymbol{\lambda}$$

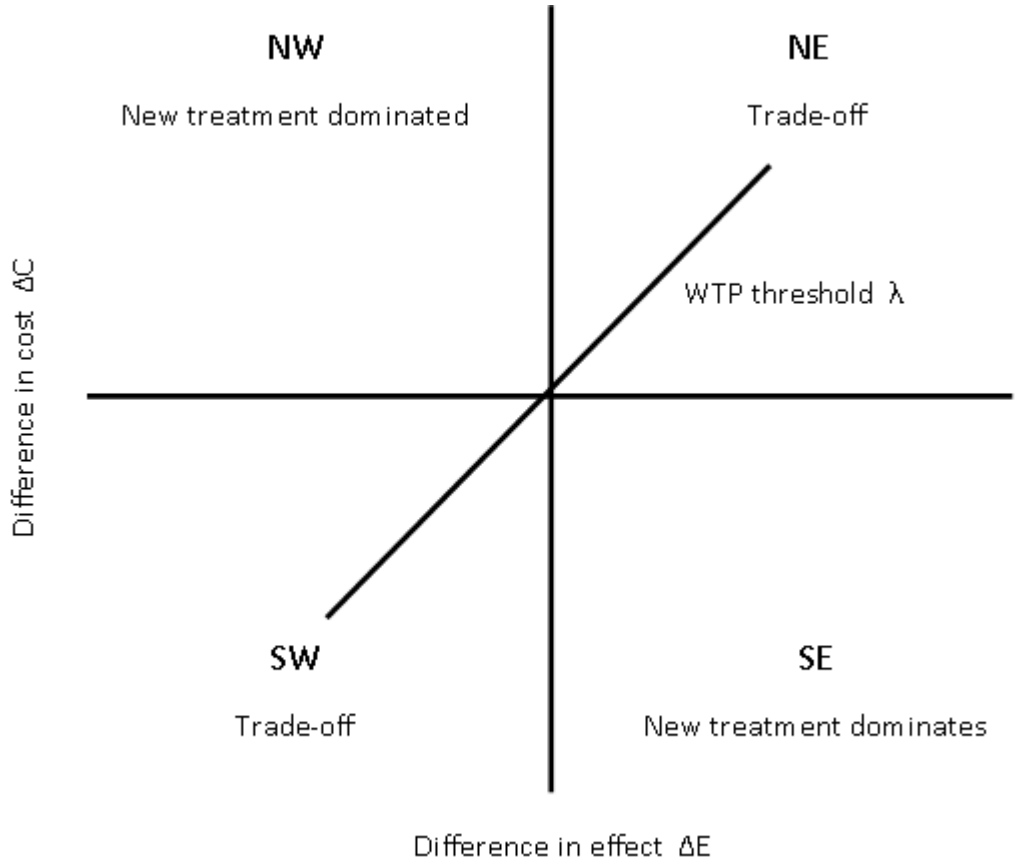


Figure 1: Cost-effectiveness plane

It is important to note that negative ICERs that fall within different quadrants have vastly different interpretations. A negative ICER can have the same value, but depending on whether it falls in the SE or NW quadrant can be the difference between the treatment being dominated (where effects are negative in the NW quadrant), to the new treatment dominating the old treatment (where the costs are less in the SE quadrant). This is why negative ICERs are generally not reported; instead they are reported in relation to what quadrant they fall, or in terms of dominated or dominates.[91] Negative ICERs are an issue when bootstrapping cost and effect pairs to analyse uncertainty around the point estimates, as the pairs will be ordered from low-to-high in a distribution when estimating confidence intervals.[92] To overcome this problem, the decision rule can be rearranged into linear functions net monetary benefit (NMB) and net health benefit (NHB) given below in Equation 3 and Equation 4. Estimating NMB or NHB is also useful when comparing three or more comparators as each comparator can be ranked and selected based on which comparator provides the most NHBs within the maximum threshold.

Equation 3: Net monetary benefit (NMB)

$$\lambda * \Delta E - \Delta C > 0$$

Equation 4: Net health benefit (NHB)

$$\Delta E - \frac{\Delta C}{\lambda} > 0$$

NMB allows more meaningful presentation of cost-effectiveness results, but relies on the WTP threshold (λ) being known. In cases where λ is unknown or unspecified, a range of values can be estimated. Because of the use of specific measures of effects, one of the biggest limitations of CEA is the difficulty in quantifying the opportunity cost (or the benefits forgone) of the displaced programmes covered under the same budget.[68]

2.1.4 Cost-utility analysis (CUA)

Cost-utility analysis (CUA) is often referred to as a variant of CEA because the only difference is that CUA uses a generic measure of health gain. Many authors of economic evaluations do not always distinguish between the two, particularly in the USA.[77] Thus, it is common to see variation in the use of the terms in the literature. Drummond and colleagues[68] characterise CUA as a special case of CEA which is expressed as a 'cost per healthy year gained.' The most common measure of years in full health is the QALY.[38] The QALY is a year of life adjusted for its quality or its value. QALYs are calculated by weighting length of life by health-related quality of life (HRQoL). The QALY is essentially what 'utility' refers to in cost-utility analysis. A utility is a generic measure of health gain and is valued to reflect population preferences. In this sense, 'utility' refers to preferences individuals or society has for a particular set of health outcomes or health states.[68] A year in perfect health is considered equal to 1'[88] and death is considered 0. There are health states considered worse than death so negative utility values are possible. A terminal illness that causes a lot of pain, immobility, or a decreased quality of life that the patient deems worse than death might give rise to a negative utility value. QALYs are used as the primary outcome in CUAs for a couple of reasons. First, they are generic, thus facilitate the comparison of very different programmes or interventions on a single effectiveness measure. Second, they are weighted by the population's preferences hence, they not only prioritise interventions that extend length of life, but those that improve overall quality of life.

2.1.4.1 Eliciting preferences

There are two components for estimating quality adjustment of QALYs; a description of the possible health states being measured, and a valuation of those health states.[38] A generic preference-based measure such as the EuroQol EQ-5D,[93] the Health Utilities Index (HUI),[94] or the SF-6D[95] have health state descriptive systems that are accompanied by a set of health state utility values (health-utilities) that were elicited using preference-based valuation techniques.[38] More recently, the Child Health Utility 9D (CHU9D)[96] has been developed specifically for children. Briefly, the CHU9D is the only HRQoL measure that has been developed specifically for children and has been valued by adolescents. All other HRQoL outcomes for children are missing one or both of these circumstances (further detail in section 4.2.1). There are a number of methods for obtaining these preferences; the standard gamble (SG), the time trade-off (TTO), and the rating scale and its variants being the most common.[77] These methods allow for the valuation of the health states described in by the generic preference-based measures mentioned above (i.e. EQ-5D, HUI, and SF-6D).

The SG technique derives directly from expected utility theory, in which a rational individual will make decisions, or act in such a way to maximise their utility. The SG presents participants with two choices, a certain outcome or a gamble. For example, participants might be asked to imagine that they have a chronic disease where they experience limited mobility, some pain, and some problems with performing usual activities. They are then presented with a gamble, they can either stay in their current health state or take a gamble in which they have a 70% chance of being cured or a 30% chance of dying immediately. The probability of a cure is then varied until the participant is indifferent between their current health state and probability of a cure. This point of indifference represents the utility the participant places on the cure.[97]

The TTO method was developed by Torrance and collegues[98] and involves asking participants to state their choice between two certain outcomes at different lengths of time. Choice A might be life in full health for 8 years followed by death, and Choice B is life in a particular health state (like the one described above) for 10 years followed by death. The participant must choose which is preferred. If it is Choice A, the times are divided by one another and that is the preference given to health state B (i.e. 8/10 = 0.8).

If the answer is Choice B, the time in Choice A is shortened until, it is selected (e.g. four years in perfect health is preferred to the health state in B; 4/10 = 0.4).

Rating scales and visual analogue scales are one of the simplest methods for obtaining preferences. Participants are presented with a scale, and they are asked to rank a number of health outcomes on the scale, with intervals between the outcomes representing the differing preference for those health states. There are measurement biases associated with these types of scaling tasks when compared to choice-based tasks such as SG or TTO. These include end-of-scale bias in which participants tend to avoid placing outcomes at the high and low end of the scale, and context bias where participants tend to evenly space outcomes regardless of if their preferences align.[68] NICE recommends using a utility measure that uses a choice-based method to elicit the public's preferences.[64]

Because the QALY is a generic HRQoL measure, it is possible to compare programmes with very different objectives to one another because effectiveness outcomes are all being valued in the same way. A variant of the QALY is the disability-adjusted life year (DALY) which is commonly used in developing countries. A DALY is 'a measure to adjust life years lived for disease related disability, age and time preference.'[88] Other alternatives to the QALY are the healthy years equivalent[99] and the saved-young-life equivalent[100] HRQoL measures attempt to quantify and measure all possible health states. The more detailed a questionnaire (more dimensions and levels), the more possible resulting health states. These health states will have been valued using population preferences obtained from SG, TTO, or other methods. There are many different HRQoL measures available and the same individual filling in different questionnaires can end up with markedly different utility values, depending on the questionnaire and the method used to value the population's preferences. Thus, in order to facilitate comparability between evaluations for decision making, NICE recommends use of a single measure,[64] the EQ-5D.[93]

2.1.4.2 EuroQoL EQ-5D

The EQ-5D consists of the following five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. There is the 3L and 5L version where the 'L' stands for levels, which describes varying levels of problems within each dimension.

Participants are asked to select the level of severity for each of the five dimensions, which makes up their unique health state. Preferences from the relevant population can then be applied to value the QALY. Country specific value sets are available from the EuroQol website.[101] Box 1 below gives an example of the EQ-5D-3L.

Box 1: Example of EQ-5D-3L from EuroQol website

By placing a tick in one box in each group below, please indicate which statements best describe your own health state today.

Mobility
I have no problems in walking about □
I have some problems in walking about □
I am confined to bed □

Self-Care
I have no problems with self-care □
I have some problems washing or dressing myself □
I am unable to wash or dress myself □

Usual Activities (e.g. work, study, housework, family, or leisure activities)
I have no problems with performing my usual activities □
have some problems with performing my usual activities □
I am unable to perform my usual activities □

Pain/Discomfort
I have no pain or discomfort □
I have moderate pain or discomfort □
I have extreme pain or discomfort □

Anxiety/Depression
I am not anxious or depressed □
I am moderately anxious or depressed □
I am extremely anxious or depressed □

2.1.4.3 Calculating QALYs

Since the EQ-5D incorporates the two components of a QALY, the health state description and its valuation, QALYs can now be calculated. If utilities are plotted on a graph; utility values are plotted along the y-axis and time runs along the x-axis. In a very simple example, the EQ-5D is measured at baseline, and at 1 year ($t_1 = 0$ and $t_2 = 1$, where t represents time). The utility at each time point is 1, perfect health ($u_1 = 1$ and $u_2 = 1$, where u represents the utility value). The area under the curve (AUC), in this case a flat

horizontal line, is the QALY gained over that time. The AUC is calculated as the product of the time difference and the average of the two measures as given in Equation 5.[102]

Equation 5: Area Under the Curve

$$\boldsymbol{AUC = (t_2 - t_1) \times \frac{(u_1 + u_2)}{2}}$$

In this simple example the AUC is 1, so the QALY gained is one, see Figure 2.

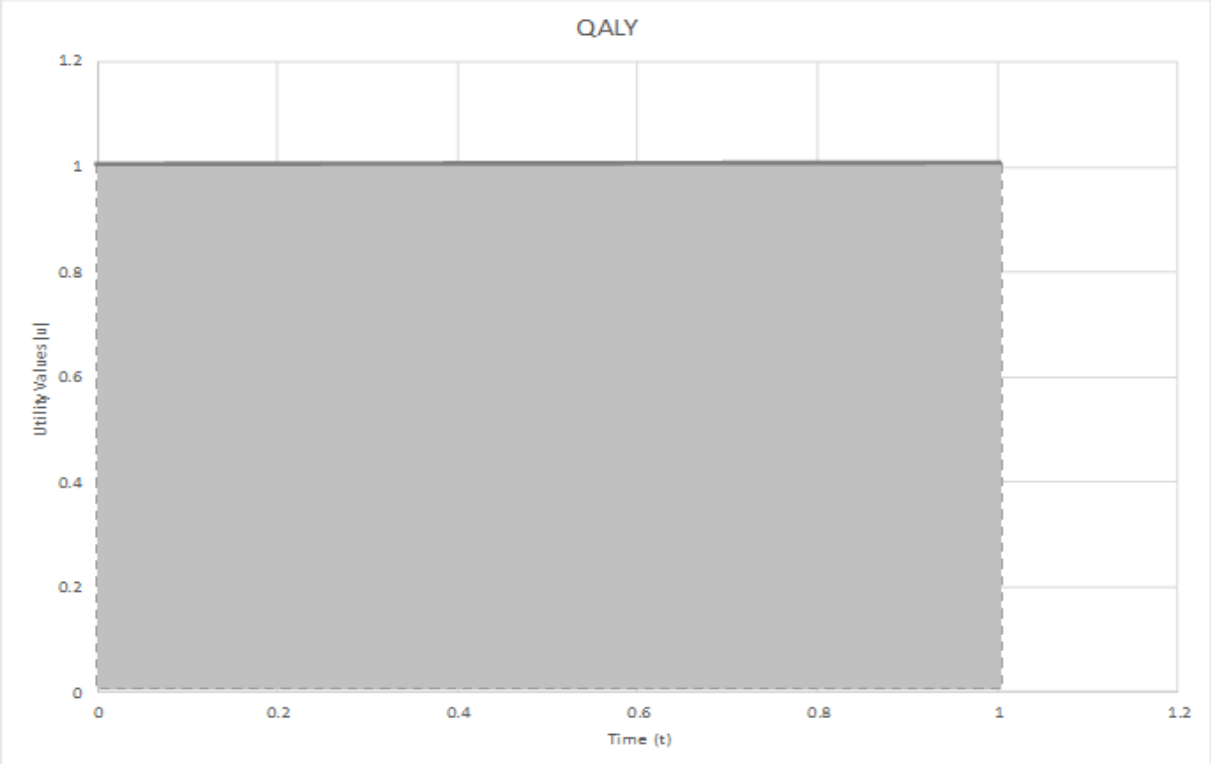


Figure 2: Simple example of individual QALY gain

A more complex example is displayed in Figure 3. Note the grey area is the QALY gained without the intervention (in the control) and white area between the two series represents the QALY gains from the intervention, which is simply the difference between the two.

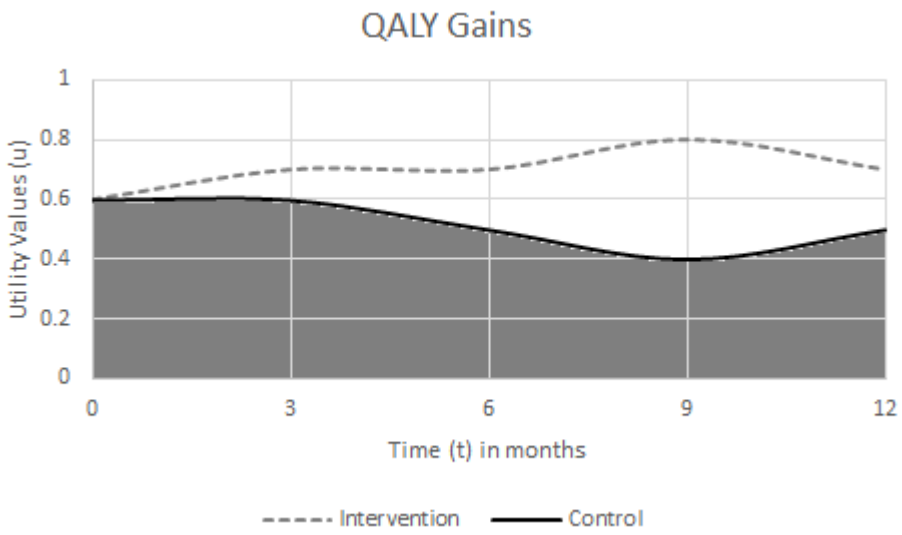


Figure 3: QALY gains from a comparison of two alternatives

Once QALYs are calculated for each alternative, the resulting ICER is expressed as a cost per QALY in CUA. CUAs are now the most common form of economic evaluation in the UK,[68] partly due to the official requirements set out in the NICE reference case.[64] The reference case sets out the methods to be used in health technology appraisals submitted to NICE, as a way to promote consistency and quality in determining cost-effectiveness of health technologies. Specifically, the reference case states that, 'health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults'[64] (p 29). The use of a generic QALY measure provides a uniform 'yardstick' measure for which all health technologies regardless of disease area can be compared. Additionally, NICE has an established cost-effectiveness threshold based on improvements in QALYs, making for ease and consistency[b] in the healthcare decision-making process.[64] This is in line with the extra-welfarist view, which has a sole focus on maximising health utility through a QALY framework.[103]

Many major funding bodies in the UK now require or expect to see an economic evaluation built into primary research study grant applications. The National Institute for Health Research (NIHR) was created in 2006 under a 2005 English Government strategy to improve research in the health field and consolidate existing research programmes, one

[b] Funding allocation decisions are not based-solely on the cost-effectiveness threshold of £20,000 to £30,000 per QALY. Other factors play a role such as patient safety and ethics, and in some cases a higher threshold is warranted such as at the end of life.[64]

of which was the HTA programme.[104] Three years into the existence of the NIHR, the Public Health Research programme was introduced to fund research and generate evidence of the delivery on non-NHS interventions that improve public health and reduce inequalities. The NIHR stipulates that most primary research projects applying for funding are expected to include an economic evaluation.[105] The Chief Scientist Office (CSO) is a major funding body for Scottish health research. Through their contribution to the NIHR funding pool, researchers in Scotland are able to apply for most research programmes funded by the NIHR including the HTA and Public Health Research programmes. CUAs are now built into many of these funding applications due to these requirements and expectations of determining not only effectiveness, but cost-effectiveness as well.

2.1.5 Cost-benefit analysis (CBA)

Cost-benefit analysis (CBA) differs from CEA and CUA in that all effects or benefits as well as costs are valued in monetary terms. An intervention is considered worthwhile if all the benefits (valued in monetary units) exceed the costs (i.e. there is a positive net benefit.) CBA addresses the question of whether an intervention is worthwhile to society rather than restricting it to the health services' budget.[38] It is often considered the gold standard as it is the most comprehensive form of economic evaluation.[106] With monetised benefits readily compared to costs, CBA allows decision makers to directly address if it is worthwhile expanding the healthcare budget, as opposed to how to best allocate an existing budget as is the case with CEA and CUA.[68] A decision to expand a programme from a CEA or CUA has an opportunity cost in terms of benefits forgone to other health technologies covered in the same programme.[68] CUA is based on the notion that those who gain, could compensate the losers. Additionally, CBA's measure of benefit is more comprehensive including non-health benefits.[38] The results of a CBA might be presented as a ratio of costs to benefits, or a simple sum of the net benefit (or loss) of one programme over another.

CBA has a long history outside of health in sectors such as the environment and transport.[68] In fact, CBA can facilitate the comparison of healthcare technologies to programmes from multiple sectors of the economy such as the education, environment, and transport sectors. Because outcomes are expressed in monetary units, it is possible to determine net monetary gains to society. For example, the net benefit of a surgical

procedure can be compared to the net benefit of an educational programme to improve maths learning in schools, which can also be compared to the net benefit of improved public transport links if CBA methodologies are used in each case. This comparison is not possible with CEA or CUA because health benefits and costs are measured with different units so to attempt to compare a cost per QALY outcome to a monetary transport outcome would be like attempting to compare apples to oranges. The main challenge with CBA is placing a monetary value on human life and health benefits, and this is a main reason why CEA and CUA have been utilised more in the health sector.[106] Additionally, CBA requires more time and resources to conduct (not to be confused with cost savings analysis – see section below). This is due to a larger burden placed on measuring a broader spectrum of outcomes (health and non-health) which also need to be valued in monetary terms. Lack of standardisation in elicitations methods, stated-preference biases (see section below), and the considerable measurement burden (as CBA needs to be tailored for each intervention) are some of the practical barriers of CBA.[107]

2.1.5.1 Valuing benefits in monetary terms

There are a number of ways to place a monetary valuation on human health. They are broadly divided into two main categories: the human capital approach and approaches based on individual observed or 'stated preferences.' The human capital approach estimates the present value of an individual's future earnings. Benefits are valued in terms of how the health changes impact an individual's labour productivity. This approach has been favoured in legal applications that require estimates of damages.[108] An example may be a pay out to a former employee who suffered a work place accident that prevented them from returning to work. The human capital approach does not directly measure an individual's 'willingness-to-pay' to avoid ill health or their 'willingness-to-accept' as compensation.[108] The second approach to valuing health benefits uses observed preferences that are revealed in markets, or asks individuals to state their preferences in monetary terms. Where functioning markets do not exist (e.g. health care), individuals can express their hypothetical willingness-to-pay (or accept) health outcomes. This is a conventional economic concept in which an individual's WTP for a good is an indicator of the strength of their preference for such good or attribute of the good.[72] As well, the hypothetical nature of the task is similar the SG and TTO approaches mentioned in section 2.1.4.1 for eliciting utility preferences. The techniques used to elicit such

preferences come under the broad heading of 'stated preference' methods or contingent valuation where the valuation poses a set of contingencies to determine the individual's WTP for the desired benefit. There are a number studies in the literature that assess individual's WTP for a health technology, however, few comprehensive CBAs that incorporate these values are published.[68, 72] Another method for eliciting and valuing preferences are discrete choice experiments (DCE). A DCE is an attributed-based technique for collecting stated preferences involving a sequence of hypothetical scenarios or choice sets for the respondent to choose from.[109] Depending on the complexity, DCE choice sets can quickly multiply leading to more cognitive burden for the respondent, and the possibility of 'irrational' stated preferences which cannot be used.[109] As stated above, there is a lack of standardisation in eliciting these stated preferences as well as biases that go along with asking an individual to state their preferences. This is a challenge when undertaking CBA and is one potential reason why few comprehensive CBAs have been published in the literature.

CBA should be distinguished from a related technique, cost-savings analysis, which involves the comparison of costs and benefits that are easily converted into monetary units with other effects ignored. Cost-savings analysis is and continues to be more commonly used in the evaluation of social welfare services.[110] An example is comparing the costs of an intervention to the savings generated from reductions in crime. This type of analysis is less sound than CBA because it does not attempt to value all relevant outcomes.[110]

2.1.6 Cost-minimization analysis (CMA)

Cost-minimization analysis (CMA) assumes two or more alternatives are equivalent in the health benefits produced, and thus CMA is simply a costing exercise to determine which programme costs less. CMA has often been criticised for failing to explore uncertainty around determining equivalence in treatment outcomes of two different treatment options and in 2001 Briggs and O'Brien[111] announced the 'death of the cost-minimisation analysis.' CMA was historically recommended for trials finding no statistically significant differences in effectiveness because of its simplicity and ease of interpretation.[112] With the 'death of CMA' Briggs and O'Brien[111] argued researchers should instead conduct CEA or CUA to estimate the joint density of cost and effect differences and present

uncertainty on cost-effectiveness acceptability curves. More recently Dakin and Wordsworth[113] conducted a literature review to examine how the use of CMA has changed since 2001, if CMA was appropriate in non-inferiority trials, and if CMA gives biased results. Through examples of simulated and trial data, they found CMA does bias measures of uncertainty, and even when the bias is negligible in non-inferiority trials, where there is a large difference in cost, it is still necessary to collect and analyse data on costs and effects to assess this bias. They went on to conclude,[113] 'The remit of CMA in trial-based economic evaluation is therefore even narrower than previously thought, suggesting that CMA is not only dead but should also be buried.' CMA has since fallen out of health economics textbooks as a recommend form of full economic evaluation.[38, 68, 77]

2.1.7 Cost-consequence analysis (CCA)

Cost-consequence analysis (CCA) is not considered by Drummond et al.[77] as a full economic evaluation because the trade-offs between costs and consequences have not been made explicit. It has since fallen out of the latest version of Drummond 'blue book.'[68] However, public health guidance issued by NICE in 2012, stated more emphasis would be placed on CCA and CBA than has been in the past due to local governments being responsible for implementing public health programmes and having a larger remit than the health services sector.[79] NICE began focusing on public health in 2005 in order to avoid ill health and promote healthier lifestyles. The first public health guidance was issued in 2006 relating to smoking interventions and referrals.[114] The significance of decision-making in public health contexts will be covered further in section 2.2.1.

CCA was developed from scepticism that all relevant considerations could be summarised in a single outcome such as incremental cost per unit of effect or a net benefits approach.[38] Instead all relevant costs and effects are presented in a table, but there is no single resulting figure to enable ranking of different treatment options; decision rules are left up to the decision maker. Decision makers may be more interested in seeing disaggregated costs and outcomes of the two or more alternatives because there may be multiple objectives of the programme.[77] Presenting an array of costs and outcomes leaves the decision maker to decide on the trade-offs between costs and effects. This is keeping in line with the traditional notion of economic evaluation as an aid to decision makers.

The main disadvantage of CCA is that the basis for the decision may be unclear or not made explicit.

Sections 2.1.3 to 2.1.7 described the different types of economic evaluation methodologies and gave examples of outcomes that distinguish the different types and situations where certain methods may be more appropriate. CUA, a form of CEA, is the most commonly used type of economic evaluation in the UK due to NICE guidance specifically calling for this type of evaluation in the reference case. CBA is the most comprehensive form of economic evaluation, but due to complications with valuing health outcomes in monetary terms, comprehensive CBAs with stated preferences are still rarely published. CMA should no longer be used due to problems and biases that present from attempting to determine total equivalence in effectiveness of two or more alternatives. Finally, CCA is not always considered a full economic evaluation, however it provides decision makers the option of deciding themselves the appropriate trade-offs that need to be made in terms of costs and benefits. The final subsection of section 2.1 describes decision making in the UK and internationally. There are additional methods that are becoming more popular in the economic evaluation context such as multi-criteria decision analysis (MCDA), and the use of natural experiments in population health, but these will be covered later in section 2.2.1.2.

2.1.8 The role of economic evaluation in healthcare decision making in the UK and internationally

Economic evaluation methodologies were developed to aid decision making in the context of prevalent market failure in health care markets. In the last 20 years, annual growth rate of public health spending exceeded GDP growth in all OECD countries.[115] While this has led to improved health outcomes, there is concern over the sustainability of the trend. Rising health expenditure is mainly due to new technologies, rising incomes, and population aging.[115] NICE is the only public body to specifically state a cost-effectiveness threshold of £20-£30,000/QALY[116] that is used in aiding decision making of the potential cost-effectiveness of new health technologies. This threshold has been maintained since appearing in NICE's methods guidance since 2004.[117] However, it is important to note that cost per QALY is not the only criterion considered when making decisions on whether to accept or reject new health technologies. The origins of the

threshold figures are not based on empirical evidence.[118] Appleby and colleagues state, 'the uncomfortable truth is that NICE's threshold has no basis in either theory or evidence' (p358).[119] This has led some to consider the threshold may be too high.[119] In 2008, the House of Commons Health Select Committee stated,

> 'The affordability of NICE guidance and the threshold it uses to decide whether a treatment is cost-effective is of serious concern. The threshold is not based on empirical research and is not directly related to the NHS budget. It seems to be higher than the threshold used by PCTs [primary care trusts] for treatments not assessed by NICE. Some witnesses, including patient organisation and pharmaceutical companies, thought NICE should be more generous in the cost per QALY threshold it uses, and should approve more products. On the other hand, some PCTs struggle to implement NICE guidance at the current threshold and other witnesses argued that a lower threshold should be used. We recommend that the threshold used by NICE in its full assessments be reviewed; further research comparing thresholds used by PCTs and those used by NICE should be undertaken. An independent body should determine the threshold used when making judgements of the value of technologies to the NHS.' (p6)[120]

The lack of evidence around this value-based threshold poses problems for primary care trusts struggling to implement new guidance from NICE, while on the other hand patients and drug providers are arguing the threshold is too low. As per the recommendation of the House of Commons Health Select Committee, Claxton and colleagues[118] have attempted to value the threshold based on technical fact rather than informal judgement. The aim of the work was to re-estimate the NICE threshold using routinely available data. The work encountered major technical challenges as well as challenges from fellow academics.[121] The final estimate which is closer to £13,000/QALY is surrounded by considerable uncertainty, however one could argue the same of the informal judgment made around the original value of the threshold. However, now that the precedent of the £20-£30,000/ QALY has been set and practiced for nearly two decades, real life implications for a drastic lowering of the threshold may not be acceptable to the healthcare system or the public. Sir Andrew Dillon, NICE's chief executive, argues that the use of the new threshold would mean the NHS would not be able to provide most new treatments as he does not believe drug companies would be willing to lower their prices in an unprecedented way.[122] He believes the balance between accepting new costly treatments and displacing other effective healthcare treatments from the NHS has been achieved with the current threshold, and it would be up to a debate in the government,

NHS, NICE, and the public to determine if the current threshold should be adjusted. There are the practical issues to consider as well including if NICE would need to reissue all guidance that was now over the £13,000/QALY threshold. There would also be issues with implementation of all the new guidance in the NHS. The concept that it is difficult to break away from the status quo, as described by Charles Value Chapin in section 2.1.2, is echoed here in a modern day example. Despite this, the new lower threshold should not be forgotten completely, and wider discussion of the potential outcomes should continue. In the words of Charles Value Chapin, "let us earnestly study to do that which pays best."

Outside of the UK, the World Health Organization (WHO) has been using average per capita income as a means for establishing cost-effectiveness thresholds in low and middle income countries.[123] Cost-effectiveness is determined as cost per DALY averted, and those interventions which cost less than three times the average per capita gross domestic product (GDP) is considered cost-effective. Those that cost less than the average per capita GDP is considered very cost-effective. Marseille and colleagues[124] argue this approach has major shortcomings. Ultimately, the value placed on the threshold should come from the collective values of the society.

Australia was the first country to use an element of HTA in routine decision-making regarding pharmaceuticals in 1993.[125] Since January of 1993, economic analyses were a requirement to support applications to list new pharmaceuticals on the Australian schedule of pharmaceutical benefits. HTA submissions are considered by the Pharmaceutical Benefits Advisory Committee in Australia. Canada soon followed in issuing its first set of guidelines in November 1994.[126] HTA submissions are considered by the Canadian Agency for Drugs and Technologies in Health. For an updated table of country-specific pharmacoeconomic guidelines please see the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) website at: https://www.ispor.org/PEguidelines/index.asp.

In the US, the Panel on Cost-Effectiveness in Health and Medicine released its first guidance in 1996 recommending the use of QALYs as a standard metric for identifying and assigning value to health outcomes.[127] The Second Panel on Cost-Effectiveness in Health and Medicine[128] convened 20 years later to update the recommendations. In the second

set of recommendations, QALYs were still recommended, but in terms of interpreting results in comparison to a threshold,

> 'Comparison with 1 specific threshold should be avoided (unless appropriate for the decision context); analysis should instead highlight how clinical or policy recommendation might change with consideration of a range of thresholds.'[128]

Additionally, the Patient Protection and Affordable Care Act,[129] a landmark health policy reform in the US, created a Patient-Centered Outcomes Research Institute to conduct comparative-effectiveness research, but prohibited this institute from developing or using cost-per-QALY thresholds.[130] The Act states,

> 'The secretary shall not utilize such an adjusted life year (or such a similar measure) as a threshold to determine coverage, reimbursement, or incentive programs…'[129]

Reasons for this resistance are complex, but likely rooted in many individuals' fear of rationed care. Many individuals in the US have a distrust of the government meddling in affairs of individuals, including healthcare, despite the existence of government sponsored healthcare programmes such as Medicaid and Medicare, which contribute to a considerable proportion of healthcare spending. Individuals in the US are likely to minimise the underlying problem of resource scarcity and the need to explicitly ration care.[128] This might be due to the current ability to 'choose' the heath care they desire based on their ability to pay. Any attempts to reform the system would be viewed by many as the government imposing limits on their individual liberties. The lack of concern for making sure everyone has access to affordable health care may not stem from a lack of altruism, but a genuine belief that the status quo in America is the most efficient way to deliver health care at the highest quality.

Some countries, such as USA, outright disagree with use of a threshold – not public money spent on health care system. Other countries see the benefit of a value yet there remains an ongoing debate over the appropriateness of this due to a lack of theoretical and empirical evidence.[116] It is unclear whether they should represent normative values or real resource constraints within a health care system.[37] Even if the use of a threshold is welcome, there are differing views of what it represents. Vallejo-Torres and colleagues[116] write, 'The two main conceptual perspectives include the view that the threshold should

reflect 1) society's monetary valuation of health gains, or 2) the opportunity cost resulting from the disinvestment required to adopt a new technology.' The author believes both perspectives are valid, because society understands rising healthcare costs are unsustainable, so there should be a monetary valuation of health gains based on society's collective values. At the same time society should also understand that this valuation is equivalent to the disinvestment in health or social care services elsewhere in the healthcare system. This opportunity cost will result in health losses for individuals elsewhere who are often less visible than the families who might campaign to make a new (and expensive) drug available on the NHS. The cuts to mental health and social care are often in the news, but it is rarely equated to the opportunity cost of investing in new medical technologies elsewhere in the health care budget (assuming social care fell within this remit). The public may be outraged when NICE decides to reject a new cancer drug; however, what is left out of the conversation is the trade-off that would have been made if they decided to accept the new drug. According to a WHO report[131] mental illness cost the UK economy £110 billion in 2008 but only accounted for 10.48% of the 2008/9 NHS budget. This equates to roughly £10.1 billion spending on mental health services of the approximate £96.4 billion NHS budget in 2008/9.[132] There is a clear discrepancy between what mental illness is costing the UK economy in health service use and labour productivity and what is actually being spent by the NHS on mental health services.

Health care costs are rising and the trend is unsustainable. There needs to be a systematic process in place for deciding how society's scarce healthcare resources will be allocated. The use of economic evaluation is one such way to do this. Specifically, the use of a threshold, even if value-based, provides a means for making allocation decisions in a consistent fashion. As NICE explicitly states a cost-effectiveness threshold, this threshold is used in the economic evaluation described in Chapter 4. As will be seen in section 2.2.1, the threshold as it relates to PHIs is still the same £20,000 to £30,000 per QALY as stated in methods guides for the appraisal of health technologies in the healthcare setting.

2.1.8.1 Vehicles for conducting economic evaluation

Economic evaluation can be conducted in various ways. These include alongside trials; in a decision modelling context; or a mix of both methodologies. The mixed methodology approach might include an economic evaluation alongside a trial with a long-term model

to extend the within trial time horizon. Or a pre-trial decision analytic model may be employed to inform the potential cost-effectiveness of a new health technology and to determine if a large-scale trial-based economic evaluation is required. More recently, emerging methodologies for economic evaluation such as multi-criteria decision analysis (MCDA) and economic evaluation alongside natural experiment designs are developing further vehicles for conducting economic evaluation (further information in section 2.2.1.2).

Economic evaluation alongside pragmatic randomised controlled trials (RCT) have traditionally been accepted as the best vehicle for economic studies.[68, 91, 133] The RCT is considered the 'gold standard' in effectiveness studies because the design leads to high levels of internal validity. Randomisation is a powerful tool that minimises selection bias between groups. Thus, there are three distinct advantages to conducting trial-based economic evaluation.[68] The first, as previously described is the internal validity. Second, given the high cost of conducting an RCT and collecting clinical data, the marginal cost of collecting economic data is minimal.[134] Finally, the RCT data (with economic data collected alongside) may be the most recent and rapidly available relevant evidence for conducting economic evaluation.[68]

It is important to recognise however, that all available evidence should be utilised in health care decision making; to rely solely on a single RCT as the vehicle for economic evaluation could pose potential bias in decision making.[135] Sculpher and colleagues[135] argue that any economic analysis aims to inform two key questions. First, whether to adopt a new health technology given existing evidence, and second, to determine if more evidence is needed to support the decision in the future.

Additionally, trial-based economic evaluations may suffer from design issues, they are often conducted 'alongside' clinical trials which means the economic analysis is not typically the primary purpose of the study.[135] The sample size calculations are often based on the primary clinical outcome and may be potentially underpowered for the economic analysis. Because economic evaluations are often considered a secondary aim of the research, health economists must maintain regular contact and engagement with the trial coordinator to ensure timely and correct collection of health economic data.

Trial-based economic evaluation often only compares two alternatives; the new health technology in question compared to the standard existing technology. However, there may be more alternatives than those included in the trial and failure to incorporate all relevant alternatives could bias decision making.[135] Trials are also very expensive to run, and thus follow-up may be limited to a year or two. The trial time horizon may be truncated in terms of the relevant time horizon for the economic evaluation,[135] which is often the participant's lifetime. Also, the results observed in strictly controlled environments are not always replicated in the real world. Examples include the Hawthorne Effect, in which individuals behave differently simply because they are being observed, or there may be practical implications of implementing an intervention in a less controlled real world situation. Therefore, even though the RCT is considered the 'gold standard' there are limitations to its use. These are all important considerations for decision makers who are presented with economic evidence based primarily on a single RCT.

The next vehicle for economic evaluation is the use of decision analytic modelling. A decision analytic model 'uses mathematical relationships to define a series of possible consequences that would flow from a set of alternative options being evaluated.' (Briggs et al., 2006).[136]

> 'Decision models provide a structure within which evidence from a range of sources can be directed at a specific decision problem for a defined population and context. Being clear about this distinction between measurement (undertaken in trials and other primary studies) and decision making (which needs an analytical structure within which to direct the evidence at the decision problem being addressed) emphasises that models and trials are complements, not substitutes.' (Sculpher et al., 2006)[135]

Sculpher and colleagues[135] therefore argue that trials should been seen as a source of inputs into, versus a vehicle for economic evaluation. One of the downsides to conducting a trial is the expense and time involved; decision modelling can make use of existing evidence at a much lower cost (the cost of the analyst's time). It can address many of the limitations of trial-based economic evaluations such as consideration of all relevant alternatives, appropriate time horizon, evidence synthesis, and management of uncertainty.[136]

Models are often criticised for being 'black boxes' whereby the inner workings are not transparent therefore making it difficult to assess the validity of the model results. For models to be useful, decision makers need to be confident in the results. This confidence comes from transparency in terms of model structure, parameters, and assumptions and validation, or how well the model represents reality.[137]

One of the first stages in developing a decision model is the conceptualization of the model. This involves specifying the decision problem and its components[136] followed by model conceptualization which incorporates these components through a choice of a particular analytic method.[138] These methods might include a decision tree, Markov model, discrete event simulation, or dynamic transmission model. The next step involves identifying and synthesising available evidence through a systematic approach. The final steps include dealing with uncertainty and assessing if there is value in undertaking additional research.[136] Uncertainty is often a major consideration in decision making and how the analyst handles uncertainty is important. A decision maker who adopts a do nothing approach in response to an evaluation with too much uncertainty is still a decision not to implement the new health technology. What if that was the wrong decision? A way to quantify the value of acquiring additional information to inform a decision problem is through value of information analysis.[139] The potential benefits of further research (reduced uncertainty) are compared to the costs of further investigation to help with prioritisation of research recommendations (e.g. invest in further research because uncertainty is large and the value of making the wrong decision it too great, or use those funds elsewhere because there is currently enough information to make an informed decision).

With healthcare spending increasing worldwide many countries have adopted use of economic evaluation to aid decision making about which health technologies to fund, as well as, to combat the unsustainable increases in spending. The UK specifically adopts a £20,000 to £30,000 per QALY threshold to aid decision making. The WHO uses a multiple of the average per capita income to determine cost-effectiveness in low and middle-income countries. The US chooses not to set a threshold, instead preferring to adopt a range of thresholds that are considered in relation to clinical and policy recommendations; there is no requirement to abide by any of these thresholds. Other country specific guidance can be found by visiting the ISPOR website at:

https://www.ispor.org/PEguidelines/index.asp. Finally this subsection, concluded by describing the various vehicles for economic evaluation. They can be conducted alongside trails, in a decision analytic framework, or a mix of both methodologies and the advantages and disadvantages of each were described.

2.1.9 Summary

Section 2.1 introduced fundamental concepts of economics, health economics, and a brief history of discipline. The various methods of economic evaluation were introduced; the three full economic evaluation methodologies being CEA, CUA, and CBA, with CCA and CMA being considered partial economic evaluations. The final section detailed the role of economic evaluation in decision making in the UK and internationally. This was described in terms of cost-effectiveness thresholds and trial-based, model-based, and mixed economic evaluation methodologies. This section introduced the general methods of economic evaluation in a typical clinical trial hospital based economic evaluation. The next section details economic evaluation of PHIs, which often take place outside of the hospital setting where a wider perspective is more appropriate than that of the health services perspective. RoE falls into this category.

2.2 Economic evaluation of population health interventions

What is population health? The Population Health Intervention Research Initiative for Canada has defined population health as,

> '…policy and program interventions that operate within or outside of the health sector and have the potential to impact health at the population level.'[21]

PHIs are characterised similarly as population or community-oriented programmes intended to *promote, protect,* and *prevent* ill health.[60] They may be delivered in the community, workplace, or school and are usually considered different from health service and clinical interventions which are intended to *treat* illness in individuals. However, it is recognised that public health agencies and health care services must work together closely to provide early intervention.[60] The terms population and public health are often used synonymously. In those who make a distinction between the two, it is usually to

define public health as the actions of local public health departments to prevent disease and promote healthy behaviours, whereas population health is defined more broadly as the health outcomes of a group of individuals, including the distribution of such outcomes within the group.[140, 141] In any case, a PHI aims to improve outcomes at the population level. Policy changes aimed to improve health, might include taxes to reduce consumption of foods and substances related to ill health such as cigarette, alcohol, and sugar taxes. Or policy changes might directly target food producers, such as requirements to lower salt in processed food, or requirements to clearly label certain products such as monosodium glutamate (MSG) on food packaging. Other government initiatives might include increasing green space or providing cycle and walking paths to increase active transport and physical activity. Interventions may take place at the school level affecting the school-age population. In each of these examples, the intervention is aimed at the population versus specific individuals.

The rest of this section is split into four subsections. The first subsection describes evidence for the cost-effectiveness of PHIs and appropriate methodologies for evaluating population health programmes. The next subsection deals with the economics of prevention and introduces the prevention paradox while making a case for investing in preventive programmes such as a PHI. The third subsection details to current challenges of performing economic evaluation of population health programmes, and how this thesis aims to address these challenges. The final subsection summarises the themes introduced in section 2.2.

2.2.1 Cost-effectiveness of PHIs

In addition to improving outcomes, PHIs have the potential of being cost-effective through efficiencies that are achieved through providing health intervention at the population level. These efficiencies can be achieved by spreading the cost (and potential savings) over an entire population as well as by reaching a whole population with one initiative. As costs are spread out over the population, so too are the outcomes of the PHI which may result in minimal changes at the individual level (more on this concept described in section 2.2.2.1). A WHO report estimates that population level approaches cost on average five times less than individual intervention.[131] Additionally, investing in ‘upstream’ preventive activities aimed at the population is more effective at reducing

health inequalities than 'downstream' approaches.[142] Upstream refers to the prevention of the causes of ill health before healthcare services are needed, while downstream refers to the services received to treat ill health. Public health policy makers understand the tension of breaking away from the medical model to divert funds into the investment of more upstream prevention activities.[142] Often times the evidence-base is less strong for upstream approaches, policy makers can be short-sighted and target driven, and there is too much pressure from the current patients needing treatment to address the prevention of the condition.[142]

> "Medicine is failed prevention." – Sir Michael Marmot[143]

It is worth noting however, that while prevention can save lives and increase net health benefit, it does not always save money. A microsimulation model for chronic disease prevention targeting diseases of obesity and physical inactivity was developed as a WHO and Organisation for Economic Co-operation and Development (OECD) initiative.[144] It found that school-based interventions are likely to have only modest effects, and they might not even be meaningful for another 40 or 50 years. The model did look at other interventions that were found to be cost saving in the long-term such as fiscal measures, food advertising and regulation, and food labelling.[144] However, it is also worth pointing out that there seems to be an expectation that population health or preventative measures *need* to be evidenced as cost saving in order to justify the investment.[145] Medical intervention such as surgery, or new drugs are not subjected the same expectation.[145] We do not normally expect the effects of a drug to continue to last long after the drug is stopped; and therefore eventually save money to the health service. Once the drug is stopped, the effects stop, and if continued effectiveness is desired, then continued investment in the drug is required.

Return on Investment (ROI) originates from a business context and is the direct financial return received from an investment.[146] Calculation of ROI is given in Equation 6. If an initial investment of £50 returned £75, the ROI would be a 50% ROI.

Equation 6: Return on Investment

$$ROI = \frac{£\,Gained - £\,Cost\ of\ Investment}{£\,Cost\ of\ Investment}$$

By viewing population health initiatives simply in terms of ROI, potential health and wellbeing gains over the long-term could be disregarded because they fail to save money in the short-term.[23] Based on this thinking, many valuable public goods should also be abandoned, such as public libraries, parks, and museums.[30] There is a need to correct this misguided notion that prevention must be cost-saving in the long run, and instead invest more resources into preventive measures that prevent patients from developing the resource-intense chronic diseases that require long-term contact with the health service. To illustrate this point further, NICE advice on judging whether public health interventions offer value for money states that, 'public health interventions cannot, however, be viewed solely in terms of value for money because of the broader and longer-term impact they have on general wellbeing – not only for individuals but also the wider community'(p. 1).[23] Now that the concept of prevention being cost-effective has been introduced, the next section discusses the appropriate types of economic evaluation for the evaluation of PHIs.

2.2.1.1 Economic evaluation methodologies for population health

Section 2.1 described the types of economic evaluation, in this subsection each type will be discussed in terms of its use in population health. As there is a distinction between population health and healthcare interventions, there is also a distinction between economic evaluations of both types of interventions. This has been formally recognised by NICE when developing separate guidance for technology appraisals of PHIs.[79] The guidance recognises the differences in the nature and scope of population based interventions which require different methods for technology appraisal particularly, in relation to perspective, type of economic evaluation and discount rate used for both costs and effects.[79] The guidance places more emphasis on the use of CCA and CBA methodologies than it has in previous methods manuals, however the use of QALYs and CUA will still be required routinely as a 'yardstick' measure of effectiveness comparable across health and disease areas. The guidance also points out that all NICE programmes should include the use of a common method of economic evaluation that allows comparison between programmes. Indeed, in some cases of population/public health interventions almost all benefits are health benefits, and therefore if inclusion of further analysis such as CCA or CBA is unlikely to change a decision (because there is a clear indication of cost-effectiveness or ineffectiveness), their use is not required.[79] The main

limitation of using CUA in population health is its narrowness, or its inability to capture a broad ranging set of non-health outcomes. Many PHIs result in non-health benefits that would not be captured in a narrow cost per QALY outcome. An example is increased labour productivity as a result of a workplace intervention. Healthier people take less sick leave and are more productive in the workplace; however, this outcome would be missed in a CUA. A published CUA of a public health intervention is the economic evaluation of the Football Fans in Training programme.[147] The physical activity programme was run across Scotland in football stadiums and included a classroom-based heathy diet and lifestyle component. A CUA was employed as the programme was primarily focused on improving the health outcomes of the men, and the evaluation was also in line with NICE recommendations. However, important spillover effects were not captured in the economic evaluation, such as the impact that the men's lifestyle changes had on changing their partner and/or family's lifestyle, which included healthier family behaviours.[147]

The use of CEA in population health has many of the same disadvantages to CUA; an additional disadvantage is that a non-QALY outcome does not provide the advantages of using a common 'yardstick' measure that QALYs provide. An advantage is that if there is not enough data to estimate QALYs, a natural unit such as a disease specific outcome, or cases averted may be used which might capture more appropriate health benefits of the intervention. An attempt can be made to capture more health outcomes by conducting multiple CEAs of various health and non-health outcomes that are available. However, interpretation of CEA ICERs can be difficult and place more burden on decision makers to interpret cost-effectiveness of different outcomes (e.g. deciding appropriate cost-effectiveness thresholds for each outcome).

Population health economists argue that economic evaluation should not be equated with CUA; CCA and CBA may be better frameworks to capture and value health and non-health outcomes with broader aims.[36, 79] CBA encompasses all cost and benefits, therefore incorporates societal interests. Also, expressing benefits in monetary terms avoids interpretation difficulties of non-aggregated outcomes such as those in CCA, or in the case described above with the use of multiple CEAs of different outcomes. However, there are concerns over the monetary values that survey participants place on outcomes[148] as well as individual preferences not being expressed through the market.[107] Indeed, CBA is often mentioned by experts as an alternative to CUA,[107] but it is still not

often used in economic evaluation of PHIs due to practical and methodological reasons.[36] Lack of standardisation in elicitations methods, stated-preference biases, and the considerable measurement burden (as CBA needs to be tailored for each intervention) are some of the practical barriers of CBA.[107] To address some of these issues New Economy, a research support group based in Manchester, developed in depth guidance on how to conduct CBA in a local public services context where analytical and research resources may be limited.[149] The guidance is also supported by an example excel-based CBA model and unit cost database with more than 600 unit cost estimates. While this is a useful resource and a good starting point to encourage CBA in the public services and population health context, analysts may still be limited by the unit cost estimates that are available in the database. The database would benefit from cost estimate contributions from reputable sources (e.g. government, academic, etc.) as its use grows.

Finally, the advantages of employing CCA in population health is that welfare and quality of life can measured more broadly with this methodology. Relevant outcomes do not need to be converted in any way as they are reported in their natural form, and broader outcomes that decision makers might find useful can be included, such as spillover effects into other sectors of interest. CCA does have its disadvantages, the difficulty in aggregating outcomes mentioned previously is one. It also takes more time and resources to measure broad outcomes versus a single QALY measure. Individuals may rank outcomes differently resulting in allocation decisions that may be less transparent and systematic, which is why NICE still requires CUA, but considers alternate forms of economic evaluation due to the stated advantages. CCA is still a relatively uncommon type of economic evaluation method used in recent literature of PHIs. This is partly explained by the disadvantages mentioned above, and also may partly be due to lack of familiarity with the method as it has since fallen out of the latest version of the Drummond 'blue book'[68] as mentioned in section 2.1.7.

In 2014, NICE published an updated manual for developing NICE guidelines which incorporated reference case guidance for interventions with outcomes in NHS, public health/public sector, and social care settings.[150] Table 1 replicates Table 7.1 provided in this updated guidance. In the table, CMA is included as a type of economic evaluation that NICE would consider, however the guidance specified that this is rarely used because it is unusual to find two interventions that provide exactly the same health benefits.

Table 1: NICE reference case side-by-side comparison summary. Replicated from Developing NICE guidelines: the manual (PMG 20)[150]

Element of assessment	**Interventions with health outcomes in NHS settings**	**Interventions with health and non-health outcomes in public sector and other settings**	**Interventions with a social care focus**
Defining the decision problem	The scope developed by NICE		
Comparator	Interventions routinely used in the NHS, including those regarded as current best practice.	Interventions routinely used in the public sector, including those regarded as best practice.	Interventions routinely delivered by the public and non-public social care sector.[1]
Perspective on costs	NHS and PSS.	Public sector – often reducing to local government. Societal perspective (where appropriate). Other (where appropriate); for example, employer.	
Perspective on outcomes	All direct health effects, whether for people using services or, when relevant, other people (principally family members or informal carers.)	All health effects on individuals. For local government and other settings, non-health benefits may also be included.	Effects on people for whom services are delivered (people using services and/or carers.)
Type of economic evaluation	Cost-utility analysis.	Cost-utility analysis. Cost-effectiveness analysis. Cost-consequences analysis. Cost-benefit analysis. Cost-minimisation analysis.	
Synthesis of evidence on outcomes	Based on a systematic review.		
Time horizon	Long enough to reflect all important differences in costs or outcomes between the interventions being compared.		
Measuring and valuing health effects	QALYs: the EQ-5D is the preferred measure of health-related quality of life in adults. [2] 'Social care QALY' with parallel evaluation based on capability measures where an intervention results in both capability and health or social care outcomes. ASCOT instruments may be used as measures of social care quality of life and ICECAP instruments may be used to measure capability		
Measure of non-health benefits	Not applicable.	Where appropriate, to be decided on a case-by-case basis.	Capability measures where an intervention

			results in both capability and health or social care outcomes.
Source of data for measurement of quality of life	Reported directly by people using services and/or carers.		
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population.		
Discounting	The same annual rate for both costs and health effects (currently 3.5%). Sensitivity analyses using rates of 1.5% for both costs and health effects may be presented alongside the reference-case analysis. In certain cases, cost-effectiveness analyses are very sensitive to the discount rate used. In this circumstance, analyses that use a non-reference-case discount rate for costs and outcomes may be considered.		
Equity considerations: QALYs	A QALY has the same weight regardless of the other characteristics of the people receiving the health benefit.		
Equity considerations: other	Equity considerations relevant to specific topics, and how these were addressed in economic evaluation, must be reported.		
[1] Social care costs are the costs of interventions which have been commissioned or paid for in full, or in part by non-NHS organisations. [2] Guidance from The social care guidance manual (PMG10)[151]			

2.2.1.2 Emerging methodologies for facilitating economic evaluation of PHIs

In response to the requirements for evaluation of PHI's there has been an emergence of new evaluation approaches. Natural experiments are one such emergent research design, which have long existed as an observational study design. To date there is no comprehensive guidance for conducting economic evaluation alongside natural experiments; it is an emerging field of study. Natural experiments are defined as,

> 'Naturally occurring circumstances in which subsets of the population have different levels of exposure to a hypothesized causal factor in a situation resembling an actual experiment. The presence of a person in a particular group is typically non-random; yet for a natural experiment, it suffices that their presence is independent of (unrelated to) potential confounders.' (Porta, 2014 p 193)[152]

Natural experiments are not a new experimental design methodology; an early example is the investigation of the distribution of cholera cases in London by John Snow, often deemed "the father of epidemiology."[152] By tracing the source of the drinking water for those who recently developed cholera and those who did not, he was able to identify the source of the contaminated water supply. It was a single water pump that he disabled, which therefore prevented any further cases in that epidemic.

A key difference between natural experiments and RCTs is that the researcher cannot control the group allocation or input into the data that is collected. It is a type of observational study that is reliant on the use of secondary data, which might come from routinely collected sources, surveys, administrative, or census data. Natural experiments may arise from policy changes, e.g. a policy affecting only Scotland where England would be considered a natural comparison group. The non-randomisation element is a threat to internal validity, which is when one can draw inference from the observed outcomes of a study and infer they were actually caused by differences in relevant explanatory variables.[153] Systematic reviews assessing the agreement between non-randomised intervention studies and RCTs of the same clinical question find differences in the results of the two different study types.[154, 155] Non-randomised intervention studies tended to overestimate treatment effects, and caution is needed when interpreting results of natural experiments due to the potential presence of residual bias.

Despite these biases and threats to internal validity, natural experiments have several positive aspects that are attractive when considering economic evaluation. Although researchers cannot directly input into data collection, they can exploit the use of existing data, which saves on time and resources needed to collect new data. If data linkage is available, researchers could access a multitude of existing health and resource use data, both within the hospital and GP setting, and including prescriptions. This is often more accurate than self-recall and over a longer time horizon than RCTs which are usually limited to short follow-ups of not more than a year. Routinely collected data sources are also likely to reduce the loss to follow-up and low response rates that are observed in trials. Finally, the observed data is practical, real-world, data. There are concerns that effects observed in RCTs are not always replicated in real life due to implementation issues as well as the potential bias from the Hawthorne Effect.

Another emerging methodology for conducting economic evaluation of PHIs is multi-criteria decision analayis (MCDA). MCDA has been defined as, 'a set of methods and approaches to aid decision-making, where decisions are based on more than one criterion, which make explicit the impact on the decision of all the criteria applied and the relative importance attached to them.'[156] Recently, health economists have been turning to MCDA with the aim of improving transparency in decision making with more explicit scores and weights for different consequences considered in a decision problem.[157] A main critique of CUA within PHIs is that benefits that go beyond the QALY are not captured. PHIs will likely take place in a community setting rather than a hospital setting giving rise to non-health benefits in education, transport, housing, labour, and criminal sectors. MCDA attempts to aid decision makers considering multiple criteria in an explicit and transparent manner.[158]

If MCDA were to be adopted by an organisation like NICE for example, a decision would have to be made if a generic or appraisal specific approach will be taken. A generic approach would involve pre-specified criteria and subsequent weights. It may improve comparability, however there is the risk that relevant benefits specific to certain appraisals will not be captured, as is the case currently with the QALY. An appraisal specific approach would not use generic criterion or weights, however, this would involve significant cognitive burden on decision makers as they will need to identify and provide preference weights of all relevant benefits.[157] To date, MCDA methods have not been widely adopted in health care decision-making.[159] In 2014, an Emerging Good Practices Task Force was established by ISPOR to develop good practice guidelines for conducing MCDA in health care decision-making. The first task force illustrated the many different types of MCDA methods available for different decision-making contexts and provided a list of steps in the value measurement process (see Figure 4).[160] The second task force reports guidance on how to implement MCDA in healthcare decision-making including a checklist () with accompanying guidance.[161] However, specific guidance relating to how MCDA should be used in HTA still requires further research.

Table 2 – Steps in a value measurement MCDA process.

Step	Description
Defining the decision problem	Identify objectives, type of decision, alternatives, stakeholders, and output required
Selecting and structuring criteria	Identify criteria relevant for evaluating alternatives
Measuring performance	Gather data about the alternatives' performance on the criteria and summarize this in a "performance matrix"
Scoring alternatives	Elicit stakeholders' preferences for changes within criteria
Weighting criteria	Elicit stakeholders' preferences between criteria
Calculating aggregate scores	Use the alternatives' scores on the criteria and the weights for the criteria to get "total value" by which the alternatives are ranked
Dealing with uncertainty	Perform uncertainty analysis to understand the level of robustness of the MCDA results
Reporting and examination of findings	Interpret the MCDA outputs, including uncertainty analysis, to support decision making

MCDA, multiple criteria decision analysis.

Figure 4: List of steps to perform MCDA from Thokala et al.[160] ISPOR task force

Table 1 – ISPOR MCDA Good Practice Guidelines Checklist.

MCDA step	Recommendation
1. Defining the decision problem	a. Develop a clear description of the decision problem b. Validate and report the decision problem
2. Selecting and structuring criteria	a. Report and justify the methods used to identify criteria b. Report and justify the criteria definitions c. Validate and report the criteria and the value tree
3. Measuring performance	a. Report and justify the sources used to measure performance b. Validate and report the performance matrix
4. Scoring alternatives	a. Report and justify the methods used for scoring b. Validate and report scores
5. Weighting criteria	a. Report and justify the methods used for weighting b. Validate and report weights
6. Calculating aggregate scores	a. Report and justify the aggregation function used b. Validate and report results of the aggregation
7. Dealing with uncertainty	a. Report sources of uncertainty b. Report and justify the uncertainty analysis
8. Reporting and examining of findings	a. Report the MCDA method and findings b. Examine the MCDA findings

MCDA, multiple criteria decision analysis.

Figure 5: MCDA checklist from Marsh et al.[161] ISPOR task force

Natural experiments and MCDA although still in development in a HTA context, provide novel opportunities for the economic evaluation of PHIs. As these methods are further developed and guidance issued, some of the first economic evaluations using these study designs are likely to start appearing in the literature. Earlier, this subsection introduced the concept of 'prevention being better than cure' as well as the potential cost-effectiveness of preventive PHIs. CBA and CCA may be more appropriate methodologies for evaluation of PHIs, however, their use in the recent literature is still low due to them being resource intense to conduct (CBA), and placing more cognitive burden on decision makers (CCA) as well as for other reasons. Additionally, difficulties evolving from the status quo, primarily the use of CUA, may also be contributing to their restricted use.

NICE specifically requires CUA in addition to considering other methodologies such as CBA and CCA, but states if most of the benefits of PHIs are health benefits and the additional analyses are unlikely to change the decision, they will not require CBA or CCA. The next subsection sets up the economic case for investing in preventive measures.

2.2.2 Economics of prevention

PHIs are preventive by nature; they aim to prevent ill health and illnesses that need treatment in medical centres. In addition to the health advantages, there are economic advantages to preventing illness and disease. On the most basic level, one case of disease prevented, saves all health and social care resources that would have been spent in treating the disease. Most prevention efforts are aimed at changing unhealthy behaviours and lifestyles (such as smoking, being overweight, obese, or inactive) that are associated with increasing risk of disease. An economic approach to prevention looks at health behaviours functioning much like goods consumption functions in market places. Many external influences impact on individual choices such as cost, opportunity, incentives, and constraints and economists see individual health behaviours being influenced in much the same way as choices for goods consumption are influenced by market forces.[162] However, sometimes markets fail to operate efficiently. There are market failures that create an economic rationale for government intervention as a means to increasing societal welfare. Individual health behaviours may lead to costs beyond the individual that society bears such as diseases and fatalities related to second hand smoke or traffic fatalities from driving under the influence of alcohol. Economists call these externalities and prices will not reflect these impacts in the free market.[162] Perfect or sound information about lifestyle choices may not always be available, or individuals may not be able to make rational choices due to addictive behaviours for example, substance abuse. Additionally, individuals may be myopic, choosing to enjoy an unhealthy lifestyle today and highly discounting their future risks. They may also plan and fail to make a future change.[162]

Government intervention is therefore acceptable if it increases social welfare even at the expense of individual choice. Examples include indoor smoking bans, taxes on alcohol, tobacco, sugary drinks, and lowering the drink drive limit, as was recently implemented in Scotland. These population initiatives aim to improve population health and as an added benefit reduce health expenditure by encouraging healthy behaviour change that leads to

reductions in overall disease prevalence. These policy initiatives are not always causally linked to reductions in disease prevalence, but they certainly do help to reduce the unhealthy behaviour.

2.2.2.1 The Prevention Paradox

The prevention paradox was coined by Geoffrey Rose, 1981[163] as,

> 'a measure that brings large benefits to the community offers little to each participating individual.'

Rose refers to a 'mass strategy,' which today might be referred to as a population health approach. A mass strategy involves endeavouring to lower the distribution of risk over the entire population. The individual gains little, but the small individual benefits add up to significant community level benefits. This is opposed to a 'high-risk' strategy in which those at highest risk are identified and offered intervention.[163] Figure 6 depicts the differences between a population level and high-risk approach to prevention.

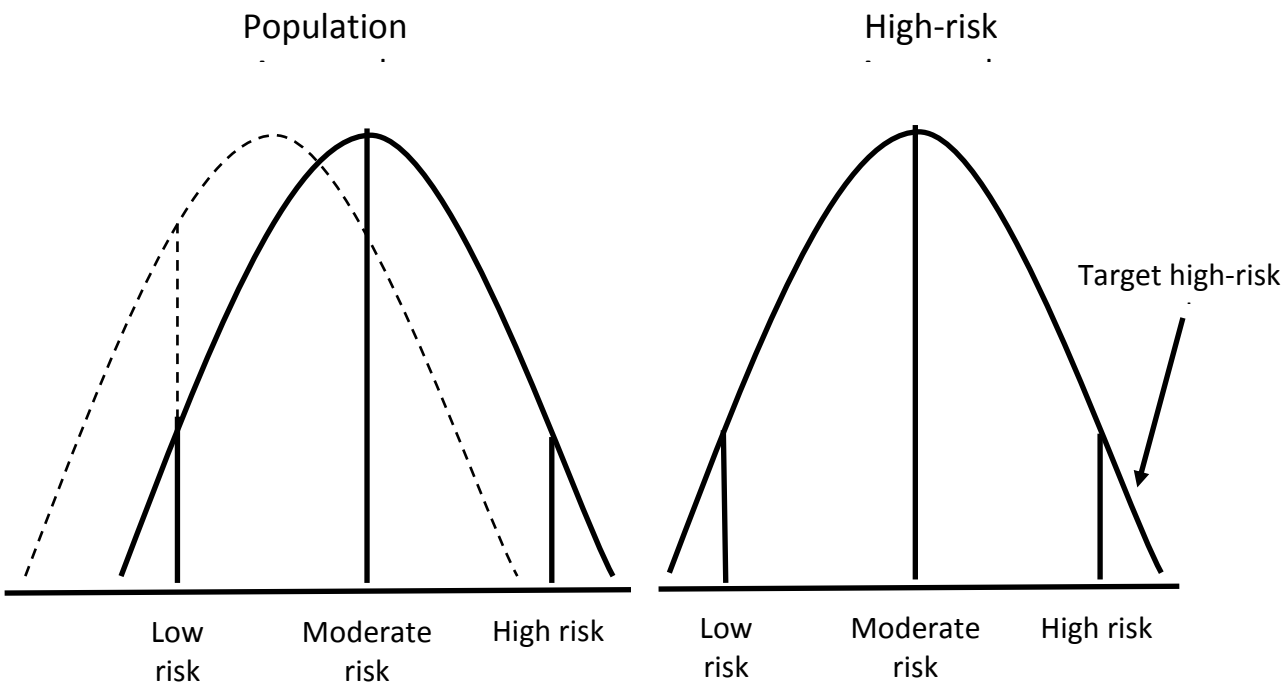


Figure 6: Shifting the distribution of risk in the population approach versus a targeted high-risk approach

As can be seen in Figure 6, the population approach aims to shift the entire distribution of risk, therefore shifting everyone out of the high-risk category, but also shifting a larger proportion of the population into the moderate and low risk categories. The high-risk

approach only targets those at highest risk, which leaves the majority of the rest of the population in moderate risk, with only a small proportion in low-risk. Conversely, if the intervention poses any potential harm or unwanted side-effects, the population approach would subject more of the population to these unwanted side-effects. If the risk is high and a treatment poses a small risk of harm, then a trade-off is usually taken if the benefits outweigh the risks. But if the risk is low or moderate to begin with, the trade-off of benefits may not always outweigh the risks so it is important that population level approaches do not cause any harm and are safe[163] because of the potential for a subsection of the population being exposed to unnecessary harm.

One of the most successful examples of a population health approach to prevention is the North Karelia Project in Finland. In the 1970s, mortality from coronary heart disease in Finland was the highest in the world, particularly in the eastern part where North Karelia is located.[164] The project, established in 1972, was the first large-scale community-based prevention programme for cardiovascular disease. From 1972-2012, a 40-year period, smoking prevalence, serum total cholesterol, and systolic blood pressure declined immensely[164] due to the efforts of the project to change the population's lifestyle. This correlated with an 82% decrease in coronary heart disease mortality in working-age men, and an 84% decrease in mortality for working-age women.

In the Netherlands, a study aimed to compare population and high-risk strategies to prevention in regards to their impact on population health between 1970 and 2010.[165] Twenty-two preventive programmes were identified during that period and classified as either population or high-risk if they specifically targeted groups based on their risk of disease. The study found considerably larger health gains from population approaches such as tobacco control and road safety measures, versus high-risk approaches such as hypertension detection and cancer screening.[165]

These examples clearly demonstrate the benefits of adopting a population approach to intervention. If the entire population is exposed to a safe intervention, then the entire distribution of risk in that population can be shifted placing more people in low and moderate risk than would be the case if only the high-risk group was targeted. However, the prevention paradox states the individual benefits may be insignificant or non-existent, but those small benefits added up over the population could result in significant

community or population benefits. This is particularly relevant for RCTs of PHIs or public health interventions as measures are collected at the individual level. Often effect sizes are small and/or diminishing, and this could partly be explained by the prevention paradox. PHIs evaluated in a cluster RCT framework typically only include a subset of the whole population, and therefore the sum of the combined effect may not reach significance, as the entire population was not included in the study. Investing in preventive and effective measures of the population have the potential to bring large community benefits, as well as being cost-effective. However, as economic evaluation moves from a narrow NHS setting to a broader public sector/societal perspective, challenges start to emerge.

2.2.3 Challenges of economic evaluation of PHIs

Section 2.2 focused on the appropriate types of economic evaluation methodologies for population health initiatives, as well as introducing key concepts of the economics of prevention such as the prevention paradox. A recent synthesis of methods guidance for undertaking economic evaluation of PHIs identified only four guidance documents (with a fifth being identified during the publication process).[166] Amongst the guidance identified, there was heterogeneity in approaches to deal with the challenges of evaluating PHIs and variations were unjustified. The author suggested the lack of consensus may be due to insufficient development of methods to evaluate PHIs.[166] This current section highlights the main challenges of conducting economic evaluation of PHIs. A pivotal piece of work by Weatherly and colleagues[36] identifies four key challenges of conducting economic evaluation of public health interventions. Difficulties arise in moving from the strict clinical setting of CUA alongside clinical trials from the health care perspective, to a broader public sector perspective of non-clinical and often cluster RCTs. Economic evaluation has long been recognised as necessary in clinical health care settings, but until recently, there seems to have been less appreciation for it in public health even though there is a clear need for it.[167] Weatherly et al.[36] identifies key methodological challenges as: i.) attribution of effects, ii.) measuring and valuing outcomes, iii.) identifying intersectoral costs and consequences and iv.) incorporating equity considerations. The following details each of these challenges separately.

2.2.3.1 Attribution of effects

There is a preference for economic evaluation to be conducted alongside the gold standard RCT comparing all relevant alternatives.[64] However, relatively few complex interventions delivered at the population level have been evaluated in an economic evaluation RCT framework.[36] The relatively short trial follow-up poses problems in economic evaluation as health economists are interested in lifetime costs and benefits of an intervention and prevention programmes may impact on health over the long term. The value of prevention is more difficult to assess than evaluating treatment of established disease because the long time horizon introduces considerable uncertainty around potential benefits of prevention.[168] Weatherly et al.[36] recommends when extrapolating outcomes beyond the trial follow-up, the outcomes used in the trial should be the same as those available in long-term observational studies. With a broader spectrum of public health intervention outcomes, observational studies over the life-course may not exists for all outcomes of interest, particularly paediatric outcomes which involve a longer lifetime follow-up. Longer time horizons coupled with fragile causal relationships in preventive care, results in higher levels of uncertainty that need to be modelled appropriately.[169] This may require a variety of analytic assumptions to be made and these should be reported clearly to allow for replication and/or further extrapolation of the findings.[170]

This is perhaps where the reliance on RCT as the 'gold standard' for conducting economic evaluation needs to be re-evaluated as mentioned in section 2.1.8.1 when recognising that the economic evaluation alongside RCT should not be the sole vehicle for economic evaluation and the use of other emerging methodologies (section 2.2.1.2) and modelling may be appropriate. Because of the prevention paradox, population or public health interventions evaluated alongside an RCT may show small or diminishing effect sizes that might disappear at follow-up. This is particularly troublesome when attempting to model some time into the future, as assumptions need to be made about these highly uncertain long-term outcomes, which in many cases may be inappropriate.

2.2.3.2 Measuring and valuing effects

Because public health interventions typically take a broader public sector perspective,[79] challenges arise when attempting to measure the non-health outcomes in the various

sectors the intervention might impact (i.e. crime, education, or transport). Ideally, a societal perspective should be taken and all costs and benefits should be measured and valued, no matter to whom they are accruing.[77] However, in many cases this is impossible or impractical to do so.[167]

Additionally, there may be spillover effects into different sectors that weren't targeted by the intervention (i.e. a more informed public); some of these effects may fit well within the QALY framework and some may not.[36] It has been suggested that QALYs may be less appropriate in public health interventions aimed at addressing behaviour change and inequalities in health.[167] The extra-welfarist view, with a sole focus on maximising health utility through a QALY framework may not be appropriate for population health because it may result in underestimates of the benefits gained.[103] There have been proposals for a broader utility measure such as the 'capability-QALY,'[171] the 'super-QALY'[172] or the WELBY (wellbeing adjusted life years)[173] as a way to capture non-health benefits.[174] Other approaches might include a return to welfare economics by adopting CBA, CCA, behavioural economics, taking a capabilities approach, [103, 145] or using MCDA.[107] While these alternatives exist to incorporate non-health outcomes, experts cannot agree on a single preferred alternative method as revealed in a recent qualitative study that interviewed experts on their views of the incorporation of non-health outcomes in economic evaluation.[107]

Identifying and measuring the appropriate additional non-health outcomes in PHIs is time-consuming and with NICE still requiring a cost per QALY as a 'yardstick' metric to compare all other evaluations, researchers may be tempted to forego additional analyses using other perhaps more appropriate methodologies (such as CCA or CBA) due to resource constraints. The lack of consensus for a preferred alternative may also defer researchers from investing their research resources in explorative methodologies that have less concrete evidence-bases as they may not be accepted by certain funders and journal editors. Careful planning at the outset of an economic evaluation needs to take into consideration, the time, effort, and resources required to identify and measure the additional health and non-health outcomes that are relevant to a wider public sector or societal perspective. This planning will feed into the appropriate choice of type of economic evaluation that will be performed.

2.2.3.3 Identifying intersectoral costs and consequences

The impact of public health interventions can be wide ranging with costs and benefits falling into different sectors of industry. Just as measuring these spillover effects proves challenging, so too does identifying them in the first place. There may be knock-on effects where expenditure in some sectors may reduce expenditure in others,[36] e.g. expansion in public transport may increase physical activity, thereby reducing obesity related illness and health care utilisation. Therefore, it is important to clearly state the perspective of the analysis as the costs and consequences included in the analysis should reflect the perspective of the analysis. Ideally, the evaluation should span several sectors as many agencies may be involved in provision of PHIs, particularly early childhood intervention, and benefits may be felt across several domains.[175]

In their review of economic evaluation of public health interventions, Weatherly et al.[36] identifies numerous costs that fall outside of health care services including: productivity losses, out-of-pocket, social care, criminal justice, voluntary, education, housing, environment and transport. Compelling arguments for differential discount rates of costs and benefits exists,[176] however current NICE guidance suggests a discount rate of 3.5% applied to both cost and health effects with a sensitivity analysis using a 1.5% rate.[150] Claxton et al.[177] propose a compensation test for interventions with multisectoral effects which involves preferences based on net benefits falling on different sectors. Outcome valuation would then be based on shadow prices of existing budget constraints. A final approach to consider is a general equilibrium approach where consequences of different interventions across all sectors of the economy are considered simultaneously.[36] A macroeconomic approach may be more suitable for capturing spillover effects whereas a traditional microeconomic approach focusing only on the healthcare sector may mis-specify benefits and underestimate costs.[178]

As is the case in the previous section, identifying intersectoral costs and consequences is time consuming and with little consensus, one might wonder when to stop collecting and measuring intersectoral costs and consequences as population health initiatives may have small but broad societal effects. A balance must be struck between available research resources and comprehensive measurement of intersectoral costs and consequences.

2.2.3.4 Incorporating equity considerations

Reducing inequalities in health are often an aim in public health interventions. However, health economists rely on the value judgement that 'a QALY is a QALY is a QALY' i.e. all QALYs are weighted equally regardless of to whom they accrue. There are notable exceptions to this rule; namely NICE's decision in January 2009 to consider giving greater weight to QALYs achieved in late stages of terminal disease at the end of life.[179] However, typically, the goal of economic evaluation is to maximise health over the population and not necessarily capture how health is distributed in terms of equity.[60] But there is evidence that a majority of the general public prefer to give greater weight to health gains accruing in children, the severely ill and to a lesser extent, the materially deprived.[180]

Clearly, equity considerations conflict with the strict economic evaluation goal of maximising health gains. Economic evaluation of public health interventions usually fails to identify and measure impacts on health inequality let alone value them.[180] From a review conducted by Weatherly et al.,[36] none of the interventions identified had outcomes that were explicitly equity-weighted. Furthermore, there is no field in the NHS Economic Evaluation Database abstract template to record equity considerations. More research is needed in public and stakeholder views on equity-weighting so that guidance can be issued to policy makers about how much efficiency, or total population health, should be sacrificed to pursue equity goals.[36]

Theoretically, providing a population health programme that is available to everyone should in itself, help to reduce inequalities. Practically however, this is not always the case when the most deprived fail to take up the intervention and those who are least deprived end up benefiting more and therefore increasing inequality. Inequality in health is much researched, but practical solutions are still yet to be agreed and implemented as evidenced by the increasing inequalities in health in the UK since the 1980 Black Report.[181]

To summarise this subsection, the four main challenges of conducting economic evaluation of PHIs are i.) Attribution of effects; ii.) Measuring and valuing effects; iii.) Identifying intersectoral costs and consequences; and iv.) Incorporating equity

considerations. In this thesis, an economic evaluation of PHI, RoE, was undertaken which aimed to address these challenges (see Methods in Chapter 4). The first challenge relates to more complex policy driven PHIs that cannot be evaluated in an RCT context. RoE is delivered at the school level, so it was possible to evaluate its effectiveness in a cluster RCT design. The other issue of the small and sometimes diminishing effects, related to the prevention paradox, is an important challenge to note when evaluating public and PHIs. The second challenge of measuring and valuing the broad ranging effects is something that needs to be dealt with in the design stage of an economic evaluation. In the RoE trail, efforts were made to capture any spillover effects that the school programme potentially had on parents by measuring parental EQ-5D. Even so, the QALY may not always be the best measure of effect in population health, so a separate CEA was conducted on a child-specific behaviour outcome measure, the Strengths and Difficulties Questionnaire (SDQ). Briefly, the SDQ is a behavioural screening tool which is being used recently as a measure of SEW (more detail to follow in section 4.2.1.4). To keep in line with current NICE guidance, the main analysis was a CUA with a cost per QALY outcome. The third challenge of identifying intersectoral costs and consequences is as challenging as the second, as it can be difficult to determine all appropriate costs and outcomes that the intervention may affect outside of the health sector. To address this, the RoE economic evaluation took a broader public sector perspective which included health, social service, and local education authority costs and resource use. The final challenge of incorporating equity considerations conflicts with the extra-welfarist goal of CUA to maximise health benefit regardless of who accrues the benefits. Determining how to measure equity considerations in a PHI is still a challenge. As RoE was give on a whole class basis, the most deprived children are able to access the benefits of the intervention as equally as less deprived children. Sensitivity analysis exploring the effects of RoE by deprivation is a starting point for exploring if and how RoE affects inequalities.

2.2.4 Economic evaluation in school settings

Up until this point, this chapter has focused mainly on reviewing the methods for economic evaluation as applied in healthcare and population health settings. This thesis is specifically concerned with economic evaluation of school-based interventions which is a subset of PHIs. As the methods for economic evaluation of PHIs are less well established than those of clinical hospital-based evaluation, the methods for school-based

interventions have even less consensus. The use of a threshold for cost-effectiveness in a healthcare setting was covered in section 2.1.8, in the UK and worldwide. Decision-making in population health settings in the UK still follow the same £20,000 to £30,000 threshold, with additional analyses such as CCA and CBA being considered.[79] Generally, PHIs tend to be cost effective; Owen and colleagues found 85% of the 200 public health interventions analysed to be cost-effective at the £20,000 threshold.[182] This would support the idea that the threshold could be lowered in the population health setting so as not to increase the opportunity cost of prioritising one intervention over another unnecessarily (see below for the author's interpretation of the value of a cost-effectiveness threshold). Perhaps the £13,000/QALY estimate by Claxton and colleagues mentioned in section 2.1.8 is not so far off as it would prioritise more preventive, community and PHIs.

However, as will be noted in Chapter 3, there is no consensus as to what constitutes a cost-effectiveness threshold in an education decision-making context. This is perhaps an opportunity to learn from agencies like NICE when establishing a transparent and consistent decision-making process (more on estimating a cost-effectiveness threshold outside of the health sector in section 5.7). As stated previously, the author of this thesis believes the value of the threshold should represent the opportunity cost resulting from investing in a new cost-effective programme. While the new investment will bring about benefits to those impacted by the programme, disinvestment may be required in another area of the education budget to adopt the new programme, resulting in potential losses to other areas (opportunity cost). Examples of these disinvestments might take the form of cuts to areas of education which are not considered core academic subjects, such as the arts, languages, sports and recreation, and "softer skills" such as SEL programmes. These other areas are at risk of disinvestment without their cost-effectiveness being established because it is likely that only new programmes will be evaluated for cost-effectiveness if economic evaluation of school-based programmes are deemed necessary in the future (as has been the case for new pharmaceuticals and health technologies in the healthcare sector).

As mentioned previously, a number of practical challenges have limited the use of CBA in healthcare contexts such as measurement burden and challenges and biases with preference elicitation. In the education setting, similar practical challenges, insufficient

training of researchers, and a lack of demand from policy-makers has limited the use of CBA in education.[183] Additionally, as a threshold has not been established in the education context, there is scope to incorporate flexibility to consider other methods of evaluation such as MCDA, CCA, and CBA. The QALY may not be the most appropriate outcome to base an education cost-effectiveness threshold, as health outcomes are less likely to take priority over education outcomes. Examples of effectiveness outcomes for CEA in education depend on the programme objective; generic examples of effectiveness outcomes might include: the number of students completing the programme, dropouts, graduates placed in appropriate jobs, college placements, and test scores.[184] However, as there is little consensus over a threshold in education, more research is required to determine how decision-making in education should be made consistent and transparent.

There are also calls from academics and policy makers to expand the use of economic evaluation in education policy.[185] Early examples of CEA in education focus solely on the head-to-head comparison of cost-effectiveness ratios (not ICER) of two competing alternatives while considering other criteria without focusing on an actual threshold.[184] CEA in education settings that use intervention specific outcomes, such as the SDQ, may be more attractive to decision-makers who may have more knowledge of descriptive measures versus preference-based measures. As alluded to earlier, CBA in its current form has limitations for practical application in the healthcare literature, the same is the case in education. CBA does not easily deal with issues of equity distribution, ethical consistency, and educational appropriateness, instead focusing on efficiency.[185] At any rate, all of the different types of economic evaluation require the use of child outcomes to measure effectiveness, and as the development of child outcomes has typically lagged behind adult outcomes, research in this area is necessary. Child outcomes will be explored further in Chapters 4-6.

2.2.5 Summary

Section 2.2 introduced and defined the concept of a PHI. The economics of prevention were discussed setting up the economic case for investing resources to prevent disease and ill health versus spending those resources later on in treatment. The concept of the prevention paradox was introduced and how PHI has the potential to impact population outcomes on a large scale. The next section covered the challenges of conducting

economic evaluation of population and public health interventions. Finally, this section concluded with reflections of decision-making in population health and education settings and the potential use of a cost-effectiveness threshold and flexibility to allow researchers to explore the use of other methodologies (in addition to CUA) in the education setting.

2.3 Conclusion

The fundamental concepts of scarcity and opportunity cost combined with market failure in the healthcare setting has led to the study of health economics. Scarce healthcare resources means that societies need to decide the most efficient way to use those resources understanding there is an opportunity cost to every decision. In the brief history (section 2.1.2), Chapin's idea that institutions find it difficult to break away from traditions of the past[82] was a theme that reoccurred throughout this chapter. In 1917, he hypothesised that if healthcare institutions were to start over with a new health care budget, they would probably end up with a different allocation of resources based on current knowledge of costs and effectiveness. However, starting over is difficult to do, and examples of this notion are the reluctance of NICE to consider a lower cost-effectiveness threshold, as well as a reluctance to break away from the standardised method of economic evaluation, the CUA.

The traditional methods of economic evaluation introduced in the first half of this chapter are fairly well developed and standardised. The second half focused on PHIs and how those traditional methods could be applied to deal with broader perspectives, inclusion of non-health outcomes, and other challenges. Additionally, the prevention paradox sets the case for large population-level benefits that can be achieved through PHI, however the individual will see little or no benefit themselves which can be problematic in smaller contexts such as RCT design frameworks. CBA and CCA might be a more appropriate way to deal with these challenges, and emerging methodologies such as MCDA and economic evaluations alongside natural experiment study designs may become more prominent in the future. However, in the UK, there is still a major focus on the CUA. Arguably, we may be coming full circle in reconsidering CBA for PHIs, as one of the first records of economic evaluation was a CBA of England's Public Health Act in 1875. The key concepts of this thesis were introduced in this chapter including those necessary for economic evaluation of RoE, a PHI, which is delivered at school. The next chapter unpicks the appropriate

methodologies of conducting economic evaluation in school settings. To reiterate from Chapter 1, schools are ideal settings for PHI as they have the ability to reach the majority of young people making the delivery of school-based PHIs efficient. Schools play an important role in developing and maintaining children's SEW, which is one reason why this thesis considers economic evaluation of a school-based SEL programme. Additionally, the theoretical basis, advantages, and challenges of conducting economic evaluation in this setting are explored. A systematic review of school-based economic and non-economic evaluation methodologies follows in Chapter 3.

3 School-based intervention evaluation methodologies: a systematic review

3.1 Introduction

CUA is a useful and preferred tool[150] for health care evaluation as the cost per QALY outcome provides a uniform measure to make comparisons across health areas. It can also be used in non-health settings where the intervention may give rise to health benefits, such as the community or school setting. The downside is that the current QALY framework does not take into account any non-health outcomes. CUA of a school-based programme (and economic evaluation in general) is a novel concept due to challenges associated with identifying, measuring, and valuing child health and broader, non-health outcomes. Schools are an ideal setting to promote children's health and wellbeing because school-based interventions can reach most children. However, decision-makers in education may be less willing to invest their limited resources in the promotion of health and wellbeing versus activities related to core education, even more so when they are under pressure to demonstrate added value.[186] In order to make informed decisions about how to best allocate funds, education decision-makers need complete information about the relationship between expenditures and pupil outcomes of interest, which includes details of how services are delivered.[187]

School-based health economic evaluation is uncommon partly due to the fact that until recently there were not many validated paediatric outcome measures, much less preference-based HRQoL measures. Selecting appropriate outcomes for children need to take into account the differences between children and adults and go beyond the assumption that children are simply small adults; very few studies address the appropriate choice of paediatric outcomes in clinical trials.[188] There are well validated and accepted preference-based measure for adults, but until recently, there has been less research into child preference-based measures of health.[96] Schools are continually being constrained by budget cuts, so economic evaluation could prove a useful decision making aid for prioritisation of school programmes.[185] Funding cuts leave teachers with fewer resources and less time to provide a comprehensive education. This scarcity means that only the most cost-effective school programmes should be prioritised, and the current preferred method for determining cost-effectiveness from a healthcare perspective is the

CUA. In education settings, the preferred outcome of economic evaluation still needs to be established (this idea is discussed further in Chapters 5, 7, and 8).

The combined effect of the challenges relating to economic evaluation of PHIs and a lack of research into child preference-based measures results in a novel and 'uncharted area' for economic evaluation methodology of school-based PHIs. Thus, because of the lack of clear methods guidance for undertaking economic evaluation in a school setting, a systematic literature review was conducted to identify evidence of economic and other evaluation methodologies that currently exist for the evaluation of school-based programmes. This review was conducted to gain an understanding of the methods of evaluation currently being practiced in school-based PHIs and to help inform the appropriate methodology for the economic evaluation of school-based interventions.

This chapter consists of four main parts. The first part describes the aim and the research question, the second part describes the methods used to conduct the review, the third part presents the results followed by the discussion and conclusion of the review. At the time of undertaking, this was the first study to systematically review school-based evaluation methods. Mounting pressure on educational decision makers to increase student achievement while constrained by education budgets, means that economic evaluation is an ideal tool to aid prioritisation.[29] However, the application of these methodologies is limited. The types of evaluation methods of school-based programmes will be explored to further understand if and how economic evaluation of school-based interventions is currently being conducted, and what types of preference-based child utility measures (if any) are being utilised.

3.1.1 Aim

The aim of this chapter is to systematically review all available evidence on evaluation methodologies for programmes that are delivered in a school or pre-school setting. Given the dearth of evidence anticipated, a broad definition of 'evaluation methodologies' was adopted to include economic and non-economic methodologies, which might include cost analyses or non-economic evaluation associated with a generic preference-based utility measure. This wide selection was deliberate because an initial scoping review revealed that the literature on school-based economic evaluation is much more limited than

clinical trial-based economic evaluation, and thus a broader search would minimise the risk of not identifying key studies as well identify key methodologies used to evaluate school-based intervention. School-based economic evaluation is a novel area; its use is limited in the education economics literature and the use of economic evaluation terminology differs slightly between the education and health economic literature. For example, 'cost-effectiveness analysis' in the education literature may be focused solely on which programmes 'can achieve particular objectives at the lowest cost'[184] which differs from the interpretation of CEA in health care which aims to maximise health within a certain threshold or budget. The education economics literature may also use the term 'cost-effectiveness ratio' to refer to average costs and effects per unit of outcome (which is not that same as an ICER) as well as to also refer to 'additional or marginal' costs and effects (which is similar to an ICER).[184] Because of these differences in terminology, it is unknown what methodologies are currently used, economic and non-economic (in addition to those outlined earlier). The decision was made to intentionally leave the type of school-based program or intervention undefined in order to maximise the identification of any existing evidence related to the aim of this review.

3.1.2 Research Question

This systematic review aimed to answer the following research question:

What evidence[c] currently exists around economic and other evaluation methodologies of school-based interventions and/or programmes?

3.2 Systematic Review Methods

The systematic review was conducted in line with recommendations from the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in healthcare.[189] Inclusion criteria were defined (section 3.2.1) and a search strategy was developed, piloted and revised several times through an iterative process before the final strategy

[c] Evidence is referred to in its broadest form and includes economic and non-economic evaluation evidence of studies that attempt to collect costs or resource use for the intervention, and/or a generic HRQoL measure. Please refer to Box 2.

was conducted to provide a comprehensive review of the existing literature (section 3.2.2).

The review team consisted of the author and two supervisors. All members of the team were involved in the design and final approval of the review. The author solely conducted the piloting, searching and reviewing, so to help mitigate bias in the review process, a series of validity checks were performed by all reviewers which is detailed in section 3.2.3.

Appropriate data were extracted, appraised (section 3.2.4), and synthesised through use of a narrative synthesis as detailed in section 3.2.5. The narrative synthesis was performed using guidance from the Economic and Social Research Council (ESRC) guidance for conducting a narrative synthesis.[190]

3.2.1 Inclusion Criteria

An initial search strategy was developed and subsequently revised after it was piloted with a specialist subject librarian. It was then revised further after input from the other two reviewers (PhD supervisors). After three iterations of piloting and revising the strategy, a final search strategy was agreed. The search strategy inclusion criteria were determined using a Participants, Interventions, Comparators, Outcomes[191] and Study type (PICOS) framework[189] (see Box 2). The search defined participants as children who took part in any programme or intervention given at the school or during pre-school. The type of intervention was not defined; it was left open to capture a range of economic or non-economic evaluation methodologies for school-based programmes. Comparators were also left undefined to capture a broader range of studies. The lack of a comparator was not an explicit exclusion criterion as the goal of this review was to identify all evaluation methodologies of school-based programmes.

The following range of standard economic evaluation terms were adapted from the Scottish Intercollegiate Guidelines Network (SIGN) economic filter[192] for the purpose of this review: utility, quality of life, health related quality of life, quality adjusted life year, disability adjusted life year, net benefit, cost, resource use, fund, benefit, effect, contingent valuation, WTP, and human capital. The truncation wildcard (*) was used at the end of root words to represent any number of characters. For example, cost* would

represent cost, cost*s*, cost*ing*, or cost*ings*. The (?) wildcard was used to represent zero or one character, most commonly used to return studies in both British and American English. For example, p?ediatric would return pediatric and paediatric. Finally, the study type for the search included any sort of economic evaluation, cost analysis, or non-economic evaluation associated with HRQoL outcomes including: CUA, CEA, CBA, decision analytic models, and partial economic evaluations – cost and/or outcome descriptions, and cost analyses. The frameworks for economic evaluation differ mainly by outcome measurement, therefore, CUAs were classified as such if they reported a cost per QALY or cost per DALY. CEAs were classified as such if they reported any cost per natural unit of effect other than QALY or DALY. Modelling studies were classified as CUA or CEA if they met classification conditions described above. CBAs were classified as such if they reported a cost-benefit or benefit-cost ratio, reported outcomes in monetary units, or reported ROI results. All other approaches were deemed partial economic evaluations and were classified as cost analyses if they had a component of cost, but did not meet the classification criteria set out for full economic evaluations as described above. Studies were classified as non-economic evaluation if only HRQoL was estimated, or there was no actual cost element reported.

Additionally, other evaluation and analytical methodologies were considered such as: social return on investment (SROI), social impact assessment (SIA), health impact assessment (HIA), MCDA, discrete choice experiments (DCE), and studies using stated preference survey methodologies. Study protocols that indicated a planned economic evaluation were also included and classified into one of the five categories described above (CEA, CUA, CBA, cost analysis, or non-economic). The trial timeline indicated in each protocol was checked, and if the study was due to be complete at the time of thesis write-up, a search was conducted to find any main trial publications.

Box 2: PICOS Criteria

Review objective
To examine current evaluation methodologies (economic, costing, and non-economic studies associated with generic HRQoL) that have been used to assess school-based interventions or programmes and collate the retrieved evidence from these studies.

Participants
Children under the age of 18.

Interventions
Any intervention or programme delivered at a school or pre-school setting.

Outcomes
A range of economic costs and outcomes: utility, quality of life, health related quality of life, quality adjusted life year, disability adjusted life year, net benefit, net present value, cost*, resource use, fund*, benefit*, effect*, contingent valuation, willingness-to-pay, and human capital

Study type
Full and non-economic evaluations: CUA, CEA, CBA, cost analyses, and non-economic evaluations

3.2.1.1 Exclusion Criteria

Studies were excluded if the programme or intervention was not delivered at the school or pre-school setting. Due to limited time and resources, studies were also excluded if they were not in English language or if they were abstract only or the full paper was unavailable through the University of Glasgow Library.

3.2.2 Database and Search Strategies

A range of databases were selected to ensure a comprehensive search of the literature would be conducted. No date restrictions were placed on the eight databases that were systematically searched provided on the following page:

CINAHL - Cumulative Index of Nursing and Allied Health Literature
The Cochrane Library
ERIC - Education Resources Information Centre
MEDLINE (Ovid)
PsychINFO
Web of Science
NHS EED, HEED, and HTA database (CRD York)
EconLit

Using guidance from the CRD,[189] these databases were selected to reflect an extensive literature base to capture relevant studies. Searching commenced in June 2015 and the final strategy was carried out in July 2015. Relevant medical subject headings (MeSH) were searched to identify any evaluation methodology of school-based interventions. The general strategy is outlined in Box 3 which was adapted to each database to account for differences in MeSH terminology between databases.

Box 3: General search strategy adapted to each database

1. (emotion* OR social*) AND (learn* OR wellbeing OR "well being")
2. (improve OR develop) AND (health OR academ* OR mental* OR physical*)
3. crim* OR ("criminal justice") OR famil*
4. 1 OR 2 OR 3
5. school* OR educat* OR academ*
6. (primary OR secondary OR elementary OR junior OR middle OR high) AND (school*)
7. child* OR adolescent OR p?ediat*
8. program* OR intervention OR curriculum OR course
9. 3 AND 4 AND 5 AND 6 AND 7
10. (economic*AND eval*) OR ("cost effective*") OR ("cost benefit") OR ("cost utility") OR ("cost consequence") OR model* OR ("decision tree") OR ("health impact assessment") OR ("return on investment") OR ("social return on investment") OR (social impact assessment") OR ("discrete choice") OR ("stated preference") OR ("multi-criteria decision analysis")
11. utility* OR ("quality of life") OR ("health related quality of life") OR ("quality adjusted life year") OR ("disability adjusted life year") OR ("net benefit*") OR cost* OR ("resource use") OR fund* OR benefit* OR effect* OR "contingent valuation" OR "willingness-to-pay" OR "human capitol"
12. 10 AND 11
13. 9 AND 12

Details of each database search including database specific MeSH can be found in Appendix 1. Evidence is not just found in the published literature that is indexed by electronic databases. Grey literature makes up a huge body of evidence produced by governments, academics, businesses, and industry that have not been published in an academic journal that would be identified in an electronic database search. In order to

identify published, unpublished, and grey literature not indexed by an electronic database, additional relevant articles were identified by:

- Visually scanning reference lists
- Hand searching specified search terms in key educational, psychological, economic journals, and conference proceedings - Education Economics, Economics of Education Review, and International Journal of Education Economics and Development
- Key author search – e.g. Heckman, James was identified as an influential economist in early childhood education programmes
- Key website search – Personal Social Services Research Unit (PSSRU), NICE, Collaborative for Academic, Social, and Emotional Learning, Community Guide, and the Joseph Rowntree Foundation were identified from suggestions from the college subject librarian.

These additional searches help to identify evidence that is very recently published and has not yet been indexed by electronic databases, as well as, compensating for poor or inaccurate database indexing.[193] Results from the database searches were exported to EndNote X7 reference managing software or manually added if identified from sources other than an electronic database. Duplicates were excluded, and remaining records were screened at title and abstract by applying the inclusion criteria set out in Box 2. Remaining records had their full-texts screened applying inclusion and exclusion criteria. Validity checks were performed as described below in section 3.2.3. The studies that remained after the full-text screening were included in the systematic review. After an initial review, eight (10%) of the top studies included in review were selected for a Web of Science citation search (described further in section 3.2.4). A citation search involves reviewing the citations of a relevant study to identify additional studies to include in review while providing an additional check that the search strategy is thorough. At the time of thesis write-up all protocols included in review had studies that had been completed. A further search for the publications of these studies was conducted to determine if they had been published and explore potential reasons for not being published.

3.2.3 Mitigating bias

To mitigate bias throughout the review process, two validity checks were performed by the entire review team. The first validity check took place at the screening of title and abstract stage. Twenty randomly selected records that were being screened at title and

abstract were sent to each member of the review team. Using inclusion criteria specified in Box 2, each reviewer independently assessed each record for inclusion in the next stage of the review (full-text screening) and the results were checked and discussed. When assessing records at the title and abstract stage for inclusion in the full-text screening stage, two filters were created, Filter 1 and Filter 2. Studies were included in Filter 1 if they met some but not all of the inclusion criteria, because all criteria may not be apparent from title and abstract. Filter 2 included studies that appeared to have met all inclusion criteria from title and abstract, and were very likely to be included in review. These records were still reviewed at full-text screening stage.

The second validity check took place at the full-text screening stage. Ten randomly selected records from the pool of papers whose full text were being reviewed were sent to each member of the review team. Each reviewer then independently assessed each record for inclusion in the final review using the inclusion criteria specified in Box 2. The results of the independent assessments were then checked, discussed, and agreed upon amongst the review team. These validity checks helped ensure the selection process was systematic and is found in Appendix 2.

3.2.4 Data Extraction and Study Appraisal

A data extraction form was created using CRD guidance[189] to extract general information as well as economic specific data from each article included for review (Appendix 3). General information extracted included the following: study identification features, study characteristics, participant characteristics, intervention and setting, outcome measures, follow-up, results, and conclusions. Additionally, economic data was extracted from all full and partial economic evaluations which included: type, costs, perspective, time horizon, description of competing alternatives, resource use, effectiveness data, preference based measure used, measure of benefit, discounting, currency, ICER, analysis of uncertainty, and key model parameters.

Several study appraisal tools to evaluate the conducting and reporting of economic evaluations exist. A systematic review of quality assessment tools spanning from 1991 to 2012 identified ten such instruments.[194] Some of the most commonly used appraisal tools identified by the review are the British Medical Journal (BMJ) checklist as described by

Drummond et al.,[195] the Quality of Health Economic Studies (QHES) List,[196] and the Consensus on Health Economic Criteria (CHEC) List.[197] Scoring of checklists allows studies to be compared and ranked quantitatively by quality, but many checklists do not have a scoring system to qualify them as a quality assessment tool. To add up and score items on most checklists (that are not intended as a quality assessment tool) would require the assumption that each item has equal weight. This then does not take into account the relative importance of each criterion.[194] The QHES checklist is the only assessment tool for economic studies to apply a grading system that weights different criteria based on their level of importance. However, there has been no evidence generated to validate this scoring system, or describe its generalisability,[194] thus it is still subjective by nature. The CHEC List does not have any items related to modelling studies, however it is useful in that it is easily adaptable to different study designs, as it was designed to assess clinical trials and observational studies. The Drummond checklist was recommended in Cochrane reviews and was one of the more commonly used checklists adapted and used by the BMJ and other journals. It is simple and brief, as it consists of only 10 items. Like the CHEC list, it is less well suited to evaluating modelling studies as it did not distinguish between economic evaluation alongside clinical trials and decision analytic modelling methods; it also does not have a grading or scoring mechanism. It became outdated as it did not capture any of the new analytical techniques to support economic evaluation such as multiple imputation (MI) for missing data, extrapolation, and methods for pooling data. The systematic review of quality assessment tools concluded that 'the choice of an appropriate checklist should be made with the understanding that quality assessment tools will continue to evolve over time and must improve in reliability and validity for all decision makers over time.'[194]

The most recently developed and consolidated assessment tool is the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist.[170] This checklist aims to consolidate and update the previous checklists mentioned above into a single useful reporting standard. The CHEERS checklist consists of 24 items and has been co-published in 10 health economics and medical journals.[170] The checklist was developed in 2013 for researchers reporting economic evaluations and editors or peer reviewers assessing them for publication. It is more recent and can be seen as an improvement to the Drummond checklist for critical appraisal of published articles.[77, 195] CHEERS is a modern checklist that incorporates items that represent the advances in analytical methods used in current

economic evaluation. Because of its recency and improvement upon the other checklists available, the CHEERS checklist was used to assess the quality of reporting of the articles included for review (Appendix 4). The choice for using the CHEERS checklist was also justified due to its recommendation by ISPOR,[198] the BMJ,[199] The EQUATOR Network,[200] Value in Health,[170] and other notable organisations within and outwith health economics. A copy of the CHEERS checklist is given below in Figure 7.

The studies included in the review were heterogeneous and made up of a variety of evaluation methodologies; thus, not every item of the CHEERS checklist was suited to the studies included in review. Because the checklist is intended for full economic evaluations, partial (cost analyses) and non-economic evaluations were identified as such in the first item of the checklist. This was to avoid misinterpreting a partial economic evaluation as having poor quality reporting simply because the items did not apply. It should be emphasized that the CHEERS checklist is intended to assess the *quality of reporting*, the items are a *minimum* amount of information to be included when reporting an economic evaluation today to improve reporting and health care decisions.[170] Counts from the CHEERS checklist were given by item and by study, but these counts were for descriptive purposes only as well as to identify the eight studies included for the citation search referred to in section 3.2.2. To add up or score the items makes the assumption that each item has equal importance. Therefore, even though the counts are given, they will not be used to make judgements of the quality of how studies were conducted. Some items are more important for assessing quality, and many of the studies included for review pre-date the CHEERS checklist.

Section/item	Item no.	Recommendation	Reported on page no./line no.
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 analyzed 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.	
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.	
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.	

Figure 7: CHEERS Checklist From Husereau et al. Value in Health 16 (2013) 231-250

Section/item	Item no.	Recommendation	Reported on page no./line no.
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.	
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.

Figure 5 (cont.): CHEERS Checklist From Husereau et al. Value in Health 16 (2013) 231-250

3.2.5 Data Synthesis

Data synthesis is a process that involves collation, combination, and summarising of the findings of the individual studies included in a systematic review.[189] It can be done quantitatively by formal pooling of results with statistical techniques such as meta-analysis.[201] This is often done when pooling effectiveness results from multiple studies. Sometimes however, formal pooling of results is inappropriate due to heterogeneity in how studies were conducted and within the effectiveness measures used to report results. In these cases a narrative approach may be taken.[189] As mentioned above, the study design, population, and outcomes of the included studies were heterogeneous, therefore undertaking meta-analysis was deemed inappropriate. Additionally, the purpose of the review was not to look at any one type of school-based intervention and

pool their effectiveness results, it was to explore the different methodologies used for school-based evaluation.

Even when meta-analysis is possible, certain aspects of narrative synthesis are required to interpret the evidence.[189] Guidance on the conduct of narrative synthesis[190] was used to define narrative synthesis as 'an approach to the systematic review and synthesis of findings from multiple studies that relies primarily on the use of words and text to summarise and explain the findings of the synthesis.' Narrative synthesis is a form of storytelling and telling a trustworthy story is the aim of a narrative sythensis.[190] It is not to be confused with narrative review which is a phrase sometimes used to describe a more traditional literature review that is not systematic or transparent in their approach to synthesis.[202]

The current review question dictated inclusion of an extensive range of research designs producing varied findings for which quantitative approaches to synthesis are inappropriate.[190] To further justify this method, CRD guidance suggests a narrative approach to data synthesis when formal pooling of results is inappropriate.[189] Therefore, a narrative synthesis was conducted which was descriptive in nature and objective in summarising findings. Historically, there has been little consensus on how to carry out narrative synthesis and the elements to establish credibility.[189, 190] A general framework comprising the following elements can be applied to help maintain transparency and add credibility to the process:[190]

- Developing a preliminary synthesis of findings of included studies
- Exploring relationships within and between studies
- Assessing the robustness of the synthesis.

The preliminary synthesis comprised of tabulated details of study characteristics including study type (methodology), years of publication, country of origin, age of participants, and intervention type. Studies were then grouped by methodology: economic evaluation which included CEA, CUA, CBA; cost analyses which only included costs or did not directly value benefits instead assuming cost savings; and non-economic evaluation studies which did not include costs but may have included a generic HRQoL measure. CEAs, and CUAs included trial and model-based studies and were classified as such if the results were expressed as an ICER. Relationships between studies within and across groups were

explored, identifying patterns and exploring heterogeneity between studies. A concept map was developed to give a visual representation of the patterns and relationships identified between included studies. Finally, robustness was assessed through critical and systematic review of the quality of reporting of the included studies so that an overall assessment of the strength of the evidence could be made. A descriptive synthesis of the CHEERS checklist items was conducted and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist was completed.

3.3 Results

The results of the systematic review are organised into three subsections as described in the methods section, the preliminary synthesis; exploration of relationships within and between groups; and the assessment of the robustness of the synthesis. The counts of search records identified from each database are summarised in Table 2. Further details of the database searches are in Appendix 1. The rest of the study selection process is detailed in Figure 9 which is a flow chart as recommended in the PRISMA statement.[203] The validity checks to mitigate bias in the data selection process are found in Appendix 2. Completed data extraction form and CHEERS checklist are found in Appendix 3 and 4 respectively and completed PRISMA checklist in

Appendix 5. Study characteristics are summarised in Table 3.

Table 2: Number of records identified by each database searched

Database	No. of Records Exported to Endnote
CINAHL	16
The Cochrane Library	419
ERIC	264
MEDLINE using OVID interface	99
PsychINFO	28
Web of Science	200
HTA Database, DARE, and NHS EED	166
EconLit	149
Total	1,341

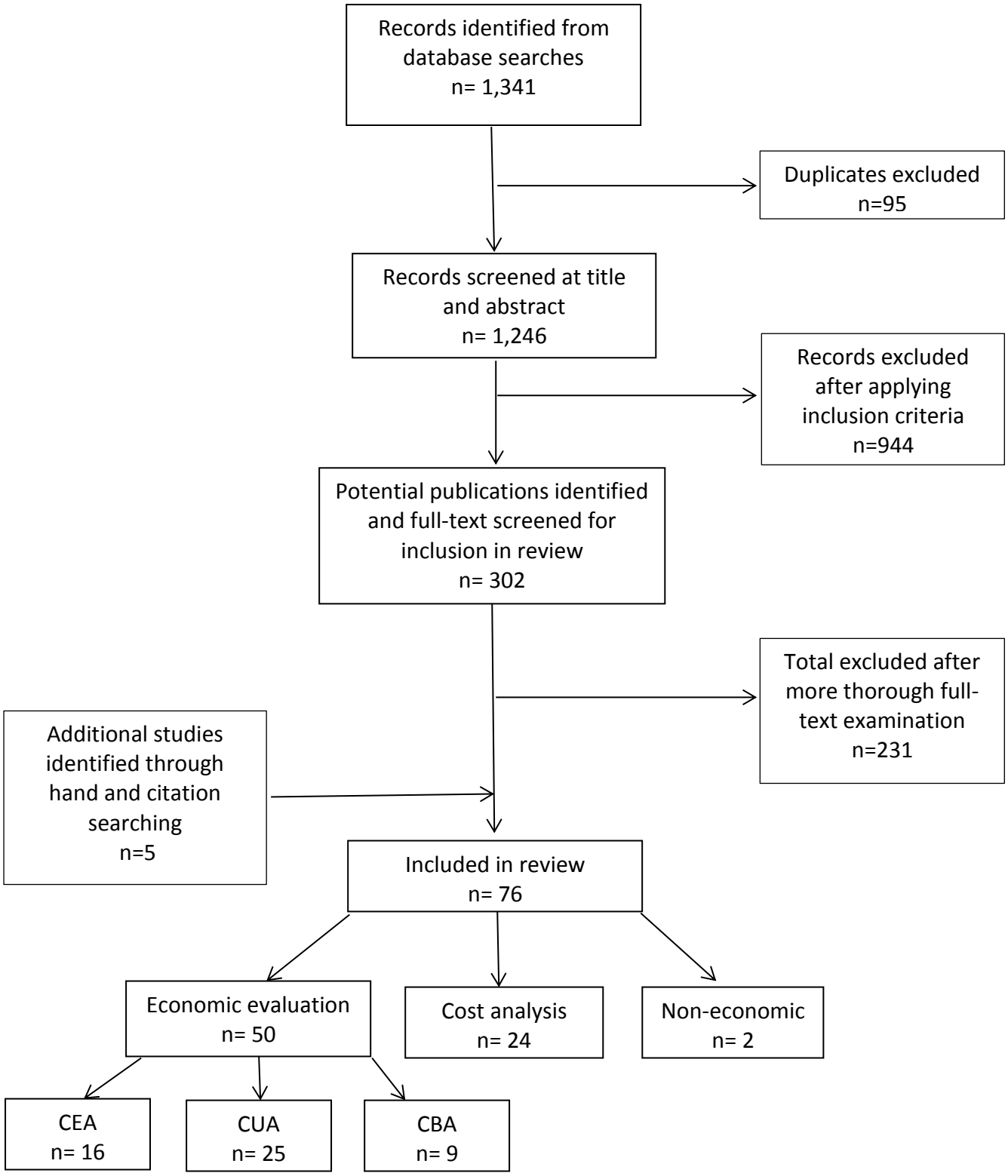


Figure 8: Flow chart of study selection process

3.3.1 Preliminary synthesis

The methodology groupings from the preliminary synthesis are displayed in Figure 9. Economic evaluations comprised CEA, CUA, and CBAs. CEA and CUAs were classified as such if they reported an ICER and according to classification criteria set out in section 3.2.1. Modelling studies were also included in those classifications if they reported an ICER. The next group was cost analyses which included all studies that mainly focused on the cost side of an evaluation and did not formally combine costs and effects to meet the definition of a full economic evaluation as defined in Chapter 2. Studies included in this category may not have directly measured outcomes, or made assumptions about expected outcomes being cost savings. They may have reported all the necessary elements to conduct an incremental analysis but failed to complete and report ICERs. The final category was non-economic evaluations, which included studies that measured HRQoL, but did not measure any costs.

The preliminary synthesis study characteristics are displayed in Table 3. Economic evaluations outnumbered cost analyses and non-economic evaluations. When breaking down economic evaluations by type, CUAs were the most common (n=25) followed by cost analyses (n=24). The next most common study type was a CEA (n=16), followed by CBA (n=9), followed by non-economic evaluations (n=2). The database searches were conducted in 2015 and the majority of the included studies were published within the last five years. There were 44 studies published between 2010 and 2015. Next, there were 16 studies published from 2005 to 2009. There were 14 studies published from 2000 to 2004. Finally, there were two studies published before 2000 included for review. Most studies included for review originated in the USA; 34 out of the 76 included. The next most common country of origin was the UK with 12 studies included for review. The following lists the rest of the countries with their respective counts indicated in parentheses: Australia (5), Germany (3), New Zealand (3), Canada (3), the Netherlands (3), Italy (2), Egypt (2), Japan (2), China (2), Sweden (1), Tanzania (1), Kenya (1), Zimbabwe (1), and the Philippines (1). Classifying the study's country of origin by continent results in most studies originating in North America (48%), followed by Europe (28%), Australasia (10%), and Africa (7%) and Asia (7%) (Table 3). The following lists the age categorisations with their respective counts: preschool (n=6), primary school (n=22), secondary school (n=21), combination of age groups (n=18), age not specified (n=9). The type of

programme or intervention that was most common was focused on increasing physical activity and fruit and vegetable intake or preventing obesity (n=18). Next were school-based screening or vaccination programmes to prevent infectious diseases (n=11). The rest of the studies fell into the following categories: mental health/ SEW (n=8), asthma (n=7), illicit substance abuse/misuse (n=7), sexual health (n=6), early intervention/parenting programmes (n=4), food insecurity (n=4), dental health (n=3), and generic education programmes (n=8).

Table 3: Study characteristics

Study Type	No. (%)
CEA	16 (21%)
CUA	25 (33%)
CBA	9 (12%)
Cost analysis	24 (31%)
Non-economic	2 (3%)
Country of Origin (classified by continent)	
Continent	
North America	37 (48%)
Europe	21 (28%)
Australasia	8 (10%)
Africa	5 (7%)
Asia	5 (7%)
Year of Publication	
Pre-2000	2 (3%)
2000 to 2004	14 (18%)
2005 to 2009	16 (21%)
2010 to 2015	44 (58%)
Age	
Pre-school (age > 5)	6 (8%)
Primary-school (ages 5 to 11)	22 (29%)
Secondary school (ages 12 to 18)	21 (27%)
Age not specified	9 (12%)
Combination of age groups covered	18 (24%)
Intervention Type	
Physical activity/nutrition education/obesity prevention	18 (24%)
Infectious disease screening/prevention/vaccination	11 (14%)
Mental health and wellbeing	8 (11%)
Asthma	7 (9%)
Illicit substance abuse/misuse	7 (9%)
Sexual health	6 (8%)
Early intervention/parenting	4 (5%)
Food insecurity/nutrition	4 (5%)
Dental health	3 (4%)
Generic/education programmes	8 (11%)

3.3.2 Exploring relationships within groups and between studies

The following section 3.3.2 details the findings and relationships identified through the synthesis of the studies included in the review. As can be seen in Figure 9, the studies were grouped into three broad methodologies, with a further three categorisations specifying the type of economic evaluation. Each methods group is discussed in terms of the findings and relationships found within the group. Following that, findings across the groups are discussed. Finally, a concept map follows which gives a visual representation of the findings detailed in this section.

3.3.2.1 Economic Evaluations

Cost-effectiveness analyses

There were 16 studies included for review that were classified as CEAs. The majority originated from North America (n=9),[204-212] four were from Europe,[213-216] two from Africa,[217, 218] and one from Asia.[219] The majority of CEAs were published fairly recently; 12 were published in the last 5 years.[204, 205, 207, 209, 211-217, 219] The rest were published within the last 15 years.[206, 208, 210, 218] The majority of CEAs were aimed at primary school children (n=7),[204, 210-213, 215, 217] four were a combination of school age groups,[206, 207, 209, 214] two were not specified,[218, 219] and three were aimed at secondary school children.[205, 208, 216] The types of interventions varied from infectious disease screening/vaccination (n=4),[208, 217-219] to physical activity/obesity prevention (n=4),[210-212, 215] mental health/SEW (n=3),[206, 213, 214] substance abuse/misuse (n=2),[205, 216] generic education programmes (n=1),[207] dental health (n=1),[204] and asthma (n=1).[209] Only two studies report the results of their intervention as being 'not cost-effective.'[206, 219] The authors made these judgements based on context-specific WTP thresholds. The rest report relatively low costs per unit of effect, but as the outcomes vary and there is no established threshold for CEAs, no further comment is made on the 'cost-effectiveness' of the remaining studies.

Cost-utility analyses

There were 25 studies included for review that were classified as CUAs. Most studies originated from Europe (n= 8)[220-227] and North America (n=8),[228-235] followed by Australasia (n=7).[236-242] There was one study originating from both Africa[243] and Asia.[244] The majority of studies were published within the last 5 years (n=18), and the rest within

15 years.[221, 225, 228, 230, 233, 234, 237] Most CUAs were aimed at secondary school children (n=10).[220, 222, 223, 225, 233, 234, 236, 242-244] Five were aimed at primary school children,[224, 227, 229, 237, 240] six covered a combination of age groups,[228, 230, 232, 235, 238, 239] and two were aimed at preschool children.[221, 226] There were two studies where the age was not specified.[231, 241] The type of intervention that was most common within this group was physical activity/nutrition/obesity prevention programmes.[224, 226, 231, 237-240] There were five programmes aimed at sexual health,[220, 223, 236, 242, 243] three programmes dealing with infectious disease screening/prevention,[232, 244, 245] illicit substance misuse,[225, 233, 234] and generic or education programmes.[228, 230, 241] There were two aimed at mental health[222, 235] and wellbeing, and one asthma[229] and dental health intervention.[227] Only one study directly measured HRQoL using the Shona-language version of the EQ-5D.[243] The rest of the studies use different methods for estimating and modelling QALYs. Cost estimates vary as there was a large variation in types of currencies used (as evaluations were included from around the globe). Many authors were hesitant to comment on their study's cost-effectiveness, using terms such as 'may be cost-effective,'[222, 224] 'appears,'[236] 'seems,'[221] 'could be,'[220] or 'relatively'[240] cost-effective. Or they may have stated the results were cost-effective, but with a caveat of uncertainty in effectiveness estimates.[225, 231] There were studies that stated outright they were cost-effective[230, 233, 234, 238, 244] or not cost-effective.[228, 229, 237, 239, 242, 243] Comparisons of cost-effectiveness between studies included in this group is cautioned as costs and effect estimation methodologies varied widely.

Cost-benefit analyses

There were nine studies included for review that were classified as CBAs. Six originated from North America.[27, 28, 246-249] There was one from Europe,[250] Asia[251] and Australasia.[252] There was a wide range of publication dates, ranging as early as 1985[246] to 2014.[252] Three studies looked at the same intervention, but had different analyses based on continued follow-up.[27, 28, 246] Four studies were aimed at pre-school children,[27, 28, 246, 247] one at primary school[252] and two at secondary school,[248, 250] and the rest were aimed at a combination of age groups.[249, 251] Four of the studies were evaluating early intervention/parenting interventions.[27, 28, 246, 247] There was one of the following studies: nutrition education,[252] asthma,[249] illicit substance misuse,[250] sexual health[248] and food insecurity/nutrition.[251] All CBAs reported that their interventions were cost saving or had

a positive net present value. All but two reported their results in a benefits-cost ratio as given in Equation 7.[249, 253] The amount saved per every $1 invested ranged from $2.65[248] to $12.90.[28] The results were subject to uncertainty around the valuation of benefits in each study. None of the studies stated if they used a human capital approach or a stated preference approach for valuing benefits, thus although they have been classified as a CBA, they are more likely to fit into the category of cost-savings analysis as mentioned in section 2.1.5.

Equation 7: Benefits-cost ratio

Benefits-cost ratio = $(\frac{Benefits}{Costs})$

3.3.2.2 Cost analyses

There were 24 studies included for review that were classified as cost analyses. Twelve studies, or half, originated in North America.[254-265] Eight originated in Europe[266-273] and two from Africa[274, 275] and Asia.[276, 277] All studies were published post 2000, with exception to one which was published in 1966.[254] Eight studies were aimed at primary school students,[255, 261, 262, 264, 265, 269, 275, 278] six at secondary school,[257, 266-268, 272, 279] five were a combination of age groups[256, 259, 260, 273, 276] and five weren't specified.[254, 258, 270, 271, 274] There were six studies related to physical activity/obesity prevention,[255, 258, 261, 264, 276, 278] four screening/vaccine programmes,[259, 262, 274, 275] three of the following: mental health and SEW,[267, 268, 280] asthma,[257, 272, 279] food insecurity/nutrition,[270, 271, 273] and generic health/education programmes.[254, 265, 269] There was one dental health intervention,[256] and one illicit substance misuse prevention programme.[266] Three studies reported cost-effectiveness ratios (CER) which fail to take the incremental costs and effects between groups into account.[255, 264, 278] One study[267] was nearly classified as a CUA, but did not report an ICER so was not classified as such. This study did however directly measure HRQoL from adolescents using the EQ-5D. A few studies report negative results[260, 267, 268, 272, 273] and the rest report 'appropriate' use of funds, cost savings, or cost-effectiveness.

3.3.2.3 Non-economic evaluations

There were two studies included for review that were classified as non-economic evaluations. They both originated from North America[281, 282] The years of publication ranged from 2006[281] to 2008.[282] One study was aimed at primary school children,[281] and

the other was aimed at combination of age groups (Kindergarten to grade 8).[282] One was an asthma[281] programme and the other a generic health programme looking at the effect of school-based health centres on students' HRQoL.[282] The HRQoL measures used were the Pediatric Quality of Life Inventory (PedsQL) 4.0[282] (detailed further in section 4.2.1) and the Pediatric Asthma Caregiver's Quality of Life Questionnaire.[281] The latter study found few improvements in health outcomes including the Caregiver's QoL, however the former found significant improvement in student-reported HRQoL over the comparison group.

3.3.2.4 Relationships across groups

All studies were preventative by nature in that the programme or intervention being evaluated aimed to prevent future problems or prevent current problems from escalating. They could all be considered PHIs. The most common types of programmes were obesity prevention and screening programmes which is perhaps a reflection of how childhood obesity is a pressing issue facing children all over the globe as well as the school being an ideal setting for mass screening programmes. The groups that included the oldest studies were the CBA and cost analysis groups. Therefore, it could be said that these types of methodologies have been used the longest in school-based evaluation, or they simply were the most common type of methodology employed that would have been published during that time. CUAs and CEAs are more recent school-based economic evaluation methodologies to appear in the literature. They all start to appear after the year 2000, so within 15 years of the search being conducted. The 'youngest' type of publication methodology is the protocol, as they were all published within the last five years. The need for transparency around the conduct of RCTs has contributed to this recent phenomenon.[283]

Not all defined PICOS criteria (see Box 2) were identified in the literature. 'Contingent valuation' and 'human capital' were terms specified under the outcomes criterion that were not identified in any of the studies screened. Likewise, many study designs that were defined were also not identified in the literature such as, SROI, SIA, HIA, MCDM, DCE or any stated preference methods.

There was great uncertainty across groups particularly in measuring and valuing effectiveness or benefits. This is particularly true in the CUA and CBA groups. Only one study in the CUA group directly measured HRQoL, therefore as the rest were modelled from other sources, huge uncertainty remains around the effectiveness results. As there was no valuation of benefits (either using a human capital approach or contingent valuation) stated in any of the CBAs, benefits were estimated as likely cost savings due to prevention of future problems. This places a lot of uncertainty on the effectiveness of studies as results are dependent on many assumptions, mainly that the intervention will prevent these future problems from occurring and subsequent costs. These methods deviate from the CBA methods described in section 2.1.5 as there is little attempt to value benefits observed from the studies, and are more likely to be classified as cost-savings analysis despite being labelled otherwise. In terms of costs, the main cost drivers are the cost of providing the intervention in the short-term. If the study modelled a longer-term time horizon, these costs may be spread over a longer period, but then the uncertainty around the extrapolation increases.

The concept map detailed in Figure 10 below gives a visual representation of the relationships identified between and within groups. It also identifies some of the main findings of this synthesis. Figure 10 demonstrates the hierarchical relationships of the studies included. It starts by indicating that all 76 studies included were preventative school-based programme evaluations. Downward arrows describe the methodological groupings. The green boxes indicate economic evaluations and findings are outlined in blue. The relationships between groups are indicated by arrows pointing to the findings highlighted in the diagram. The number of studies which directly measured HRQoL is indicated as well as those that incorporated children's preferences. The types of evaluation methodologies that had uncertainty in effectiveness estimates and evaluated obesity prevention programmes are indicated by arrows to each finding.

Study Protocols

Of the eight study protocols included, half originated in the UK.[213, 226, 227, 266] All studies were published within the last five years. All state that the study design will be a RCT, and two are pilot studies.[226, 266] Not all studies explicitly stated the type of economic evaluation that was planned; four stated CEA[212, 213, 217, 241] and one stated CUA,[226] the rest did not give enough detail. In general, much detail was missing, only one study reported

that they would use discounting[241] or justify why it was not needed. Only two[226, 266] of the eight studies are taking place in a pre-school or secondary school, the rest are based in a primary school setting. Three studies state they will use an outcome measure that directly measures HRQoL in children.[226, 227, 241] Two will use the PedsQL[226, 241] (detailed further in section 4.2.1) and two will use the CHU9D.[227, 241]

The search for the published protocol studies that were included in review returned three publications[284-286] and one report.[287] Two studies were still ongoing with NIHR publication dates in April 2017[227] and September 2018.[213] One study had recently finished in December 2016, but no publication had been found at the time of write-up.[241] Only one study could not be accounted for.[276] The protocol was published in 2010, so plenty of time had passed for the results of the study to have been published. This protocol did not provide many details of the methods that would be undertaken, so a potential reason for it not being published may be due to poor reporting or poor study quality. Furthermore, there is a publication bias in which editors are more favourable to publishing studies with positive results, so the study may have been victim to this bias if the results were not positive.

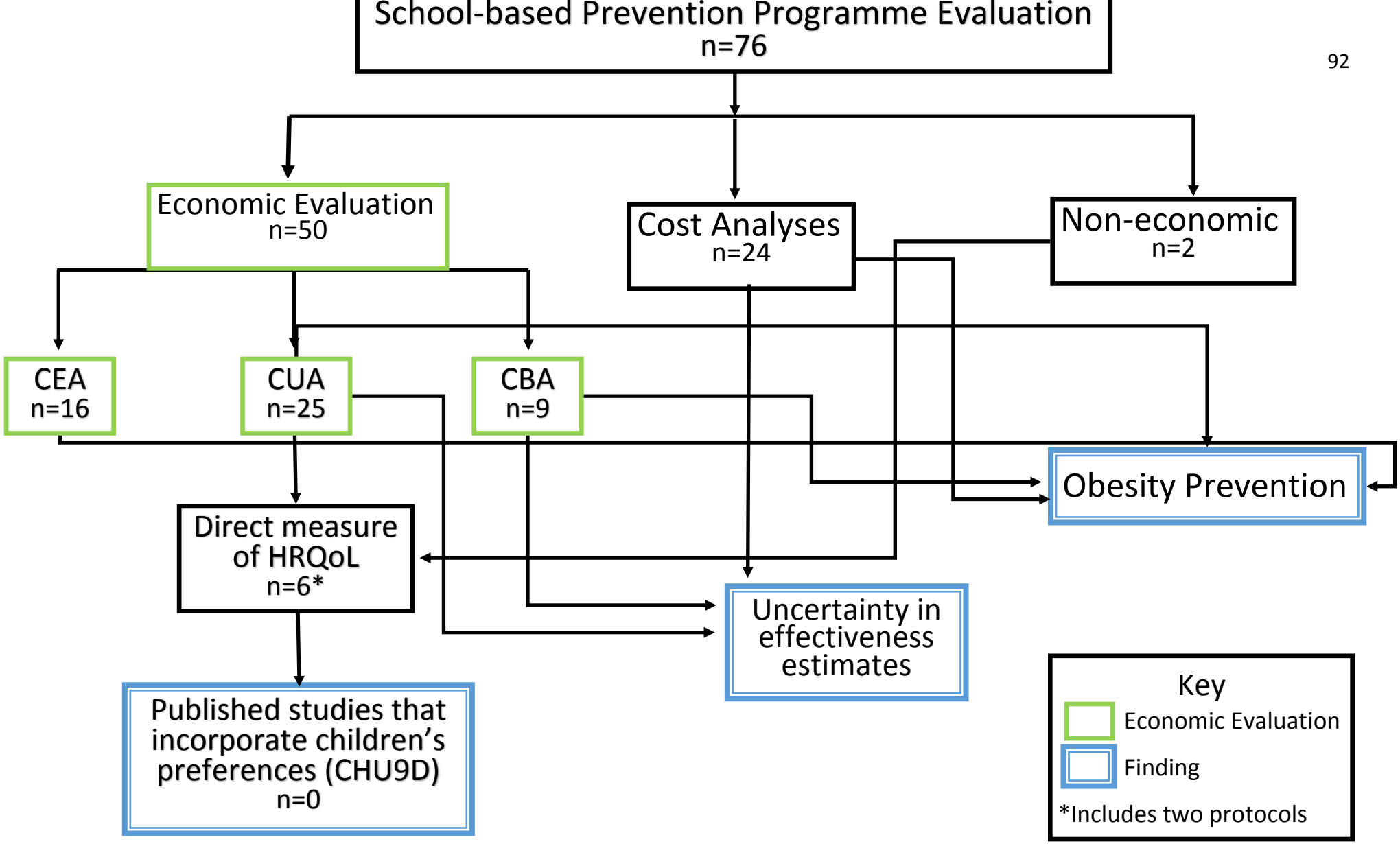


Figure 10: Concept map representing relationships and findings from the narrative synthesis

3.3.3 Assessing the robustness of the synthesis.

The counts of the CHEERS checklist are presented in Table 4 and Table 5. Table 4 presents the counts of checklist items by study, in ascending order. Table 5 presents the counts across all studies by item number. These counts are reported for descriptive purposes only. Table 4 shows that none of the studies reported all 24 checklist items. The most reported was n=21, shared by three studies.[236-238] The fewest number of items was n=2.[281] Table 5 shows that all studies include CHEERS checklist item number 3, which says, *'Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.'* The item that had the fewest studies report it was item number 20a which says, *'Single study-based economic evaluation: Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).'*

Many studies predate the publication of the CHEERS checklist, so the counts of items should not be used as a judgement of quality for individual studies. Fifty-four of the 76 studies were published prior to 2013 when CHEERS was published. Of the remaining 23, nine studies reported 15 or more of the CHEERS items. These counts are reported for descriptive purposes only; any inference of quality should be made with caution as there is publication lag time to consider as well. However, when describing the quality of reporting for the studies as a whole, the economic evaluations would not hold up to today's reporting standards. Many reviewers and editors would require authors to report all of the CHEERS checklist items as a minimum. The studies that reported 18 or more of the CHEERS checklist items all came from high-income, primarily English speaking countries: Australia, New Zealand, Germany, USA, UK, and the Netherlands. This indicates that geography may be impacting how a study was conducted, or how well it was reported. Again it should be noted that many of the studies reviewed were not considered full economic evaluations, so they would not normally need to be assessed by CHEERS. All studies in this current review were assessed by CHEERS for consistency.

Because the overall level of reporting, methodology used, and study design variation, assessing the robustness of the review findings is challenging. In terms of the systematic review process followed, the synthesis is robust. However, the different methodologies

included in review, and the accompanying challenges of analysing those heterogeneous studies, resulted in an inherently less robust review. This trade-off was made to keep the scope of the review broad in order to identify the most wide-ranging types of methodologies used to evaluate school-based programmes. To address this trade-off, efforts have been made to follow a review process that is systematic, as detailed in section 3.2.

Table 4: CHEERS checklist totals by study

Study No.	Author, Year	Total
17	Bruzzese, J; 2006	2
1	Abt, C; 1966	3
22	Curtale, F.; 2005	3
13	Boyle, J; 2007	4
53	Newbury-Birch, D.; 2013	5
64	Shemilt, I; 2004	5
9	Beets, M; 2014	6
72	Wade, T; 2008	6
78	Young, T; 2003	6
3	Ansell, J.; 2002	7
14	Brassard, P; 2006	7
27	Foster, J; 2013	7
44	Li, Y; 2010	7
4	Atherly, A; 2009	8
23	Eckermann, S; 2014	8
62	Salisbury, C.; 2002	8
6	Barber, S; 2013	9
19	Chestnutt, I; 2012	9
29	Gelli, A; 2009	9
30	Gelli, A; 2011	9
32	Gesell, S; 2013	9
33	Glewwe, P; 2001	9
38	Joseph, C; 2007	9
54	Nishura, H; 2014	9
5	Babey, S; 2014	10
10	Belfield, CR; 2005	10
15	Brooker, S; 2010	10
21	Crowley, D; 2014	10
24	Ford, T; 2012	10
68	Tai, T; 2010	10
34	Guay, M.; 2003	11
36	Hoeflmayr, D; 2006	11
58	Quach, J.; 2013	11
63	Scherrer, C; 2006	11
7	Barnett, S; 1985	12
18	Carabin, H.; 2000	12
45	Liping, M; 2013	12
2	Anderson, R; 2014	13
26	Foster, E; 2010	13
39	Kesztyues, D; 2013	13
67	Stallard, P; 2013	13
71	Vijge, S; 2008	13
74	Wang, L; 2001	13
75	Wang, L; 2003	13
16	Brown, H; 2007	14
20	Cooper, K.; 2012	14
28	Frick, K; 2004	14
35	Heckman, J; 2010	14
11	Bertrand, E; 2011	15
50	Moodie, M; 2010	15
52	Muenning, P; 2014	15
56	Pearson, A; 2014	15
57	Philipsson, A.; 2013	15
69	te Velde, S; 2011	15
73	Wang, L; 2000	15
76	Wang, L; 2008	15
8	Barrett, J; 2015	16
41	Kowada, A; 2012	16
51	Muenning, P; 2007	16
77	Wang, L; 2011	16
25	Foster, E; 2006	17
47	Miller, T.; 2013	17
55	Noyes, K; 2012	17
60	Reynolds, A; 2011	17
61	Rush, E; 2014	17
31	Gerald, J; 2010	18
37	Hollingworth, W.; 2012	18
59	Rein, D; 2012	18
65	Shepherd, J; 2010	18
66	Simon, E; 2013	18
70	Tengs, T; 2001	18
43	Levaux, H; 2001	19
40	Konig, H; 2004	20
12	Blakely, T; 2014	21
48	Moodie, M; 2009	21
49	Moodie, M; 2013	21

Table 5: CHEERS checklist totals by item

Item No.		Description	Total
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	61
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.	55
Intro 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.	76
Methods Target pop and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	22
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	72
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	50
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	56
Time Horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate	41
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	42
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.	59
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.	12
	11b	*Synthesis-based estimates:* Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	20
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	4
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.	10
	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.	19

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.	44
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	14
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	23
Analytic methods	17	Describe all analytical 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 (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	10
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.	24
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.	24
Characterising uncertainty	20a	*Single study-based economic evaluation:* Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	3
	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.	21
Characterising 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.	11
Discussion Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	63
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 non-monetary sources of support.	58
Conflict of interest	24	Describe any potential for conflict of interest of 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.	36

3.4 Discussion

3.4.1 Summary of findings

This systematic literature review identified a wide range of evaluation methodologies for school-based interventions and programmes. The majority of studies identified were economic evaluations and cost analyses. However, the two studies included in the non-economic evaluations provided useful information on HRQoL; including what and how measures were being used. Around two-thirds were classified as full economic evaluations. The quality of reporting of these economic evaluations varied, and most would not be up to today's standards of reporting as defined by the CHEERS checklist.[170] Geography did seem to impact on quality of reporting as the studies that reported the most checklist items all originated from high-income countries.

The review did not identify any studies originating from South America. Instead, the majority originated from North America and Europe. The review only identified two studies that were published pre-2000; a CBA[246] and a cost analysis.[254] This demonstrates the relative novelty of the school-based economic evaluation. The wide range of methods employed and the varying quality of reporting also demonstrate the novelty of a school-based economic evaluation, as there is no standardised global guidance for this type of evaluation. Protocols make up some of the more recent additions to the literature as every protocol reviewed, was published within the last five years from when the search was undertaken. Two protocols stated they would use the CHU9D as a direct measure of HRQoL for children, however, to date the full studies have not yet been published to verify the use of this measure (and if adolescent values were used to value utilities).

Only four published studies were identified that directly measured HRQoL. Two used the EQ-5D,[243, 267] an adult measure which was used on adolescents, and the other two were non-economic evaluations using the PedsQL[282] and a disease specific caregiver's quality of life instrument.[281] The use of an adult measure in an adolescent population is concerning because it ignores the developmental changes in adolescents which means the values they place on certain health attributes may differ from adult values.[288] Similarly, a systematic review of paediatric CUAs, found most evaluations used an adult preference-based measure despite NICE guidance stating they should be developed specifically for

children.[289] The review also found the measurement of QALYs to have the greatest variability between studies.[289] Additionally, there may not be a common set of health attributes that are applicable across all age groups.[290] It is an interesting finding that the only studies to incorporated child-specific HRQoL were non-economic evaluation studies. In CUA, it has been deemed preferable to directly measure HRQoL from participants versus relying on an indirect method for obtaining utility (e.g. mapping or 'crosswalking' algorithms,[291] more detail in Chapter 6). There is a paucity of evidence in the published literature of CUA of school-based interventions that directly measure HRQoL using appropriate, child-specific measures. The review identified zero published studies that directly measured HRQoL in children (using a measure designed specifically for children) and which used children's preferences; the only measure to fulfil both of these conditions, at the moment, is the CHU9D. None of the published studies reviewed incorporate this measure. Therefore, use of the CHU9D in this context is an important and novel contribution to the literature as it would be the first school-based economic evaluation to incorporate the CHU9D with adolescent values (currently there are no values from younger children available).

There were eight studies included in review that intended to promote SEW in schoolchildren. This finding provides justification for leaving the intervention type open, as to restrict it to SEW interventions would have limited the results included in review. There was one study protocol identified that had a similar aims to RoE to improve the SEW of primary school students. It is the Incredible Years programme which is a classroom management programme which also uses the SDQ as the primary outcome measure (which is the case with RoE).[213] This study is still ongoing, so it will be interesting to see the results of this CEA once they are available. An important difference between this study and RoE is that there was no utility outcomes included which means there can not be a direct calculation of QALYs without the use of a mapping or 'crosswalk' function. The methodologies used in RoE will be a novel contribution to the existing evidence base.

Most programmes being evaluated were aimed at primary and/or secondary school students. Few studies were aimed at pre-school children, but those that were, were all classified as CBAs. The most common types of interventions or programmes being evaluated at the school level, were obesity prevention and screening/vaccination programmes. The rising levels of global childhood obesity is reflected in this finding.[292]

The next most common programme type was screening or vaccination programmes. Schools are an important setting for health promotion because of their ability to reach a large number of children at once;[293] thus, the school has long been thought as the ideal place to provide mass screening/vaccination, and determining the cost-effectiveness of such strategies is warranted. This is a reason intervention at the school level is considered a PHI, and if effective, may have the potential to bring about school-wide benefits because of its reach. The next most common intervention type addressed mental health and wellbeing, which is where RoE would be categorised. This finding relates to the rising awareness of the need to promote and look after children's SEW.[5]

The review did not identify certain PICOS items such as 'contingent valuation' and 'human capital.' Likewise, many study designs that were defined were not identified such as, SROI, SIA, HIA, MCDM, DCE or any stated preference methods. Contingent valuation, human capital, DCE and stated preference are all search terms that relate to CBA. The fact that they are missing from this review is potentially down to the practical and time-consuming issues of CBA that were mentioned in section 2.2.1.1 on appropriate methodologies for economic evaluation of PHIs. MCDM is an emerging methodology for PHIs, so its absence from the literature is understood. The other evaluation designs such as SROI, SIA, HIA must be unusual or inappropriate in some way for this context. The lack of these analytical methodologies and study designs are important findings from this review.

The main cost drivers in the evaluations reviewed were the costs of delivering the programme. Some studies included hospital and medical costs if they were relevant, and some collected very detailed health and resource use cost data. The main causes for uncertainty in results were around the effectiveness of the interventions. Particularly in the CUAs and CBAs the effectiveness estimates were based on many assumptions. QALYs were modelled from other sources in all but one of the CUAs. The CBAs relied on assumptions for the accurate valuing of health benefits in monetary terms as not a single CBA performed a contingent valuation or used a human capital approach. In many of the cost analyses, benefits were valued as cost savings that were based on assumptions of the intervention's effectiveness. As thresholds in other countries may not be as clearly defined as in the UK, there was a reluctance for some authors to comment directly on the

intervention's cost-effectiveness. This reluctance could also have resulted from uncertainty in the effectiveness results, which was a common theme across studies.

3.4.2 Limitations

The author undertook the review of all studies, data extraction, and synthesis. Having only one reviewer, the supervisory team attempted to mitigate any bias (section 3.2.3) by contributing to the search strategy and performing validity checks on samples during the evidence gathering process.

The review question was quite broad which meant a wide range of evaluation methodologies were included. This adds to the difficulty in evaluating quality with a single appropriate yet comprehensive tool. The CHEERS checklist was an appropriate assessment tool; however, because of the broad range of evaluations included, the non-economic evaluations and cost analyses had items that did not apply. The broad range of methodologies also posed difficulties in evidence synthesis, as the included studies were heterogeneous. However, heterogeneity was dealt with through narrative synthesis and followed a systematic process that included preliminary synthesis, exploring relationships within and between groups, and assessing the robustness of the synthesis. The initial scoping review was more focused; the inclusion criteria was much narrower only focusing on economic evaluation of school-based interventions. This initial review identified few studies, so a broader approach was taken to make sure a comprehensive review of all available evidence was conducted.

The review found no use of alternative methods MCDA, SROI, SIA, or HIA, even though they have been suggested as appropriate alternatives for capturing broader outcomes.[157, 294-296] As mentioned previously in section 2.2.1.2, MCDA approaches have not been widely adopted in healthcare decision-making and this systematic review has demonstrated that this is also the case in education decision-making. As further research into how MCDA should be used in HTA for healthcare settings is still required, the same is true for the use of MCDA in education decision-making contexts. SROI has been suggested for PHIs as they can allow the measurement of broader outcomes; however, no record of the use of these methods in school-settings has been identified. During the scoping review to identify appropriate methodologies to include in the search strategy,

SIA and HIA were identified, however it is noted that these methods have not appeared in any recent literature. Thus, it is less surprising that these methods were not identified in this review. The lack of results indicating methods such as MCDA or SROI in school-based evaluations are important findings in themselves, as it speaks to need for further development or guidance for applying these methods. This is particularly relevant to MCDA as ISPOR has issued several recent guidance documents.[160, 161] There may be a research time-lag in the use of MCDA, or the existing guidance may not be sufficient for researchers new to the method to confidently conduct MCDA.

3.5 Conclusion

Evaluation methodologies of school-based programmes are varied and widespread. This systematic review, revealed the types and state of economic evaluation of school-based interventions as well as the non-economic evaluation methodologies, which comprised mainly of costing studies. Economic evaluation is a relatively novel concept in the school setting despite the need for efficient resource allocation in budget constrained education boards.[19] Thus, the quality of methods used in the identified economic evaluations was not quite up to the standards that might be expected in the clinical trials-based medical literature. Few studies directly measured HRQoL in children leading to uncertainty in the intervention's effectiveness estimates. In most CUAs, QALYs were not estimated from utilities directly collected from the children, but were modelled based on estimates from other sources, usually taken from adult studies. No published studies were identified that directly measured HRQoL in children which was also valued by children using the CHU9D. This is an important avenue for future research that this thesis intends to address; whether it is worth considering children's values in decisions that will ultimately affect them.

Improvements can be made in the quality of reporting of economic evaluations of school-based programmes as low quality of reporting was prevalent. As a minimum, economic evaluation should report each of the applicable CHEERS checklist items and this review did not identify any studies that reported on each item. As the methods for school-based economic evaluation develop, the quality of reporting should improve as well.

The purpose of this systematic review was to gain and understanding of how economic evaluation (and other evaluation methodologies) of school-based programmes are currently being conducted and the types of preference-based child utility measures that are currently being utilised through a comprehensive review of existing evidence around evaluation methodologies. The review revealed relatively few high quality existing studies and zero published studies that incorporated children's preferences in CUA. The review also revealed that alternative methods, which have been suggested for evaluation of PHIs such as MCDA and SROI, are not being implemented in a school-based evaluation context. The next two chapters aim to address this paucity of existing evidence in the literature by describing the methods and results of a CHEERS compliant CUA which directly measures children's HRQoL and incorporates adolescent values into the calculation of QALYs.

4 RoE Economic Evaluation Methods: a Case Study

4.1 Introduction

Economic evaluations of school-based PHIs are relatively uncommon, especially those that aim to improve children's social and emotional wellbeing (such as RoE), as was demonstrated in the previous chapter. Yet there is growing consensus of the value of investing in children's health.[297, 298] By improving the overall health and wellbeing of children, they may perform better in school; reduce the use of costly healthcare services; and ultimately be better prepared and successful in adulthood in terms of labour and employment outcomes.[297, 298] Additionally, social, emotional, and psychological health affect physical health and can also protect children against emotional and behavioural problems, violence, crime, teenage pregnancy and drug misuse.[5, 10] Beyond the health and social benefits to the individual, such outcomes have long-term economic impacts to society which need to be evidenced in order to justify investment in such interventions.

Economic evaluation has been typically used to aid allocative decision-making in the health sector, as healthcare costs continue to rise and NHS resources are consistently under pressure. The education sector faces many of the same financial constraints as the health sector and stands to benefit from consistent and transparent allocative decision-making. In order to address some of the shortcomings of economic evaluation of school-based PHIs identified in the previous chapter, this chapter presents the methods of a thorough economic evaluation of the RoE programme. This case study will demonstrate and illustrate key components of a school-based economic evaluation, which will be the first of its kind in this specific context; therefore providing a novel example of the advantages and challenges that remain for economic evaluation in the education sector. The chapter describes in detail, the methods used for the main trial economic evaluation of the RoE programme. It is broadly split into three sections. The first section details economic evaluation in child health interventions introducing some of the main paediatric outcome measures in use, including the outcomes used in this economic evaluation. The second section provides background and contextual information to the RoE trial. The final section provides a detailed description of the analytic methods which were used in the economic evaluation of the RoE programme.

4.2 Economic evaluation in child health

Interventions aimed at children have great potential of being cost-effective because of the longer time-frame over which health benefits can be gained.[186] A child's development is marked by physical, emotional, and cognitive changes; the trajectory of this development has distinct vulnerable periods where appropriate care is essential for growth and healthy development.[37] Poor health in childhood may lead to adverse effects in adulthood, such as limited educational attainment and labour market opportunities, as well as poorer health outcomes.[186] Early intervention is more effective as it aims to prevent problems from developing verses merely treating the problems once they manifest. They aim to provide the appropriate care during those vulnerable developmental stages. Advantages gained from early intervention are better sustained when they are continued with high quality learning experiences as depicted in Figure 11 below.[299]

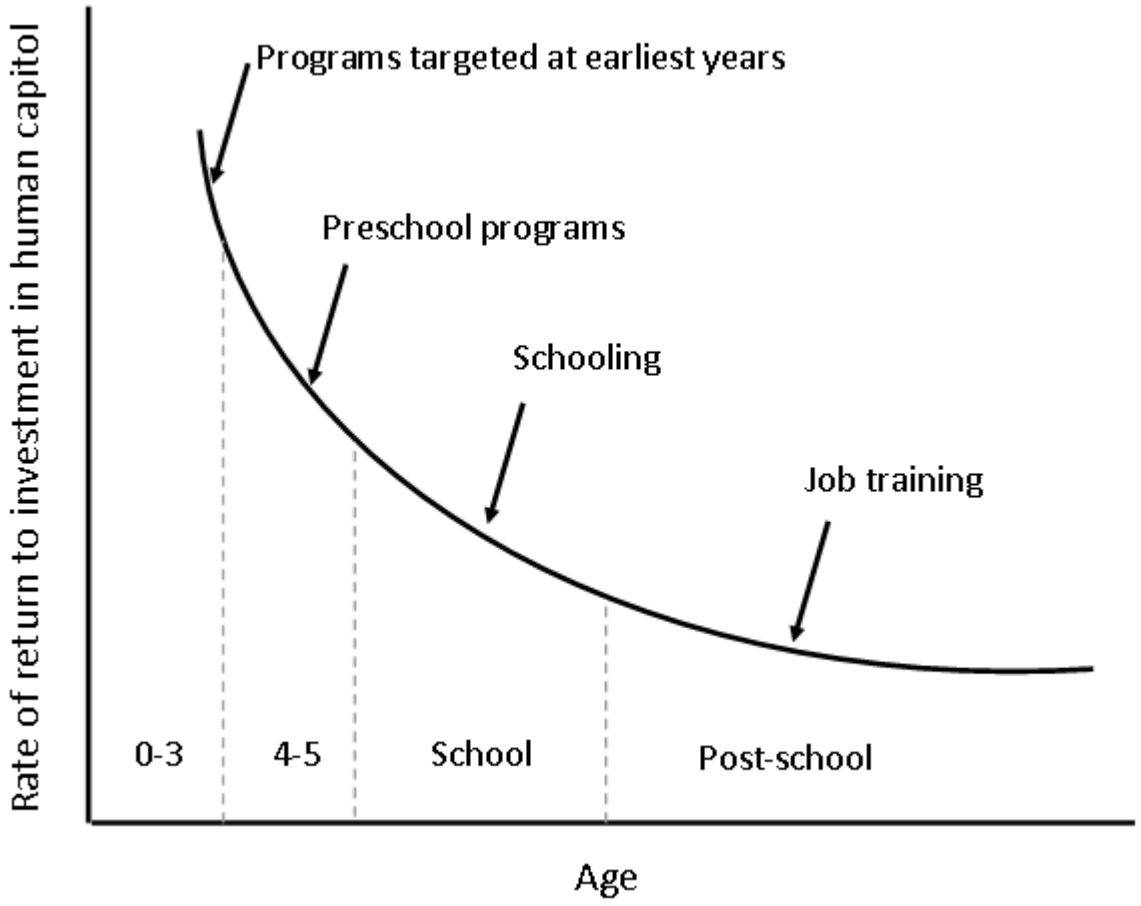


Figure 11: Marginal increase in investment at different stages of the life cycle - adapted from Heckman, 2008[299]

The methods for economic evaluation that were described in Chapter 2 were developed for the economic evaluation of health technologies intended for adults. Considering the differences between children and adults, simply applying the same methods of economic evaluation may not be entirely appropriate. Traditionally, outcome measures developed for adults were administered to children without alteration. However, modifying an adult measure risks compromising the validity and psychometric properties of the instrument.[37] That is why it is important to develop child specific outcome measures for use in economic evaluation. The following describes paediatric outcome measures that have currently been developed specifically for children. The first subsection describes generic measures, which is then followed by preference-based measures. Next, the CHU9D, a generic preference-based measure specifically developed for children, is formally introduced. Finally, in the last subsection of 4.2.1, the SDQ, a child SEW outcome measure is detailed.

4.2.1 Paediatric Outcome Measures

4.2.1.1 Generic measures of outcome

Generic patient-reported outcome measures (PROMs) are appropriate for anyone as a means for reporting their health. They are useful in that they can be applied to measure health in conditions that do not have a specific outcome measure, as well as, to make comparisons across conditions. If two or more different conditions are measured using the same generic PROM, they can easily be compared as they will have the same unit of outcome in common. A recent systematic review of generic PROMs for children identified 29 such measures.[300] These types of measures may often provide measurements for each domain or dimension separately, as opposed to a single score summarising them all. This could result in possible conflicts, where an intervention may result in improvements in one domain, but deteriorate in other domains when comparted to the control or other alternatives. Additionally, generic PROMs have not been valued with society's preferences, so they cannot be used to make adjustments in quality of life. In other words they cannot be used in CUA because the 'Q' in QALY is missing.

4.2.1.2 Preference-based generic measure of HRQoL

To put the 'Q' in QALY, a generic health measure must be preference-based; or the health state descriptive system must be accompanied by a set of utility values that were elicited using preference-based valuation techniques as described in section 2.1.4.1. There are challenges when evaluating paediatric QoL; direct elicitation of preferences may be preferred, however, the child's ability to complete a SG or TTO may be restricted by cognitive and age limitations.[301] The systematic review on generic PROMs for children mentioned above, identified six preference-based measures: 16 Dimensional (16D),[302] Assessment of Quality of Life Mark 2 6D adolescents (AQoL-6D),[303] Child Health Utility 9D (CHU9D),[96] EuroQol 5D Youth (EQ-5D-Y),[304] Health Utilities Index 2 and 3,[305] and Comprehensive Health Status Classification System – Preschool (CHSCS-PS).[306] The 16D was adapted from the 15D adult measure. It is intended for children aged 12-15 and has been valued by 15-18 year olds using a visual analogue scale. The AQoL-6D uses adult valuations that were calibrated with TTO responses from 15-18 year olds. The youth version of the EQ-5D (EQ-5D-Y) is available for children 7-12 years, but there is no UK valuation set available. The existing social value sets for EQ-5D cannot be assumed appropriate preference weights for paediatric populations[307] as the EQ-5D-Y is a distinct instrument from the standard adult version. Thus, this missing value set is a limitation to use of the EQ-5D-Y in economic evaluation. Likewise the HUI-2 and CHS-CS-PS have not been valued by children or adolescents.

Additionally, there is also the PedsQL first mentioned in section 3.3.2.3 as it was identified as an HRQoL measure used in the literature of school-based evaluation. It is a brief, 23 item, HRQoL measure for children and adolescents aged 8-12 which can be completed by children themselves or by parent proxy.[308] While it was developed with children and for children, there are currently no paediatric values available to estimate child health utility.[309] However, Khan and colleagues[309] have developed mapping algorithms which map to EQ-5D to estimate health utility. There is ongoing research by Stavros and colleagues[310] to develop a preference-based index for the PedsQL, however this was not yet available at press time. The CHU9D is the only preference-based measure to have been developed specifically for children versus being adapted from an adult measure, as well as being valued by adolescents to estimate health utilities.

For paediatric economic evaluation, NICE advises use of a standardised and validated preference-based health related quality of life measure that has been designed specifically for use in children.[64] This advice is sound due to the risks of compromising validity and psychometric properties when modifying adult measures for use with children. Since the CHU9D is the only preference-based measure that was not adapted from an adult measure, it is the only measure that can meet NICE's specifications of being 'designed specifically for use in children.'[64] The advantage the CHU9D brings to CUA contribute to, and warrant further research into the limited paediatric outcomes evidence. More on this point will be explored in Chapter 6, which validates algorithms that map from the SDQ to the CHU9D to facilitate CUA in line with NICE guidance.[64]

Another important consideration is whose values should be used to value health states? Typically, a sample from the general population is used to value generic preference-based measures because in most cases decisions are made on a societal basis and members of society are all contributing to the funding of the health care system through taxation, especially in countries such as the UK where there is a national health service. But some might argue for the preferences of the patients being used as they are the actual recipients of what is being evalutated.[68] It may be difficult for a member of the general public to value a health state they have never actually experienced. However, using a representative sample of the general population reflects the societal preferences of the population of interest. Should these representative samples include children and adolescents? Children are not typically considered rational, informed, or autonomous, and legally they are treated differently than adults and do not participate in the labour force.[37] Thus, their preferences are not deemed to be relevant to societal decision-making. Additionally, it is debatable whether or not children have the cognitive development to actually complete SG or TTO direct preference elicitation tasks, and whether it is ethical to do so as they may be subjected to questions about death.[311] On the other hand, adult values have been deemed inconsistent[312] and irrational so there is scope for incorporating children's preferences. Ideally, there would be one preference-based HRQoL measure (such as the EQ-5D) that is appropriate for all ages to fill in, and has been valued by the general population including children. However, this is not possible due to the significant differences between the two groups as well as cognitive and age limitations; thus, the reason child-specific measures have been developed. A study by Ratcliffe and colleagues[288] found when applying adolescent and adult values to

the same CHU9D health states, the differences would likely impact the assessment of cost-effectiveness. Differences between child and parent proxy preferences for health outcomes poses risks for delivery and evaluation of paediatric programmes as children's attitudes towards interventions are linked to their compliance and adherence.[312] If the differences in preferences are apparent between, children, adolescents and adults, the question remains, whose values should be used? Given the shortcomings of proxy preferences, adolescent values were used in the base-case analysis because it is justified that the preferences of young people should be incorporated in decisions that ultimately affect them.

4.2.1.3 The Child Health Utility 9D

The CHU9D is a relatively new generic preference-based health-related QoL instrument suitable for use with children ages 7-17.[313-315] It was developed by Kathrine Stevens, at the University of Sheffield, who carried out over 70 interviews with children, from two schools in Sheffield, to determine what dimensions of HRQoL would be included in the descriptive system. This descriptive system was then piloted with 150 children in schools, and a further 95 children from a clinical population from the Sheffield Children's Hospital which helped to further refine the descriptive system. The final descriptive system comprised of nine dimensions with five levels each. The nine dimensions are worried, sad, pain, tired, annoyed, school work/homework, sleep daily routine, and ability to join in activities. The five levels range from 1 to 5 and represent increasing severity. The full descriptive system is given in Table 6.

Table 6: CHU9D descriptive system

Dimension	Level	Description
	1	I don't feel worried today
	2	I fell a little bit worried today
	3	I feel a bit worried today
	4	I feeling quite worried today
	5	I feel very worried today
	1	I don't feel sad today
	2	I feel a little bit sad today
	3	I feel a bit sad today
	4	I feel quite sad today
	5	I feel very sad today
	1	I don't feel annoyed today
	2	I feel a little bit annoyed today
	3	I feel a bit annoyed today
	4	I feel quite annoyed today
	5	I feel very annoyed today
	1	I don't feel tired today
	2	I feel a little bit tired today
	3	I feel a bit tired today
	4	I feel quite tired today
	5	I feel very tired today
	1	I don't have any pain today
	2	I have a little bit of pain today
	3	I have a bit of pain today
	4	I have quite a lot of pain today
	5	I have a lot of pain today
	1	Last night I had no problems sleeping
	2	Last night I had a few problems sleeping
	3	Last night I had some problems sleeping
	4	Last night I had many problems sleeping
	5	Last night I couldn't sleep at all
	1	I have no problems with my daily routine today
	2	I have a few problems with my daily routine today
	3	I have some problems with my daily routine today
	4	I have many problems with my daily routine today
	5	I can't do my daily routine today
	1	I have no problems with my work today
	2	I have a few problems with my work today
	3	I have some problems with my work today
	4	I have many problems with my work today
	5	I can't do my work today
	1	I can join in with any activities today
	2	I can join in with most activities today
	3	I can join in with some activities today
	4	I can join in with a few activities today
	5	I can join in with no activities today

The CHU9D has demonstrated itself as a practical and valid measure for use in economic evaluation of child and adolescent healthcare programmes.[314, 315] Valuation of the CHU9D was directly elicited from an adult and adolescent population. Preference weights were derived from 300 members of a UK adult population using a SG technique for use in children 7-11.[313] Subsequently, through collaborative work with Julie Ratcliffe at Flinders University, Australia, preference weights were since derived from best-worst scaling DCE interviews of 590 Australian adolescents aged 11-17.[311] This means that the CHU9D can be valued using adult and adolescent preference weights.

At the time of writing, there were over 150 research studies currently applying the CHU9D in clinical trials, observational, and cohort studies across the world and there were Chinese, Spanish, Welsh, Danish, Italian, and Dutch versions available in addition to the original British English version.[316] It is a self-complete measure with a proxy completion available for younger children, and the recall period is today/last night. It is the only paediatric generic preference-based measure of HRQoL exclusively developed with children and for children (i.e. it did not start out as an adult measure that was adapted for use with children). Current research collaboration between the London School of Hygiene and Tropical Medicine, Great Ormond Street Hospital and the Royal College of Art is developing an app for iPad to collect CHU9D health state data from children through use of animation.[317] The five levels of each dimension are represented through animation and children pick which animation is most like them today. The three-stage project, CHILDSPLA, is still ongoing. In stage one the app was developed with primary school children and children in hospital. In stage two, the app was tested in multiple schools with multiple age groups ranging from 4-14 years. The research is currently in stage three which involves the development and testing of a method to elicit health state preferences from children.[317] This would mean that preference weights from younger children may soon be available in addition to the preference weights from adolescents and adults.

4.2.1.4 Social emotional wellbeing and condition specific measures of outcome

Mental wellbeing in children and adolescents in the UK has been declining over the past 30 years.[318] There have been increases in the number of young people reporting frequent

feelings of depression or anxiety,[319] in parent-reported behaviour problems, and in conduct disorders.[320] The frequent use of social media by young people as well as the increase in cyber bullying is often purported as a potential cause of this recent trend. Impacts of emotional and behavioural or mental health problems in childhood can impact adult outcomes including educational failure, unemployment, unhealthy lifestyles, and problems with interpersonal relationships.[321] SEW encompasses all of these problems (or lack thereof), however problems arise when attempting to meaningfully measure SEW.

> "[…] not everything that can be counted counts and not everything that counts can be counted" – William Bruce Cameron (1963)

A NICE guidance report on SEW in education[5] highlights the lack of valid methods for measuring SEW of primary school children and monitoring those changes over time. A suggested measure that has gained popularity among clinicians is the SDQ.

The SDQ [322, 323] is a widely used and validated behavioural screening questionnaire which can be used for children aged 4 to 17.[324-326] An additional early-years SDQ can be completed by parents or educators for ages 2-4.[326] There are three versions of the questionnaire that can be completed by the teacher (ages 4-17), parents (ages 2-4 and 4-17) and self-completed by the pupils (ages 11-17).[326] The use of all three informants (teacher, parent, and child) is considered ideal so that the results can be triangulated. Using just one informant can be problematic because parents tend to be good at identifying externalising and conduct problems, but less so at identifying emotional problems. Children are better at reporting emotional symptoms accurately, but under-report conduct problems, and teachers are somewhere in between (Minnis H 2016, oral communication, 14th October).

The SDQ consists of five symptom scales (emotional, conduct problems, hyperactivity, peer problems and prosocial) with five items each. Four of the scales represent negative attributes of the child's behaviour (total difficulties), while the fifth (prosocial scale) represents a positive attribute of the child's behaviour. The total difficulties score is the sum of the four negative attribute symptom scales. The 25 item behavioural and emotional assessment tool is much shorter and less cumbersome than other instruments such as the Child Behaviour Checklist.[327] The SDQ is also less dated with a focus on identifying children's strengths rather than solely focusing on their deficits as with the

traditional yet well-established Rutter Questionnaire.[328] The SDQ was developed by Robert Goodman who worked closely with Michael Rutter as many of the questions are similar and the two measure are highly correlated; the main difference of course being the addition of prosocial behaviours.[322]

SDQ Scoring algorithms converted into Stata (StataCorp LP, College Station, Texas, USA) syntax are available on the SDQinfo website.[326] They involve assigning a score from 0-2 (0=not difficulties, 2=many difficulties) for each item of the questionnaire and summing the total for each scale. The total difficulties and prosocial scores can be assigned to one of three general clinical thresholds; 'normal,' 'borderline' and 'abnormal.' These are general bandings and may be adjusted depending on the population which may vary by country, age and gender.[326] Table 7 gives the banding for interpretation of the teacher completed SDQ scores.

Table 7: Bandings for interpretation of Teacher Completed SDQ

	Normal	Borderline	Abnormal
Emotional Symptoms Score	0-4	5	6-10
Conduct Problems Score	0-2	3	4-10
Hyperactivity Score	0-5	6	7-10
Peer Problems Score	0-3	4	5-10
Prosocial Behaviour Score	6-10	5	0-4
Total Difficulties Score	0-11	12-15	16-40

The bandings are not considered a diagnostic threshold, rather they tend to be used as a screening tool to refer children who score in the 'borderline' and/or 'abnormal' ranges to Child and Adolescent Mental Health Services (CAMHS) for further examination (Minnis H 2016, oral communication, 14th October). The SDQ is also used as a clinical outcome measure to examine change over treatment. However, some experts in child and adolescent psychology are unsure how sensitive the SDQ is to change (Minnis H 2016, oral communication, 14th October). This could have important consequences when using the SDQ as a primary outcome measure of effectiveness in a RCT. Because the bandings are not considered diagnostic, changes in scores cannot be compared to any clinically meaningful differences when using the SDQ as an outcome measure in an RCT. This poses challenges when interpreting effectiveness as no consensus has been reached on what a clinically meaningful change in the SDQ represents (Minnis H 2016, oral communication, 14th October). More discussion on this topic is given in section 5.7. The SDQ is certainly

popular; it is freely available, has been translated into over 80 languages, and there have been over 4,000 published articles from over 100 countries that use the SDQ.[326] In the UK, it is being routinely collected in the Millennium Cohort Study as well as by CAMHS. However, as the NICE guidance alluded to earlier, there is a lack of valid measures for measuring primary school children's SEW.

Given the points brought up above (bandings cannot be used as a diagnostic threshold, it may not be sensitive to change, and few valid measures exist to measure SEW), the SDQ may not be appropriate to measure something that is very difficult to quantify. SEW is an abstract and subjective concept. The five SDQ subscales certainly do cover most of the major aspects of child and adolescent mental health, but SEW is not simply the absence of mental health problems, i.e. SEW involves a child flourishing and is not equivalent to mental health. In the absence of specific and validated measures of SEW, the SDQ is appropriate in attempting to quantify and measure this difficult area. In a randomised controlled trial and economic evaluation context, it is important to include both a generic preference-base quality of life measure, as well as a condition specific measure of outcome. This is because of the difficulties that arise when trying to quantify changes in a non-generic preference-based outcome, such as the SDQ, in terms of other education outcomes covered under the same budget, e.g. increases in test scores. This is where a generic 'yardstick' measure can be quite useful, and to address the limitations of a generic outcome, condition-specific outcomes can be included which may be able to measure the intervention's effectiveness more accurately. However, when using a condition specific outcome measure in CEA, there may be difficulties in interpretation of unit changes in scores and what values should be attributed to such changes (i.e. cost per unit increase/decrease in SDQ). CEA ICERS are more difficult to interpret because unit changes in condition specific measures have rarely been valued. This places more burden on the decision-maker to determine these values and can lead to less transparency and consistency in decision-making.

Mapping algorithms have been published that allow SDQ scores to be converted into CHU9D utility values,[329] but more information including primary analysis will follow in Chapter 6. The base-case analysis, detailed later, uses the CHU9D to measure QALYs (to be compliant with NICE guidance), while a sensitivity analysis CEA, uses the SDQ to

examine the cost-effectiveness of the primary outcome measure of the RoE trial. The following section describes the main trail of RoE.

4.3 Roots of Empathy

The economic evaluation was conducted alongside: *A cluster randomised controlled trial evaluation and cost-effectiveness analysis of the Roots of Empathy school-based programme for improving social and emotional wellbeing outcomes among 8-9 year olds in Northern Ireland*; which was funded by the National Institute of Health Research Programme (Project Reference: 10/3006/02).[55] The funder did not have any role in the identification, design, conduct, and reporting of the analysis. Professor McIntosh reports that she is a member of the funding board of the NIHR PHR programme and all other co-authors have nothing to disclose and no conflicts of interest.

RoE is a universal school-based SEL programme that was developed in Canada over two decades ago by Mary Gordon.[53] It is one of the few SEL programmes that has an existing evidence base regarding its effectiveness.[40, 43, 45] A recap of the RoE programme is provided here for clarity. The programme is delivered on a whole-class basis over one academic year (October to June) and consists of 27 lessons, which are all based around a monthly classroom visit from an infant and parent, usually recruited from the local community. During these monthly visits, children learn about the baby's growth and development through interaction and observation of the baby and parent over the course of the year. The intervention is highly structured and any adaptation or tailoring of either the content or method of delivery is discouraged by the RoE organisation.

Each month a trained RoE instructor, who is not the class teacher, visits the classroom three times for: a pre-family visit; the visit of the parent and infant; and a post-family visit. Instructors undergo a total of four days intensive training that is delivered directly by a specialist RoE trainer from Canada. The specialist trainer also provides on-going mentoring support via regular telephone calls to all instructors. In addition, on-going support is also available to each instructor through each Health and Social Care Trust's lead RoE coordinator. Each RoE lesson takes place in the classroom with the teacher present but not actively involved in delivery. The programme provides opportunities to discuss and learn about the different dimensions of empathy, namely: emotion

identification and explanation; perspective-taking; and emotional sensitivity. The parent-infant visit serves as a springboard for discussions about understanding feelings and infant development and effective parenting practices.

At the heart of the programme is the development of empathy in young children. The psychological definition of empathy is ambiguous[330] with few coming to a consensus,[331] but it is largely agreed to consist of three processes: 1.) an emotional simulation process; 2.) a conceptual, perspective-taking process; and 3.) an emotion-regulation process.[332] The Oxford Concise Medical Dictionary provides a clear, simple, and easy to understand definition as, 'the ability to imagine and understand the thoughts, perspective, and emotions of another person.'[333] It is through the development of empathy that RoE seeks to improve children's social and emotional understanding, promote prosocial behaviours, and decrease aggressive behaviours. Because the baby cannot verbally communicate his/her needs, wants, and emotions, children must learn to identify these through observations of the baby's behaviour. This allows children to not only become better at identifying emotions of their peers, but within themselves as well. If and when children learn empathy, they have the foundation for developing positive and prosocial interactions. This social and emotional development has potential implications for a child's future and longer-term outcomes.

4.3.1 The RoE trial

The RoE programme's reach is now worldwide, but it has only recently been introduced to the UK, thus the RoE trial aimed to evaluate the immediate and longer-term impacts of the programme on SEW outcomes and its cost-effectiveness. The trial was conducted in primary schools in four of the five Health and Social Care Trust areas in Northern Ireland and given to Year 5 pupils (8-9 years old). The trial was led by Professor Paul Connolly, Head of the School of Education and Interim Dean of Research at Queen's University Belfast. The economic evaluation was led by Professor Emma McIntosh, Deputy Director of the Health Economics and Health Technology Assessment research group at the University of Glasgow. Information given in this section (pertaining to the main trail) was completed by the main trial research team and reported elsewhere in the end of study report currently in press for peer reviewed publication by the NIHR Journals Library (Ref: 10/3006/02).[55]

The research team identified and synthesised data from seven eligible evaluations of RoE that had been conducted to date (synthesis led by Dr Sarah Miller, August 2016). A summary of the synthesis is given. Of the seven eligible studies, only one was a (cluster) randomized controlled trial. The pooled data from these studies suggests that Roots of Empathy is effective in leading to small improvements in prosocial behaviour (standardised mean difference (SMD) = +0.13) and reductions in aggressive behaviour (SMD = -0.18). There is no evidence to suggest it is effective in improving other SEL outcomes amongst children, in this case empathy and emotional regulation. Only one evaluation studied the longer-term impact of the programme and it suggests that after three years the intervention group had poorer prosocial behaviour compared to the control group (SMD=-0.12, 95% CI [-0.17, -0.07]). With respect to aggressive behaviour three years post intervention, the intervention group were displaying only slightly less aggressive behaviour compared to the control group (SMD=-0.06, 95% [-0.09, -0.03]) and although statistically significant, this effect was much reduced compared to the effect observed at immediate post-test (SMD=-0.25). There were no evaluations to examine the potential cost-effectiveness of RoE so the following economic evaluation is highly original and a significant contribution to the RoE evidence base.

4.3.2 RoE trial aims

Given the limited existing evidence base for RoE, particularly economic evidence, the aims of the overall trial evaluations were to:[55]

A. Evaluate the immediate and longer-term impact of the RoE programme on social and emotional wellbeing outcomes among 8-9-year-old pupils.

B. Evaluate the cost-effectiveness of the programme from a public sector perspective over trial time horizon of 45 months (3.75 years or 3 years follow-up after intervention completion).

The trial aimed to answer the following research questions:[55]

1. What is the impact of the programme at post-test and up to three years following the end of the programme on a number of specific social and emotional wellbeing outcomes for participating children?

2. Does the programme have a differential impact on children depending on: their gender; the number of siblings they have; and their socio-economic status and/or the socio-economic profile of the school?

3. Does the impact of the programme differ significantly according to variations in implementation fidelity found?

4. What is the cost-effectiveness of the programme in reducing cases of aggressive behaviour and increasing prosocial behaviour among school-aged children?

The final aim (B) and research question (4) for the RoE trial were the focus of this economic evaluation and the following methods reported in this chapter. Addressing the fourth research question is not only relevant to decisions on school policy, but health policy as well. The other aim and research questions were tackled by the main trial team whose findings are available elsewhere, currently in press with the NIHR Journals Library.[55] As research question 2 is relevant to PHIs in terms of reducing inequalities, the findings from the report are summarised here briefly. Pre-specified subgroup analyses were undertaken to explore whether the programmed worked better according to the socio-economic background of the child's family which was measured using multiple deprivation rankings for the child's home address. Given that there were 27 tests in total and only two interaction terms were found to be significant, the findings may have occurred by chance and should be considered with caution. SDQ total difficulties or prosocial scores were not found to be significantly impacted by deprivation level in the multilevel model analyses.[55]

4.3.3 Data Collection

The data collection for the RoE trail was led by the RoE trial team and full methods detailing data collection are available elsewhere.[55] Seventy-four primary schools were recruited to the trial between March and June 2011. Schools were randomly allocated to receive the RoE intervention (n=37), or to the waitlist control group (n=37), which did not receive RoE and continued with their regular curriculum and usual classroom activity. This comparator was selected because in the absence of RoE, usual classroom activity would be what would take place normally. Schools allocated to the intervention group, received the RoE programme in their selected year 5 class for one academic year (2011/2012).

Schools placed on the waitlist received the programme in 2012/2013, but on the understanding that RoE would not be delivered to their current Year 5 cohort (control group) as they progressed through the remainder of the trial follow-ups.

Pre-test (or baseline) data collection from the children, parents and teachers took place in October 2011 across all participating schools prior to the first sessions of RoE being delivered in the intervention schools. Consent forms were sent home with children prior to baseline data collection. Post-test (or immediately after intervention completion) data were collected in June 2012. Follow-up data collection took place annually at 12 months (June 2013), 24 months (June 2014), and 36 months (June 2015). At the final sweep of data collection, children were 11-12 years of age and at the end of their first year in secondary school. Outcomes collected for the RoE trial but not included in the economic evaluation included the: Child Behaviour Scale, Infant Facial Expression of Emotions Scale, Emotion Recognition Questionnaire, Interpersonal Reactivity Index, Child Anger Management Scale, and the Revised Olweus Bully/Victim Scale.

Teachers were asked to complete a questionnaire for each participating child at each time point, which included the SDQ and the Child Behaviour Scale. Parents were contacted via post and asked to complete a questionnaire and return it to the research team in a freepost envelope. The questionnaire included the SDQ and asked parents about background information on family composition, parental education, and employment. Fieldworkers administered questionnaires to the children on a whole-class basis. Fieldworkers were fully trained and coordinated by the research team. Included in the children's questionnaire was the CHU9D as well as other secondary outcome measures: emotion regulation, empathy, recognition of emotions, understanding of infant crying, and bullying. Children were asked not to confer, and this was ensured by the teacher and RoE fieldwork present. Each question was read aloud to the class and any words or phrases that were difficult were explained. If a child was absent, efforts were made to return to the school at a later date.

4.3.3.1 Economic evaluation outcome measures

The primary outcomes for use within the economic evaluation were the SDQ and CHU9D, which were collected at each data collection time point as described above. Due to a low

response rate from parents, the teacher complete version of the SDQ was used in the analysis. As teachers completed the SDQ, they acted as a proxy for child behaviour outcome, as the self-complete version was only available for older children aged 11-17. Other outcomes that were incorporated within the economic evaluation included age (as measured by year in school), gender, deprivation level, and number of siblings. These were all collected from the trial and deprivation was measured by the Northern Ireland Multiple Deprivation Measure 2010 (NIMDM) which is a relative measure of deprivation.[334] Additionally, in order to try to capture broader outcomes of the programme, parent's quality of life was measured via EQ-5D. Unfortunately, it was only available at 24 and 36-month follow-up due to issues with the trial design (see section below and section 4.4.2.2). Response rates were low and it was subsequently dropped from analysis.

4.3.3.2 Resource use

Resource use and costs of the intervention were also collected from the trial. Due to issues with the trial design, resource use was only collected and available from 24 and 36-month follow-up. As cost and outcome data should be consistent over the relevant time horizon, the 24-month resource use questionnaire asked parents to recall resource use for their child since the beginning of the trial period. The long recall is a recognised limitation and more detail about the resource use and cost data collection is given in section 4.4.2.2 and 4.4 which details the within trial economic evaluation of the RoE programme.

4.4 RoE main within-trial analysis methods

The last section (4.3) gave background to the RoE trial, aims, and data collection. This final section details the full economic evaluation methods using all data from the three years of follow-up from the RoE trial. The first section (4.4.1) gives a brief overview of the methods that were employed, followed by a detailed description of the costs, outcomes, missing data, analyses, and sensitivity analyses performed.

4.4.1 Overview

The base-case analysis of the RoE economic evaluation took the form of a CUA, which was based on the incremental cost per QALY gained. Various sensitivity analyses were performed including a CEA, which was based on the incremental cost per one-unit decrease of the total difficulties score and incremental cost per one-unit increase of the prosocial behaviour subscale of the SDQ. Health economics data were collected at five time points:

1. Pre-test (baseline)
2. Post-test (after intervention completion)
3. 1-year follow-up from post-test
4. 2nd year follow-up form post-test, and
5. 3rd year follow-up from post-test.

The analysis had a time horizon of 3.75 years (45 months) which equates to three years follow-up after intervention completion. This time horizon is appropriate within the limits of resource constraints, as it is one of the longest cluster RCT follow-ups of RoE to be performed. The study took a public sector perspective with NHS, PSS, local government authority, and family costs included. Costs were derived from resource use questionnaires that were developed by the author and supervisor specifically for this trial, which were sent home to parents. Costs were also derived from the actual cost to deliver the RoE intervention. Costs and QALYs were discounted using NICE's public health guidance discount rate of 1.5%.[79] QALYs were determined from the CHU9D which was completed by children in their classroom. Missing data on costs and QALYs were handled using MI with chained equations.[335] Regression methods were used to obtain incremental cost and effect estimates. Multiple regression methods that ignore clustering (e.g. the within school clusters as in this trial) can lead to biased coefficients and especially biased standard errors.[336] Multilevel models have been proposed as a method to address issues surrounding clustering in economic evaluation[337] and their use was explored. Upon recognition of the model being a poor fit for costs in this particular dataset, regression with robust standard errors was conducted to adjust standard errors by indicating that observations within schools may be correlated, but are independent between schools.

ICERS were estimated by dividing the difference in mean costs between groups by the difference in mean effects between groups. The uncertainty surrounding the ICER was investigated by use of a nonparametric bootstrap of 1,000 iterations. This uncertainty was then presented on the cost-effectiveness plane and summarised on the cost-effectiveness acceptability curve (CEAC).These estimates of ICERs were considered with respect to the £20,000 to £30,000 per QALY threshold generally accepted by NICE to determine cost-effectiveness in the UK. To allow for uncertainty a series of sensitivity analyse were performed. All analyses were conducted as intention-to-treat analyses and in Stata/SE 14.1 (StataCorp, College Station, TX, USA). Table 8 below gives an overview of the data collected for use in the economic evaluation. A completed CHEERS checklist for the RoE economic evaluation can be found in Appendix 6.

Table 8: Data from the RoE trial collected for the economic evaluation

Data Type	Description of Data	Time Points
Costs of Intervention	Fees, training, personnel, and materials to run RoE	Pre-test
NHS/PSS Resource use	NHS/PSS Service use including staff time and parent self-report children's medications	F2, F3
Cost to Society	Time off work to care for child and police visits	F2, F3
HRQoL	Child Health Utility 9D (CHU9D) questionnaire	Pre-test, post-test, F1, F2, F3
Trial Primary Outcome	Strengths and Difficulties Questionnaire (SDQ)	Pre-test, post-test, F1, F2, F3
Demographics for Subgroup Analysis	Gender, school, Multiple Deprivation Measure 2010, number of siblings	Pre-test

4.4.2 Costs

Costs of the RoE programme were made up of the following:

Equation 8:Total cost of RoE programme

$$\boldsymbol{C_T = C_{Int} + C_{NHS} + C_{Soc}}$$

Where C_T is the total cost made up of: C_{Int}, the cost of the intervention including personnel, training, materials, fees and other cost; C_{NHS}, NHS resources used including service use, staff time, and medications; and C_{Soc}, societal costs such as parental time off work, charity, and police costs.

4.4.2.1 Costs of the Intervention

All costs were reported in price year 2014 British Pounds (GBP). A number of costs were incurred in 2011 when the intervention ran. Where required, costs were inflated to the base year 2014 using the Hospital and Community Health Services (HCHS) price index (see Table 9).[338] The HCHS is a weighted average of two separate inflation indices: the pay cost index and the health service cost index.[338] The total cost of the intervention was made up of the following cost categories: key point people, administrative support, instructor time, instructor training materials, instructor materials, instructor fee, and other costs.

Table 9: Hospital and Community Health Services Index

Year	Pay and Prices index (1987/88=100)
1995/96	166
1996/97	170.6
1997/98	173.5
1998/99	180.4
1999/00	188.6
2000/01	196.5
2001/02	206.5
2002/03	213.7
2003/04	224.8
2004/05	232.3
2005/06	240.9
2006/07	249.8
2007/08	257.0
2008/09	267
2009/10	268.6
2010/11	276.7
2011/12	282.5
2012/13	287.3
2013/14	290.5
***2014/15**	286.8

*** Estimate only, an average of the three previous years**

Personnel costs (salary costs) were classified by NHS Band and were taken from the 2011 Health Service pay scale.[339] Personnel costs included: four key point people (Band 7) who are Health Trust employees who co-ordinate RoE in each of the four participating Trusts, four administrative support part-time workers (Band 3), and a RoE instructor for each school (Band 6). Salaries were based on mid-spine points for each respective band range (including 25% oncosts) and adjusted for time spent on RoE activities (see footnotes in Table 11). Key point people underwent 28 hours of self-directed learning as training over three to five days and spent an average of 13 hours a week on RoE related activities. Administrative support salaries were 50% full time equivalent, or 18.75 hours per week. RoE instructor costs were split into training, time spent preparing for the 27 sessions and time spent delivering each session. Each instructor received 30 hours of training. Time spent preparing and delivering the sessions varied; the average time spent preparing and delivering all 27 sessions was 24 hours for preparation and 24 hours for delivery. Additionally, there were instructor training materials, instructor materials for delivering

the programme, and an instructor fee paid to each instructor. This fee would become an annual fee if the programme were to be continued along with all other personnel costs described previously.

Fees paid to the RoE programme in Canada for use of the programme in the UK are reported in a cost category referred to as 'other costs' in Table 11. These included programme support costs, materials shipping, training and mentoring expenses, and ongoing mentoring. The programme fees were originally purchased in 2011 Canadian dollars and converted to GBP price year 2011 using purchasing power parities (PPP) reported by the OECD[340] (see Table 10) and inflated to the current price year (2014) using the HCHS index. The RoE intervention was given to 33 schools with 764 pupils receiving the intervention. Please see Table 11 for a list of component costs that make up the total cost of providing the RoE programme in a Northern Ireland context.

Table 10: OECD Purchasing Power Parities (PPP)

OECD PPP

	2011	2014
Canada	1.2399	1.2612
United Kingdom	0.6997	0.7081

Table 11: Component costs of the RoE programme

Cost Item	Unit cost per hour (personnel)/Item cost	Quantity
Key point person [a]	£18.26	Varies
Administrative support [b]	£6.61	18.75 per week
Instructor time [c]	£15.29	Varies
***Instructor training materials**	£1,027.97	1
***Instructor materials**	£456.88	1
Instructor fee [d]	£171.33	1
***Other costs**		
Programme support costs	£5,710.94	1
Materials Shipping	£2,569.92	1
Trainers/mentoring expenses	£3,426.57	1
Mentoring	£5,139.85	1

*Annuitized cost

[a] One key point person per Trust at mid-point Band 7 salary range £35,600

[b] One part-time support worker per Trust at mid-point Band 3 salary range £12,900

[c] One instructor per school at mid-point Band 6 salary range £29,800

[d] 300 CAN$

Annuitization was carried out to spread fixed costs of the intervention over the anticipated five-year life span of the RoE intervention. Annuitization is typically performed for capital costs such as buildings and equipment, however other costs such as training and materials may also be annuitised if they are incurred at the start of the programme, yet have a useful life longer than the initial period.[341] Training and development incur costs at the beginning of a programme, but the effects of training often last much longer than the initial period. Training, materials and other programme costs were one-time costs that were annuitised over the expected life of the intervention.[341] The base-case assumption of the expected life of the intervention was assumed to be five years, at which point training would need to be repeated and materials replaced. Therefore, costs were annuitised over five years at a discount rate of 1.5%. The equivalent annual cost was estimated using the annuitisation formula given in Equation 9.

Equation 9: Annuitization formula

$$K = E * \left[\frac{1-(1+r)^{-n}}{r}\right]$$

where K= the initial outlay
E=the equivalent annual sum
n= the expected life of the asset
r = the rate of interest or discount rate

A number of sensitivity analyses were conducted around this assumption such as use of varying discount rates (3.5% and 5%) and the useful life of the training and materials (3 years). Table 12 is the discount table used to calculate the equivalent annual sum for the annuitized costs. A scenario with no annuitization or discounting was also performed.

Table 12: Discount table for Annuitization

n	1.5%	3.50%	5%
1	0.985221675	0.966183575	0.952380952
2	1.955883424	1.899694275	1.859410431
3	2.912200417	2.801636981	2.723248029
4	3.854384648	3.673079209	3.545950504
5	4.782644973	4.515052375	4.329476671
6	5.697187165	5.32855302	5.075692067
7	6.598213956	6.11454398	5.786373397
8	7.48592508	6.873955537	6.463212759
9	8.36051732	7.607686509	7.107821676
10	9.222184552	8.316605323	7.721734929

4.4.2.2 Resource Use

Resource use was identified through early discussions with the trial managers and their contacts with the school to identify likely resource use. Resource use was then measured over the length of the trial and was made up of the following data collection: i.) NHS resource use including service use and staff time, and parent self-report children's medications; and ii.) societal costs such as social worker, school nurse, parent's time and potential contacts with the police. These broad ranging costs were considered from a public sector perspective as per NICE public health guidance.[79]

Resource use that was expected to differ between groups was collected at the second and third year follow-ups (24 and 36 months). A series of complications arose due to changes in co-investigators during the trial, and thus resource use (and parental EQ-5D) were not collected at pre-test, post-test, or at the 12-month follow-up. To account for resource use over the entirety of the trial period, resource use questionnaires at the 24-month follow-up asked parents to recall health and social care resource use from 'when their child started Primary 5,' which relates to the beginning of the study. At the final follow-up (36 months), resource use questionnaires asked parents to recall their child's resource use from the past 12 months. While the long recall periods are not ideal, it was decided that some data on resource use was better than none. Resources were valued using UK national unit costs.[338]

Specifically, health and social care resource use collected included the number of contacts with various NHS services, children's medications, time off work or daily activities parents needed to take due to their child being off school, and any contacts children had with the police. The time off work or other leisure activities was collected in order to approximate the opportunity cost of how parents choose to spend their time. As not all parents may need to take off work, the average British wage was applied as a unit cost to represent parent's time equally. The NHS services that were collected were visits to: general practitioner (GP), school nurse, accident and emergency (A&E), social worker, speech therapist, occupational therapist, physiotherapist, educational psychologist, education welfare officer, psychiatrist, counselling/therapy, dentist, optician, hospital inpatient and outpatient stays, and any other services that were not included could be written in. See

Figure 12, which shows the initial health and social care services questionnaire sent home to parents at the 24-month follow-up.

Unit costs were assigned to resource use using the PSSRU Unit Costs of Health and Social Care 2014,[338] NHS Reference Costs 2013/14,[342] and Office for National Statistics (ONS) median weekly earnings.[343] See

Table 13 for unit cost and source information for RoE resource use. Up to four medications could have been self-reported and unit costs for those were obtained from the British National Formulary (BNF) for children.[344] Occasionally parents reported over-the-counter medications which were considered societal costs. These were assigned unit costs using a market value from a national pharmacy, Boots. Up to two 'other service uses' could have been self-reported by parents; these were assigned unit costs in the same manner as described above. Occasionally parents listed contacts with charitable services so these were considered societal costs. Once all resource use had been assigned a unit cost, two sample t-tests with equal variances were performed to test for significant differences in resource use between groups. Finally, total cost was calculated for each group and discounted by 1.5% in the base case. Sensitivity analyses were performed varying the discount rate to the traditional 3.5% rate recommended by NICE.[64] Additionally, t-tests were performed on each service use to determine if the intervention had any impact on resource use between the groups.

Your child's use of Health and Social Care Services

1. What health and social care services has your child used since he/she started Primary 5?

***Note 1:** please enter '0' if service has not been used*

Service	Total number of contacts
General Practitioner (GP)	
School Nurse	
Accident and Emergency (A&E) Visit	
Social Worker	
Speech therapist	
Occupational therapist	
Physiotherapist	
Educational Psychologist	
Education welfare officer	
Psychiatrist	
Counselling/therapy	
Dentist	
Optician	
Hospital inpatient stay	Number of nights:
Hospital outpatient stay	
Other _____	
Other _____	

2. Please list below your child's use of any medication taken since he/she started P5?

Name of medication	How long did your child take this medication for? (e.g. 1 week)	Daily Dosage
1.		
2.		
3.		
4.		

3. Since your child started P5, have you had to take time off work or your usual daily activities due to your child being off school? *(For example, time off due to child's illness, behavioural problems, attending appointments etc.)* Yes ☐ No ☐
If yes, please state how many days:

4. Since your child started P5, has he/she been in contact with the **police**?

Yes = 1, No = 2 ☐

If yes: How many contacts with the police? Contacts ☐
(Note: contact= interview or stay of some hours)
How many court appearances? Appearances ☐

THANK YOU

Figure 12: Resource use questionnaire

Table 13: RoE resource use, unit costs, and sources for unit costs

Variable	Unit Cost	Source
GP	£46.00	PSSRU 2014 pg. 195. Per patient contact lasting 11.7 minutes, with qualifications
School Nurse	£63.00	PSSRU 2014 pg. 85. Nursing average cost per contact. School-based children's health care services- group.
Education Welfare Officer	£27.00	PSSRU 2014 pg. 155. TAC meeting attended by education welfare officer
A&E	£72.00	NHS Reference costs 2013/14. Type 1 admitted, emergency medicine any investigation with category 5 treatment
Social Worker	£41.00	PSSRU 2014 pg. 99
Speech therapist	£89.00	PSSRU 2014 pg. 85. Average cost per group session
Occupational Therapist	£113.00	PSSRU 2014 pg. 85. Average cost per group session
Physiotherapist	£81.00	PSSRU 2014 pg. 85. Average cost per group session
Educational psychologist	£41.00	PSSRU 2014 pg. 156
Psychiatrist	£228.00	NHS Reference costs 2013/14. CAMHS, Children and adolescents, national average unit cost
Counselling/therapy	£81.00	PSSRU 2014 pg. 85. Average cost per group session
Dentist	£65.00	PSSRU 2014 pg. 197. Unit cost/hour
Optician	£21.10	Northern Ireland sight test fee (children don't pay) MOS/294
***Police**	£325.00	PSSRU 2014 pg. 149. Police cost for criminal offence (statement and interview), cost to others
Hospital Stay (no. nights)	£326.00	PSSRU 2014 pg. 111. Inpatient specialist palliative care, average cost per bed day
Hospital Outpatient visit	£189.00	PSSRU 2014 pg. 85. Paediatrics average cost per attendance
Other Service use (x2)	varied	Varied: PSSRU or NHS Reference costs
Medication (x4)	varied	Varied: BNF or *Boots market prices for over the counter drugs
***Time off work (days)**	£104.00	£518 median weekly earnings April 2014 http://www.ons.gov.uk/ons/rel/ashe/annual-survey-of-hours-and-earnings/2014-provisional-results/stb-ashe-statistical-bulletin-2014.html

*Indicates societal cost

4.4.3 Outcomes

The primary child outcomes for the trial were increases in prosocial behaviour and decreases in difficult behaviour as measured by the teacher rated version of the SDQ. Given the primary outcomes of the trial, the SDQ was a logical choice for measuring those outcomes. Thus, the cost-effectiveness analysis (a sensitivity analysis) was based on incremental changes in both total difficulties and prosocial behaviour scores of the SDQ.

In order to gain further understanding and background context of the SDQ, an informal expert interview was conducted with Professor Helen Minnis of Child and Adolescent Psychiatry. The questions focused of gaining a clearer understanding of how the SDQ was currently being used in practice, if it was a reliable measure of SEW, and what the changes in scores meant from a clinical standpoint. The interview was recorded and transcribed by the author; however, no formal qualitative analysis was conducted. The transcript was used for general background information and to add to the discussion of the SDQ as an outcome for SEW in CEA.

The trial was a large cluster-randomised controlled trial with over 1,000 pupils taking part in the study. The size and rigour of a large randomised controlled trial provided a sufficient source of effectiveness data as both clinical effectiveness (SDQ) and HRQoL (CHU9D) were collected and available for analysis. Because the two main benefits collected in this trial were health benefits, a CUA was conducted and inclusion of a further CCA or CBA was not required.[79] Other secondary outcomes collected for the main trial were all related to measuring SEW, and thus non-health benefits were not collected. Because RoE is a school-based PHI, broader non-health benefits could be expected to arise such as improved education outcomes. This is a limitation from the perspective of the education decision maker as they may be interested in RoE's potential effect on education outcomes. However, this trial was funded for and focused on analysing the potential health benefits arising from the programme.

4.4.3.1 Quality Adjusted Life Years (QALYs)

In the RoE trial, HRQoL was measured using the CHU9D which is the first generic preference based measure specifically designed for use with children to estimate QALYs for economic evaluation of programmes/interventions for young people.[345] All other generic HRQoL measures for children were originally developed for adults and adapted for children, or developed for children, but require use of a mapping or crosswalk function to adult values to estimate health utility. The CHU9D was designed with children, specifically for children, and has been valued by adolescents and adults without requiring the use of a mapping function to estimate child health utility.

Because the CHU9D is the only HRQoL measure developed specifically for children and valued by children, the adolescent values tariff was deemed the more appropriate tariff to apply to health state profiles in the base-case analysis as it incorporates adolescent values into the decision making process. Sensitivity analysis was performed which applied the adult values tariff. Utilities were converted to QALYs using the AUC method described by Matthews et al[102] and given in Equation 5. In this context, QALYs should be interpreted in the same way as the outcome of any PHI. RoE QALYs reflect the quality of life gains achieved from the intervention's aim to increase social and emotional understanding, empathy, promote prosocial behaviours, and decrease aggressive behaviours.

4.4.4 Missing Data

Health and resource use costs for children were measured using parental self-report. Health and resource use questionnaires (Figure 12) were posted home to parents who were asked to return the completed questionnaire in a freepost envelope. Health and resource use data was available for the second and third year follow-ups only as mentioned previously. A descriptive analysis of missing data was first undertaken to identify an appropriate analysis method to deal with the missing data. The missing data analysis follows recommendations set out by Faria and colleagues[346] for handling missing data in CEA. Missing data mechanisms are often categorised using Rubin's framework for missing data:[347]

- Data missing completely at random (MCAR) assumes missing data do not depend on the observed and unobserved data values, the missing data is independent. The observed data is a representative sample of the overall population.

- Data missing at random (MAR) is a less restrictive assumption than MCAR. Missing data depend only on the observed data and not the unobserved missing data. Any systematic differences between observed and unobserved data can be explained by differences in observed data.

- Data are not missing at random (NMAR) when the probability that data are missing depends on unobserved values. For example, individuals with worse outcomes may be more likely to be missing. There is no way to identify with certainty if data are NMAR because it depends on the unobserved data that are missing.

If data are MCAR a complete-case analysis is valid. In complete-case analysis, only individuals with complete data at each follow-up are included in the analysis. This is an inefficient use of the data because any individuals with missing follow-up data are dropped from the analysis.[346] Available-case analysis makes more efficient use of data by calculating costs and QALYs by treatment group at each follow-up point. They are then summed by treatment group over the whole time horizon of the study. A limitation is that different samples of costs and QALYs may be used which can lead to non-comparability and affect the covariance structure.[348] The MAR assumption is a less restrictive assumption as missing data depend only on the observed data and not the unobserved missing data. MI is an appropriate analysis strategy for dealing with MAR data. Data are unlikely to be MCAR if the proportion of missing data varies widely by group. Therefore, descriptive analysis of percentage of missing values by group and in total was undertaken along with range, mean, and standard deviation of the observed data. If a variable was found to have over 80% of its values missing at any one time point, the variable would be dropped from further analysis and MI on those variable would not be performed.

4.4.4.1 Missing Data Patterns

Patterns of missing data were explored using the Stata/SE 14.1 (StataCorp, College Station, TX, USA) 'misspattern' command on total costs and QALYs at each time point. Data follow a non-monotonic pattern when data may be missing for an individual in one follow-up but then they return in subsequent follow-ups. Here the MCAR assumption would be inefficient because data from subsequent follow-ups would not be utilised and all non-complete cases would be dropped.

4.4.4.2 Association between missing and baseline variables/observed outcomes

Logistic regression was undertaken to explore if baseline covariates were associated with the probability of data being missing. A dummy variable indicating missing data was created for overall costs and QALYs. Logistic regression was conducted with baseline covariates including gender, year group, multiple deprivation, and number of siblings. A significant association between a baseline covariate and missing data indicates that data are not MCAR.

Dummy variables were also created for costs and QALYs at each time point to explore association between missing data and observed outcomes. Each indicator variable was then regressed on all other costs and QALYS observed in each year (i.e. missing baseline QALYs were regressed on costs and QALYs in each subsequent follow-up). Data were assumed to be MAR in which MI is an appropriate method of analysis to deal with MAR data.

4.4.4.3 Multiple Imputation

MI first arose in the early 1970s to address the problem of survey nonresponse in educational testing (Rubin, 1976).[349] Since then it has gained popularity as a flexible statistical technique for handling missing data. Missing data within CEA poses particular analytical challenges due to complex data structures such as correlated cost and effect endpoints and right skewed cost distributions; MI has been proposed by several authors as an appropriate method to deal with missing data specifically in CEA.[133, 346, 348, 350, 351]

MI consists of three steps:

1. Imputation step: an imputation model is used to predict plausible values for missing observations from the observed values. M imputations are generated allowing uncertainty to be reflected in both the imputation model and missing data (m=number of completed datasets generated). Originally, Rubin[352] recommended five imputations to achieve sufficient, valid inference. Shafer[353] proposes little to no value of using more than 5 to 10 imputation unless the percentage of missing information is unusually high. However, due to advances in computational feasibility a rule of thumb has been proposed that 'the number of imputations should be similar to the percentage of cases that are incomplete.'[335]

2. Completed data analysis step (estimation): each completed data set is analysed separately using the desired analysis method. This is performed after data have been imputed.

3. Pooling step: estimates obtained from each completed dataset are combined using Rubin's rules[352] to generate a single mean estimate of the quantity of interest with its standard error.

MI was employed using chained equations to handle missing cost and QALY data. Costs were imputed at the total cost level and QALYs imputed at the index score level for each time point. Missing data on resource use costs was particularly high so 75 imputations (m=75) were performed as it was computationally feasible to do so in Stata. Predictive mean matching (PMM) was used for continuous, restricted range, and skewed cost and QALY variables. PMM is useful as it avoids predictions that lie outside the bounds of each variable,[335] however it can produce predictions that closely match observed values. The uncertainty in these values is incorporated into the mean costs and QALY estimates using Rubin's rules.

MI was implemented separately by allocation (intervention and control) as recommended as good practice.[346] Covariates included in the imputation model were the same as those used during the estimation step and included: gender, year in school, intervention

allocation, number of siblings, school, trust, and deprivation level. After imputation, three passive variables were created in Stata to allow total costs, total QALYs, and QALY decrements to be classified as imputed variables to be analysed during the estimation stage. The total costs and QALYs variables generated were the sum of the imputed costs and QALYs at each time point. The QALY decrement was defined as the maximum QALYs that could possibly be accrued within the timeframe minus the actual QALYs gained.

4.4.5 Analyses

Regression methods were used to estimate the incremental difference in cost and QALYs while simultaneously adjusting for baseline characteristics which were the same covariates used in the imputation model. Generalised linear models (GLMs) were selected due to their advantage over ordinal least squares and log models in that they model both mean and variance functions on the original scale of cost.[91] They also take into account the typically skewed nature of cost and QALY data.[354] As cost data are typically right-skewed, a right-skewed gamma distribution is appropriate. As QALYs are typically left-skewed, the QALY decrement (described above) was analysed with a gamma distribution. Thus, both costs and QALYs were analysed with a GLM model specifying a gamma family and identity link. Cost and QALY decrements were adjusted for the following covariates: gender, year in school, intervention allocation, number of siblings, school, trust, and deprivation level. Baseline HRQoL was also included to adjust for any imbalance of HRQoL between groups.[355]

Mean costs and QALYs for each group were presented using the method of recycled predictions.[91] Incremental costs and QALYs along with their respective robust standard errors were reported from results of the GLM model. The ICER was estimated and uncertainty surrounding the estimates of cost and effects for RoE and usual classroom activity were investigated through the use of a nonparametric bootstrap of the cost and effect pairs for 1,000 iterations.[356] This approach employs re-sampling techniques to generate a distribution of estimates; in this case the distribution of mean costs and mean outcomes for each group. This provided an estimate of the extent of the uncertainty surrounding the costs and effects individually.

This uncertainty was then presented graphically on the cost-effectiveness plane and a 95% confidence interval (CI) of the bootstrapped ICER was calculated. Results were summarised using a CEAC to reflect the probability of RoE being cost-effective at various WTP thresholds. CEACs are an alternative to confidence intervals around ICERs and were originally developed in the context of a decision problem involving two interventions.[357] They provide a graphical representation of a range of values (WTP thresholds) where the probability of the intervention is at optimal cost-effectiveness.[358] The thresholds varied from £0 to £50,000 per QALY reflecting the range generally accepted to be considered cost-effective by NICE (£20,000 to £30,000/QALY).

4.4.5.1 Clustering within Economic Evaluation

RoE was a cluster randomised controlled trial, so randomisation took place at the cluster (school) level versus at the individual level. It is therefore important take the effects of clustering into account in the economic analysis.[336] Cluster randomisation tends to reduce statistical power and precision[359] because in the case of RoE, individual pupils from the same school will be more similar than pupils from other schools. This non-independence is referred to as the intracluster (or intraclass) correlation coefficient (ICC).[360] The ICC could be thought of as the proportion of variance due to between-cluster variation; or the correlation between members of the same cluster.[336] For sample size calculation in the trial, an ICC of 0.05 was assumed.

Clustering was accounted for by use of a multilevel model (MLM)[337] and the true ICC was estimated. It was anticipated that use of a MLM may not actually be the best fitting model for this analysis (due to only having collected cost at two time points) to which the ICC was examined to determine if clustering had a design effect on the economic outcomes. If the ICC was lower than 0.01, then a more practical approach to reflect clustering would be employed by reporting robust standard errors[361] for the GLM regressions.

A simple MLM of cost was fit, but due to issues with the design of the trial (i.e. resource use was only collected at second and third year follow-up), the data did not fit this type of model as there were only two time points for cost. The ICC was estimated for cost and it was low at 0.0055. The low ICC was assumed to have a minimal design effect for this

outcome so robust standard errors were reported within the GLM regressions to account for clustering in the uncertainty estimates.

4.4.6 Sensitivity Analyses

A series of one-way sensitivity analyses were undertaken to allow for, explore, and assess the uncertainty around the cost-effectiveness results. Thorough exploration through sensitivity analysis strengthens the external validity and generalisability of the results. All sensitivity analyses were derived from the base-case analysis described above and a description of each variation is provided in Table 15. To answer the fourth research question for the main trial (section 4.3.2), outcomes were varied by conducting CEA on the primary outcome, the SDQ. The SDQ was scored using the predictive algorithm converted into Stata syntax available on the SDQinfo website[326] (and in Appendix 7) in StataSE 14 (StataCorp LP, College Station, Texas, USA). This involved assigning a score from 0-2 (0= 'Not True' or no difficulties; 1= 'Somewhat True' or some difficulties; and 2= 'Certainly True' or many difficulties) for each item of the questionnaire and summing the total for each scale. Totals from all scales (excluding prosocial behaviour) were then summed to generate the total difficulties score. As the SDQ comprises two components, the total difficulties score and the prosocial behaviour score; CEA was conducted on both. For the CEAs, differences in effect were measured as the difference in scores from year 3 to baseline by group, see Table 14. There is no established WTP threshold for changes in the SDQ outcome measure, therefore the probability of the SDQ being cost-effective within a £20,000 to £30,000 threshold will not be reported.

Table 14: ICER for cost-effectiveness analyses on SDQ

	Total Cost (mean)	Baseline Score (mean)	Score at final follow-up (mean)	Difference in Score	ICER
RoE	a	c	e	(e-c)	
Control	b	d	f	(f-d)	(a-b)/((e-c)-(f-d))
Difference	(a-b)			(e-c)–(f-d)	

There is currently a mapping algorithm available to map SDQ scores to CHU9D utilities.[329] In order to explore the validity of the mapping algorithm, a sensitivity analysis CUA was

performed using utility scores mapped from the SDQ. The final outcomes related sensitivity analysis involved a CUA using the values from the adult values tariff to estimate utilities.

The cost of the intervention was a main cost driver so annuitisation assumptions around the useful life of the intervention were varied to account for no annuitisation and annuitisation over a shorter useful life of three years versus five in separate sensitivity analyses. The discount rate was also varied to reflect a more traditional discount rate of 3.5% versus the 1.5% public health discount rate. Missing resource use and HRQoL data from the trial was particularly high, thus sensitivity analysis was conducted to explore the uncertainty surrounding the MAR assumption and use of MI. An available-case analysis was conducted assuming data were MCAR to assess the impact MI had on the incremental costs and QALYs. A limitation of available-case analysis is that different samples of costs and QALYs may be used which can lead to non-comparability and affect the covariance structure, therefore the results of this sensitivity analysis should be interpreted with caution.

Finally, to further explore whose values in health should be considered,[288] the base-case assumption using the adolescent values tariff to value child health utility was switched to the adult values and all sensitivity analyses were re-run with the updated base-case assumptions. This second set of sensitivity analyses are referred to as adult values (AV). SA4 which reported adult values is the base-case in this set of results (AV0). A simple 'scaling up' exercise was performed to demonstrate the prevention paradox. The effects from the trial were scaled up to represent the total QALY gain that might be expected over the population if RoE were rolled out throughout Northern Ireland. The incremental QALY gain from the base case was multiplied the total number of children aged 5-9 in Northern Ireland. This estimate was taken from census estimates from 2014.[362]

Table 15: List of sensitivity analyses

Sensitivity Analysis (SA)	Element	Description of Variation
0	Base-case	Multivariate analysis of cost and QALY Public Sector perspective, 1.5% discount rate, child health utility adolescent values, MAR assumption, and multiple imputation
1		SDQ Total Difficulties (CEA)
2		SDQ Prosocial Behaviour (CEA)
3		CHU9D mapped from SDQ
4		CHU9D estimated from adult values tariff (UK)
5	Costs	Training and material costs not annuitised
6		Training and material costs annuitised over 3 years
7	Discount Rate	Use of more traditional 3.5% discount rate for costs and outcomes
8	Missing Data	Available case analysis assuming MCAR
Adult Values (AV)		
0	Base-case	Multivariate analysis of cost and QALY Public Sector perspective, 1.5% discount rate, child health utility adult values, MAR assumption, and multiple imputation
1	Outcomes	SDQ Total Difficulties (CEA)
2		SDQ Prosocial Behaviour (CEA)
3		CHU9D mapped from SDQ
4		CHU9D estimated from adolescent values tariff (UK)
5	Costs	Training and material costs not annuitised
6		Training and material costs annuitised over 3 years
7	Discount Rate	Use of more traditional 3.5% discount rate for costs and outcomes
8	Missing Data	Available case analysis assuming MCAR

4.4.7 Summary

Currently there is a paucity of evidence in the literature of high quality school-based economic evaluations that include outcomes designed specifically for children and incorporate their preferences. This methods chapter described the first economic evaluation of school-based PHI, RoE, to address the evidence gaps in the literature. The chapter started by describing the differences between economic evaluation of child and adult interventions, and how there is a need to develop outcome measures specifically for children. The CHU9D is the only preference-based HRQoL measure that was developed specifically for children, which was also valued by adolescents (with the elicitation of younger children's values currently ongoing). Another child specific measure, the SDQ, was also described, in detail as it was used in CEA.

The next section (4.3) provided background and contextual information to the main cluster randomised controlled trial of RoE. A review of the existing evidence of RoE's effectiveness found that only one evaluation was a cluster RCT design, with follow-up at three years. This evaluation took place in a different contextual setting to Northern Ireland and none of the existing evidence included an economic evaluation. The main trial aims and research questions were stated and data collection detailed. The final section described in detail, the methods of the economic evaluation of the RoE programmed. The section started with an overview, followed by detailed descriptions of the costs, outcomes, how missing data was handled, analyses, and sensitivity analyses performed. The next chapter reports the results of this novel economic evaluation.

5 RoE Main Trial Results: a Case Study

5.1 Introduction

The previous chapter outlined economic evaluation in child health including paediatric outcome measurement, the RoE trail, and the methods for conducting the economic evaluation. This chapter reports the results of the economic evaluation of RoE. The next section (5.2) provides descriptive results from the main trial which has been replicated from the original end-of-study report.[55] Sections 5.3 through 5.6 report the results under the same headings which were described in the methods; costs, outcomes, missing data, and cost-effectiveness which describes the results of the base-case analysis and sensitivity analyses. The discussion of these results, the limitations of this study, and the conclusion follows.

5.2 RoE Main Trial Descriptive Results

This section provides a description of the data collected from the main trial. This section was originally described elsewhere[55] and replicated here for clarity. Data collection is presented in Figure 13, which is a flow diagram of teacher, pupil, and parent responses through the trial. Seven schools withdrew before the start of the trial. Of the 1,182 pupils tested at pre-test, 902 remained in the study at the final third-year follow-up (76.3% retained). Fewer parents returned data about their child; 686 returned data at pre-test (58.0% of the sample of children tested) which reduced to 373 at the end of the study (31.6% of the sample of children tested).

In total 1,278 pupils aged between eight and nine years were recruited into the study, n=583 in the control group and n=695 in the intervention group. Table 16 describes the sample characteristics at baseline, showing a breakdown by gender, Health and Social Care Trust, geographic area (urban vs. rural) and primary school type (controlled, Catholic maintained, integrated, or other).

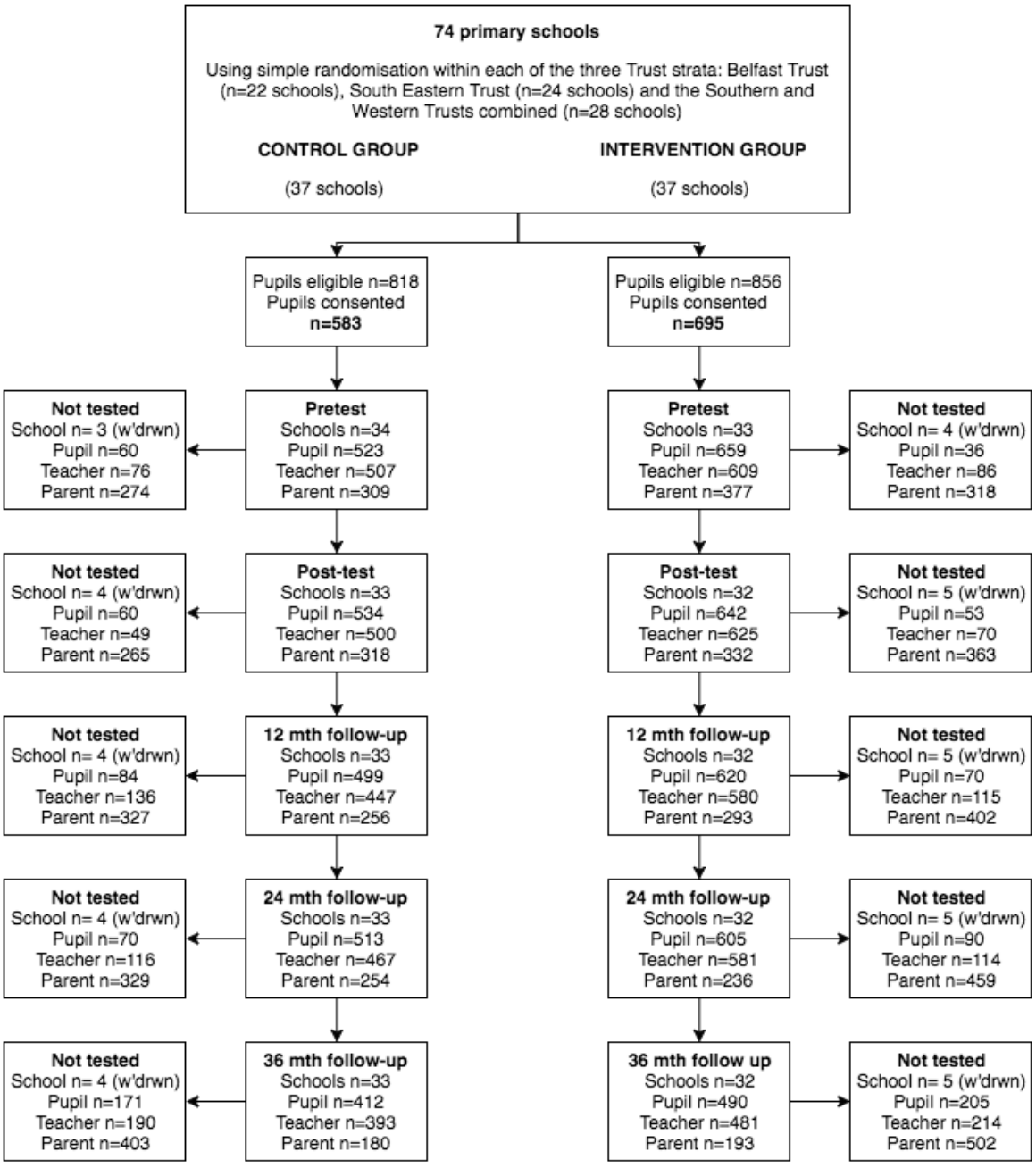


Figure 13: Flow diagram of recruitment and testing of children

Table 16: Baseline sample characteristics

	Control N (%)	Intervention N (%)	Total N (%)
Gender			
Male	310 (24.3)	347 (27.2)	657 (51.4)
Female	273 (21.4)	348 (27.2)	621 (48.6)
Class			
P4	43 (3.4)	38 (2.9)	81 (6.3)
P5	528 (41.3)	611 (47.8)	1139 (89.1)
P6	12 (.94)	46 (3.6)	58 (4.5)
Trust			
Belfast	145 (11.4)	201 (15.7)	346 (27.1)
South Eastern	150 (11.7)	222 (17.4)	372 (29.1)
Southern	181 (14.2)	171 (13.4)	352 (27.5)
Western	107 (8.4)	101 (7.9)	208 (16.3)
Area			
Urban	330 (25.8)	363 (28.4)	693 (54.2)
Rural	253 (19.8)	332 (26.0)	585 (45.8)
School type			
Controlled	189 (14.8)	242 (18.9)	431 (33.7)
Catholic Maintained	286 (22.4)	360 (28.2)	646 (50.6)
Integrated	85 (6.7)	77 (6.0)	162 (12.7)
Other	23 (1.7)	16 (1.3)	39 (3.1)
Total	583 (45.6)	695 (54.4)	1278 (100)

The main trial analysis found initial positive effects on prosocial (effect size, g=+0.20, p=.045) and difficult behaviour (g=-.16, p=.06) at the post-test time point. These initial positive effects disappeared at all subsequent follow-ups. For all other secondary outcomes, there was no statistically significant difference between scores in the intervention and control at any subsequent follow-up point. The next section reports the resource use and costs of the RoE main trial economic evaluation.

5.3 Costs

All costs reported in this section are subject to the base-case assumptions i.e. they were discounted by a rate of 1.5% and fixed costs were annuitized over 5 years.

5.3.1 RoE intervention costs

A summary of intervention costs can be found in Table 17. A detailed further breakdown of costs including unit costs is available in

Table 18. The total instructor time cost was £37,200. Per instructor: training materials were £1,030, materials for delivering RoE were £457, and fees were £172.

Table 17: Summarised cost of the Roots of Empathy Intervention

Total Costs Annuitized 5 years 1.5%		**Per School**
Key Point People	£51,419.28	£4,056.54
Admin Support	£25,793.46	
Instructor Time	£37,231.17	
Instructor Training Materials	£7,092.94	**Per Pupil**
Instructor Materials	£3,152.42	
Instructor Fee	£5,653.83	£175.22
Other Costs	£3,522.59	
Total Cost	£133,865.69	

Table 18: Detailed cost breakdown of intervention costs

Cost Item	**Unit Cost (2014)**	**Quantity**	**Total***
Number of pupils	-	764	764
Salaries			
Key point person	£35,600	4	-
Administrative support FTE	£12,900	4	-
RoE instructor	£29,800	33	-
RoE activities			
Key point person training	£18.26	112	£2,045
Key point person time spent on RoE	£18.26	2704	£49,374
Administrative support	£6.61	3900	£25,793
Instructor training	£15.29	30 hours	£15,137
Instructor preparation time	£15.29	varied	£11,982
Instructor delivery time	£15.29	varied	£10,112
Instructor fees	£171.33	33	£5,654
Materials			
Instructor training	£1,027.97	33	£33,923
RoE	£456.88	33	£15,077
Other costs			
Programme support	£5,710.94	1	£5,711
Materials shipping	£2,569.92	1	£2,570
Training and mentoring expenses	£3,426.57	1	£3,427

Mentoring	£5,139.85	1	£5,140

*** Exact figures not shown, totals were rounded**

5.3.2 Resource Use

Overall, resource use did not differ significantly between groups. One item did demonstrate a significant difference; average dentist costs in the control was £24 more than the intervention group. Some resource use items had a large amount of missing data and were subsequently dropped from the analysis. These dropped resource use items included days off work due to a child being home from school, other resource use, and medications. More information on missing data is given in section 5.5. Mean resource use before MI is given in

Table 19.

MI and regression of the mean total cost (including intervention and resource use costs) for RoE was £1,190 and the mean cost for the control group which was £1,030 (

Table 24). The incremental cost was £160 (95% CI: £14-£307) significantly higher for RoE (p-value = 0.032). The additional cost of the intervention is the main cost driver in this incremental cost.

Table 19: Mean resource use costs by group and differences between groups

	RoE			Control			Difference between groups*		
Resource Use Item	Mean cost (£)	Std. Err.	95% CI	Mean cost (£)	Std. Err.	95% CI	Mean cost (£)	Std. Err.	95% CI
GP	£ 49.83	4.09	41.80 to 57.87	£ 56.12	4.43	47.42 to 64.81	£ 6.28	6.03	-5.54 to 18.10
School Nurse	£ 13.41	1.91	9.65 to 17.17	£ 18.94	3.52	12.03 to 25.85	£ 5.54	3.87	-2.05 to 13.12
A&E Visit	£ 15.54	1.71	12.17 to 18.90	£ 18.31	2.07	14.25 to 22.37	£ 2.77	2.66	-2.45 to 8.00
Social Worker	£ 3.48	2.94	-2.29 to 9.25	£ 1.27	0.58	0.13 to 2.40	-£ 2.21	3.20	-8.50 to 4.08
Speech therapist	£ 2.52	1.73	-0.88 to 5.91	£ 1.68	0.76	0.19 to 3.18	-£ 0.83	1.99	-4.74 to 3.07
Occupational therapist	£ 1.01	0.47	0.08 to 1.94	£ 0.97	0.51	-0.04 to 1.98	-£ 0.04	0.70	-1.41 to 1.33
Physiotherapist	£ 8.44	5.03	-1.44 to 18.31	£ 12.80	4.69	3.60 to 22.00	£ 4.37	6.94	-9.26 to 17.99
Educational Psychologist	£ 2.93	0.84	1.27 to 4.58	£ 3.31	1.06	1.22 to 5.40	£ 0.38	1.34	-2.25 to 3.01
Psychiatrist	£ 3.05	2.17	-1.21 to 7.32	£ 10.97	9.53	-7.75 to 29.68	£ 7.92	9.17	-10.07 to 25.90
Counselling/therapy	£ 10.13	2.88	4.46 to 15.79	£ 16.77	6.55	3.91 to 29.63	£ 6.65	6.84	-6.76 to 20.06
Dentist	£ 125.94	7.35	111.50 to 140.37	£ 149.99	8.69	132.92 to 167.06	£ 24.05	11.30	1.86 to 46.23
Optician	£ 14.26	1.20	11.89 to 16.62	£ 16.02	1.34	13.39 to 18.67	£ 1.77	1.80	-1.75 to 5.29
Police	£ 3.87	1.52	0.89 to 6.86	£ 2.79	1.47	-0.10 to 5.69	-£ 1.08	2.14	-5.27 to 3.11
Hospital stay	£ 17.94	8.91	0.45 to 35.44	£ 17.92	4.70	8.70 to 27.15	-£ 0.03	10.53	-20.68 to 20.63
Hospital outpatient	£ 48.09	8.42	31.56 to 64.63	£ 56.51	10.20	36.48 to 76.53	£ 8.41	13.11	-17.31 to 34.14
***Two-sample t-test with equal variances**									

5.4 Outcomes

The mean QALY gain in the RoE group was 2.97 versus 2.95 for the control. The incremental QALY gain of 0.0146 (95% CI: -0.023 to 0.0522) was not statistically significant (p-value = 0.448). The results of the base-case CUA as well as sensitivity analysis are reported in

Table 24.

5.5 Missing Data

38% of resource use questionnaires were returned for the second year follow-up and 29% were returned at the final third year follow-up, see Table 20. Variables that were dropped due to having over 80% of their values missing were other resource use (97%), medications (86%), and days off work due to a child being home from school (88%).The dropped variables were all self-report free-form text variables (questions 1, 2, and 3 in Figure 12: Resource use questionnaire). See Table 20 for the descriptive missing data analysis which details percentage of missing values by group and in total, range, mean, and standard deviation of the observed data.

Missing data followed a non-monotonic pattern (see Figure 14) because cost or QALY data may be missing for an individual in one follow-up, but then they return in subsequent follow-ups. The grey shading represents observed data, while the black represents missing data for one or more individuals along the horizontal axis. The cost (a) and QALY (b) variables at each time point lie along the vertical axis. The chequered pattern demonstrates how data for an individual may be missing at one time point, but then observed at a subsequent time point.

Table 20: Variable descriptions and missing data percentages

Variable	Description	Missing values, %			Range	Mean	SD
	Total (n=1,254) RoE (n=672) Control (n=582)	Total	RoE	Control			
	Baseline variables						
Gender	Male or Female	0%	0%	0%	0,1	51.45% Male	
YearGroup	Year in School at trial entry	0%	0%	0%	4,5,6	89% P5	
MD-rank	Northern Ireland Multiple Deprivation Measure	2%	3%	0%	1 to 889	414.13	245.9
Siblings_PT0	Number of siblings at baseline	1%	1%	0%	0 to 7	1.01	1.26

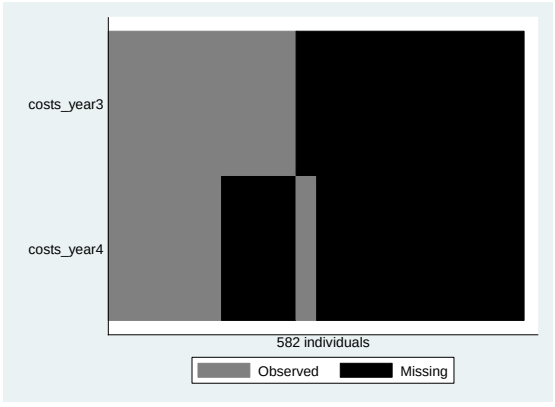
Table 20 continued: Outcome variable descriptions and missing data percentages

	Outcome variables for health related quality of life						
utility0	CHU9D at pre-test	13%	10%	16%	0.3261 to 1	0.84	0.12
utility1	CHU9D at post-test	12%	11%	13%	0.3261 to 1	0.85	0.11
utility2	CHU9D at 1 year follow-up	14%	12%	16%	0.4582 to 1	0.84	0.1
utility3	CHU9D at 2 year follow-up	14%	15%	13%	0.3261 to 1	0.85	0.1
utility4	CHU9D at 3 year follow-up	31%	31%	31%	0.3929 to 1	0.87	0.1
	Outcomes for cost-effectiveness						
total_QALYs	Total QALYs over 3.75 years[a]	45%	43%	48%	1.70 to 3.61	3.09	0.26
total_costs	Total costs over 3.75 years[a]	76%	78%	75%	77 to 10580	£899.04	£841.93

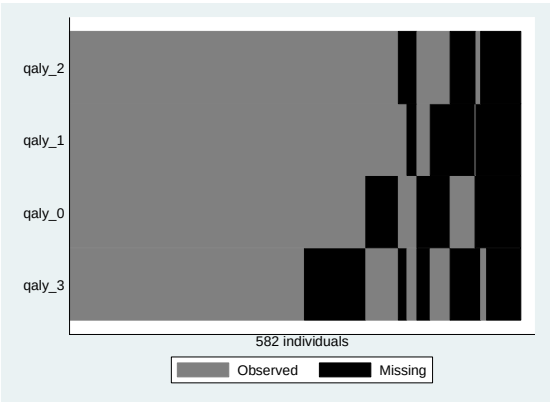
[a*] Total QALY and costs refers to the sum of QALYs and costs over the 3.75 year trial period discounted at a 1.5% annual rate.

Table 20 continued: Resource use variables for cost

Variable	Missing values, %			Range	Mean	SD
	Total	RoE	Control			
Intervention cost	0%	0%	0%	£175.22	£175.22	
GP_3	62%	66%	57%	0 to 706	£ 96.07	102.56
School Nurse_3	62%	66%	57%	0 to 1209	£9.65	£64.75
A&E_3	62%	66%	57%	0 to 345	£29.26	£ 53.22
Social Worker_3	62%	66%	57%	0 to 1416	£4.43	£65.43
Speech therapist_3	62%	66%	57%	0 to 1025	£4.89	£52.51
Occupational Therapist_3	62%	66%	57%	0 to 261	£2.44	£18.86
Physiotherapist_3	62%	66%	57%	0 to 1555	£12.89	£107.42
Educational psychologist_3	62%	66%	57%	0 to 393	£ 6.52	£33.31
Psychiatrist_3	62%	66%	57%	0 to 5252	£10.74	£237.53
Counselling/therapy_3	62%	66%	57%	0 to 2332	£20.35	£137.51
Dentist_3	62%	66%	57%	0 to 1247	£253.53	£138.02
Optician_3	62%	66%	57%	0 to 202	£27.00	£32.96
Police_3	62%	66%	57%	0 to 623	£4.47	£46.62
Hospital Stay_3	62%	66%	57%	0 to 1564	£21.12	£127.31
Hospital Outpatient visit_3	62%	66%	57%	0 to 2902	£88.30	£277.06
GP_4	71%	72%	70%	0 to 652	£44.09	£74.42
School Nurse_4	71%	72%	70%	0 to 595	£39.07	£67.80
Education Welfare Officer_4	71%	72%	70%	0 to 102	£0.48	£5.78
A&E_4	71%	72%	70%	0 to 204	£15.83	£35.80
Social Worker_4	71%	72%	70%	0 to 465	£2.10	£25.48
Speech therapist_4	71%	72%	70%	0 to 84	£0.46	£6.19
Physiotherapist_4	71%	72%	70%	0 to 3064	£16.77	£175.15
Educational psychologist_4	71%	72%	70%	0 to 155	£1.48	£10.61
Psychiatrist_4	71%	72%	70%	0 to 1293	£7.58	£84.40
Counselling/therapy_4	71%	72%	70%	0 to 919	£15.88	£88.97
Dentist_4	71%	72%	70%	0 to 614	£110.88	£75.35
Optician_4	71%	72%	70%	0 to 79	£13.46	£14.06
Police_4†	71%	72%	70%	0 to 307	£4.98	£38.87
Hospital Stay_4	71%	72%	70%	0 to 5241	£30.00	£289.11
Hospital Outpatient visit_4	71%	72%	70%	0 to 1787	£51.83	£191.40



a



b

Figure 14: Pattern of missing data in a. costs and b. QALYs. Black shading represents missing data grey represents observed data.

5.5.1 Logistic Regression

Deprivation level and number of siblings at baseline were found to be significant predictors of missing cost (Table 21). Gender, age, deprivation, and number of siblings were all significant predictors of missing QALYs which can rule out the MCAR assumption (Table 22). For regressions that explored the association between missing data and observed outcomes, at least one covariate produced statistically significant results (Table 23) indicating the data are unlikely to be MCAR and thus assumed to be MAR. As the results from the missing data patterns and logistic regression both indicated data to be MAR, MI was performed as a method to address the missing data in the dataset.

Table 21: Association between missing cost and baseline variables

Logistic regression — Number of obs = 1,091; LR chi2(7) = 114.62; Prob > chi2 = 0.0000; Log likelihood = -557.5685; Pseudo R2 = 0.0932

cost_m	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Gender	.0400518	.148199	0.27	0.787	-.2504129	.3305166
YearGroup	-.0825119	.2192114	-0.38	0.707	-.5121582	.3471345
MD_rank	.0026005	.0003382	7.69	0.000	.0019375	.0032634
siblings_pt4	.3372088	.0556304	6.06	0.000	.2281753	.4462423
Group	-.197672	.1590615	-1.24	0.214	-.5094268	.1140827
blqaly	.4837079	.6466744	0.75	0.454	-.7837506	1.751166
School	-.0025925	.0041737	-0.62	0.535	-.0107729	.0055879
_cons	-2.66438	1.262331	-2.11	0.035	-5.138503	-.1902574

Table 22: Association between missing QALY and baseline variables

Logistic regression — Number of obs = 1,091; LR chi2(7) = 29.63; Prob > chi2 = 0.0001; Log likelihood = -703.74733; Pseudo R2 = 0.0206

qaly_m	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Gender	.3739303	.1282619	2.92	0.004	.1225416	.625319
YearGroup	-.4455213	.2010916	-2.22	0.027	-.8396536	-.051389
MD_rank	.0007216	.0002662	2.71	0.007	.0001999	.0012434
siblings_pt4	.109397	.0496154	2.20	0.027	.0121525	.2066414
Group	.1136906	.1353783	0.84	0.401	-.1516461	.3790272
blqaly	.3074137	.5512154	0.56	0.577	-.7729485	1.387776
School	-.0009335	.0035208	-0.27	0.791	-.0078342	.0059672
_cons	1.832976	1.121123	1.63	0.102	-.3643847	4.030337

Table 23: Example of regression output for association between missing and observed values

```
Logistic regression                             Number of obs     =      1,254
                                                LR chi2(5)        =     476.40
                                                Prob > chi2       =     0.0000
Log likelihood = -592.79415                     Pseudo R2         =     0.2866
```

cost3_m	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
cost4_m	2.872562	.1644156	17.47	0.000	2.550313	3.19481
qaly0_m	.1634797	.2281219	0.72	0.474	-.283631	.6105904
qaly1_m	.1453802	.3238573	0.45	0.654	-.4893684	.7801289
qaly2_m	.6772399	.3228683	2.10	0.036	.0444296	1.31005
qaly3_m	.3820004	.1791242	2.13	0.033	.0309235	.7330773
_cons	-2.490252	.254197	-9.80	0.000	-2.988469	-1.992035

5.6 Cost-effectiveness

The ICER was £11,000 per QALY gained (CI: -£95,500 to £147,000), see

Table 24, SA0. This is below that standard £20,000 to £30,000 threshold that is generally accepted as cost-effective in the UK. Uncertainty around this estimate was explored through bootstrapping. The CE plane is presented in Figure 15. The majority of the bootstrap estimates lie within the NE quadrant demonstrating that RoE is a more costly, but more effective intervention. However, because there are a few bootstrap estimates in the NW quadrant, there is some uncertainty about whether RoE is more effective than usual classroom activities. This uncertainty is also demonstrated in the non-significant incremental mean QALY gain of 0.0146 (CI: -0.0230 to 0.0522) and overall ICER uncertainty of £11,000 per QALY gained (CI: -£95,500 to £147,000 reported in

Table 24. There is little uncertainty surrounding the difference in costs, as demonstrated on the CE plane where all points lie above the x-axis, indicating RoE is more costly than usual classroom activities. The CEAC is presented in Figure 16, which demonstrates that at a cost-effectiveness threshold of £20,000, RoE had an 84.6% probability of being cost-effective. This probability rises to 89.9% at a threshold of £30,000.

Table 24: Cost-effectiveness results (adolescent values)

	Mean Costs				Mean Effects						
Analysis	**RoE**	**Control**	**Incremental cost (95% CI)**	**Robust Std. Err.[a]**	**RoE**	**Control**	**Incremental Effects (95% CI)**	**Robust Std. Err.[a]**	**ICER (£ per QALY)**	**95% CI of bootstrapped ICER**	**Probability of being cost-effective[b] (%)**
SA[c]0	£1,190	£1,030	£160 (£14 to £307)	74.6	2.97	2.95	0.0146 (-0.0230 to 0.0522)	0.0192	£11,000	-£95,500 to £147,000	84.6 (89.9)
SA1	£1,170	£1,060	£107 (-£38 to £252)	73.7	1.17	0.627	0.541 (0.0718 to 1.01)	0.239	£197[d]	£77 to £471	[e]
SA2	£1,190	£1,040	£154 (£12 to £297)	72.4	-0.547	-0.574	0.0274 (-0.349 to 0.403)	0.192	£5,630[f]	-£23,400 to £29,100	
SA3	£1,180	£1,040	£143 (-£21 to £306)	82.9	3.04	3.02	0.0150 (-0.00398 to 0.0339)	0.0967	£9,540	£4,160 to £30,300	93.1 (97.4)
SA4	£1,180	£1,030	£153 (£14 to £292)	70.9	3.09	3.07	0.0160 (-0.0143 to 0.0462)	0.0154	£9,570	-£87,800 to £107,000	83.1 (90.1)
SA5	£1,260	£1,030	£230 (£83 to £380)	74.5	2.97	2.95	0.0146 (-0.0230 to 0.0522)	0.0192	£15,800	-£137,000 to £202,000	76.4 (85.4)
SA6	£1,200	£1,030	£172 (£26 to £319)	74.6	2.97	2.95	0.0146 (-0.0230 to 0.0522)	0.0192	£11,800	-£103,000 to £156,000	82.6 (89.5)
SA7	£1,130	£968	£161 (£22 to £301)	70.8	2.85	2.83	0.0134 (-0.0229 to 0.0497)	0.0185	£12,100	-£103,000 to £137,000	83 (89.4)
SA8	£1,130	£895	£236 (£54 to £417)	92.6	2.96	2.96	0.00587 (-0.0429 to 0.0546)	0.0249	£40,200	-£218,000 to £157,000	78.6 (86.7)

[a] Adjusted for 66 clusters in school

[b] At £20,000 per QALY (£30,000 per QALY)

[c] Sensitivity Analysis (SA) (see Table 15 for description), results reported to 3 significant figures

[d] ICER per unit decrease in SDQ total difficulties score

[e] No cost-effectiveness threshold for change in SDQ defined

[f] ICER per unit increase in SDQ prosocial behaviour score

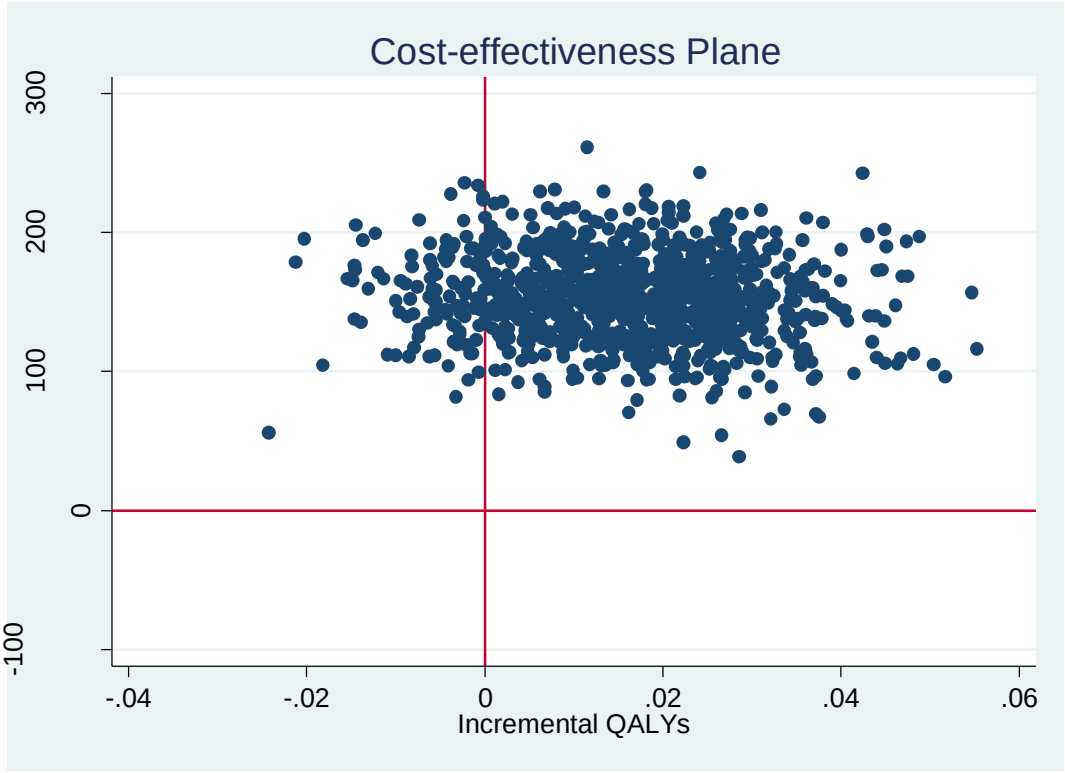


Figure 15: Cost-effectiveness plane representing 1000 bootstrapped cost and QALY pairs

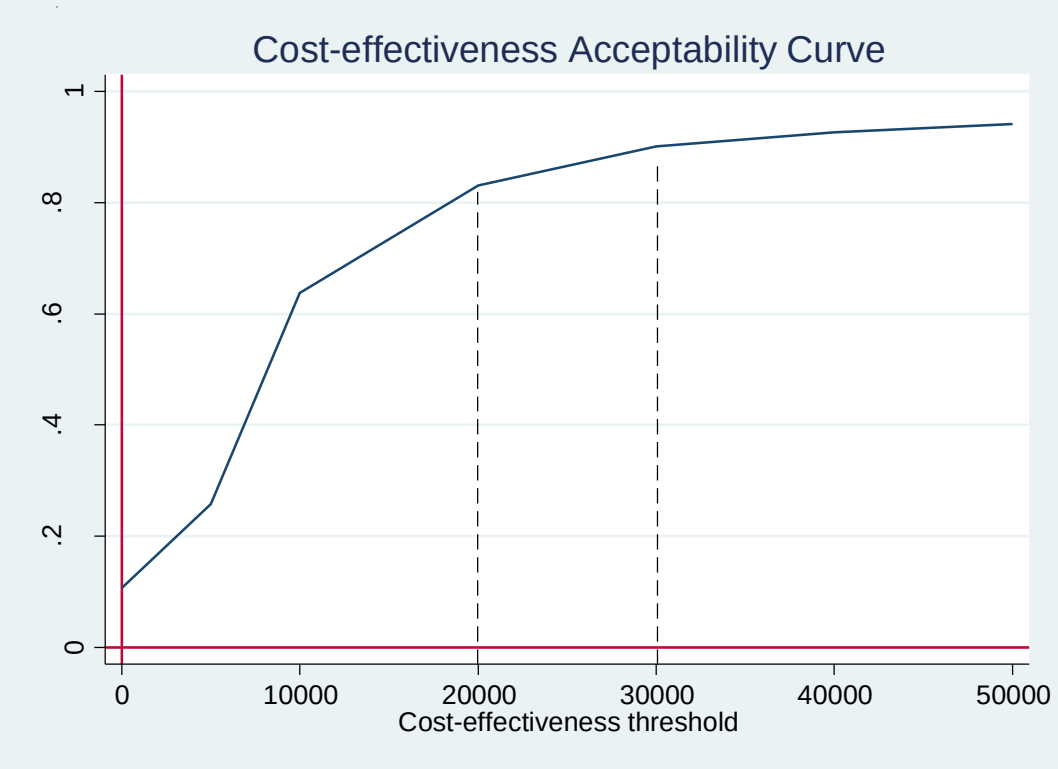


Figure 16: CEAC showing probability of RoE being cost-effective compared to usual classroom activities. The dashed lines indicate the probability of RoE being cost-effective at the defined threshold.

5.6.1 Sensitivity Analysis

The planned sensitivity analyses described in Table 15 are replicated here for clarity.

Sensitivity Analysis (SA)	Element	Description of Variation
0	Base-case	Multivariate analysis of cost and QALY Public Sector perspective, 1.5% discount rate, child health utility adolescent values, MAR assumption, and multiple imputation
1		SDQ Total Difficulties (CEA)
2		SDQ Prosocial Behaviour (CEA)
3		CHU9D mapped from SDQ
4		CHU9D estimated from adult values tariff (UK)
5	Costs	Training and material costs not annuitised
6		Training and material costs annuitised over 3 years
7	Discount Rate	Use of more traditional 3.5% discount rate for costs and outcomes
8	Missing Data	Available case analysis assuming MCAR
Adult Values (AV)		
0	Base-case	Multivariate analysis of cost and QALY Public Sector perspective, 1.5% discount rate, child health utility adult values, MAR assumption, and multiple imputation
1	Outcomes	SDQ Total Difficulties (CEA)
2		SDQ Prosocial Behaviour (CEA)
3		CHU9D mapped from SDQ
4		CHU9D estimated from adolescent values tariff (UK)
5	Costs	Training and material costs not annuitised
6		Training and material costs annuitised over 3 years
7	Discount Rate	Use of more traditional 3.5% discount rate for costs and outcomes
8	Missing Data	Available case analysis assuming MCAR

SA refers to 'Sensitivity Analysis' which used adolescent values and AV refers to 'Adult Values.' The CEA of the main trial outcome measure, the SDQ total difficulties score (SA1), resulted in an ICER of £197 per one-unit decrease in the total difficulties score (CI: £77 to £471). In this sensitivity analysis, the cost difference of £107 between arms was not significantly different (CI: -£38 to £252). For the incremental effects, the difference between the arms was a significant decrease in total difficulties score of 0.541 (CI: 0.0718 to 1.01). SA2 was a CEA of the SDQ prosocial behaviour score. The ICER was £5,630 per unit increase in SDQ prosocial behaviour score (CI: -£23,400 to £29,100). For SA2, costs were significantly higher in the RoE group, but the difference in effect was not significant. SA3 used a mapping algorithm to map from the SDQ to the CHU9. Neither the incremental costs or effects were statistically significantly different resulting in an ICER of £9,540 per QALY gained (CI: £4,160 to £30,300). The final sensitivity analysis that varied outcomes was SA4, which used adult values to estimate child health utilities; this sensitivity analysis resulted in an ICER of £9,570 per QALY gained (CI: -£87,800 to £107,000).

SA5 and SA6 varied how costs were annuitised. SA5 did not annuitize any costs which resulted in an ICER of £15,800 per QALY gained (CI: -£137,000 to £202,000). SA6 annuitised training and material costs over a shorter three-year period compared to the five years in the base-case. The resulting ICER was £11,800 per QALY gained (CI: -£103,000 to £156,000). SA7 varied the discount rate to a more traditional 3.5% and the available-case analysis (SA8) explored the uncertainty around the MAR assumption by only analysing the available data and not performing MI. The ICER for SA7 was £12,100 per QALY gained (CI: -£103,000 to £137,000) and SA8 was £40,200 per QALY (-£218,000 to £157,000). All results of the adolescent values are reported in

Table 24. All resulting ICERs fell within the considered 'cost-effective' range except SA8, the available-case analysis. In all analyses RoE had significantly higher costs except in SA1 and

SA3, likewise there was a lack of statistically significant difference in effects in all analyses except SA1.

Table 26 reports the results of the sensitivity analyses re-run with adult values. AV0, the base-case using adult values to estimate child health utility, is same as SA4, ICER £9,570 per QALY gained (CI:-£87,800 to £107,000). Additionally, the results of AV1 and AV2 have not changed from SA1 or SA2 because the outcome measure of effect was the SDQ, ICER £197 per unit decrease in total difficulties score (CI: £77 to £471 and £5,630 per unit increase in prosocial behaviour score (CI: -£23,400 to £29,100) respectively. Neither the incremental costs nor effects were significantly different in AV3, which mapped SDQ scores to utility values resulting in an ICER of £9,700 per QALY gained (CI: £4,210 to £30,800). AV4 is the same as SA0, ICER £11,000 per QALY gained (CI: -£95,500 to £147,000). AV5 did not annuitise any costs resulting in an ICER of £13,900 per QALY gained (CI: -£125,000 to £151,000). AV6 annuitised costs over 3 years resulting in an ICER of £10,300 per QALY gained (CI: -£93,700 to £114,000). AV7, where costs and outcomes were discounted at 3.5% had an ICER of £9,660 per QALY gained (CI:-£94,500 to £113,000). Finally, AV8 the available-case analysis, had the highest ICER estimate of £19,600 (CI: -£149,000 to £145,000). Cost-effectiveness planes and CEACs for all sensitivity analyses are detailed in Appendix 8.

Finally, to demonstrate the prevention paradox, the total potential QALY gains that could be expected if RoE reached all children in Northern Ireland aged 5-9 are presented in Table 25. The results of the simple scaling up exercise indicate that nearly 1,800 additional QALYs could be gained if RoE reached this entire population of children in Northern Ireland.

Table 25: QALY gain over population demonstrating the prevention paradox

		Source
Incremental QALY estimate from trial	0.0146	RoE economic evaluation
Population estimate of children 5-9 in Northern Ireland 2014	121850	Office for National Statistics published 29 October 2015
QALY estimate for population	1779.01	RoE economic evaluation

Table 26: Cost-effectiveness results (adult values)

	Mean Costs				Mean Effects						
Analysis	**RoE**	**Control**	**Incremental cost (95% CI)**	**Robust Std. Err.[a]**	**RoE**	**Control**	**Incremental Effects (95% CI)**	**Robust Std. Err.[a]**	**ICER (£ per QALY)**	**95% CI of bootstrapped ICER**	**Probability of being cost-effective[b] (%)**
AV[c]0	£1,180	£1,030	£153 (£14 to £292)	70.9	3.09	3.07	0.0160 (-0.0143 to 0.0462)	0.0154	£9,570	-£87,800 to £107,000	83.1 (90.1)
AV1	£1,170	£1,060	£107 (-£38 to £252)	73.7	1.17	0.627	0.541 (0.0718 to 1.01)	0.239	£197[d]	£77 to £471	[e]
AV2	£1,190	£1,040	£154 (£12 to £297)	72.4	-0.547	-0.574	0.0274 (-0.349 to 0.403)	0.192	£5,630[f]	-£23,400 to £29,100	
AV3	£1,180	£1,040	£143 (-£20 to £306)	82.9	3.04	3.02	0.0147 (-0.00404 to 0.0335)	0.00957	£9,700	£4,210 to £30,800	92.7 (97.4)
AV4	£1,190	£1,030	£160 (£14 to £307)	74.6	2.97	2.95	0.0146 (-0.0230 to 0.0522)	0.0192	£11,000	-£95,500 to £147,000	84.6 (89.9)
AV5	£1,250	£1,030	£222 (£83 to £362)	70.9	3.09	3.07	0.0160 (-0.0143 to 0.0462)	0.0154	£13,900	-£125,000 to £151,000	75.2 (84.2)
AV6	£1,190	£1,030	£165 (£25 to £304)	70.9	3.09	3.07	0.0160 (-0.0143 to 0.0462)	0.0154	£10,300	-£93,700 to £114,000	82.1 (88.6)
AV7	£1,120	£965	£154 (£17 to £290)	69.5	2.96	2.95	0.0160 (-0.0128 to 0.0446)	0.0146	£9,660	-£94,500 to £113,000	82.5 (89.8)
AV8	£1,130	£894	£238 (£58 to £419)	92.2	3.09	3.08	0.0121 (-0.0271 to 0.0514)	0.0200	£19,600	-£149,000 to £145,000	77.3 (86.3)

[a] Adjusted for 66 clusters in school

[b] At £20,000 per QALY (£30,000 per QALY)

[c] Sensitivity Analysis (SA) (see Table 15 for description), results reported to 3 significant figures

[d] ICER per unit decrease in SDQ total difficulties score

[e] No cost-effectiveness threshold for change in SDQ defined

[f] ICER per unit increase in SDQ prosocial behaviour score

5.7 Discussion

Over the trial period, the base-case analysis indicated that the RoE intervention incurred a mean additional cost of £160 (95% CI: £14 to £307) per pupil. Utility, as measured by the CHU9D and combined with duration to calculate QALYs showed no significant QALY difference between groups (incremental effect 0.0146 (CI: -0.0230 to 0.0522) over the three year follow-up post-intervention completion. Although the use of directly measured child health utility in a CUA framework is infrequent, QALY gains in other areas of child health research are often small and insignificant.[363, 364] However, economic evaluation methods still use such estimates to explore the probability of cost-effectiveness when combined with the cost of achieving these gains. When applied across a population even small QALY gains can be highly cost-effective. The simple scaling up exercise presented in Table 25 demonstrated small QALY gains to the individual could potentially result in nearly 1,800 QALYs gained over the population of children in Northern Ireland. A recent study looking at a family-based childhood obesity treatment used the EQ-5D youth version to measure QALYs.[363] They reported a non-significant QALY gain of 0.03 (95%CI: -0.04 to 0.10). Another recent study for an asthma intervention in children used adult EQ-5D QALY estimates.[364] They found a difference in mean QALYs of -0.00017 (95% CI: -0.00051 to 0.00018). These non-significant results are reported here to demonstrate that non-significant QALY gains are not unusual in paediatric PHIs and additionally demonstrate that the evidence produced from the RoE trail economic evaluation is the first of its kind in a SEW context.

This research adds to the current evidence available for the effectiveness and cost-effectiveness of RoE. Compared to the evidence currently available, this study used other outcome measures such as mental health, empathy, perspective taking, and SDQ showing that RoE is effective immediately post intervention.[40, 43-45] However, most evaluations of RoE had no follow-up after post-test and the only published study that did follow-up pupils (three year after post-test), similarly found no significant differences in effect after three years of follow-up.[40] Two interpretations of these results are possible: 1.) RoE, like other child PHIs, are not effective at follow-ups post-test, or 2.) RoE is effective at follow-ups post-test, however we have not been able to accurately measure and evidence its effectiveness. The latter point may be due to measures not being sensitive to change, incorrectly

identifying appropriate outcomes to measure, or its effectiveness is not quantifiable in the mid-term with future outcomes demonstrating meaningful differences (i.e. a sleeper effect detailed in section 7.3.1). Although QALY differences between the arms of this RCT were not statistically significantly different, the majority of the incremental points lie in the northeast quadrant (Figure 15) indicating a more costly, yet more effective intervention. This leads to a high probability of RoE being cost effective within the £20,000 to £30,000 per QALY threshold.

Because of the uncertainty demonstrated in the 95% CIs around costs, effects, and cost-effectiveness, sensitivity analyses were performed to explore this uncertainty further. The CEA of the SDQ total difficulties score (SA1) was the only effect that was statistically significantly different at the final follow-up between groups. This perhaps reflects that the SDQ is the most sensitive for detecting changes in SEW, the main outcome RoE intends to improve. The CHU9D is appropriate for a QALY framework, however many of the dimensions would not have been affected by RoE, e.g. pain and daily routine. Therefore, its appropriateness for detecting change in SEW is questioned. It does however capture a generic health improvement. Its nine dimensions worried, sad, pain, tired, annoyed, school work/homework, sleep, daily routine, and ability to join in activities, capture an overall improvement in functioning. One of the hypothesized health outcomes of RoE is to decrease aggressive and bullying behaviour, so if fewer children are being bullied that may be evidenced in the worried, sad, pain, annoyed, sleep, and ability to join in activities dimensions of the CHU9D. The CHU9D is the only HRQoL instrument designed for children and valued by adolescents, which was a main reason for selecting this outcome to measure QALYs in children. Other HRQoL measures for children exist however, they are usually either adapted from an existing adult measure (16D),[302] they are valued using adult values (EQ-5D-Y[304] and HUI-2[305]), or they have not been valued at all but mapped to an adult measure (PedsQL[309]). This is partly because it has typically been very difficult elicit children's health preferences due to ethical and cognitive difficulties. Time-trade off would involve asking children about death and the ethics of such an activity is questioned. It is also a cognitively challenging task the may not be appropriate for children. The base-case analysis used

adolescent values to value health state utilities as they were thought to be more the appropriate values to apply to our population of children (section 4.2.1.2). However, as can be seen in

Table 24 and

Table 26, differences do exist and they do impact on ICERs. The adolescent values result in mean health state utilities that are consistently lower than those of the adult values; in the base-case analysis the adult values resulted in an ICER £1,400 less than the adolescent values tariff. As was demonstrated in Ratcliffe, et al.,[288] these differences could impact on policy decisions and it is important to consider whose values in health matter in decision making. In this study, the differences between the two value tariffs would not be enough to impact on a decision of whether or not RoE would be deemed cost-effective, except in the available-case analysis where the minor difference of 0.01 QALY impacts the ICER massively (£19,600 versus £40,200). Interestingly, the comparison of the probabilities of SA8 and AV8 being cost-effective are indeed very similar. So depending on what criteria are being used, the differences between the adult and adolescent values may not impact on the decision in this instance. It is important to note the £20,000 to £30,000 per QALY gain threshold is from an NHS and PSS perspective.

If RoE were to be rolled out to schools across Northern Ireland, it is likely the cost of providing the programme will largely fall on schools or local education authorities and their WTP for the programme may be very different from current threshold supported by NICE. In fact, recent Education Endowment Foundation (EEF) guidance states that all EEF funded evaluations must now include a cost evaluation where schools are assumed to be paying all costs to provide the intervention, even if the EEF provides funding for the intervention during the evaluation phase.[365] This guidance provides an important finding in itself as it answers the question ‘Who should pay for funding preventive PHIs that may generate

multisectoral benefits (e.g. health, education, economic, and social)?' The guidance makes an explicit assumption that schools should pay for school programmes even if it seems appropriate theoretically for the other sectors that stand to benefit from the intervention to contribute to the funding. Schools should be aware of this and not depend on funding from other sectors; however, calls should be made to redistribute the burden of funding from one sector to multiple sectors if multiple sectors stand to gain from the intervention.

Remme and colleagues[35] suggest a cofinancing approach in which multiple sectors dedicate parts of their budgets, based on their current marginal productivity, to jointly finance interventions (such as PHIs) that generate multisector benefits including health and non-health benefits. While the article puts forward a stylised example of how cofinancing could work in theory, there are still many practical issues identified that make this approach an area for further research. For example, each sector may not have a single payer to make allocation decisions about funding other sectors; rather there may be multiple payers with differing budget constraints.[35] There is no established cost per QALY threshold from the education sector perspective, so while RoE is arguably cost-effective from an NHS perspective, the same cannot be said from an education sector perspective, and education decision-makers may ultimately need to decide whether or not to continue funding RoE from their own budgets. There are two methods for estimating a cost-effectiveness threshold that have been explored in health that could potentially be used in other sectors to help determine their WTP for sector-specific outcomes and/or the proportion they would be willing to cofinance. The first, mentioned previously in section 2.1.8, involves estimating marginal productivity through econometric analyses of routinely available health expenditure and outcome data.[118] The second uses a 'bookshelf' analogy to demonstrate how a cost-effectiveness threshold could be estimated from cost-effectiveness evidence available in the literature.[366] Imagine a bookshelf with the tallest books lined up from the left, representing an intervention's effectiveness. The width of the books represent their costs, so the length of the bookshelf represents an exhausted budget as the tallest books will be included starting from the left and shorter books added until funding runs out. To estimate the threshold would be to identify the least cost-effective intervention included on

the bookshelf.[366] These approaches to estimate non-health-sector cost-effectiveness thresholds are promising in theory; however, they rely on substantial data that may not be available.[35] Also, they rely on a generic sector-specific outcome (such as the QALY) being available to make cost-effectiveness comparisons within each sector. From the work of this thesis, a generic, education-specific outcome measure has not been identified in economic evaluation of school-based interventions (see Chapter 3). Identifying an appropriate education-specific generic outcome measure; estimating the education sector's potential cost-effectiveness threshold; and exploring potential cofinancing options are thus areas for further research.

Mapping utility scores from the SDQ (SA3 and AV3) may underestimate the uncertainty around the ICER estimate. While ICER point estimates were similar to base-case point estimates, SA3 and AV3 had the tightest confidence intervals and highest probabilities of being cost-effective despite neither cost nor effect coefficients being statistically significant. The use of this mapping algorithm will be explored further in Chapter 6. The use of annuitisation and the assumptions around the useful life of the intervention do impact on the cost-effectiveness results. SA5, where there was no annuitisation, resulted in an ICER of £15,800 versus £11,000 in the base-case SA0. There is less of a difference between the base-case and when costs are annuitised over three years, ICER £11,800. In this study, the choice between a 1.5% and 3.5% discount rate minimally affects the cost-effectiveness results when using the adult values (£90 difference in ICERs) and adolescent values (£1,100 difference in ICERs). The available-case analysis (SA8) demonstrated the most conservative estimate; with greater incremental costs and lower incremental QALYs resulting in the highest ICER estimate and lowest probability of cost-effectiveness for RoE. As was described earlier in the methods for the missing data analysis (section 4.4.4), a limitation of available-case analysis is that different samples of costs and QALYs may end up being used which can lead to non-comparability and affect the covariance structure.[348] Thus, the results from the available case analysis should be interpreted with caution.

Because a threshold for cost-effectiveness does not currently exist for units of effectiveness outside of the QALY used in healthcare, it makes it difficult to determine if costs and benefits accruing outside of the QALY framework are cost-effective (e.g. the value of

changes in SDQ scores or increases in educational attainment).[367] To complicate matters further, there is still no consensus among clinicians as to what constitutes a meaningful change in SDQ. In order to value changes in the SDQ, it first must be established what these changes represent, and if they follow a linear pattern. When decision-makers are presented with an ICER from a CEA, they will have to rely on their own experience of the CEA effectiveness outcome and value judgements to decide on an appropriate cost-effectiveness threshold. The interpretation of the cost-effectiveness of RoE is dependent on a number of factors because a.) an attempt was made to capture wider societal benefits of which no current threshold for cost-effectiveness exists; b.) there are few other school-based economic evaluations of similar aims to compare to; and c.) RoE is delivered at the school and if local authorities are making funding decisions they may be more interested in non-health related benefits (i.e. educational attainment) versus quality of life. Ultimately, whoever is making the funding decision about RoE will need to decide which threshold will guide their decision-making and what other factors to consider. From the analyses presented, RoE has demonstrated its cost-effectiveness across many assumptions and values of the threshold. The research question asked what the cost-effectiveness was of the programme in reducing aggressive behaviour and increasing prosocial behaviour. SA1, SA2, AV1, and AV2 were conducted to answer this question and cost-effectiveness ICERs were presented. However, the interpretation of the results are dependent on the decision-maker's WTP for unit improvements on each of the scales.

Even within the realms of a stated cost-effectiveness threshold per QALY gained, there is contention. The lack of a theoretical and empirical basis for the estimation of the current threshold is still debated[116] as mentioned in section 2.1.8. Claxton and colleagues[118] suggest the current threshold is too high and it should be much lower because £13,000 of NHS resources adds one QALY to the lives of NHS patients. This new lower threshold (£12,936 to be exact) was estimated from use of routinely collected NHS data. The research found the NHS spends too much on approving new drugs and the consequence of these decisions is the opportunity cost forgone which relates to actual NHS patients who bear these costs.[118] This was contested in a critique by Barnsley and colleagues,[121] but perhaps the current threshold is too high. If a new lower threshold were applied to all new HTAs, the effect would be to prioritise less costly public health programmes and interventions in current

health and social care decision making. RoE was evaluated in relation to the currently accepted thresholds, if the new lower threshold was applied, a majority of the sensitivity analyses would meet this new criterion including both base-case analyses using both adolescent and adult values.

The CEA of the SDQ total difficulties score (SA1) was the only effect that was statistically significantly different at the final follow-up, with the RoE group demonstrating lower difficulties. This perhaps reflects that the SDQ is the more sensitive for detecting changes in SEW, as it was selected as the primary outcome in a similar classroom based CEA of a teacher management programme to increase child and teacher mental health and wellbeing.[213] The ICER for the SDQ was based on a one-unit decrease/increase in scores (total difficulties and prosocial behaviour), but there is still uncertainty around the meaning, or the value of a one-unit decrease/increase in scores. As reported previously, there are bandings in place to help with interpretation of SDQ results. However, those bandings are not based on any diagnostic thresholds and are instead meant be used to recommend referral for further examination.[368] Because the SDQ has not been valued by the preferences of the public, it is difficult to assess the opportunity cost of other programmes covered under the same budget that do not use the SDQ as a measure of outcome, in other words it is like comparing apples to oranges. Additionally, there is not a consensus upon a minimally important difference in the SDQ. In personal communication with an expert colleague from Child and Adolescent Psychiatry (Minnis H 2016, oral communication, 14th October), Professor Minnis noted a typical change you might see in the SDQ total difficulties score is 0.3 to 0.35 in an RCT. This would reduce the score from “borderline” to “normal,” however, this estimate unfortunately is not based on a lot of data, further demonstrating this lack of agreement. Essentially, the incremental unit decrease/increase in scores that were used to calculate the SDQ ICERS were arbitrary as no consensus has yet been reached as to what incremental should be used for the SDQ. Despite this apparent arbitrariness, this method for calculating SDQ ICERs has been employed elsewhere in the literature.[369] This issue cannot be ignored as future interest in using the SDQ as a primary outcome measure is likely to increase, and this is particularly relevant for economic evaluation using the SDQ. Further research into determining a clinically meaningful difference in scores is warranted.

There are three forms of the SDQ, parent-complete, teacher-complete, and self-complete by the child.[326] The perfect study would have information from all three informants as parents are good at identifying externalising problems (such as Attention Deficit and Hyperactivity Disorder and conduct problems) but less so at identifying emotional problems (Minnis H 2016, oral communication, 14th October). Children are more likely to record depression and anxiety symptoms accurately, but under-report conduct problems (Minnis H 2016, oral communication, 14th October). And teachers are somewhere in between; some will argue they are less biased with identifying behavioural type problems (Minnis H 2016, oral communication, 14th October). The economic evaluation used the SDQ teacher complete form, thus being exposed to a potential risk of bias for this particular outcome. The parent informant version had been sent home to parents during the trial, however due to low response rates the data had been dropped from the analysis. Self-report is recommended for children aged 11-17, therefore it would not have been appropriate to include this measure, as children were too young to fill it in on their own. Even though the base-case analysis was a CUA, thorough discussion of the CEA using the SDQ was provided as the SDQ as an outcome in economic evaluation is novel and it was the primary outcome in the main trail for detecting changes in SEW.

The health and medical fields have long used CUAs to aid policy decision making. Without such analyses, decisions are at risk of being made based on emotional appeal, absolute intervention cost, and political pressure.[370] This CUA and accompanying sensitivity analyses provide initial evidence that school-based PHIs are feasible, are likely to be cost-effective according to current thresholds, and can be employed to aid decision making.

5.8 Limitations

Data on resource use would have ideally been collected at each data collection time point. It was recognised that recall bias was likely with the long recall periods for estimating resource use expenditure; however, the alternative was to completely forego collecting any resource use for the trail. The lack of resource use being consistently collected was the main limitation within this CUA, which also had a limiting effect on the choice of analytical methods employed.

The available resource use data was also limited by large percentages of missing data. Variables with the highest percentages of missing data may have been impacted by a survey design effect as they were all questions that were self-reported using free form text. Therefore, a detailed descriptive analysis was employed to determine the appropriate assumptions around the missing data and missing data were subsequently handled using MI. Future evaluation work of school-based PHIs should be mindful of potentially large amounts of missing data, particularly data that is collected from parents by post.

Wider non-health benefits such as educational outcomes and spillover effects such as increases in quality of life at home were not captured in this study, but they would have added further understanding of the cost-effectiveness of RoE. Until 2012, CUA was NICE's main method for determining cost-effectiveness of public health interventions.[23] It wasn't until the 3rd edition of the NICE public health guidance'[79] that more emphasis was placed on CCA and CBA to ensure all relevant benefits (health, non-health, and community) were taken into account and aid local authorities or other organisations to judge whether or not an intervention is value for money. The attempt to collect wider societal costs and benefits was also hindered by the high percentage of missing data. The only method for capturing wider outcomes available was through contact of the children's parents by post. In this trial, this method proved difficult and was prone to producing missing data. Other more routine data sources might have provided more reliable societal costs and benefits and these should be considered for future research.

Cost estimates in this trial may not be generalizable to contexts that differ from that of the current trial (e.g. resource use implications, RoE fees, and healthcare organisation). The costs estimated in this evaluation were specific to the costs incurred during the trail and information was not available about how implementation throughout Northern Ireland might impact on these costs. Estimating the cost of rolling out RoE across Northern Ireland may not be a simple 'scaling up' exercise (such as multiplying the intervention cost per child by the population of school children) because it is unknown how the fee structure (those paid to the RoE organisation in Canada) might change depending on scale. This however, does not seem to apply to the effectiveness of RoE as evaluations conducted in multiple countries have found RoE to be effective.[40, 43, 44]

Educational outcomes such as attainment would have benefited this analysis because local authorities will most likely end up making and funding the decision to implement RoE. They will need to decide whether RoE represents value for money, and they are more likely to be interested in comparing costs in terms of educational outcomes in addition to health outcomes. The thresholds stated are what NICE considers to be cost-effective from an *NHS perspective*. Cost-effectiveness thresholds do not exist outside the health sector,[367] nor has a method been devised to apportion costs (who should bear them) when more than one government department or sector is involved.[30] This is particularly an issue when one sector benefits from a public health intervention while the other is required to fund it. NICE does not make any recommendations for how costs should be apportioned, rather the methods chosen should be transparent and justified.[30] This trial was funded by the NIHR and delivered through the Public Health Trusts in Northern Ireland. In the event that the funding decision about RoE is transferred to local authorities, the collection of educational outcomes would have aided the decision-making process. Additionally, there is overwhelming evidence that education is linked to health and other outcomes[371] so the presence would have provided further information to aid a decision.

It would have been useful to explore the longer-term impacts of ROE by modelling potential impacts over the child's lifetime. However, there is a paucity of longer-term evidence using the main outcomes of our analysis, the SDQ and CHU9D, especially the CHU9D which is a relatively new generic HRQoL measure. Additionally, the lack of statistically significant difference in effects (in terms of any other outcome measured in the trial) at the third year follow-up meant that any potential longer-term benefits would have significant assumptions and uncertainty attached. The RoE trial did provide one of the longest follow-ups of any RoE evaluation identified, so the single trial was a sufficient source of immediate and mid-term data.

5.9 Conclusion

This study shows that, within current commonly accepted thresholds for the value of a QALY, RoE is likely to be a cost-effective school-based population health intervention. Even when considering a much lower QALY threshold of £13,000, over 80% of the sensitivity

analyses would still be cost-effective. To my knowledge, this was the first school-based economic evaluation to incorporate children's preferences through use of the CHU9D. It also adds to the growing body of cost-effectiveness evidence incorporating the SDQ. A growing pool of incremental costs per change in SDQ may help with the estimation or valuation of those incremental changes, which is an area for further research. There is a plethora of SEL classroom programmes available, but none of them have been rigorously analysed in terms of cost-effectiveness as RoE has. Furthermore, no other evaluations of RoE have been fully costed within a formal economic evaluation so this work represents the first cost-effectiveness evidence of a full-scale cluster randomised controlled trial of the programme. These findings are novel in this context; however, this novelty presents difficulties for allocative decision-making as there are few other school-based programmes that have been evaluated in a cost per QALY framework to act as comparators, much less in a SEL context.

From an NHS perspective, RoE is likely to be cost-effective immediately after intervention and for up to three years post-intervention. However, important additional analyses relating to the total budgetary impact of rolling out this intervention, assumptions about RoE intervention life span, and longer-term quality of life benefits are required to draw definitive conclusions relating to its longer-term cost-effectiveness. In addition, future studies are needed to compare RoE interventions with alternative interventions aiming to achieve the same SEW gains.

The following chapter describes a methodological work that examines the use of mapping or 'crosswalking' from a behavioural screening tool, the SDQ, to a generic preference-based HRQoL measure, the CHU9D. There is interest in such algorithms as they allow for the calculation of QALYs when no utility measure is available or collected. The methods research that follows applies two previously developed mapping algorithms using RoE data.

6 Appropriate Outcome Measurement: a Mapping Validation Study

The importance of looking after children's SEW is clearly a priority as highlighted in chapter 1. The school setting is an ideal place to reach children and offer intervention to improve SEW. However, measuring SEW in a school environment is highly challenging as it is recognised that a lack of valid methods exist for primary schoolchildren.[5] A recent review of eleven mental health outcome measures found none to have sufficient psychometric evidence to reliably measure severity and change over time in key groups.[372] Despite this, the use of the SDQ[322] has been viewed positively by staff in pre-school establishments[373] and is currently being used in school-based settings to assess SEW.[213, 374-376] It is also widely used in CAMHS throughout the United Kingdom[377] providing a source of routinely collected SDQ data.

The SDQ is a favoured primary outcome measure of SEW in school-based interventions however, due to its measurement properties, i.e. lack of a value-based outcome, its applicability in economic evaluation (i.e. CUA) is limited. The significance of the SDQ's inclusion in school-based economic evaluation was detailed in section 4.2.1.4). Briefly, the SDQ is widely used behavioural screening tool whose use in SEW is gaining popularity.[368, 373, 378] It's specific properties may make it more sensitive to change as compared to a generic measure. As was discussed in section 5.7, interpretation of incremental changes in SDQ in terms of the incremental costs places more burden on decision-makers because there is no explicitly stated threshold with which to compare. One potential solution is to apply mapping or 'crosswalk' algorithms to convert data from a non-preference-based generic or condition-specific measure, e.g. SDQ, to a generic preference-based measure. This would allow evaluations of school-based interventions that collected the SDQ to estimate health state utility values. Mapping is an option recommended by NICE for estimating EQ-5D utility values when EQ-5D data are unavailable.[379] However, as stated in 2011 technical support guidance, 'in most cases, mapping should be considered at best a second-best solution to directly collected EQ-5D values, as the use of mapping will lead to increased uncertainty and error around the estimates of health-related utility.'[291] There is now a large body of literature that have used functions to map between non-preference based and generic

preference-based measures for the purposes of estimating health utilities for use in economic evaluation.[380]

Currently there are two mapping algorithms available to convert SDQ data into preference-based utility values. The first uses all five SDQ subscales to map from the SDQ to CHU9D.[329] This algorithm was used in SA3 and AV3 in Chapters 4 and 5. The second uses only three of the SDQ subscales to estimate CHU9D child health utility. For studies that only have SDQ data available, these mapping algorithms provide an additional tool for the facilitation of CUA; however, its use and applicability for economic evaluations within a school-based context is under-researched. In particular, how SEW is valued within CUA of school-based interventions and which tools are best placed to do this valuing.

Use of non-traditional economic outcomes such as the SDQ may provide a useful starting point for health economists to determine long-term health impacts of PHIs as the SDQ is now established in long-term cohort studies[381, 382] as well as being recently mandated for use in Australia's specialised CAMH services as a consumer-oriented outcome assessment tool. Furber et al.[329] outlines that national and international data coordination efforts (e.g.[383, 384]) have led to the creation of large SDQ data sets, which represent thousands of episodes of care in CAMH services across Australia and the United Kingdom. Transforming SDQ scores to utility values would facilitate CUA of routine CAMHS data, open up school-based SDQ data to this possibility, as well as provide the opportunity to estimate longer-term QoL impacts from long-term cohorts which include the SDQ.

This chapter details a methodological mapping validation study that was conducted while the RoE trail was ongoing.[385] This work is the final of the three empirical studies conducted to help answer the overarching research question, 'How should the cost-effectiveness of school-based, population health interventions aimed at children be determined?' As discussed earlier, the different types of economic evaluation rely on appropriate and valid outcome measurement to determine effectiveness. With the research in child outcomes lagging behind adult outcomes,[37] outcomes research as related to economic evaluation outcomes is necessary to contribute to the limited evidence-base. This early stage methodological work was planned using non-randomised baseline data available during the

interim of the trial. Outcome data from pre-test, post-test, and the first follow-up of the RoE trial were analysed to examine the suitability of mapping the SDQ to the CHU9D within a CUA framework using previously published mapping algorithms.[329]

6.1 Study aims

This study aims to contribute to the outcomes evidence base for economic evaluation of school-based PHIs by testing and validating previously published mapping algorithms[329] to translate SDQ scores to utility values. Given this aim, the research question asks:

1. Can SDQ scores elicited within an educational context be mapped using published algorithms to preference-based CHU9D utilities with a view to incorporating such utilities within an economic evaluation framework?

Utility mapping methods have been conducted to transform SDQ scores into CHU9D values;[329] beyond that, there are no completed economic evaluations using these two measures together or indeed externally validating the algorithms. This empirical sub-study within the RoE economic evaluation was planned to explore the relationship between these two measures, as well as externally validate the SDQ mapping algorithm developed by Furber et al[329] against the self-completed CHU9D utility scores from the RoE trial.

6.2 Methods

This section describes the outcomes and methods used to address the research question above by describing the analyses undertaken. Data incorporated into this analysis were non-randomised and those collected at pre-test (baseline October 2011), post-test (after intervention completion June 2012), and at 12-month follow-up (June 2013). Data collection methods from the full trial given in Chapter 4.

6.2.1 Strengths and Difficulties Questionnaire

There are three forms of the SDQ questionnaire available as described in section 4.2.1.4; this study utilised the teacher complete proxy version. The SDQ was scored using the predictive algorithm converted into Stata syntax available on the SDQinfo website[326] (Appendix 7) in Stata 11.2 (StataCorp LP, College Station, Texas, USA). This involved assigning a score from 0-2 (0= ‘Not True’ or no difficulties; 1= ‘Somewhat True’ or some difficulties; and 2= ‘Certainly True’ or many difficulties) for each item of the questionnaire

and summing the total for each scale. Totals from all scales (excluding prosocial behaviour) were then summed to generate the total difficulties score.

SDQ scores can be classified into four provisional bands that reflect the distribution of the general population's scores; these bandings were based on a large UK community sample provided elsewhere.[386] The provisional bandings categorise SDQ scores into four groups: 'close to average' (80% of the population), 'slightly raised' (10%), 'high' (5%) and 'very high' (5%). The teacher complete four-band categorisation for SDQ scores is given below in Table 27. Previous versions of these cut-points included a three-band categorisation which combines the highest two categories (High and Very High) shown in Table 27.

Table 27: SDQ domain score four band categorisation*

Teacher Complete	Close to Average	Slightly Raised	High	Very High
Total Difficulties Score	0-11	12-15	16-18	19-40
Emotional Problems Score	0-3	4	5	6-10
Conduct Problems Score	0-2	3	4	5-10
Hyperactivity Score	0-5	6-7	8	9-10
Peer Problems Score	0-2	3-4	5	6-10
Prosocial Behaviour Score±	6-10	5	4	0-3

***From http://www.sdqinfo.org/py/sdqinfo/b3.py?language=Englishqz(UK) 'scoring instructions for SDQs for 4-17 year olds**

±Higher values preferred in this subscale. Column titles for this subscale are as follows: Close to Average, Slightly Low, Low, Very Low.

6.2.2 Child Health Utility 9D

There are currently two tariffs available to value the CHU9D as described in section 4.2.1.3. The adult values tariff was developed using preferences of 300 members of the UK adult population using a standard gamble technique.[96] To incorporate adolescent values in decision making, an adolescent values tariff has also been developed using preferences from 590 Australian adolescents aged 11-17.[311] A best-worst scaling discrete choice experiment was used to derive preferences from this population. For this study each tariff was applied to CHU9D scores to calculate utility values, for comparative purposes. For the adult values tariff, coefficients from the ordinary least squares (OLS) parsimonious model (model 5)[96] were used as decrements to calculate utility. For the adolescent values tariff, rescaled conditional logit estimates were used.[311]

The two OLS regression based algorithms developed by Furber et al.[329] were applied to transform SDQ scores into utility values. The dataset used to develop the mapping algorithms assessed CHU9D by parent proxy, an important difference to this current study in which children self-completed the CHU9D. The CHU9D was developed and intended to be completed by children. Both algorithms using three and five SDQ subscales are replicated in Equation 10 and Equation 11 below from Furber et al.[329]

Equation 10: Algorithm using five SDQ subscales[329]

$$Utility = 0.880 + (-0.019 \times emotion) + (-0.009 \times conduct) + (-0.001 \times hyper) \pm (0.008 \times peer) + (0.005 \times prosocial)$$

Equation 11: Algorithm using three SDQ subscales[329]

$$Utility = 0.918 + (-0.018 \times emotion) + (-0.12 \times conduct) + (-0.009 \times peer)$$

6.2.3 Analysis

All analyses performed were on the entire non-randomised sample which included data from pre-test, post-test and 12-month follow-up from post-test. A descriptive analysis (mean, standard deviation (SD)) was performed to describe the sample in terms of gender, grade, deprivation rank (measured by the Northern Ireland multiple deprivation measure 2010), and mean scores from the SDQ and its subscales as well as the CHU9D estimated from both tariffs and both mapping algorithms. Missing data were modelled through MI via chained equations as recommended by good research practice guidelines.[133, 348, 353, 387] Tables of frequency were graphed for CHU9D and SDQ level responses for a visual representation of the spread and nature of the data. When assessing the agreement between prosocial behaviour, total difficulties and utility measures, variables were plotted in pairs to check for approximate linearity, outliers, and subgroups. Pearson's correlation coefficient was used to assess the strength of relationship between utility (adult values tariff), total difficulties and prosocial behaviour. *T*-tests were performed to test for pairwise differences in utility values created from the adult values tariff,[96] the adolescent values tariff,[311] and both mapping algorithms.[329]

6.3 Results

Questionnaires were returned by teachers in 67 schools at baseline, 65 schools after intervention and 64 schools at 12-month follow up. After data cleaning and MI, a total of 1,254 child participants were included in the analysis making up 3,762 observations. At baseline, a majority of the pupils (88.9%) were recruited in Primary 5 (approximately 9 years old); however, some Primary 4 (6.5%) and Primary 6 (4.6%) pupils were also included. Table 28 presents the characteristics of these participants. The sample was made up of 51.5% boys, and median deprivation rank was 430 which is comparable to median population rank of 445. As the sample deprivation rank is less than the median rank it can be said the sample median is more deprived than the population median rank, but the extent to which the sample is more deprived cannot be inferred from the rankings.

The mean (SD) for SDQ total difficulties and prosocial behaviour scores were 12 (3.2) and 8.3 (2.1) respectively, which are classified as 'slightly raised' and 'close to average.' Please refer

to Table 27 which provides the four bands to aid interpretation of the SDQ scores. The mean (SD) for SDQ subscales emotion, conduct, hyperactivity and peer problems was 1.5 (2.0), 2.3 (1.0), 4.1 (1.3) and 4.1 (0.9). Emotion and hyperactivity subscales were classified as 'close to average' and conduct and peer problems were 'slightly raised.' The frequency of responses for each symptom scale is reported in Figure 17.

The mean (SD) utility scores were 0.84 (0.11) and 0.80 (0.13) based on the adult and adolescent values tariffs respectively. With both scoring algorithms, approximately 5.72% of participants were classified in full health (i.e. utility = 1). In all dimensions of the CHU9D except 'tired,' no problems were most commonly reported. Figure 18 reports the frequency of responses to all levels.

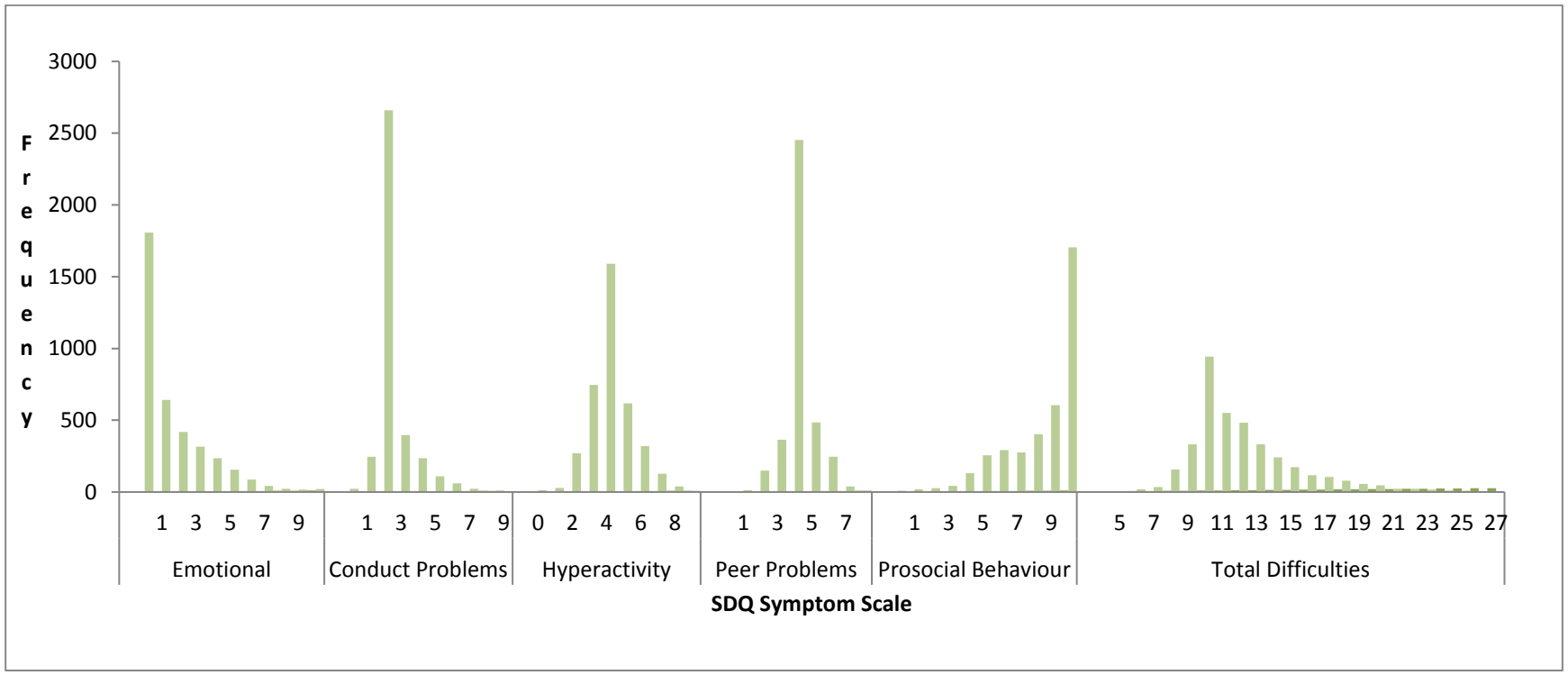


Figure 17: SDQ response frequency

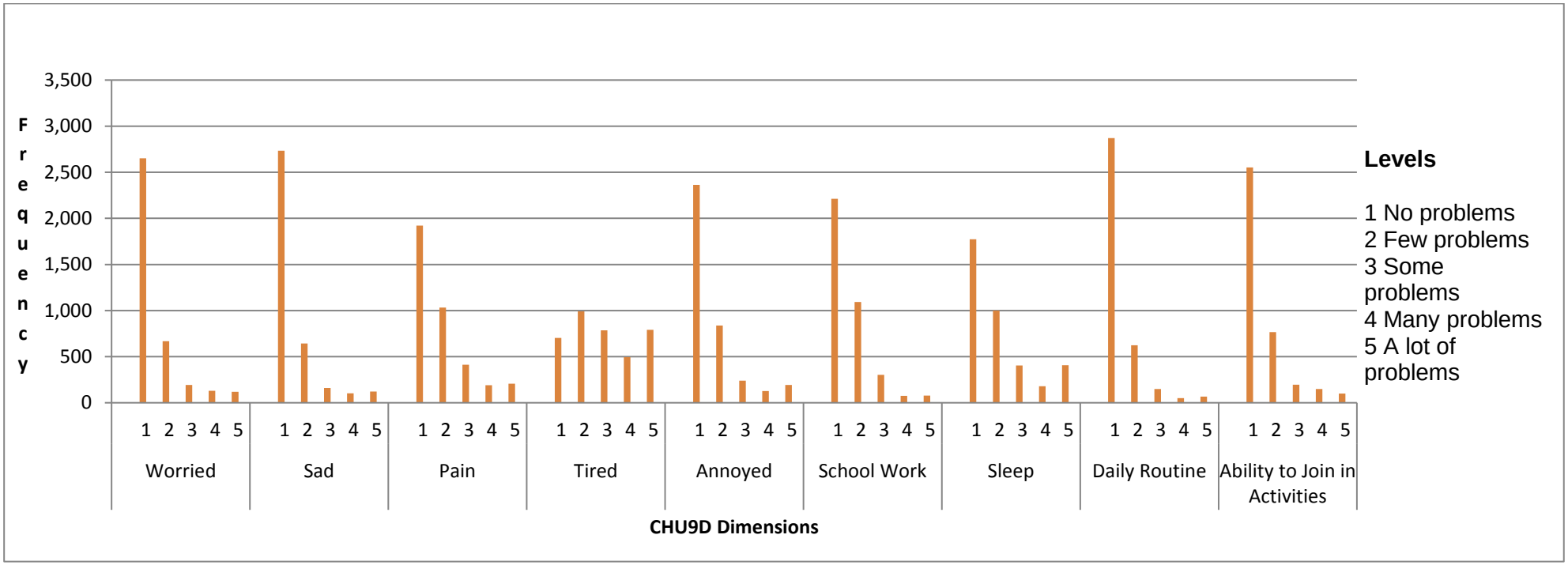


Figure 18: CHU9D response frequency by level

Table 28: Characteristics of participants

Characteristics	*Participants‡* (*n* = 1254)	*British Community Sample†*
Gender		
Boys, n (%)	646 (51.5)	
Girls, n (%)	608 (48.5)	
Grade level		
P4 (≈8 years old), n (%)	81 (6.5)	
P5 (≈9 years old), n (%)	1115 (88.9)	
P6 (≈10 years old), n (%)	58 (4.6)	
NIMDM deprivation rank,* median (SD)	430 (245.9)	
SDQ Total Difficulties, mean (SD)	12 (3.2)	6.6 (6.0)
SDQ Prosocial subscale, mean (SD)	8.3 (2.1)	7.2 (2.4)
SDQ Emotion subscale, mean (SD)	1.5 (2.0)	1.4 (1.9)
SDQ Conduct subscale, mean (SD)	2.3 (1.0)	0.9 (1.6)
SDQ Hyperactivity subscale, mean (SD)	4.1 (1.3)	2.9 (2.8)
SDQ Peer Problems subscale, mean (SD)	4.1 (0.9)	
CHU9D Original tariff, mean (SD)	0.84 (0.11)	
CHU9D Alternative tariff, mean(SD)	0.80 (0.13)	
CHU9D Algorithm using five SDQ subscales, mean(SD)	0.84 (0.05)	
CHU9D Algorithm using three SDQ subscales, mean(SD)	0.83 (0.04)	
‡Participants had responses at 3 time points for a total of 3,762 observations		
***Lower rank=higher deprivation**		
†From British sample 8,208 teachers of children aged 5-15		
http://www.sdqinfo.org/norms/UKNorm1.pdf		

The mean (SD) utility values for the mapping algorithms using five and three of the SDQ subscales were 0.84 (0.05) and 0.83 (0.04). Table 29 reports the *t*-tests results from the pairwise comparisons. Each method for estimating utility produced statistically significantly different results except the adult values tariff and mapping algorithm using five SDQ subscales in which no statistically significant difference was detected (p=0.69) (95% CI: -0.003, 0.004).

There were weak but statistically significant correlations between all combinations of CHU9D (adult values tariff), total difficulties, and prosocial behaviour. Pearson's rank correlation coefficient showed significant correlations between: total difficulties and CHU9D (r = -0.08, p<0.01); total difficulties and prosocial behaviour (r= -0.27, p<0.01); and prosocial behaviour and CHU9D (r= 0.04, p=0.02).

Table 29: Differences in utility values

Difference in pair	*n*	*Mean*	*SD*	*t*	*df*	*p*	*95% CI*
Adult vs Adolescent	3762	0.036	0.051	43.926	3761	0.000	0.035 to 0.038
Adult vs 5 SDQ subscales	3762	0.001	0.116	0.402	3761	0.688	-0.003 to 0.004
Adult vs 3 SDQ subscales	3762	0.010	0.115	5.360	3761	0.000	0.006 to 0.014
Adolescent vs 5 SDQ subscales	3762	-0.036	0.136	-16.10	3761	0.000	-0.040 to -0.031
Adolescent vs 3 SDQ subscales	3762	-0.026	0.135	-12.022	3761	0.000	-0.031 to -0.022
5 SDQ vs 3 SDQ subscales	3762	0.009	0.011	53.209	3761	0.000	0.009 to 0.010

6.4 Discussion

According to the bandings set out in Table 27, the prosocial behaviour, emotional, and hyperactivity SDQ subscales were considered 'close to average' in comparison to a large UK sample.[386] Using that same sample as a comparison, total difficulties, conduct, and peer problems subscales were classified as 'slightly raised.' This is somewhat unexpected as the sample comprises a general school population in Northern Ireland, and all subscales would be expected to fall within the 'close to average' band. In terms of economic evaluation, this outcome on its own is less useful because the 'value' associated with unit changes in SDQ scores is unknown as discussed in section 5.7. For the CHU9D, all dimensions had most respondents classified in the 'no problems' category, with the exception of 'tired' (see Figure 18). The SDQ total difficulties scores is 'slightly raised' in this sample compared to a generally healthy 'no problems' quality of life scores; these differences demonstrate that the two descriptive systems do not overlap entirely. This is due to differences on a conceptual basis; the SDQ is a behavioural screening tool designed to assess emotional and behavioural function, while the CHU9D assesses the child's broader functioning and HRQoL. Mapping functions rely on statistical association and this is less strong when the descriptive systems of the two measures are not measuring the same thing.[173] However, when comparing single dimensions of the two measures in terms of frequency of responses (Figure 17 and Figure 18), there is some overlap. Worried and Sad dimensions of the CHU9D overlap the Emotional symptom scale of the SDQ well indicating some overlap in the two descriptive systems. Furber and Segal[388] conducted a recent study to assess the suitability of the CHU9D as a routine outcomes measure in a CAMHS setting. They also found the CHU9D and SDQ correlated

moderately, with significant correlations between the CHU9D worried item and three SDQ items: many worries, low confidence, and many fears. The CHU9D sad item was also significantly correlated to the SDQ often unhappy item.[388] In terms of overlap between the two measures, the authors found that linear regression of the nine items of the CHU9D explained 31.5% of the variance in the SDQ total difficulties score.[388] Their results are in line with this current study which does not demonstrate large overlap between the two measures' descriptive systems. A key difference to note was that their study was conducted in a CAMHS setting while this current study was conducted in a school setting which indicates the relationship between the two measures is consistent across both settings.

It is important to note that despite all of the correlations between the SDQ and CHU9D being significant, they were not very strong. Using a rule of thumb whereby correlations of 0.70 to 1.0 represent strong correlation and 0.30 and below represent weak correlations, all of the current correlations were considered weak and thus the statistical significance of the correlation may simply be a result of the large sample size.

The mean utility generated for adult values CHU9D was 0.84, which compares with the range of mean values reported in previous studies (0.803-0.86).[388-390] The studies varied in context, setting, and age groups, but were included for comparison as so few studies have published CHU9D outcomes. The mean utility from adolescent values CHU9D was lower than the adult values tariff which is consistent with recent Chinese and Australian studies that applied both tariffs to their samples.[288, 389] Ratcliffe and colleagues[288] have compared the adult and adolescent value tariffs using the responses to a web-based survey of 500 Australian adolescents, aged 11-17. They found differences in adult and adolescent values for identical health states may have enough significance to impact on health care policy decision making.[288] Differences between the instruments may be due to differences in descriptive systems, size and nature of the samples, and the valuation methods used to develop each scoring algorithm.[288] Nevertheless, the Chinese version CHU9D found utilities generally discriminated well in relation to self-reported health status, regardless of which tariff was employed.[389] As noted throughout this thesis, the author proposes the use of adolescent values in decision-making that ultimately affect them.

The SDQ alone cannot provide insight into resource allocation decision making, i.e. whether the programme is a worthwhile use of educational resources (or indeed an argument for investing health care resources). As discussed previously, there are issues around the understanding of clinically meaningful differences in SDQ scores; the value of those differences; and whether or not there should be a pre-specified threshold to determine cost-effectiveness. The use of the SDQ alone in a CEA context places more burden on the decision maker in terms of determining cost-effectiveness (as outlined in section 5.7) and for these reasons, it is less useful in economic evaluation. Yet, the SDQ is a common primary outcome measure in many paediatric PHIs. For economic evaluation, the CHU9D is useful because it has value associated with incremental change. The advantage the CHU9D brings to the evaluation of paediatric interventions is that they can now be assessed using a preference-based measure combined with costs, and judgements can made in relation to their relative cost-effectiveness. It is now possible to compare paediatric programmes from a range of areas that aim to improve different aspects of children's health and wellbeing by including a generic HRQoL measure such as the CHU9D. Changes in effectiveness as measured using the SDQ and mapped to CHU9D, can now be compared in terms of their costs required to achieve those changes in outcomes. For example, a cost per three-point change in the SDQ could not readily be compared to a cost per three-point increase on a national exam. Having a uniform measure of QoL that has been valued by the population, allows comparison of programmes in terms of both cost and effects because they have been measured on the same generic scale.

To my knowledge, this is the first study to apply the preliminary mapping algorithms[329] to an external dataset, contributing to the sparse evidence base of an appropriate and validated alternative for conducting CUA using the SDQ. The caregiver version of the SDQ was used in development of these algorithms as opposed the teacher-rated version used in the current study. Additionally, parent completed proxy report CHU9D was used,[329] as opposed to child completed CHU9D in the current study. The validity of applying the mapping algorithms to different versions of SDQ and CHU9D is questioned (i.e. the validity of mapping from teacher complete SDQ to child complete CHU9D when the algorithms were developed using parent complete SDQ and CHU9D). However, the CHU9D was intended to be completed by children and there are multiple valid versions of the SDQ. Recently, due to the increased use of mapping methods to generate generic

preference-based health utilities, guidance on a set of preferred reporting items for mapping studies has been published. The MAPS (Mapping onto Preference-based measures reporting Standards) Statement is a set of 23 items deemed essential in order to increase clarity and promote complete and transparent reporting of mapping studies.[391] The original study that developed the mapping algorithms was published before these guidelines became available, however it would have benefited from reporting to a uniform standard.

Utilities derived from the four different approaches were all significantly different, the only pair that was not was the adult values tariff and five SDQ subscale algorithm. This is an interesting finding because the population from which the algorithm was developed was sampled from a CAMHS setting. These children would be expected to have lower QoL than a general school-aged population. Also, these algorithms were developed using the adolescent values tariff and it is of note that in our results, the five SDQ subscale algorithm better predicts the adult values utilities. Nonetheless, this study adds to the evidence and generalisability of the mapping algorithm using all five of the SDQ subscales.

By applying the mapping algorithms to an external dataset, this research contributes to the existing evidence base around the suitability of the use of the five SDQ subscale mapping algorithm for eliciting utilities which was the aim of this study. To answer the research question, economic evaluation is now feasible in studies where SDQ data (but not preference-based utility data) have been collected and our results suggest the algorithm containing all five SDQ subscales to be superior. This is in line with recommendations;[329] however, future studies should be conducted replicating use of these algorithms to confirm these results.

These findings have practical implications as they may make conducting CUA in school-based settings more efficient, as fewer resources would be needed for data collection, speeding up the evaluation process. Additionally, it now may be possible to conduct CUA retrospectively if cost and SDQ data for school programmes are available. This provides an opportunity for a wide range of activities that could now be subject to economic evaluation with low additional resource input. There is also the potential of converting SDQ data from longitudinal datasets into utilities, which could be useful in establishing links between short-term surrogate outcomes and long-term established outcomes. As

the CHU9D is a relatively new measure, longer-term child health utility data does not yet exist to establish links between immediate and long-term child health utilities. As the SDQ has been established for longer, using this measure to estimate longer-term child health utilities, is a promising area for further research. In the future, researchers interested in mapping from the SDQ to CHU9D should use the mapping algorithm, which contains all five SDQ subscales.

6.4.1 Reflection of the overall aims and research question

As Chapter 6 concludes the final of the three empirical works making up this thesis, the overall aim and research question can now be reflected upon. The overall aims were to:

(i) determine what evaluation methods (economic and non-economic) are currently being used to evaluate school-based population health interventions;

(ii) illustrate a good practice example of a thorough cost-utility and cost-effectiveness analysis of a school-based intervention (the RoE programme) to reflect on the advantages of such practice and disadvantages that remain, such as decision-making in multisectoral settings; and

(iii) explore which outcomes are appropriate for children in the SEW and economic evaluation context to support future evaluation work in this context.

Chapter 3 addressed the first aim, finding that the methods currently being utilised to evaluate school programmes are varied and widespread with poor quality reporting of economic evaluations being noted. Of the CUAs identified, none had directly measured HRQoL using a child appropriate measure and values. For the second aim, RoE was found to be cost-effective with a base-case analysis ICER of £11,000 per QALY. There was considerable uncertainty around this estimate (CI: -£95,500 to £147,000) due to a lack of finding any statistically significant effect that lasted up to the 36 month follow-up. However, the probability of the RoE being cost-effective was high, at 85% at a WTP of £20,000/QALY from an NHS perspective. CEA using the SDQ resulted in an ICER of £197 per unit decrease in total difficulties score (CI: £77 to £471). It is unknown how this result would be interpreted in a health or education decision-making context, however this study has contributed to the growing pool of incremental costs per SDQ improvement,

which will aid the valuation of those incremental changes if at first, consensus can be reached on the clinical significance of those incremental changes. For the final aim, the SDQ is appropriate for measuring SEW (see section 4.2.1.4), but is less effective for cost-effectiveness decision-making. Converting the SDQ into child health utility, provides an option of valuing incremental changes in QoL against a generally accepted cost-effectiveness threshold. Chapter 6 validated a mapping algorithm to convert SDQ into child health utility, which can be compared to a generally accepted cost-effectiveness threshold.

The overarching research question asked, 'How should the cost-effectiveness of school-based, population health interventions aimed at children be determined?' Based on the findings from the literature review, and the empirical works on the economic evaluation and mapping validation studies, the author presents four options to be considered; CBA, CCA, CUA (first introduced in section 2.1), and MCDA (section 2.2.1.2). The first and most appropriate method theoretically is CBA. Because of the challenges identified involving multisector outcomes which are broader by nature, it makes sense that the most comprehensive form of economic evaluation is most appropriate, as it allows monetary valuation of these multisector outcomes in a final cost to benefits ratio or net benefit/loss making for clear, consistent, decision-making criteria. Practical limitations of CBA include a lack of standardisation in elicitations methods, stated-preference biases, and the considerable measurement burden which requires increased time and resources resulting in a more costly evaluation. New Economy (introduced in section 2.2.1.1) developed in-depth guidance on how to conduct CBA in a local public services context where analytical and research resources may be limited.[149] It's also supported by an example excel-based CBA model and unit cost database with more than 600 unit cost estimates; providing a resource to revive CBA in a community context where it is more appropriate than CEA/CUA. For example, in local public services there may be limited analytic and research resources; this guide and excel-based example may help facilitate the formal evaluation of programmes delivered in the community.

CCA can also take into account the varied multi-sector benefits, but places more burden on decision-makers to make trade-offs between costs and effects and does not rank alternatives. This hinders consistency in decision-making with no clear decision rules. However, disaggregated costs and benefits may be preferred by the decision-maker, so is

still worth considering. CCA requires less research burden, so the author recommends completing CCA in addition to another form of economic evaluation. If the school-based programme being evaluated gives rise to primarily health outcomes, CUA is good option because it benefits from having clear decision rules in place in terms of cost-effectiveness. However, the applicability of those decision rules to an educational setting is less clear and the current QALY framework does not take into account any non-health outcomes. In the latest NICE social care guidance manual,[151] NICE specifically states an openness to consider 'social care QALYs' if validated. The Adult Social Care Outcome Toolkit (ASCOT) used by the Department of health would be considered as a parallel evaluation as well as undertaking a capability and wellbeing approach using the Investigating Choice Experiments for the Preferences of Older people – CAPability (ICECAP). This new openness to consider other non-health QALYs demonstrates a potential openness to consider a generic education utility as a parallel evaluation when there are health and non-health benefits of school-based intervention.

The recommendation of including an 'impact inventory' from the Second US Panel on cost-effectiveness[128] could be useful in this respect as well. The impact inventory would lists all health and non-health consequences of an intervention to ensure those that occur outside of the health sector are considered regularly. If a CUA framework is to be adopted for education, further research is needed to understand what generic education utility outcome is appropriate in this setting, how it should be valued, and how the threshold for cost-effectiveness should be determined (section 5.7). A limitation to adopting this framework, is that multi-sector benefits such as health and labour market outcomes would not be captured in a generic education outcome, and thus parallel evaluations using sector-specific generic outcomes (e.g. QALY, social care QALY, and ICECAP) would be needed. Finally, MCDA is an option to improve transparency in decision-making involving multiple criteria. It does place more cognitive burden on decision-makers as they will be responsible for determining weights and scoring for the multiple criteria. There also has not been any published examples identified in the school-based literature, so methods and standardisation are still being developed.

6.4.2 Limitations

The use of mapping to derive generic preference-based indices from disease specific measures raises a fundamental concern as mapping methods assume overlap in each measure's descriptive systems.[380] One method for assessing mapping functions is to evaluate the difference between predicted and observed values by calculating the root mean squared error (RMSE).[380] The RMSE gives an indication of the size of the prediction errors between predicted and observed values. With the mapping algorithms,[329] RMSE indicated large differences between predicted and observed values at the individual level. However, the purpose of mapping is to predict differences across groups or between trial arms, not at the individual level. As was evidenced in Chapter 5, the sensitivity analysis which used the mapping algorithm to estimate utilities resulted in unusually narrow confidence intervals indicating more certainty around the results and higher probability of being cost-effective. This was an unusual result as the confidence intervals around every other sensitivity analysis were wider, indicating that the algorithm may underestimate uncertainty. A recent study by Madan and colleagues[392] found mapping algorithms that were based on raw scores overestimated QALY gains as condition specific measures may improve the condition without impacting on any other generic domains of health, which can lead to over-estimating health utility benefits. The authors conclude, that mapping algorithms should reflect within person changes and be estimated from datasets that contain repeated measures in order to avoid overestimating health utility.[392] The Furber et al.[329] study did not contain repeated measures and did not estimate the algorithms with this type of approach, so this might partly explain the over-estimation of certainty in the confidence intervals. Due to the lacking overlap between the SDQ and CHU9D descriptive systems and the potential underestimation of uncertainty, the use of the mapping algorithm is a second best option to the use of preference-based HRQoL measures, but it may be necessary in population health programmes for pragmatic reasons.

This study has demonstrated initial evidence for the justification of the SDQ in economic evaluation of school-based interventions with a view to it being mapped to a broader, more generic QALY. In settings outside of the adult healthcare sector (i.e. education, paediatric, and population health), condition-specific primary outcome measures such as the SDQ, may be the only measure of effect collected. In these instances, this study

indicates the five SDQ subscale algorithm is a useful instrument, affording health economists' the opportunity to conduct CUA. This allows decision-makers a uniform 'yardstick' measure to compare across interventions and determine cost-effectiveness.

6.5 Conclusion

The SDQ and CHU9D are able to measure outcomes in children aged 8–13 years within an educational setting, and this study has validated use of the five SDQ subscale mapping algorithm for estimating child health utility at a group level. This study adds to the currently sparse evidence base, providing an appropriate and validated alternative to conducting CUA in contexts involving children. It is now possible for researchers to perform economic evaluation of population-based interventions where traditional utility measurement methods are missing, but the SDQ is available. This allows analysts the opportunity to conduct CUA retrospectively in paediatric or school-based programmes where previously this would have been impossible due to unavailability of preference-based outcome measures. This can be achieved with few additional resources allowing decision-makers access to cost-effectiveness evidence that was previously absent, and therefore improving the decision-making process. To my knowledge, the SDQ and CHU9D have not yet been used to predict longer-term outcomes within an economic evaluation context, as the CHU9D has only become recently available. This is an important avenue for further research because issues remain as to how these childhood measures extrapolate into adulthood, and how school-based and/or preventive PHIs can demonstrate longer-term cost-effectiveness.

7 Chapter summary and discussion

7.1 Introduction

The overall aims of this thesis was to:

(i) determine what evaluation methods (economic and non-economic) are currently being used to evaluate school-based population health interventions;

(ii) illustrate a good practice example of a thorough cost-utility and cost-effectiveness analysis of a school-based intervention (the RoE programme) to reflect on the advantages of such practice and disadvantages that remain, such as decision-making in multisectoral settings; and

(iii) explore which outcomes are appropriate for children in the SEW and economic evaluation context to support future evaluation work in this context.

This thesis addressed these aims, through three separate, but interlinking empirical works: the systematic literature review and narrative synthesis of school-based evaluation methodologies; the economic evaluation of the RoE trial; and the mapping validation study which explored the appropriateness of applying previously published algorithms to predict child health utility from the SDQ. Each chapter is summarised in turn followed by a critique of the strengths and limitations of these three major works.

7.2 Chapter Summaries

7.2.1 Chapter 1

SEW, as was set out early on, has been linked to better health, wellbeing, and education outcomes in children. This is because SEW enables children to build and maintain healthy relationships and handle interpersonal situations constructively; also helping to prevent mental health problems from developing which can predict future academic, social, and labour market outcomes. Because promoting and maintaining children's SEW is gaining increased attention from academics and policy-makers, a variety of school-based SEL programmes have been developed[18] as a means to improve children's SEW as well as their success in school and life. Schools have been recognised as an ideal setting for

health promotion activities as they have the ability to reach most children efficiently. The importance of determining the effectiveness as well as cost-effectiveness of SEL programmes is justified as the need to demonstrate value for money is required when schools face increased financial pressures. This is especially important when decision-makers in education are faced with a plethora of SEL programmes to choose from. Economic evaluation in healthcare settings has been established for some time, however applying those methods to an education setting has been limited.[185] With few economic evaluations of school programmes being conducted, decision-makers are left without cost-effectiveness evidence when making decisions about implementing or continuing school based programmes. Another important question raised was regarding, 'Who should pay for implementing PHIs when multiple sectors stand to benefit from the intervention?' NICE has not issued guidance on this matter as, 'no standard method has yet been devised to apportion costs - and who should bear them - when more than one government department (or, indeed, local authority) is involved' (p 5).[30] Further reflection on this key issue follows in the Chapter 5 summary (section 7.2.5).

RoE, a SEL and PHI, was introduced briefly in Chapter 1. The overall aim and research question was put forward given the lack of established economic evaluation methods in school settings and the need for cost-effectiveness evidence to aid decision-makers in school based interventions. The overall aim was to determine: what evaluation methods are currently being used to evaluate school-based population health interventions through systematic review; the cost-effectiveness of the RoE programme through economic evaluation; and which outcomes are appropriate for children in the SEW context specifically aimed at paediatric populations in a school setting. The overall research question asked, 'How should the cost-effectiveness of school-based, population health interventions aimed at children be determined?' Chapter 1 concluded by providing a chapter outline for the rest of the thesis.

7.2.2 Chapter 2

Chapter 1 introduced key concepts of this thesis; the need to promote and maintain children's SEW, and the need to establish the cost-effectiveness of school-based SEL programmes. Chapter 2 delved deeper into these and related fundamental concepts of economics such as scarcity and opportunity cost. Important definitions of key concepts

were defined such as economics, health economics, HTA, and economic evaluation. The field of health economics developed due to market failures and the use of economic evaluation has increased to aid decision makers in determining what health services should be offered to maximise the health of the population within the financial constraints of the healthcare system. Additionally, a brief history introduced some of the earliest works in economic evaluation, before the disciple had been formally recognised. One of the earliest forms of economic evaluation was a cost-benefit type analysis of England's Public Health Act 1875 where estimates of life years gained from the improvement were compared to the cost enacting the Act. In 1917, the appropriately named Charles 'Value' Chaplin, recognised that institutions have a hard time breaking away from traditions of the past, creating inefficiencies in the health care system and this was a theme repeated throughout this chapter. In 1917, he hypothesised that if healthcare institutions were to start over with a new health care budget, they would probably end up with a different allocation of resources based on current knowledge of costs and effectiveness. However, starting over is difficult to do, and examples of this notion are the reluctance of NICE to consider a lower cost-effectiveness threshold, as well as a reluctance to break away from the standardised method of economic evaluation, the CUA. CUA is efficient and useful when the context is limited to health and health outcomes; however, PHIs cover a broader context and sometimes give rise to non-health outcomes that are not captured by the QALY. Selma Mushkin was one of the first to define health economics in 1958 before Kenneth Arrow, who is often credited with the recognition of health economics as a discipline in 1963. Herbert Klarman wrote the first health economics textbook, and was the first make a quality adjustment to life years gained from a kidney transplant.

From the 1970s onward, methods for modern economic evaluation were developed. A number of recommendation guidelines, documents, and texts have emerged since the 1990s on the design and conduct of health economic evaluation enabling the standardisation of the basic elements of economic evaluation and analytic techniques.[37] Many countries have their own country specific HTA guidance in place for conducting economic evaluation and an updated list can be found on the ISPOR website: https://www.ispor.org/PEguidelines/index.asp.

Sections 2.1.3 to 2.1.7 described the different types of economic evaluation methodologies and gave examples of outcomes that distinguish the different types and situations where certain methods may be more appropriate. CUA, a form of CEA, is the most commonly used type of economic evaluation in the UK due to NICE guidance specifically calling for this type of evaluation in the reference case. CBA is the most comprehensive form of economic evaluation, but due to various challenges, comprehensive CBAs are still rarely published. CBAs require more time and are more costly to design and implement, as there are a broader spectrum of outcomes to identify, measure, and value which will be unique to each CBA. Valuing these broader outcomes faces challenges as well because there is a lack of standardisation and biases present in stated preference methods. CMA has been criticised for failing to investigate uncertainty in determining equivalence between two different treatment options. It has been found to bias measures of uncertainty; therefore, data on costs and effects should still be collected and analysed to assess this bias. Recommendations are that CMA should no longer be used due to these problems and biases that present from attempting to determine total equivalence in effectiveness of two or more alternatives. Finally, CCA is not always considered a full economic evaluation, however it provides decision makers the option of deciding themselves the appropriate trade-offs that need to be made in terms of costs and benefits.

The final subsection of section 2.1 detailed the role of economic evaluation in decision making in the UK and internationally. This was described in terms of cost-effectiveness thresholds and trial-based, model-based, and mixed economic evaluation methodologies. The UK specifically adopts a £20,000 to £30,000 per QALY threshold to aid decision making. The WHO uses a multiple of the average per capita income to determine cost-effectiveness in low and middle-income countries. The US chooses not to set a threshold, instead preferring a range of thresholds being explored. The scepticism of the appropriateness of the use of a QALY threshold in decision-making was critiqued, as the current threshold used by NICE is not based on theory or empirical evidence. The UK is the only country to specifically state a value for the threshold. Other countries support the use of a threshold, while some, such as the USA outright disagree with a threshold.[128-130] There are differing views as to what the threshold should represent: 1.) the society's monetary valuation of health gains; or 2.) the opportunity cost resulting from disinvestment required to adopt a new health technology. This thesis proposes both

perspectives are valid as the value of the threshold should be based on collective values of society. At the same time, the public needs to be aware that valuing the threshold equates to an opportunity cost or disinvestment elsewhere in the healthcare system, often times those affected by the disinvestment are less visible, such as cuts to mental health services. Stating a cost-effectiveness threshold essentially places a value on human life, so debate and scepticism around this issue will continue. Finally this subsection, concluded by describing the various vehicles for economic evaluation. They can be conducted alongside trails, in a decision analytic framework, or a mix of both methodologies and the advantages and disadvantages of each were described.

The first half of the chapter summarised above, introduced the general methods of economic evaluation in a typical clinical trial hospital-based economic evaluation. The second half focused on PHIs and how those traditional methods could be applied to deal with broader perspectives, inclusion of non-health outcomes, other challenges, and economic evaluation in school settings. The second half introduced key concepts of economic evaluation of PHIs, which was the focus for this thesis as RoE is considered a PHI that is delivered in schools. The economics of prevention assumes health behaviours function much like goods consumption functions in market places. Many external influences impact on individual choices such as cost, opportunity, incentives, and constraints. When markets fail, such as the healthcare market, government intervention is acceptable to correct the failure (e.g. taxes on tobacco, alcohol, and sugary drinks). As unhealthy behaviours are the cause for many preventable diseases and deaths, this market is failing, and intervention to prevent disease and death are warranted. This sets up the economic case for investing resources to prevent disease and ill health versus spending those resources later on in treatment. These approaches are sometimes referred to as upstream methods of prevention.

The prevention paradox has the potential to impact population outcomes on a large scale. If the entire population is exposed to a safe intervention, then the entire distribution of risk in that population can be shifted placing more people in low and moderate risk than would be the case if only the high-risk group was targeted.

However, the prevention paradox states the individual benefits may be insignificant or non-existent, but those small benefits added up over the population could result in

significant community or population benefits. This is particularly relevant for RCTs of PHIs or public health interventions as measures are collected at the individual level. Often effect sizes are small and/or diminishing, and this could partly be explained by the prevention paradox. PHIs evaluated in a cluster RCT framework typically only include a subset of the whole population, and therefore the sum of the combined effect may not reach significance, as the entire population was not included in the study. However, simple projections can be performed to estimate the population effect by multiplying the effectiveness results of an RCT, by the total relevant population, as demonstrated in Chapter 5. While this does not produce a robust estimate, it certainly can provide and idea of the potential of PHIs.

Investing in preventive and effective measures of the population have the potential to bring large community benefits, as well as being cost-effective. However, as economic evaluation moves from a narrow NHS setting to a broader public sector/societal perspective, challenges start to emerge when conducting economic evaluation of population and public health interventions. These challenges include those previously defined by Weatherly and colleagues[36] such as attribution of effects; measuring and valuing effects; identifying intersectoral costs and consequences; and incorporating equity considerations.

Broader perspectives represent an increased research burden in terms of identifying, measuring and valuing additional health and often-times non-health outcomes. Non-health outcomes are particularly problematic in health economic evaluation as they fall outside the remit of the preferred QALY measure in the UK. NICE has recognised this by considering other forms of economic evaluation which may be more appropriate for PHIs such as CBA and CCA. There are considerable research burdens associated with CBA such increased time and resources to develop and implement a CBA. Preference elicitation is time consuming, often produces biased results, and is lacking in standardisation. Even if standardisation improved, CBA would still be time consuming because each stated preference design would need to be unique to the problem at hand. There is the in depth guidance developed by New Economy[149] (section 2.2.1) which provides resources on how to conduct CBA in a local public services context where analytical and research resources may be limited.[149] The guidance also provides an example of an excel-based CBA model and unit cost database with more than 600 unit cost estimates. This is a useful resource

to encourage CBA in the public services and population health contexts, although analysts may still be limited by the unit cost estimates that are available in the database. The methodology, CBA guidance, and Excel spreadsheet model are freely available for download from their website (http://www.neweconomymanchester.com/our-work/research-evaluation-cost-benefit-analysis/cost-benefit-analysis/cost-benefit-analysis-guidance-and-model), and the methodology has been included in supplementary guidance to the HM Treasury Green Book.[393] A development such as this is a promising attempt to 'revitalise' CBA methods for modern day economic evaluation of PHIs. CCA is less time consuming, but places more cognitive burden on decision-makers as there is no single combined outcome of cost-effectiveness presented.

Emerging methodologies such as MCDA may become more prominent in the future, once researchers have a chance to digest and implement the recent methods guidance released by ISPOR.[160, 161] The main advantages to adopting MCDA is that it gives analysts a transparent and systematic method for incorporating multiple criteria in decision-making. If a generic approach were to be adopted by NICE, evaluations could incorporate pre-specified multiple criteria that already have generic weights attached, therefore increasing comparability amongst different clinical areas. However, a generic approach may eventually face similar challenges as the QALY, in that the pre-specified criteria are too narrow to capture broader societal benefits that might arise from PHIs. Economic evaluation alongside natural experiments are also a promising alternative for evaluating national policy PHIs. However, in the UK, there is still a major focus on CUA as there are justifiable reasons for keeping the established decision-making process, which is based on an established cost per QALY threshold. Knowing the limitations of CUA for PHIs (in examples where many outcomes are non-health outcomes), health economists should not continue following the status quo, as they are not helping the growth and development of this area. This thesis proposes that instead, researchers and health economists should be pushing for funding to develop alternative methods such as CBA and MCDA and utilise them in research projects. If funders are still going to require a CUA, more time could be written into grant applications to allow completion of multiple forms of evaluation, such as CBA, CUA, and CCA together. Each method has its limitations on its own, but this approach to research alongside trials would allow a more complete picture of evidence to be presented to decision makers, as well as help advance the area

methodologically as to which use of multiple methods are suitable and appropriate in practice.

Finally, even more challenging is establishing a transparent and consistent decision-making process in the education sector by introducing economic evaluation to school-based PHIs. The education economics literature has produced some examples of economic evaluation, but it has been limited. No consensus has been reached if a decision-making threshold should be established for the education sector and what the generic outcome should be to allow consistency in decision-making. Education is not unlike health, in that resources are under constant financial pressure, so there is an opportunity to learn from the health sector. Namely, alternative methods to CUA should be considered to take account of intersectoral benefits that are likely to arise from school-based PHIs. This chapter concluded that arguably, we may be coming full circle in reconsidering CBA for PHIs, as one of the first records of economic evaluation was a CBA of England's Public Health Act in 1875.

7.2.3 Chapter 3

Because of the limited development of economic evaluation in school settings, there was a need to understand what evaluation methodologies were currently in place to evaluate school programmes to aid decision-making. In addition to the challenges relating to economic evaluation of PHIs, there is a lack of research into child preference-based measures, which are necessary when evaluating child health outcomes in a CUA framework. Together, these resulted in a novel, 'uncharted area' for economic evaluation methodology of school-based PHIs. To explore this area further, the following research question was posed, 'What evidence currently exists around economic and other evaluation methodologies of school-based interventions and/or programmes?' Because the economic evaluation of RoE was one of the first of its kind (both in terms of context and outcome measurement), a systematic literature review was conducted to identify evidence of economic and other evaluation methodologies that currently exist for the evaluation of school-based programmes. The purpose of the review was to gain an understanding of how economic evaluation (and other evaluation methodologies) of school-based programmes are currently being conducted and the types of preference-based child utility measures that are currently being utilised. There was an implicit

assumption that HRQoL would be an appropriate measure of outcome in school-settings as the majority of the economic evaluations that were identified were for programmes and interventions that aimed to improve child health in some way. Additionally, RoE aims to improve children's SEW, so because health was the focus of the economic evaluation in Chapter 4 and 5, emphasis was placed on identifying appropriate child preference-based HRQoL measures for CUA. The review revealed relatively few high quality existing studies and zero published studies that incorporated children's preferences in CUA.

Only four published studies were identified that directly measured HRQoL. Two used the EQ-5D,[243, 267] one used the PedsQL,[282] and the final a disease specific caregiver's quality of life instrument.[281] It was noted that the only studies to incorporated child-specific HRQoL were non-economic evaluation studies, clearly in CUA it would be preferable to collect utility directly versus relying on estimation, mapping or a crosswalk function. There is a paucity of evidence in the published literature of CUA of school-based interventions that directly measure HRQoL using appropriate, child-specific measures. Twenty-five of the 76 studies included for review were identified as CUAs; of those 25, only one directly measured HRQoL using the Shona-language version of the EQ-5D. The review identified zero published studies that directly measured HRQoL in children (using a measure designed specifically for children) and which used children's preferences; the only measure to fulfil both of these conditions is the CHU9D. The use of the CHU9D in this context is an important and novel contribution to the literature as RoE would be the first school-based economic evaluation to incorporate the CHU9D with adolescent values, as currently there are no values from younger children available.

The review found the evidence of evaluation methodologies of school-based programmes are varied and widespread. Economic evaluation is still a relatively novel concept in the school setting[185] despite efficient resource allocation being a high priority for budget constrained education boards. The review also revealed that alternative methods for incorporating multisector benefits such as MCDA and SROI were not being utilised in the education evaluation literature. Reasons for this might include publication time lags, or the need for further guidance and/or standardisation in these alternative methods. Few studies report follow-up longer than six months[18] and there is little evidence of cost-effectiveness and long-term effectiveness. Budget cuts to publically funded education has resulted in scarce resources needing to be maximised to their full potential. Economic

evaluation can help education decision-makers make more informed decisions about how to allocate limited funds. The lack of cost-effectiveness evidence in the area of SEW identifies a gap in the current knowledge leaving decision-makers less informed about the cost-effectiveness of new SEL programmes they might choose to implement.

The quality of reporting and methods used in the identified economic evaluations were not quite up to the standards that might be expected in the clinical trials-based medical literature. Few studies directly measured HRQoL in children leading to uncertainty in the programmes' effectiveness estimates. Improvements can be made in the quality of reporting of economic evaluations of school-based programmes as low quality of reporting was prevalent. As a minimum, economic evaluation focused on health outcomes should report each of the applicable CHEERS checklist items and this review did not identify any studies that reported on each item. As was touched upon in chapter two, economic evaluation of school-based programmes will not always have a health focus. They may give rise to benefits that span multiple sectors, and if education is the focus, more development of the methods as appropriate for an education setting should be researched further. This is because the downside of the QALY framework is that it does not take into account non-health benefits and it is not as flexible at incorporating multi-sector benefits. The chapter concluded that as the methods for school-based economic evaluation develop, the quality of reporting should improve as well.

7.2.4 Chapter 4

In order to address some of the shortcomings of economic evaluation of school-based PHIs identified in Chapter 3, Chapter 4 presented the methods of a thorough economic evaluation of the RoE programme. The chapter described in detail, the methods used for the main trial economic evaluation of the RoE programme which was the first economic evaluation of school-based PHI, RoE, to address the evidence gaps in the literature such as low-quality reporting and a lack of evidence that directly measures children's quality of life using their preferences. The evaluation was thoroughly reported to address the lack of quality reporting available in the current literature, and set an example of a standardised method (CUA) for conducting and reporting economic evaluation of a school-base PHI. The chapter started by describing the differences between economic evaluation of child and adult interventions, and how there is a need to develop outcome

measures specifically for children. The CHU9D is the only preference-based HRQoL measure that was developed specifically for children, which can also be valued by adolescents (with the elicitation of younger children's values currently ongoing). Another child specific measure, the SDQ, was also described in detail as it was used in CEA. There was discussion around appropriate measures for SEW as a lack of valid measures currently exist; the SDQ may not be entirely appropriate to measure something that is very difficult to quantify in the first place. However, in the absence of specific and validated measures of SEW the SDQ is the best available option. The importance of including both a generic preference-base quality of life measure, as well as a condition specific measure of outcome was discussed as difficulties arise when trying to quantify changes in a non-generic, non-preference-based outcome, such as the SDQ, in terms of other education outcomes covered under the same budget, e.g. increases in test scores. A generic 'yardstick' measure can be useful in this context and to address the limitations of a generic outcome, condition-specific outcomes can be included which may be able to measure the intervention's effectiveness more accurately.

Background and contextual information to the main cluster randomised controlled trial of RoE was provided. A review of the existing evidence of RoE's effectiveness found that only one evaluation was a cluster RCT design, with follow-up at three years. This evaluation took place in a different contextual setting to Northern Ireland and none of the existing evidence of RoE effectiveness included a cost component or economic evaluation. The main trial aims and research questions were stated and data collection detailed. For the economic evaluation, the aim was to evaluate the cost-effectiveness of the programme from a public sector perspective over trial time horizon of 45 months (3.75 years or 3 years follow-up after intervention completion). The research question asked, 'What is the cost-effectiveness of the programme in reducing cases of aggressive behaviour and increasing prosocial behaviour among school-aged children?' The final section described in detail, the methods of the economic evaluation of the RoE programme. The section started with an overview, followed by detailed descriptions of the costs, outcomes, how missing data was handled, analyses, and sensitivity analyses performed.

7.2.5 Chapter 5

This chapter reported the results of the economic evaluation of RoE. The evaluation found that, within current commonly accepted thresholds for the value of a QALY, RoE is likely to be a cost-effective school-based population health intervention, with an ICER of £11,000 per QALY gained (CI: -£95,500 to £147,000). The value of the threshold had been discussed and critiqued in previous chapters, with Claxton and colleagues[118] re-valuing the threshold using routinely available data. Their estimate based on technical fact versus informal judgement was closer to £13,000 per QALY. Even when considering this lower QALY threshold of £13,000, over 80% of the RoE sensitivity analyses would still be cost-effective, including the base-case analysis. This evaluation is novel for two reasons: 1.) it was the first economic evaluation of the RoE programme, and 2.) it was the first school-based economic evaluation to incorporate children's preferences through use of the CHU9D as well as adding to the growing body of cost-effectiveness evidence incorporating the SDQ. The CHU9D is appropriate for a QALY framework, because it provides a uniform 'yardstick' measure that can be compared to other programmes and interventions across education and health sectors. The CHU9D is appropriate in school-settings when the programme being evaluated generates primarily health outcomes, and it has been used in school settings previously,[389, 390] just not in a formal economic evaluation. This is particularly useful for transparent and consistent decision-making because the monetary value of a QALY has generally been accepted by British society. However, many of the dimensions of the CHU9D would not have been affected by RoE, e.g. pain and daily routine. Therefore, its appropriateness for detecting change in SEW is questioned.

The CEA using the SDQ as an effectiveness measure was more appropriate for measuring SEW. The CEA also answered the economic evaluation research question of determining the cost-effectiveness of reducing aggressive behaviour and increasing prosocial behaviour. The ICER for SDQ total difficulties score was £197 per one-unit decrease (CI: £77 to £471) and the ICER for prosocial behaviour was £5,630 per unit increase (CI: -£23,400 to £29,100). There are bandings in place to help with interpretation of SDQ results, however, these bandings are not based on any diagnostic thresholds and are instead meant be used to recommend referral for further diagnostic examination. Consequently, there is no clinical consensus for the interpretation of a one-point

decrease or increase in total difficulties or prosocial scores. The results of the CEA are undermined by this lack of clinical relevance especially when these results cannot be readily transferred to other relevant health outcomes or education attainment. This also makes valuing incremental changes in the SDQ a challenge. As more CEAs using the SDQ are undertaken and the resulting cost-effectiveness decisions documented, values for incremental changes will be revealed. There are a plethora of SEL classroom programmes available however, none of them were analysed rigorously in terms of cost-effectiveness as has been the case with this RoE economic evaluation. Because few other SEL programmes have been evaluated for cost-effectiveness, the findings from the RoE economic evaluation were novel in this context. At the same time, this novelty presented difficulties for allocative decision-making as there are few other school-based programmes that have been evaluated in a cost per QALY framework to compare the RoE results to. There are even fewer economic evaluations of SEL programmes to compare.

The probability of cost-effectiveness was high, however when considering quality of life differences at three-year follow-up, no significant differences were observed. This effect is potentially that of the prevention paradox introduced in Chapter 2. If the effects observed in the RoE trial were scaled up to reach the entire population of children aged 5-9 in Northern Ireland, then nearly 1,800 additional QALYs could be gained, thus demonstrating how small individual benefits could add up over the population. The chapter concluded that additional analyses relating to the total budgetary impact of rolling out this intervention, assumptions about RoE intervention life span, and longer-term quality of life benefits were required to draw definitive conclusions relating to its longer-term cost-effectiveness. In addition, future studies will be needed to compare the RoE intervention with alternative interventions aiming to achieve the same SEW gains.

More research is needed to determine the longer-term cost-effectiveness of RoE and other SEL programmes like it. If investment decisions are based on assumptions that the programme will be beneficial to the children in the long-run, then better links to long-term outcomes need to be established. If/when these long-term benefits are established, the author believes that funding for these types of programmes should come from the sectors that stand to benefit. This is not only the case for long-term benefits but established short-term benefits as well. It is difficult to expect the onus of investment in SEL programmes to rely solely on the education sector. EEF guidance (section 5.7) states

that all EEF funded evaluations must now include a cost evaluation where schools are expected to pay all costs of providing the intervention,[365] implicitly assuming schools should be responsible for funding all costs related to school-based programmes even if it will end up benefiting sectors outside of education. Remme et al.[35] (section 5.7) have suggested a cofinancing approach to jointly fund PHIs with multisector benefits based on other sectors' current marginal productivity or cost-effectiveness thresholds, however estimating these proportions that each sector should be responsible for is difficult to do in practice. Reasons include there being multiple payers per sector with differing financial constraints and a paucity of suitable cost-effectiveness data to estimate sector-specific cost-effectiveness thresholds in areas outside the health sector. Upon reflection of EEF guidance and difficulties estimating equitable cofinancing options, schools should be aware of that they might need to fund these types of programmes entirely themselves and not depend on funding from other sectors. However, the author considers that calls should be made to redistribute the burden of funding more equitably and more research into equitable cofinancing options should be prioritised.

7.2.6 Chapter 6

SEW is an important component of child health and wellbeing, but as was seen in Chapter 5, decision-making resulting from cost-effectiveness results of a CEA using a SEW outcome can be a challenge, due to a lack of consensus for interpretation of incremental changes in SDQ and a lack of monetary valuation for those changes. As Chapter 3 highlighted, there is also a lack of directly measured child utility measures being used in school-based economic evaluation. Chapter 6 aimed to explore a possible solution to these problems by answering the research question, 'Can SDQ scores elicited within an educational context be mapped using published algorithms to preference-based CHU9D utilities with a view to incorporating such utilities within an economic evaluation framework?' Additionally, Chapter 6 aimed to address the challenge of interpreting SDQ cost-effectiveness results by converting the SDQ into a generic HRQoL measure for which interpretation of the value of a cost per QALY is much more straightforward. Addressing this research question is not only beneficial for economic evaluation in school-settings, but more broadly in paediatric settings as the SDQ is commonly collected in clinical settings. The study found the mapping algorithm using five SDQ subscales optimal for predicting mean utility and could be used when conducting CUA where there is an

absence of a child health utility measure. This study provided validation of these mapping algorithms and contributed to the limited evidence surround the use of these two measures together in economic evaluation. The study also assessed the differences between the adult and adolescent utility values and found the adolescent values produced lower mean utilities, which is consistent with what has been found elsewhere in the literature. Adolescent values were chosen as the base-case in the economic evaluation (in Chapters 4 and 5) because the author believes adolescent values should be used when decisions are being made that ultimately affect this younger population. Because they are consistently lower than what is produced by the adult values, they also ensure a more conservative estimate, which instils more confidence in a cost-effective ICER result.

The use of mapping is a second best option compared to directly measuring health utility, because the descriptive systems of each measure do not overlap entirely; essentially, they are measuring two different outcomes. There is some evidence of overlap, e.g. in the worried and sad dimensions of the CHU9D overlap the Emotional subscale of the SDQ. This has implications on how well the mapping algorithms can perform as they rely on statistical association between the two measures. Additionally, the unusually narrow 95% CIs observed in Chapter 5, resulted in bias in the uncertainty estimates of the probability of RoE being cost-effective. However, for practical reasons, the use of mapping may be the only option to enable economic evaluation. Transforming SDQ scores to utility values (through use of the mapping algorithm suggested) would facilitate CUA of routine CAMHS data, open up school-based SDQ data to CUA, as well as provide the opportunity to estimate longer-term QoL impacts from long-term cohorts which include the SDQ. This is an important avenue for further research because issues remain as to how these childhood measures extrapolate into adulthood.

To answer the overarching research question of, ‘How should the cost-effectiveness of school-based, population health interventions aimed at children be determined?’ four alternative methods were suggested with CBA being the strongest theoretically. CBA addresses the issue of multi-sector benefits that may arise from school-based PHIs and provides decision-makers with a clear decision rule. There are practical limitations such as increased research burden, inconsistent stated preference elicitation methods, and resulting biases; however, guidance from New Economy intends to revive CBA in the

public services sector providing detailed guidance, an example excel-based CBA model, and cost database. The other three methods (CCA, CUA, and MCDA) put forward all have their strengths and weaknesses and their use for determining cost-effectiveness in school-based PHIs should be based on context-specific factors. CCA requires less research burden, so should be presented alongside with another of the three remaining methods recommended. CUA should be considered if benefits are primarily health benefits, however more work needs to be done to determine an appropriate cost-effectiveness threshold for education, and if a generic education utility measure should be developed. This would also progress aims of the cofinancing approach to determine a more equitable contribution to funding interventions with multisector benefits, as cofinancing is entrenched in the extrawelfarist framework.[35] MCDA should be considered if decision-makers are willing and able to weight multiple decision criteria and the research team is comfortable taking on newer, less standardised method of evaluation.

7.3 Limitations and Strengths

7.3.1 Critique of methods

The work of this thesis was split into three main empirical parts: the systematic literature review and narrative synthesis in Chapter 3; the case study of the RoE economic evaluation described in Chapters 4 and 5; and the outcomes validation study in Chapter 6. The following is a thorough discussion and critique of the methods used in this thesis as well as those used by others who have published similar work.

7.3.1.1 Systematic literature review

The main critique for this systematic review is that the review question chosen was quite broad which meant a wide range of evaluation methodologies were included. The broad nature of the question was justified as little was known of the types of evaluation methodologies that were currently being practiced in the school setting. Nevertheless, this posed difficulty with evaluating quality as a single, yet comprehensive tool such as the CHEERS checklist would not be appropriate for all methodologies included such as the costing studies and non-economic evaluations. It also meant that synthesising the evidence through meta-analysis would not be appropriate. However, even if the focus of the review was restricted to include only economic evaluation, meta-analysis still would

not be possible because the nature of the review question intends to explore methodology used in current practice, not to identify and synthesise an effect size from a single type of intervention. Heterogeneity in the review was dealt with through narrative synthesis and followed a systematic process that included preliminary synthesis, exploring relationships within and between groups, and assessing the robustness of the synthesis. To my knowledge, there is no existing systematic review which attempts to review methods of school-based evaluation methodologies. As was revealed from the review, obesity prevention interventions were the most common type of school-based intervention evaluated in the published and grey literature. Systematic reviews of these types of interventions do exist, however outcomes are still too heterogeneous to combine in formal meta-analysis.[394, 395] This further provides justification for use of narrative synthesis as the previous two studies cited did not provide any formal method of synthesis, rather they reported their results descriptively. An example of a systematic review which employed narrative synthesis is a study that aimed to synthesise current knowledge of shared decision making in palliative care by Bélanger et al.[396] The aim of this systematic review is similar in that it sought to gather evidence from different types of methods and approaches and thus a narrative synthesis would be appropriate for the heterogeneous outcomes and methods included in review. The authors similarly follow the same guidance for conducting a narrative synthesis by Popay et al.[190] The guidance followed involved three steps: developing a preliminary synthesis of findings of included studies; exploring relationships within and between studies; and assessing the robustness of the synthesis. This formal method for synthesising the data collected from the review instils further confidence of the robustness of the review as data synthesis is recommended for systematic review, but is often ignored when meta-analysis is not appropriate.

Initially, the scoping review conducted was much more focused; the inclusion criteria was narrower, only focusing on economic evaluation of school-based interventions. The scoping review identified few studies, so a broader approach was taken to make sure a comprehensive review of all available evidence was conducted. Economic evaluation of school-based programmes is a novel area, so the current literature may have used other, broader methodologies outside of economic evaluation such as MCDA, SROI, SIA, or HIA. As the review found, this was not the case, however, this is a finding in itself.

The other main critique of the methods for this systematic review was that only one author reviewed all studies and performed the data extraction and synthesis. The CRD guidance used to conduct the review specifically advises that a minimum of two authors should be involved in the review to minimise bias and error resulting from the conduct of the review. As only the author conducted each of the major stages, this is a potential source of bias. The review attempted to mitigate any bias with two authors (members of the supervisory team) conducting validity checks during the evidence gathering stages. The validity checks ensured that the decisions the author took screening papers to include for review, were replicated by the two supervisory team authors. The results of both validity checks found all authors coming to the same conclusions, which ensured a systematic process for deciding which articles to include for review.

7.3.1.2 RoE economic evaluation

The base-case CUA of RoE is the first school-based CUA to incorporate directly measured utility values from children using their preferences. As identified in the systematic review, only four published studies of school-based CUA directly measured health utility, the remaining 21 CUAs identified estimated or modelled health utility indirectly. Of the four that directly measured health utility, two used the adult measure EQ-5D and were published after the availability of the CHU9D, which would now be considered inappropriate.[397] Prior to the availability of the CHU9D, poor approximation through use of adult measures was considered better than no approximation of health utlities.[397] However, the main critique for the RoE economic evaluation is that CUA may not capture the full range of costs and outcomes of PHIs such as RoE. Wider non-health benefits such as educational outcomes and spillover effects such as increases in quality of life at home were not captured, but they would have added further understanding of the cost-effectiveness of RoE. Until 2012, CUA was NICE's main method for determining cost-effectiveness of public health interventions.[23] It wasn't until the 3rd edition of the NICE public health guidance'[79] that more emphasis was placed on CCA and CBA to ensure all relevant benefits (health, non-health, and community) were taken into account to aid local authorities or other organisations to judge whether or not an intervention is value for money. Since 2012 when this new guidance came out, there have still been relatively few CCAs and CBAs of PHIs to appear in the literature.[68] This could again be attributable to a reoccurring theme in this thesis that institutions have a hard time evolving from the

status quo. CUA was planned for the RoE economic evaluation because having that generic 'yardstick' measure to make comparisons to other competing sources of funding is still the main requirement set out in NICE guidance, with CCA and CBA being considered as secondary analyses.[150] Also, the outcomes collected from the RoE trial were all primarily health outcomes, and thus a QALY framework was appropriate. CBA is a resource intense evaluation method and less attractive to analysists who already have to provide a cost per QALY analysis as a requirement of many British funding bodies. However, it is a superior method theoretically in terms of capturing broader, multi-sector benefits. Researchers could consider the practical limitations of CBA and apply for additional funds to push the CBA agenda forward in order to more appropriately address economic questions in population health and education settings.

The attempt to collect wider societal costs and benefits was also hindered by a high percentage of missing data. The only method for capturing wider outcomes available was through contact of the children's parents by post. In this trial, this method proved difficult and was prone to producing missing data. In addition, the long recall period is prone to inaccuracies as there is a tendency to underreport community service utilisation with longer recall periods.[398] Even if the planned CUA is rigorous and well reported there are still unforeseen issues to be dealt with in terms of missing data. Typically, in a healthcare setting self-reported outcomes come directly from the patient/participant. In a school-setting there is the added element of complexity in that self-reported outcomes will not come directly from the participant (the child) and will instead be reliant on their caregivers who may not have much contact with the intervention/school, or be as motivated to contribute to the research.

Other more routine data sources might have provided more reliable societal costs and benefits and these should be considered for future research.[399] Data linkage to routinely collected sources of health and social care data may contribute to a more efficient research design by reducing: measurement issues, such as patient recall; time and resources used to design individual resource use questionnaires for each trial; and patient burden, or in this particular case, caregiver burden.[400] Linkage to routinely collected health service resource use may have reduced the amount of missing data in the RoE trial; the burden on caregivers; and the time and resources required to develop the questionnaire and post them out. It also could have addressed the issue of resource use

only being collected at the second and third-year follow-ups. This lack of consistency was a limitation of this CUA which then also had a limiting effect on the choice of analytical methods employed (MLMs could not be fit). The time and resources needed to extract and link data may outweigh the potential efficiency gains, but data linkage to routine resource use is something that should have been considered in the design of the economic evaluation. It is still an option to be considered for future research if necessary funding can be obtained with a small supplemental grant, because if proven to be more efficient and less prone to producing missing data, this data linkage to routine service use should be adopted more widely in economic evaluation of PHIs.

It is now possible to design RCTs using an efficient trial design relying entirely on routinely collected data.[401] The Pleasant trial used Clinical Practice Research Datalink (CPRD) to examine cost-effectiveness of an asthma intervention for children. Benefits of the efficient study design were that there was no primary data collection, and no consent needed as the data was anonymised and routinely collected. The intervention intended to optimise usual care so did not require consent for participation. Additionally, funding required to conduct the research was less than that of a typical grant funded RCT.[401] The main limitation was that utility values were not available and to ask CPRD to collect this information would nullify any cost savings to the efficiency of the trial design. Therefore, utility decrement was modelled using adult estimates. Estimating or modelling utility from adult measures adds a considerable amount of uncertainty to the effectiveness results. Consensus among health economists is still yet to be reached on the best way measure resource use; [400] there are advantages and disadvantages to traditional and efficient measurement methods.

Another methods point from this analysis that should be critiqued is the use of the SDQ in CEA. The advantage of using a condition-specific measure alongside a generic preference-based HRQoL measure is that they may be more sensitive to change. The disadvantage is that these changes along with the cost of achieving such gains produce difficulties for decision makers trying to determine cost-effectiveness. As there is no clinical consensus on clinically meaningful changes in the SDQ, placing a monetary value on these changes is even less straightforward as there is no link to an immediate health or education gain. From a local authority perspective, decision makers would need to decide how much they are willing to pay for a one-point decrease in SDQ total difficulties score. WTP could be

elicited from decision-makers assumes this is their best estimate of the opportunity cost.[35] This payer threshold may be arbitrary; however, consensus of the threshold value may be reached by determining the ranges that previous decision-makers were willing to accept as cost-effective, much like was the case with the QALY. The Incredible Years programme is a PHI which also quantified outcomes using the SDQ in CEA.[369] They report a one-point improvement in the SDQ on a 40-point scale (decrease in total difficulties) over the wait-list control to cost £1,295 (95%CI -£9,150 to £593). However, the uncertainty estimates reported are questioned as the point estimate does not fall within the 95% CI. Other analyses determined the cost per child to move out of the 'high risk' group for the SDQ ranged from £1,612 to £2,418 per child.[369] The funder in this study accepted these costs per improvement in SDQ, so this is an important case for determining the appropriate cost-threshold for point improvements in the SDQ. Moving out of the high-risk group has more meaning than a one-point improvement in the SDQ, as in a clinical setting those at 'high risk' would be assessed further for clinical diagnosis. The RoE CEA reported one-point improvements in SDQ (decreases in total difficulties score) costing £197 (95% CI: £77 to £471) which depending on contextual circumstances may or may not be acceptable to the decision-makers who will decide if RoE should be rolled-out more extensively. Publishing these costs per SDQ improvements will contribute to defining a socially acceptable threshold. This is of course assuming a one-point improvement in SDQ scores can be linked to verifiable and useful outcomes in this context. If consensus were reached to determine that a one-point improvement is actually meaningless, then any cost per unit improvement is not a worthwhile investment and could arguably be better spent somewhere else in the education system, such as hiring more teachers or providing more class options.

Educational outcomes such as attainment would have benefited the RoE economic evaluation because local authorities will most likely end up making the funding decision for RoE. They will need to decide whether RoE represents value for money, and they are more likely to be interested in comparing costs in terms of educational outcomes in addition to health outcomes. The thresholds stated are what NICE considers to be cost-effective from an *NHS perspective*. Cost-effectiveness thresholds do not exist outside the health sector,[367] nor has a method been devised to apportion costs (who should bear them) when more than one government department or sector is involved.[30] This is particularly an issue when one sector benefits from a public health intervention while the

other is required to fund it. NICE does not make any recommendations for how costs should be apportioned, rather the methods chosen should be transparent and justified.[30] This trial was funded by the NIHR and delivered through the Public Health Trusts in Northern Ireland. In the event that the funding decision about RoE is transferred to local authorities, the collection of educational outcomes would have aided the decision-making process. Additionally, there is overwhelming evidence that education is linked to health and other outcomes[371] so collection of education outcomes would have provided further information to aid a decision.

A critique of the traditional gold standard RCT design with economic evaluation alongside, is that they often fail to compare all relevant options.[135] Trials are expensive to conduct and often times only compare two alternatives head to head, but economic evaluation seeks to compare *all relevant* options. Practically, this is where modelling can be more effective at incorporating more than two relevant alternatives. The systematic literature review did identify potential CUAs that could have been synthesised and modelled over the short-term to act as comparators to RoE. Assumptions would have to be made about the comparability of the different measures used for utilities and the differing contexts in each study. Modelling over the short-term with the existing data available on cost-effectiveness in the literature could assist decision-makers as they would have a more complete information of different options. However, as few SEL programmes were identified, programmes presented would have very different aims, from obesity prevention, to vaccination. This evaluation did not address the potential longer-term impacts of RoE through use of extrapolation or modelling of potential impacts over the child's lifetime. There is a paucity of longer-term evidence using the main outcomes of our analysis, the SDQ and CHU9D, especially the CHU9D which is a relatively new generic HRQoL measure. The main reason long-term modelling was not conducted was because there was no statistically significant difference in any effect measured in the trial at the third year follow-up. If any parameters remained statistically significant, these could have been extrapolated into the longer-term; however, this was not the case. The RoE trial did provide one of the longest follow-ups of any RoE evaluation identified, so the single trial was a sufficient source of immediate and mid-term data.

An immediate post-intervention significant difference in outcomes that wane over time is common in PHIs particularly those that involve behaviour change. A lack of statistical

significance in a child outcome in the mid-term does not necessarily mean the intervention will not have significant impacts on other adult outcomes. A prime example of this is the Perry Preschool Program.[28, 246, 402] The Perry Preschool Program is a highly cited example of early intervention having long-term impacts on adult outcomes such as education, employment, earnings and crime. It is one of few intervention studies to follow the 123 participants up to age 40. It also had low attrition with over 90% of the original sample participating in age 40 interviews.[27] During preschool years, the experimental group showed significant gains in IQ over the control, however those gaps narrowed when pupils entered school and differences in IQ eventually disappeared when pupils reached the age of 8 years.[403] By age 40 however, the experimental group significantly outperformed the control group on: highest level of schooling completed, being in employment, and having fewer lifetime arrests.[402] There were methodological issues with the randomisation of this study, and unfortunately there are few other experimental study designs that have as long a follow-up to replicate these findings. However, it does provide an example of potential 'sleeper effects' of early intervention. The idea that if significant differences in child outcomes wane over time, there is the potential for other important adult outcomes to 'wake up' when the child matures. This is a possibility for RoE, but such long-term follow-up will be costly and likely unfeasible. 'Sleeper effects' provide justification for preventive childhood intervention and offer explanation for the prevention paradox and for the small and often waning effects observed in PHIs over time during one to three-year follow-ups. Establishing their validity is key for this area of prevention and further research is required exploiting the use of long-term prospective cohort studies that follow children over their lifetime.

7.3.1.3 Mapping validation study

Mapping from a condition specific outcome to a generic preference-based HRQoL outcome is a second best option to directly measuring HRQoL, but it may be necessary in population health programmes for pragmatic reasons. The fundamental concern is that mapping methods assume overlap in each measure's descriptive systems[380] and that can be a difficult assumption to make when comparing a specific to a generic measure such as the SDQ and CHU9D. The RMSE of the mapping algorithms used in this outcomes piece indicated large differences between predicted and observed values at the individual level.[329] However, the purpose of using mapping methods in economic evaluation is to

predict differences across groups or between trial arms, not at the individual level. Additionally, the sensitivity analysis that used the mapping algorithm to estimate utilities in Chapter 5 resulted in unusually narrow confidence intervals indicating more certainty around the results and higher probability of being cost-effective. This indicates that the algorithm underestimates uncertainty. The lack of overlap between descriptive systems and the algorithms' potential to underestimate uncertainty requires careful consideration.

There are other mapping algorithms available to estimate child health utility, Khan and colleagues[309] generated mapping algorithms to predict EQ-5D-Y utility scores from the PedsQL. The authors similarly conclude the use of directly measured HRQoL is preferable to mapping, but when those measures are not available mapping can produce reasonable predictions. A critique of predicting and using EQ-5D-Y utility scores in CUA is that it is not structurally different to the adult EQ-5D (the only difference is the language which is more appropriate for children) and there are still no preference weights from children available to value the resulting utility scores that have been predicted from the PedsQL. Chen et al.[404] have developed a mapping algorithm to map from non-preference-based generic HRQoL KIDSCREEN-10 index to preference-based CHU9D. Two of the original developers of the CHU9D (Stevens, K) and corresponding adolescent preferences (Ratcliffe, J) were involved in this study. There is also ongoing research by Stavros and colleagues[310] to develop a preference-based index for the PedsQL. This would provide another child specific HRQoL measure much like the CHU9D. Depending on the preference elicitation methods used, it may even provide some of the first child values for children younger than adolescence (as is currently the case with the CHU9D). More recently, a mapping study has been published to map from the CHU9D to the PedsQL to estimate QALYs in studies where only PedsQL data is available.[405] This algorithm may be useful in specific contexts where only PedsQL data is available, however when planning a child health economic evaluation, the use of directly measured HRQoL and corresponding values (the CHU9D) is preferable to estimating child health utility using a mapping algorithm.

7.3.2 Strengths of study

Having critiqued the methods used throughout this thesis, this section now highlights the strengths of this body of work. Starting with the systematic review, the novelty and breadth of the review are two key strengths. No other review (from my knowledge) attempted to collect and synthesis current evidence around evaluation methods (economic and non-economic) of school-based health interventions. This was necessary to inform methods for future economic evaluations of school-based PHIs (as well as those used in this current body of work). The breadth of the review was critiqued in the last section, however there were merits for selecting a wide scope and from a purely theoretical perspective, the comprehensive review increases confidence in relevant evidence being identified. This also minimises potential selection bias as broad selection criteria were implemented. In a recent systematic review, a majority of existing systematic reviews of economic evaluations found that searches were not extensive enough to meet minimum requirements,[406] providing further justification for the comprehensive approach taken. Other biases and errors were mitigated by implementing a novel technique of validity checks performed by the entire review team. As is often the case with doctoral candidates, limited time and resources prevent the recruitment of a second researcher to be involved in each step of the systematic review process. To address this limitation and mitigate any potential bias and error, two validity checks were performed at two key stages of the review, a third strength of systematic review. A fourth and final strength of the systematic review was the inclusion of a narrative synthesis to formally synthesise the evidence gathered. Where meta-analysis is not possible, many authors of systematic reviews simply describe the results of the evidence gathered. This work took the descriptive analysis a step further by formally synthesising the evidence, a final key strength of the review.

A major methodological component of this thesis was the economic evaluation of RoE. The key strength of this work was the novelty of the context for which the economic evaluation was performed. It was the first comprehensive CUA which employed a paediatric preference-based HRQoL measure within a school-based context. Other elements of novelty include the intervention being a PHI and relevant implications for CUA, use of child-specific outcomes developed for children, and the application of adult and adolescent tariffs in economic evaluation to compare and assess differences between

the value sets. The adolescent values are supported as they are the closest to child values available and support the notion that decision-making affecting children should incorporate their values. The adolescent tariff also produces lower utility values than the adult tariff, so it provides a more conservative cost-effectiveness result, which instils more confidence in a cost-effective result. This is the opinion of the author, and whose values should be considered in economic evaluation is an important theoretical and policy question that will continue to be debated. This research not only adds to the available effectiveness evidence of RoE, but it also provides the first cost-effectiveness evidence of the programme, despite it being around for over 20 years. As it is particularly novel in this context, education decision makers may not be entirely sure how to interpret the evidence in comparison to other competing budget constraints. The key point is that this work acts as a catalyst to start the conversation of cost-effectiveness and economic evaluation within the school context. This work acts as a first step to attempt to influence decision-making in an education setting. Dissemination will take place primarily through publication in the academic literature. This work did not investigate how decision-making in education is actually conducted in real life, and further qualitative work would aid this understanding. Previously, decisions were at risk of being made based on emotional appeal, absolute intervention cost, and the political landscape.[370] A key example of this found in the literature is the D.A.R.E (Drug Abuse Resistance Education) programme in the USA. It was a large nationwide school-based drug abuse prevention programme that averaged three quarters of a billion dollars of federal expenditure annually.[407] Numerous studies called into question the programme's effectiveness and a systematic review and meta-analysis confirmed the programme to be ineffective.[408] This example illustrates the need to determine cost-effectiveness of school-based interventions before wider implementation, which in this example was across the USA.

Another strength of this study was the inclusion of the SDQ in CEA. The SDQ is becoming more widely used in child and adolescent mental wellbeing. The CEA provided some of the first cost-per-improvement in SDQ results to allow the pool of acceptable thresholds to accumulate as well as offer a direct comparison to a known threshold of cost per QALYs gained. This work will contribute to a growing pool of evidence to aid decision makers in interpreting their own cost-effectiveness results using the SDQ. The final strength of the economic evaluation was the extensive exploration of uncertainty in the base-case estimates of cost-effectiveness, through sensitivity analysis. Sixteen sensitivity

analyses were performed to address uncertainty as well as the overarching research question of which methods are best to determine cost-effectiveness within this unique context. CUA and CEA both have their merits, and while CCA and CBA are worth exploring, within the current British context, CUA will be preferred. In terms of addressing uncertainty, even when a lower threshold of £13,000 per QALY was adopted, over 80% (10/12) of the applicable sensitivity analyses demonstrated cost-effectiveness. This increases the certainty of RoE's cost-effectiveness in a within the trial context. However, what is still uncertain is its long-term impacts on cost-effectiveness.

A key finding from the outcomes study can address this issue of estimating longer-term impacts of interventions, given they have collected the SDQ. Strengths of this work mainly relate to the implications of the findings. The mapping algorithm which incorporates all five subscales of the SDQ predicts the adult values of the CHU9D well. This finding has two key implications: 1.) analysts are now afforded the opportunity to conduct CUA in paediatric or school-based programmes in the absence of a preference-based HRQoL measure, and 2.) the opportunity now exists to estimate longer-term child health utility as the SDQ is currently being routinely collected in CAMHS and long-term cohort studies such as the Millennium Cohort Study and the Copenhagen County Child Cohort. The former impacts economic evaluation in a within trial context and the latter a longer-term, modelling context. However, it was found that the mapping algorithms may underestimate uncertainty, so this should be carefully considered in a modelling context where uncertainty compounds over the longer lifetime horizons of children.

This work of this thesis will contribute to the sparse literature currently available on economic evaluation in a school-based context. It also illustrates the need for a more formal decision-making process with regards to incorporating evidence-based education programming. While contributing important and novel findings in the area, this thesis also gives rise to important questions about how decision-making in education should be addressed:

- How are real life decisions made in education settings?

- Should CBA, the most comprehensive approach be adopted as standard?

- Should the education sector learn from the health sector and adopt a generic education outcome to provide a common 'yardstick' measure for which all programmes can be compared?

- Are there other more appropriate emerging methodologies?

- Who should pay for the initial investment of interventions that have multi-sector benefits?

- How should potential long-term outcomes of preventive interventions be accounted for?

The final chapter will address these and additional questions raised throughout by providing recommendations, areas for further research, and overall conclusions.

8 Recommendations and Conclusion

8.1 Implications for policy and practice

The systematic literature review revealed that while there are school-based economic evaluations currently available in the literature, the quality of reporting leaves a considerable amount to be desired. As a minimum, a health economic evaluation should report each of the applicable CHEERS checklist items. Granted the majority of studies identified were published prior to the publication of the CHEERS checklist. With publication time lags, authors of studies published after 2013 may not have been aware of CHEERS because of its focus on health economic evaluation, which might not be on the radar of education economists. Still no relevant reporting guidance was found for conducting an education economic evaluation. If/when economic evaluation in school settings becomes more standardised, either the CHEERS or other relevant reporting guidance could be adopted to improve the quality of reporting.

From an NHS perspective, RoE is cost-effective based on QALY outcomes, and therefore on that basis alone, would be recommended for implementation. However, as it is a school-based intervention, cost-effectiveness from the education perspective needs to be established. Development of a generic education outcome may be warranted and estimation of a cost-effectiveness threshold based on this generic education outcome (or other relevant outcome in CEA) could be estimated three different ways. These include: estimating marginal productivity through econometric analyses of routinely available cost-effectiveness data (requires routinely collected established 'education outcome'); using the 'bookshelf analogy' mentioned in section 5.7; or by eliciting WTP from decision-makers. There may be an obligation for schools to fund the entirety of these programmes for future implementation as this is assumed to be the case in Education Endowment Foundation draft guidance.[365] Given that a number of government departments (education and health) could potentially benefit from the outcomes of RoE, there may be a case for joint contribution from both sectors to fund its future implementation. However, as there is no standard method to apportion these costs,[30] this may have implications on the future implementation of the programme if schools are required to fund the entirety of the intervention.

Additionally, less is known about the competing resource constraints in schools or the cost-effectiveness of other related SEL programmes that might actually represent better value for money. There were high RoE programme costs related to maintaining fidelity to Canadian model, and the effectiveness of the programme diminished over the three-year follow-up. Other SEL programmes may provide similar effectiveness results with fewer costs, but currently this is unknown because the economic evaluation of RoE was the first school-based CUA of a SEL programme. The relative novelty of this work in this context means that few other school-based programmes have been evaluated in a cost per QALY or cost per SDQ improvement framework, meaning that few alternatives are available to compare cost-effectiveness results from a RCT study design. It is possible other study designs evaluated the SDQ, however comparability may then be compromised as other non-randomised study designs are more prone to bias.

This work has paved the way for economic evaluation in school settings by providing an example of a thorough CEA and CUA using child specific outcome measures. However, this work has also highlighted the lack of clear funding decision rules for new and existing school programmes in terms of a cost per QALY, SDQ, or other relevant education outcomes. Allocation decisions in an education setting could potentially be made more explicit when economic evaluation of school programmes is performed. Monetary analysis using CBA is a starting point as theoretically it is the most comprehensive form of economic evaluation enabling multisector benefits to be accounted for in monetary terms. If an extrawelfarist framework is to be adopted (in terms of maximising education benefits, i.e. a generic education utility), important implications for policy and practice will be to determine an appropriate generic education outcome (utility) and thresholds for cost-effectiveness from the school's perspective as mentioned previously. With an established threshold, the cofinancing approach could actually be considered in practice as the education sector would have a value per unit of education utility they would be willing to pay to help fund an intervention. This would enable the development of transparent, consistent decision-making in education as well as a more equitable method for funding cost-effective programmes in education. The work of this thesis demonstrates the importance of thoroughly evaluating new programmes or interventions for cost-effectiveness before school-wide or nationwide implementation. It is currently unknown if decision-making in education follows a clear and consistent process. In light of this, education policy could pose a requirement for all new school-based programmes to be

evaluated for effectiveness and cost-effectiveness before school-wide or nationwide implementation. This could potentially save billions in wasted resource as would have been the case with the D.A.R.E. programme (see Section 7.3.2).

8.2 Recommendations

Considering the findings, the following recommendations are made regarding the economic evaluation of school-based population health interventions.

- Economic evaluation of future school-based PHIs could clearly and consistently report a set of standardised and appropriate items relevant for an economic evaluation. If the economic evaluation is primarily concerned with health outcomes, it could as a minimum report all applicable CHEERS checklist items. If the PHI is primarily concerned with multi-sector or education benefits, then new guidance could be developed which could borrow applicable elements from CHEERS.

- Schools and/or education authorities could consider requiring economic evaluation of new programmes before wider implementation. This is especially important if substantial investment is required. Key elements of education economic evaluations are reported in the following two recommendations below.

- This thesis recommends CBA as the type of economic evaluation for school-based programmes or interventions due to it being the most comprehensive form of economic evaluation. CBA can take into account multi-sector outcomes (a key challenge for economic evaluation in this setting) as it allows monetary valuation of these outcomes in a final cost to benefits ratio or net benefit/loss. This type of analysis is feasible given appropriate time and funding, and provides results that enable clear, consistent, decision-making criteria.

- If CBA is not feasible (due to some of the limitations associated with this method as previously mentioned), then a mix of CCA, CUA, and MCDA are recommended as alternative methods to evaluate cost-effectiveness of school-based PHIs.

- CCA is recommended alongside another evaluation method, as it requires less resource input and does not present a single combined result in terms of costs and effects (therefore enabling clear and consistent decision-making criteria).

- CUA is recommended when the PHI is primarily concerned with health benefits. However, currently accepted decision rules for cost-effectiveness from an NHS perspective may not be acceptable in an education setting. Therefore, development of an appropriate generic education utility measure, its values, and cost-effectiveness threshold is recommended.

- MCDA is recommended for PHIs where emphasis for the evaluation is not based on a primary outcome but places more emphasis on incorporating multiple criteria to improve transparency in decision-making. Methods and standardisation are still in a developmental stage; however, this is an area worthy of further advancement.

- When conducting CUA on a child-focussed health intervention, utility could be directly measured using the CHU9D and valued with adolescent values. Development of preference weights for younger children is currently ongoing, and more age appropriate values could be considered for younger children in the future. However, use of the adolescent values is currently recommended for all child-focussed interventions as values of young people should be considered in decision-making that will ultimately affect them.

- Direct measurement of utility should be the gold standard, where and if this is unavailable the SDQ could be used to estimate child health utility as it can be reliably mapped to the CHU9D.

- Further investigation is needed to understand the 'real life' decision-making context in education. Decision-making could be made more transparent and consistent by adopting a set of criteria by which all funding and allocation decisions will be made. The criteria could be based on the results of one or multiple methods for economic evaluation as recommended above.

- It is difficult to expect the onus of funding school-based programmes to rely entirely on the education sector when the particular programme generates multisector benefits. Development of an equitable method for distributing the costs of funding preventive PHIs, such as cofinancing, amongst the sectors that stand to benefit, should be explored further.

- Further work is needed to establish links from short-term trials whose effects may wane over time to potential long-term adult outcomes such as 'sleeper effects.' Lifetime follow-up of an early intervention RCT would provide the strongest evidence. If this is not feasible, data linkage and long-term birth cohort studies could provide insight into these potential 'sleeper effects.'

8.3 Areas for further research

8.3.1 Practical application of appropriate methodology to evaluate PHIs

Four different methodologies have been recommended for the economic evaluation of school-based PHIs: CBA, CCA, CUA, and MCDA. CBA has been recommended as the gold standard, most comprehensive evaluation method as it has the ability to account for multi-sector benefits while providing a single combined cost-effectiveness result. However, CBA requires more time and resources in the designing and implementation of the analysis as more outcomes need to be identified, measured, and valued appropriately. Valuation takes more time as the broader outcomes are likely to be unique to each evaluation. To take this recommendation forward, researchers need to be aware of additional time and resource implications, and build this into funding applications. Funders and decision-making bodies should also encourage the use and advancement of CBA in PHI contexts.

NICE currently suggests the use of CUA in the evaluation of PHIs due to it providing a common 'yardstick' measure to compare the health outcome components of PHIs. NICE does consider that CBA and CCA might be a more appropriate way to measure non-health outcomes, but still requires a cost per QALY. In recent social care guidance, NICE has stated it will consider outcomes other than the QALY such as the 'social care QALY' or the ASCOT if validated. NICE would also consider parallel analyses incorporating capabilities

such as the ICECAP.[151] This openness to consider additional outcomes other than the QALY could signify an openness to consider CBA on its own without an accompanying CUA. CUA is not as useful in an education decision-making context if health is not the primary outcome, even if the programme is primarily concerned with health, such as RoE, decisions about cost-effectiveness and currently accepted thresholds cannot be expected to be the same in education settings. Therefore, if a CUA approach is to be adopted in the education sector further research is required to determine an appropriate generic education utility, its values, and cost-effectiveness threshold.

MCDA is an emerging methodology that holds promise for this context, as ISPOR has issued several recent guidance documents.[160, 161] A disadvantage is that it requires considerable cognitive burden on decision makers to provide weights for each individual criterion. There is the option to standardise weights, but then the multiple criteria start to become restricted and MCDA starts to act a lot like CUA where the disadvantage is that broader outcomes cannot be incorporated. The systematic literature review revealed no school-based evaluation methodologies employing MCDA suggesting a potential research time-lag in the use of MCDA, or the existing guidance may not be sufficient for researchers new to the method to confidently conduct MCDA. More research to further the development and standardisation of MCDA is needed.

Economic evaluation alongside a natural experiment design offers the advantage that no new data is required to be collected if routine sources are available. The disadvantage is that randomisation is not possible and researchers have no input into group allocation or the data that is collected. Practical application of this study design will be challenging and exploratory as there is no current guidance for the conduct of economic evaluation of natural experiments available. Development of methodological guidance for this area is justified.

A final practical issue that could be explored in further research is the utilisation of routine data sources to provide more reliable societal costs and benefits of PHIs. In the RoE trial, the attempt to collect wider societal costs and benefits was hindered by high percentages of missing data. The only method available to capture wider outcomes was through contact of the children's parents by post. This method proved difficult and was prone to producing missing and potentially inaccurate data. Further research could

explore potential cost-effectiveness of utilising routinely collected data to estimate societal costs and benefits. Currently, issues of access to routine data have hindered exploitation of this resource. Extracting the data can be time consuming and costly, and there are always concerns over patient confidentiality and data breeches. However if these issues are addressed and incorporated into the research process, benefits might include more accurate and complete data on resource use.

8.3.2 Defining resource allocation decision criteria in education

This thesis did not thoroughly investigate how the decision-making process is conducted in 'real life' education contexts and further research is required. What is lacking in the literature is reference to consistent and transparent resource allocation criteria. Clear decision criteria could be defined for educational settings to promote consistency and prevent any misuse of limited resources. Defining these criteria will depend on the economic evaluation method selected as most appropriate for school-based PHIs and other school programmes. If CBA is selected, decision criteria could be as simple as only considering programmes whose net benefit is positive. If CUA is selected, decision makers need to decide if they are willing to accept currently accepted cost-effectiveness thresholds for QALYS which don't currently take into account non-heath benefits. If they are not willing to accept this, or would like to incorporate non-health benefits, then further work is needed to develop an appropriate generic education utility, values for this utility, and an acceptable cost-effectiveness threshold. It is more difficult to define decision rules for CCA as decision-makers need to weigh up cost and benefits based on their specific needs. Therefore, to promote consistency in the decision making process, it is recommended CCA be incorporated as an additional analysis.

Decision-making in terms of SEW could be standardised if consensus were to be reached in terms of the clinical significance of incremental changes in SDQ. The economic evaluation of RoE has contributed to a small but growing pool of incremental costs per unit change in SDQ and these results will contribute to the estimation or valuation of those incremental changes, in addition to eliciting WTP directly from decision-makers. The SDQ has been mentioned over 4,000 times in the published literature, if any of those studies included or were concerned with cost-effectiveness, then valuation of this SEW outcome is warranted.

8.3.3 Determining long-term cost-effectiveness of SEL programmes

The potential long-term cost-effectiveness of RoE is still unknown. The short-term effectiveness results at 12 months and beyond were not sustained and were insignificant at the 36-month follow-up. Therefore, because all outcomes observed in the trial were not statistically significantly different at the final follow-up, extrapolation beyond this time period was not warranted. Viable extrapolation would be based on multiple uncertain assumptions, and the extrapolation time horizon is longer over a child's lifetime creating more uncertainty. The potential validation of 'sleeper effects' could mean that initial effectiveness results could be linked to long-term adult outcomes such as education attainment, health behaviours, unemployment, and crime. However, reliable long-term evidence either from a RCT or birth cohort study is needed to establish if any sleeper effects are likely.

An obstacle is the lack of sources containing long-term effects of the childhood measures used in the RoE economic evaluation. The CHU9D was only validated in 2012, which means there would only be potentially five years of CHU9D data. The SDQ however, has been routinely collected in the Millennium Cohort Study, in which participants will be reaching adulthood shortly. With the validation of the SDQ to CHU9D mapping algorithm, it may be possible estimate long-term child health utility from cohort studies that have collected SDQ scores. This is an exciting area for further research into the estimation of long-term child health utility.

8.4 Conclusion

The overarching research question asked, 'How should the cost-effectiveness of school-based, population health interventions aimed at children be determined?' The work of this thesis determined CBA to be the most comprehensive method for determining cost-effectiveness of school-based PHIs. The overall aims of this thesis were to:

(i) determine what evaluation methods (economic and non-economic) are currently being used to evaluate school-based population health interventions;

(ii) illustrate a good practice example of a thorough cost-utility and cost-effectiveness analysis of a school-based intervention (the RoE programme) to reflect on the advantages of such practice and disadvantages that remain, such as decision-making in multisectoral settings; and

(iii) explore which outcomes are appropriate for children in the SEW and economic evaluation context to support future evaluation work in this context.

This thesis was split into three empirical works to address each part of the aim. The systematic literature review revealed that current evaluation methodologies of school-base PHIs were varied, quality of reporting of health economic evaluation was poor, and no emergent methodologies such as MCDA were identified.

The case study economic evaluation provided an example of the practical application of a CUA and CEA in line with current NICE recommendations for determining cost-effectiveness using child specific outcome measures. From a health services perspective, RoE is cost-effective at £11,000 per QALY (CI: -£95,500 to £147,000). It is unknown if this result would be acceptable from an education perspective as no consensus has been reached if a decision-making threshold should be established for the education sector and what generic outcome should be used. Further research is required to understand how funding allocation decisions are made in education and how this process could be made more transparent. CEA using the SDQ resulted in an ICER of £197 per unit decrease in total difficulties score (CI: £77 to £471). It is unknown how this result would be interpreted in a health or education decision-making context, however this study has contributed to the growing pool of incremental costs per SDQ improvement, which will

aid the valuation of those incremental changes if at first, consensus can be reached on the clinical significance of those incremental changes.

Directly measuring child health utility using the CHU9D is preferable as it is the only measure to have been developed specifically for children and valued by adolescents (eliciting younger children's values is currently ongoing). However, when traditional utility measurement methods are missing, mean utility can be predicted by mapping from the SDQ as the final empirical study validated this mapping algorithm. When applying this mapping algorithm in economic evaluation, analysts should be cautious of overly narrow confidence intervals from resulting ICERS as the algorithm has a tendency to underestimate uncertainty. The validated mapping algorithm allows analysts the opportunity to conduct CUA in paediatric or school-based programmes where previously this would have been challenging due to a lack of preference-based outcome measures. This also affords the opportunity to estimate longer-term utility by utilising long-term cohort data that routinely collects SDQ outcomes. The school plays an important role in shaping our young people's futures. Economic evaluation of school-based PHIs is justified, as schools need to maximise their existing resources in order to give children the best start in life.

Appendices

Appendix 1: Description of Electronic Database Searches

CINAHL was searched using the EBSCOhost interface on 16/07/15, no date restriction was applied (1981)

S1 (MH "Psychological Well-Being") OR (MH "Well-Being (Iowa NOC)") OR (MH "Psychological Well-Being (Iowa NOC) (Non-Cinahl)+") OR (MH "Family Member Well-Being Index") OR "(emotion* OR social*) AND (learn* OR wellbeing OR “well being”)" OR (MH "Wellness") (11,047)
S2 (MH "Program Development/EC/ED/EV") OR (MH "International Classification of Functioning, Disability, and Health/EC/ED/EV") OR "2. (improve OR develop) AND (health OR academ* OR mental* OR physical*)" OR (MH "Student Health Services/EC/ED/EV") OR (MH "Community Health Centers/EC/ED/EV") OR (MH "Health Resource Utilization/EC/ED/EV") OR (MH "Child Development Disorders, Pervasive/EC/ED") OR (MH "Child Development Disorders+") OR (MH "Child Health/ED/EV") (2,966)
S3 S1 OR S2 (13,995)
S4 (MH "Schools, Middle") OR (MH "Students, High School") OR (MH "Students, Middle School") OR (MH "High School Graduates") OR (MH "Schools, Secondary") OR "(primary OR secondary OR elementary OR junior OR middle OR high) AND (school*)" OR (MH "Schools, Elementary") OR (MH "Students, Elementary") OR (MH "Child Development: Middle Childhood (6-11 Years) (Iowa NOC)") OR (MH "Unsafe Sex") OR (MH "Risk for Violence, Self-Directed or Directed at Others (NANDA)") (13,306)
S5 (MH "Schools") OR "school* OR educat* OR academ*" (4,775)
S6 (MH "Child Psychology") OR (MH "Child Psychiatry") OR (MH "Adolescent Nutritional Physiology") OR "child* OR adolescent OR p?ediat*" OR (MH "Adolescent Psychology") OR (MH "Adolescent Psychiatry") OR (MH "Pediatric Obesity") OR (MH "Physical Therapy Practice, Research-Based") OR (MH "Child Nutritional Physiology") (10,746)
S7 (MH "Course Evaluation") OR (MH "Early Childhood Intervention") OR (MH "Course Content") OR (MH "Early Intervention") OR (MH "Curriculum Development") OR (MH "Integrated Curriculum") OR (MH "Curriculum") OR "program* OR intervention OR curriculum OR course" OR (MH "Intervention Scheme (Omaha)") OR (MH "Crisis Intervention") OR (MH "Intervention Trials") (38,282)
S8 S3 AND S4 AND S5 AND S6 AND S7 (51,449)
S9 (MH "Cost Benefit Analysis") OR (MH "Costs and Cost Analysis") OR (MH "Health Care Costs") OR (MH "Cost Control") OR (MH "Cost Savings") OR (MH "Economic Aspects of Illness") OR "(economic*AND eval*) OR (“cost effective*”) OR (“cost benefit”) OR (“cost utility”) OR (“cost consequence”)" OR (MH "Health Resource Utilization") OR (MH "Substance Addiction Consequences (Iowa NOC)") OR (MH "Therapy, Computer Assisted") OR (MH "Sports Science") OR (MH "Outcomes of Education") OR (MH "Health Resource Allocation") (77,829)
S10 (MH "Health and Life Quality (Iowa NOC) (Non-Cinahl)") OR (MH "Quality-Adjusted Life Years") OR (MH "Quality of Life") OR (MH "Quality of Life (Iowa NOC)") OR (MH "Ferrans and Powers Quality of Life Index") OR (MH "Quality of Health Care") OR (MH "Quality of Care Research") OR "model* OR utility* OR (“quality of life”) OR (“health related quality of life”) OR (“return on investment”) OR (“social return on investment”)" OR (MH "Quality Assessment") OR (MH "Economic Value of Life") OR (MH "Evaluation and

Quality Improvement Program") OR (MH "Life Cycle") OR (MH "Lifelong Learning") (88,989)
S11 S9 AND S10 (7,994)
S12 S8 AND S11 (120)
S13 (MM "Health Impact Assessment") OR "(economic*AND eval*) OR (“cost effective*”) OR (“cost benefit”) OR (“cost utility”) OR (“cost consequence”) OR model* OR (“decision tree”) OR (“decision analytic model”) OR (“return on investment”) OR (“social return on investment”) OR (“budget impact analysis”) OR (social impact assessment”) OR (“health impact assessment”) OR (“discrete choice”) OR (“stated preference”) OR (“multi-criteria decision analysis”)" OR (MM "Cost Benefit Analysis") OR (MM "Costs and Cost Analysis") OR (MM "Decision Support Techniques") OR (MH "Health Care Costs") OR (MM "Decision Trees") OR (MH "Social Network Analysis (Saba CCC)") OR (MH "Decision Making") (501)
S14 "utility* OR (“quality of life”) OR (“health related quality of life”) OR (“quality adjusted life year”) OR (“disability adjusted life year”) OR (“net benefit*”) OR (“net present value”) OR cost* OR (“resource use”) OR fund* OR benefit* OR effect* OR “contingent valuation” OR “willingness to pay” OR “human capital”" OR (MM "Quality-Adjusted Life Years") OR (MM "Health and Life Quality (Iowa NOC)") OR (MM "Economic Value of Life") OR (MM "Quality of Life") OR (MH "Adolescent-Family Inventory of Life Events and Changes") (21,303)
S15 S9 OR S13 (78,268)
S16 S10 OR S14 (88,996)
S17 S15 AND S16 (8,027)
S18 S8 AND S17 (16) Exported to Endnote

The Cochrane Library was searched on 22/05/15, no date restriction was applied

#1 MeSH descriptor: [Psychology, Social] explode all trees (15257)
#2 (emotion* or social*) and (learn* or wellbeing or "well being") (5018)
#3 (improve or develop) and (health or academ* or mental* or physical*) (28403)
#4 #1 or #2 or #3 (44611)
#5 school* or educat* or academ* (128706)
#6 (primary or secondary or elementary or junior or middle or high) and (school*) (43117)
#7 child* or adolescent or p?ediat* (157987)
#8 program* or intervention or curriculum or course (154614)
#9 #4 and #5 and #6 and #7 and #8 (3001)
#10 (economic*AND eval*) or ("cost effective*") or ("cost benefit") or ("cost utility") or ("cost consequence") (31104)
#11 model* or utility* or ("quality of life") or ("health related quality of life") or ("return on investment") or ("social return on investment") (108172)
#12 #10 and #11 (15714)
#13 #9 and #12 (419) Exported to Endnote

ERIC was searched using the EBSCOhost interface on 16/07/15, no date restriction was applied

S1 (emotion* OR social*) AND (learn* OR wellbeing OR “well being”) (85,454)
S2 (improve OR develop) AND (health OR academ* OR mental* OR physical*) (41,581)
S3 school* OR educat* OR academ* (1,272,403)
S4 (primary OR secondary OR elementary OR junior OR middle OR high) AND (school*) (444,603)

S5 child* OR adolescent OR p?ediat* (1,523,815)
S6 program* OR intervention OR curriculum OR course (704,690)
S7 crim* OR (“criminal justice”) OR famil* (140,888)
S8 S1 OR S2 OR S7 (245,788)
S9 S3 AND S4 AND S5 AND S6 AND S8 (49,319)
S10 (economic*AND eval*) OR (“cost effective*”) OR (“cost benefit”) OR (“cost utility”) OR (“cost consequence”) OR model* OR (“decision tree”) OR (“decision analytic model”) OR (“return on investment”) OR (“social return on investment”) OR (“budget impact analysis”) OR (social impact assessment”) OR (“health impact assessment”) OR (“discrete choice”) OR (“stated preference”) OR (“multi-criteria decision analysis”) (202,536)
S11 utility* OR (“quality of life”) OR (“health related quality of life”) OR (“quality adjusted life year”) OR (“disability adjusted life year”) OR (“net benefit*”) OR (“net present value”) OR cost* OR (“resource use”) OR fund* OR benefit* OR effect* OR “contingent valuation” OR “willingness to pay” OR “human capital” (485,700)
S12 S10 AND S11 (86,710)
S13 S9 AND S12 (264) Exported to Endnote

MEDLINE was searched using the Ovid interface on 16/07/15, no date restriction was applied (1946 to February Week 2 2015)

1. (emotion* or social*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (456,271)
2. (learn* or well?being).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (238,234)
3. 1 and 2 (30,284)
4. (improve or develop).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (716,326)
5. (health or academ* or mental* or physical*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (2,031,573)
6. 4 and 5 (195,757)
7. (crim* or criminal justice or famil*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (723,064)
8. 3 or 6 or 7 (919,578)
9. (school* or educat* or academ*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (601,969)
10. (primary or secondary or elementary or junior or middle or high).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (4,623,768)
11. school*.mp. (149,259)
12. 10 and 11 (68,859)

13. (child* or adolescent or p?ediat*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (1,634,123)
14. (program* or intervention or curriculum or course).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (1,066,300)
15. 8 and 9 and 12 and 13 and 14 (5,336)
16(economic*AND eval* or cost effective* or cost benefit or cost utility or cost consequence or model* or decision tree or decision analytic model or return on investment or social return on investment or budget impact analysis or social impact assessment or health impact assessment or discrete choice or stated preference or multi-criteria decision analysis).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (2,247,848)
17. (utilit* or cost* or fund* or benefit* or effect*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (3,947,557)
18. (<quality of life> or <health related quality of life> or <quality adjusted life year> or <disability adjusted life year> or <net benefit*> or <net present value> or <resource use> or <contingent valuation> or <willingness to pay> or <human capital>).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (205,812)
19. 17 or 18 (4,046,146)
20. 16 and 19 (976,895)
21. limit 20 to (english language and full text and humans and "all child (0 to 18 years)" and "economics (best balance of sensitivity and specificity)") (10,087)
22. 15 and 21 (99)

PsychINFO was searched using the EBSCOhost interface on 22/05/15, no date restriction was applied (Journal coverage from 1800s to present)

S1 DE "Well Being" OR DE "Early Intervention" OR DE "Social Support" OR DE "Anxiety Disorders" (76841)
S2 DE "Health Promotion" OR DE "Health Knowledge" OR DE "Health Education" OR DE "Health Behavior" OR DE "Conduct Disorder" OR DE "Early Intervention" OR DE "Psychological Development" OR DE "Delayed Development" OR DE "Mental Health Program Evaluation" OR DE "Child Guidance" OR DE "Community Mental Health" OR DE "Child Care Workers" (66390)
S3 S1 OR S2 (131937)
S4 DE "Social Approval" OR DE "Help Seeking Behavior" OR DE "School Psychologists" OR DE "School Based Intervention" (20066)
S5 DE "Middle School Students" OR DE "Secondary Education" OR DE "Primary Mental Health Prevention" OR DE "Intermediate School Students" OR DE "School Psychologists" OR DE "Primary School Students" OR DE "Boarding Schools" OR DE "Multicultural Education" OR DE "Social Studies Education" OR DE "Junior High School Students" OR DE "Performance" OR DE "Course Evaluation" OR DE "Junior High Schools" OR DE "Junior High School Teachers" OR DE "Juvenile Delinquency" (67602)

S6 DE "Adolescent Psychopathology" OR DE "Juvenile Justice" OR DE "Juvenile Delinquency" OR DE "Adolescent Psychology" OR DE "Adolescent Pregnancy" OR DE "Adolescent Development" OR DE "Adolescent Psychotherapy" OR DE "Adolescent Psychiatry" OR DE "At Risk Populations" OR DE "Course Evaluation" OR DE "Educational Program Evaluation" OR DE "School Psychologists" OR DE "Child Psychopathology" (101035)
S7 DE "Course Evaluation" OR DE "School Based Intervention" OR DE "Early Intervention" OR DE "Family Intervention" OR DE "Curriculum Development" OR DE "Educational Objectives" OR DE "Intervention" OR DE "Group Intervention" OR DE "Curriculum" OR DE "Educational Counseling" OR DE "Behavior Change" OR DE "Health Promotion" OR DE "Psychotherapeutic Outcomes" OR DE "Psychological Assessment" (126761)
S8 S3 AND S4 AND S5 AND S6 AND S7 (70)
S9 DE "Costs and Cost Analysis" OR DE "Cost Containment" OR DE "Health Care Costs" OR DE "Health Care Economics" OR DE "Early Intervention" (economic*AND eval*) OR (“cost effective*”) OR (“cost benefit”) OR (“cost utility”) OR (“cost consequence”) OR DE "School Based Intervention" (62257)
S10 DE "Time Perspective" OR DE "Psychological Assessment" OR DE "Quality of Care" OR DE "Quality of Life" OR DE "Mental Health Program Evaluation" OR DE "Relationship Quality" (54239)
S11 S9 AND S10 (2785)
S12 DE "School Based Intervention" (10612)
S13 S9 AND S15 (28) Exported to Endnote

Web of Science was searched using the Core Collection interface on 22/05/15, no date restriction was applied

#1 TOPIC: (emotion* OR social*) Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, CCR-EXPANDED, IC Timespan=All years (1046458)
#2 TOPIC: (learn* OR well?being) Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, CCR-EXPANDED, IC Timespan=All years (621095)
#3 TOPIC: (crim* OR (“criminal justice”) OR famil*) DocType=All document types; Language=All languages; (58970)
#4 #3 OR #2 OR #1 DocType=All document types; Language=All languages; (4939163)
#5 TOPIC: (school* OR educat* OR academ*) DocType=All document types; Language=All languages; (2646697)
#6 TOPIC: ((primary OR secondary OR elementary OR junior OR middle OR high) AND (school*)) DocType=All document types; Language=All languages; (552141)
#7 TOPIC: (child* OR adolescent OR p?ediat*) DocType=All document types; Language=All languages; (605619)
#8 TOPIC: (program* OR intervention OR curriculum OR course) DocType=All document types; Language=All languages; (1292500)
#9 #8 AND #7 AND #6 AND #5 AND #4 DocType=All document types; Language=All languages; (8975696)
#10 TOPIC: ((economic*AND eval*) OR (“cost effective*”) OR (“cost benefit”) OR (“cost utility”) OR (“cost consequence”) OR model* OR (“decision tree”) OR (“decision analytic model”) OR (“return on investment”) OR (“social return on investment”) OR (“budget impact analysis”) OR (“social impact assessment”) OR (“health impact assessment”) OR (“discrete choice”) OR (“stated preference”) OR (“multi-criteria decision analysis”)) DocType=All document types; Language=All languages; (394642)

#11 TOPIC: (utilit* OR (“quality of life”) OR (“health related quality of life”) OR (“quality adjusted life year”) OR (“disability adjusted life year”) OR (“net benefit*”) OR (“net present value”) OR cost* OR (“resource use”) OR fund* OR benefit* OR effect* OR “contingent valuation” OR “willingness to pay” OR “human capital”) DocType=All document types; Language=All languages; (170748)
#12 #11 AND #10 DocType=All document types; Language=All languages; (1421934)
#13 #12 AND #9 DocType=All document types; Language=All languages; (223993)
#14 #12 AND #9Refined by: RESEARCH AREAS: (PEDIATRICS OR EDUCATION EDUCATIONAL RESEARCH) DocType=All document types; Language=All languages; (5700)
#15 #12 AND #9Refined by: RESEARCH AREAS: (PEDIATRICS OR EDUCATION EDUCATIONAL RESEARCH) AND DOCUMENT TYPES: (ARTICLE OR CLINICAL TRIAL OR REVIEW) DocType=All document types; Language=All languages; (229585)
#16 #12 AND #9Refined by: RESEARCH AREAS: (PEDIATRICS OR EDUCATION EDUCATIONAL RESEARCH) AND DOCUMENT TYPES: (ARTICLE OR CLINICAL TRIAL OR REVIEW) AND LANGUAGES: (ENGLISH) DocType=All document types; Language=All languages; (200) Exported to Endnote

Health Technology Assessment Database, DARE AND NHS EED were searched using the University of York Centre for Reviews and Dissemination interface on 22/05/15, no date restriction was applied

1 MeSH DESCRIPTOR Population EXPLODE ALL TREES (140)
2 MeSH DESCRIPTOR Schools EXPLODE ALL TREES (180)
3 MeSH DESCRIPTOR Child EXPLODE ALL TREES (4567)
4 MeSH DESCRIPTOR Early Intervention (Education) EXPLODE ALL TREES (36)
5 MeSH DESCRIPTOR Health Promotion EXPLODE ALL TREES (823)
6 MeSH DESCRIPTOR Curriculum EXPLODE ALL TREES (38)
7 #1 OR #2 OR #3 OR #4 OR #5 OR #6 (5384)
8 MeSH DESCRIPTOR Cost-Benefit Analysis EXPLODE ALL TREES (12838)
9 MeSH DESCRIPTOR Quality of Life EXPLODE ALL TREES (2248)
10 #8 AND #9 (1041)
11 #7 AND #10 (166) Exported to Endnote

EconLit was searched using the EBSCOhost interface on 17/07/15, no date restriction was applied (Journal coverage from 1800s to present)

#S1 (emotion* OR social*) AND (learn* OR wellbeing OR “well being”) (18,341)
#S2 (improve OR develop) AND (health OR academ* OR mental* OR physical*) (8,097)
#S3 crim* OR (“criminal justice”) OR famil* (51,628)
#S4 S1 OR S2 OR S3 (74,251)
#S5 school* OR educat* OR academ* (170,695)
#S6 (primary OR secondary OR elementary OR junior OR middle OR high) AND (school*) (12,661)
#S7 child* OR adolescent OR p?ediat* (1,291,718)
#S8 program* OR intervention OR curriculum OR course (96,305)
#S9 S4 AND S5 AND S6 AND S7 AND S8 (735)
#S10 (economic*AND eval*) OR (“cost effective*”) OR (“cost benefit”) OR (“cost utility”) OR (“cost consequence”) OR model* OR (“decision tree”) OR (“decision analytic model”) OR (“return on investment”) OR (“social return on investment”) OR (“budget impact

analysis") OR ("social impact assessment") OR ("health impact assessment") OR ("discrete choice") OR ("stated preference") OR ("multi-criteria decision analysis") (345,972
#S11 utilit* OR ("quality of life") OR ("health related quality of life") OR ("quality adjusted life year") OR ("disability adjusted life year") OR ("net benefit*") OR ("net present value") OR cost* OR ("resource use") OR fund* OR benefit* OR effect* OR "contingent valuation" OR "willingness to pay" OR "human capital" (502,451)
#S12 S10 AND S11 (177,671)
#S13 S9 AND S12 (149) Results Exported to Endnote

Appendix 2: Systematic literature review validity checks

Validity Check 1
Title and Abstract stage

Study Number	Excluded from Review	Context Related, but Excluded from Review	Included in Filter 1	Included in Filter 2
1		NKE		
2			NKE	
3		NKE		
4				NKE
5			NKE	
6	NKE			
7	NK	E		
8	NKE			
9	NKE			
10			NKE	
11			NKE	
12	NKE			
13	NKE			
14			NKE	
15		NKE		
16	NKE			
17			NKE	
18	NKE			
19	NKE			
20			NKE	

N – Nicki Boyer
K – Kathleen Boyd
E – Emma McIntosh

Study Information

1. Redd Z, Boccanfuso C, Walker K, et al. Expanding Time for Learning Both inside and outside the Classroom: A Review of the Evidence Base: Child Trends;2012.
2. Chavez DX, Washburn KJ, Texas State Dept. of Human Services A. Children's Antivictimization Education Project. Final Report1986.
3. Anderson P, Jane-Llopis E, Hosman C. Reducing the silent burden of impaired mental health INTRODUCTION. Health promotion international. Dec 2011;26:I4-I9.
4. Barnes RW, District of Columbia Public Schools WDCDoQA, et al. An Evaluation of the Youth Awareness Program (YAP), School Year 1983-841984.

5. Goldberg JP, Folta SC, Eliasziw M, et al. Great Taste, Less Waste: A cluster-randomized trial using a communications campaign to improve the quality of foods brought from home to school by elementary school children. Preventive medicine. 2015;74:103-110.
6. National Council on Teacher Q. State Teacher Policy Yearbook: Progress on Teacher Quality, 2007. Colorado State Summary: National Council on Teacher Quality;2007.
7. Scott S, O'Connor TG, Futh A, Matias C, Price J, Doolan M. Impact of a parenting program in a high-risk, multi-ethnic community: the PALS trial. Journal of Child Psychology & Psychiatry & Allied Disciplines. 2010;51(12):1331-1341.
8. Zahran S, Mielke HW, Weiler S, Berry KJ, Gonzales C. Children's blood lead and standardized test performance response as indicators of neurotoxicity in metropolitan New Orleans elementary schools. Neurotoxicology. Nov 2009;30(6):888-897.
9. Thiry N, Beutels P, Van Damme P, Van Doorslaer E. Economic evaluations of varicella vaccination programmes - A review of the literature. Pharmacoeconomics. 2003;21(1):13-38.
10. Rosenkranz RR, Lubans DR, Peralta LR, Bennie A, Sanders T, Lonsdale C. A cluster-randomized controlled trial of strategies to increase adolescents' physical activity and motivation during physical education lessons: the Motivating Active Learning in Physical Education (MALP) trial. BMC Public Health. 2012;12:834.
11. Bonell C, Allen E, Christie D, et al. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial. Trials [Electronic Resource]. 2014;15:381.
12. Adair LS, Fall CHD, Osmond C, et al. Associations of linear growth and relative weight gain during early life with adult health and human capital in countries of low and middle income: findings from five birth cohort studies. Lancet. Aug 10 2013;382(9891):525-534.
13. Gagliardi L. Examining the Scholastic READ 180 Program Teachers' Perceptions Regarding Local Setting Factors and Role of the Teacher Impacting the Program's Implementation in Seventh Grade at Three Middle Schools, ProQuest LLC; 2011.
14. Kushman J, Hanita M, Raphael J, National Center for Education E, Regional A. An Experimental Study of the Project CRISS Reading Program on Grade 9 Reading Achievement in Rural High Schools. Final Report NCEE 2011-4007: National Center for Education Evaluation and Regional Assistance;2011.
15. Stables GJ, Young EM, Howerton MW, et al. Small school-based effectiveness trials increase vegetable and fruit consumption among youth. Journal of the American Dietetic Association. Feb 2005;105(2):252-256.
16. Weindling AM, Cunningham CC, Glenn SM, Edwards RT, Reeves DJ. Additional therapy for young children with spastic cerebral palsy: a randomised controlled trial. Health technology assessment (Winchester, England). 2007;11(16):iii-iv, ix-x, 1-71. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/321/CN-00588321/frame.html
17. Wright R, Offord D, John L, Duku E, DeWit D. Secondary Schools Demonstration Project: Program Effects of School-Based Interventions on Antisocial Behaviour. Exceptionality Education Canada. 01/01/ 2005;15(2):27-50.
18. Apps P, Rees R. Taxation, Labour Supply and Saving: Centre for Economic Policy Research, Research School of Social Sciences, Australian National University, CEPR Discussion Papers: 590; 2008.
19. Ryzin MJ, Dishion TJ. The impact of a family-centered intervention on the ecology of adolescent antisocial behavior: modeling developmental sequelae and trajectories during adolescence. Development and psychopathology. 2012;24(3):1139-1155. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/474/CN-00848474/frame.html. 20. Spoth R, Trudeau L, Guyll M, Shin C, Redmond C. Universal

intervention effects on substance use among young adults mediated by delayed adolescent substance initiation. Journal of Consulting & Clinical Psychology. 2009;77(4):620-632.

Validity Check 2
Full text review stage

Study No.	Filter	Yes/No
1	F2	Yes
2	F1	No
3	F1	No
4	F1	No
5	F1	No
6	F1	No
7	F1	No
8	F2	Yes
9	F1	No
10	F1	No

1 Atherly A, Nurmagambetov T, Williams S, Griffith M. An Economic Evaluation of the School-Based oPower Breathingo Asthma Program. *Journal of Asthma.* 2009 2009;46(6):596-599.

2 Stirtzinger R, Campbell L, Green A, DeSouza C, Dawe I. Multimodal school-based intervention for at-risk, aggressive, latency-age youth. *Canadian Journal of School Psychology.* 2001;17(1):27-46.

3 Goldberg JP, Folta SC, Eliasziw M, et al. Great Taste, Less Waste: A cluster-randomized trial using a communications campaign to improve the quality of foods brought from home to school by elementary school children. Preventive medicine. 2015;74:103-110.

4 Gorely T, Nevill ME, Morris JG, Stensel DJ, Nevill A. Effect of a school-based intervention to promote healthy lifestyles in 7-11 year old children. *International journal of behavioral nutrition and physical activity.* 2009;6(5).

5 Kernsmith PD, Hernandez-Jozefowicz DM. A Gender-Sensitive Peer Education Program for Sexual Assault Prevention in the Schools. *Children & Schools.* 07/01/ 2011;33(3):146-157.

6 Locker D, Frosina C, Murray H, Wiebe D, Wiebe P. Identifying children with dental care needs: evaluation of a targeted school-based dental screening program. *Journal of Public Health Dentistry.* 2004;64(2):63-70.

7 Quach J, Hiscock H, Ukoumunne OC, Wake M. A brief sleep intervention improves outcomes in the school entry year: a randomized controlled trial. *Pediatrics.* 2011;128(4):692-701.

8 Rush E, Obolonkin V, McLennan S, et al. Lifetime cost effectiveness of a through-school nutrition and physical programme: Project Energize. Vol 8. Obesity Research and Clinical Practice2014:e115-e200.

9 van Beurden E, Barnett LM, Zask A, Dietrich UC, Brooks LO, Beard J. Can we skill and activate children through primary school physical education lessons? "Move it Groove it"--a collaborative health promotion intervention. *Preventive Medicine.* 2003;36(4):493-501.

10 Webster-Stratton C, Jamila Reid M, Stoolmiller M. Preventing conduct problems and improving school readiness: evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *Journal of Child Psychology & Psychiatry & Allied Disciplines.* 2008;49(5):471-488.

Appendix 3: Data Extraction

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
CEA					
Ansell, J.; Guyatt, H. L. 2002 Tanzania	Comparative cost-effectiveness of diagnostic tests for urinary schistosomiasis and the implications for school health programmes Annals of Tropical Medicine and Parasitology	To compare the costs and cost-effectiveness of self-diagnosis, use of reagent strips and parasitological examination in identifying infected individuals or 'high risk' schools for S. haematobium.	Descriptive, diagnostic testing	School-aged children 15 schools (n=2370) in Muheza district, Tanzania	Four diagnostic tests considered: 1) self-reporting of schistosomiasis during interview with public-health nurse, 2) the reagent-strip testing of urine for presence of blood, 3) the examination of urine for visible blood, and 4) filtration of a 10-ml urine sample and microscopical examination of the filter for S. haematobium eggs.
Bertrand, Elise; et al 2011 Canada	Cost-effectiveness simulation of a universal publicly funded sealants application program Journal of Public Health Dentistry	To simulate a publicly funded program of pit and fissure administration, either in the public or private sectors, and compare these hypothetical situations with the current one, i.e., a publicly funded, school-based selective program.	Markov model developed using a virtual population of 8-year-old children monitored over a time span of 10 years.	8-year-old children Children in Quebec	3 options of sealant delivery: the mixed, private, and school situation.
Crowley, D. Max; et al 2014 USA	Can we build an efficient response to the prescription drug abuse epidemic? Assessing the cost effectiveness of universal prevention in the PROSPER trial Preventive Medicine	Evaluate the cost effectiveness of universal evidence-based-preventive-interventions (EBPIs) to reduce nonmedical prescription opioid use.	Rural schools from Iowa and Pennsylvania randomised to control or intervention groups	Grade 6 children approx. 11-12 years US schools/ communities	Four programmes evaluated: Strengthen families program (n=827), All Stars (n=1936), Life Skills Training (n=1166), and Project Alert (n=1924)
Foster, E. M.; et al 2006 USA	Can a costly intervention be cost-effective?: An analysis of violence prevention Archives of general psychiatry	Examine the cost-effectiveness of the Fast Track intervention designed to reduce violence among at-risk children.	Costs estimated using program budgets of this cluster randomised controlled trial. ICERS computed to determine cost per unit improvement in the 3 outcomes measured in the 10th year of study (CEA).	School-aged children grade 1-10 US schools	High risk children identified in schools and those scoring in the top 40% were selected to receive the intervention, 91% agreed (n=3,274). Intervention delivered from grades 1 though 10. Parents offered parent training and home visiting, academic tutoring, and social skill training. Parent and child groups offered 2-hour enrichment program. Group meetings held weekly, biweekly, then monthly each year on. Social emotional learning program PATH adapted in school curriculum.
Foster, John M.; et al 2013 USA	Does Teacher Professional Development Improve Math and Science Outcomes and Is It Cost Effective? Journal of Education Finance	Examine in-service professional development program targets at chronically low-achieving schools and examine whether it improve student learning at those schools while considering cost of providing program compared to other types of school improvement interventions.	38 school districts participated and evaluated outcome data pre and post intervention	School-aged children grade 4 to 12 Appalachian Kentucky school districts	Ultimate goal is to improve student outcomes in the STEM subject areas. Training delivered for K-12 tears of math and science covering content training in algebra, geometry, physics, and biology.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
CEA					
Ansell, J.; Guyatt, H. L. 2002 Tanzania	1) number of children who were correctly identified by the test, and 2) number of infected children who were identified by the test (i.e. true positives) none	Using self report to identify high risk schools and then reagent strips to id individuals in low-risk schools was more 'cost-effective' than the gold standard urine test, but would result in 8% of cases being missed.	A cost-consequence type of analysis with various scenarios was presented for health care planners to decide which approach to take considering local resource constraints.	CEA none stated none stated	Cost per infected child: Urine filtration: $2.30; Self-reported: $1.33; and self-reported to identify high risk schools and reagent strips to identify individuals: $1.43. Cost/correct diagnosis: Urine filtration: $1.00, Reagent strips: $0.5, Visible blood $0.56, see Table 1 and 2 provide fixed and variable costs of the screening tests.
Bertrand, Elise; et al 2011 Canada	No. of children without decay on the surface of the first permanent molar. modelled over 10 years	School is most cost-effective strategy at $172 per child without decay as compared to the mixed situation. It is $868 in private compared to mixed.	Implementing a universal, school-based program of pit and fissure sealant application would improve access to preventive dental care and provide more social equity in Quebec's healthcare system. But it is ultimately a political decision.	Markov Model Health care system and parents' perspective 10 years, due to estimated efficacy of sealants.	Screening in schools, examinations in private clinics, sealant application and restoration in private clinics. Cost of staff, materials, travel by patients and their parents, and productivity loss for parents. Fees from a Fee Guide and Description used as proxy for costs of examinations, sealant application and restoration in private clinics. See table 3 for more details
Crowley, D. Max; et al 2014 USA	Cases of opioid use averted Mentions 6 year follow-up of one of the trials (PROSPER).	Authors report universal school-based EBPIs can efficiently reduce nonmedical prescription opioid use, and family programs may enhance the school-based programs.	Universal EBPIs can effectively and efficiently reduce nonmedical prescription opioid use and should be considered when developing responses to the growing national crisis.	Cost-effectiveness analysis with decision tree, classified as CEA societal not explicitly stated	Program costs estimated from previous analysis, opportunities costs estimated expenditures and any inputs from outside sources.
Foster, E. M.; et al 2006 USA	Three key long-term outcomes first is diagnosis of conduct disorder i.e. cost per case of conduct disorder averted 10 years	The intervention was not cost-effective at likely levels of policymakers' willingness to pay for key outcomes. Subgroup analysis of those most at risk showed intervention was likely to be cost-effective given specified willingness-to-pay criteria.	Intervention is cost-effective for children at the highest risk. Practical issues still remain such as the ability to effectively identify and recruit such high-risk children.	CEA Payer such as a state department of mental health. 10 years	Estimated from Fall 1991 to Summer 2003 from annual budget record or program costs from early years of the intervention. Cost of intervention estimated retrospectively.
Foster, John M.; et al 2013 USA	AMSP professional development activities' effect on student outcomes None, before and after study	AMSP professional development program appears to provide a rout to improving the quality of current teachers therefore improving student outcomes.	Replication of this type of cost-effectiveness analysis on middle school mathematics training should be carried out and if results replicated, content-focused professional development targeted to middle school teachers looks promising as a means of enhancing teacher quality of the current work force.	CEA none stated None, before and after	Table 6 from expenditure reports of the AMSP grant. Includes: admin compensations and travel, trainee stipends and meals, wages and fees, materials, supplies and overheads. Average cost per student $44.33

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
CEA						
Ansell, J.; Guyatt, H. L. 2002 Tanzania	yes Costs identified seem sufficient for each alternative. Mainly costs of delivering screening were identified with less resource use being identified.	Second visit for treatment and drug costs were considered, but only drug cost would effect an incremental if one had been calculated. Correct diagnosis, infection identified, and no. of infected children treated were the measures of effect	none none	No, assume not needed because screening under a year none USD 1995	Calculated based on cost of each scenario and effect as the proportion of number of infected children treated. Urine filtration compared to self-report: $12.13. Self report to id schools compared to self report alone: $2	none none
Bertrand, Elise; et al 2011 Canada	yes Yes, data integrated in models are specific to each option.	Staff costs, sealant, productivity loss, travel and meals. Children without decay at first molar	none Decay on first permanent molar.	yes 3% base case and 0 and 5% for sensitivity analysis CAD, no year specified	C/E ratios from health care perspective: $179/child without decay in mixed, $220 in private, $179 in school. C/E ratios for parents: $125, $68 and $84 respectively. ICER $868 per child in private compared to mixed and $172 in school compared to mixed.	Sensitivity analysis varying retention, resealing, re-restoration, decay incidence, high-risk children proportion and discounting. Yes, they are listed in table 2
Crowley, D. Max; et al 2014 USA	Yes, but mean costs and effects for each not reported no	Lumped into societal costs and estimated from previous studies Probability of youth misusing prescription opioids	none none, not CUA	yes 3% USD assumed, not explicitly stated	All are compared to controls not the next most costly/effective intervention. Life Skills: $613/prevention of one youth misusing prescription opioids, All Stars + school family program (SFP): $4,923, Life Skills + SFP: $3,959, no ICER reported for Project Alert?	Bootstrapping to construct 95% CI around each ICER (using 1,000 replications) not stated
Foster, E. M.; et al 2006 USA	No, just described as control group. No, just reported ICER	Did not detail Three outcomes, conduct disorder averted, index criminal act avoided and personal act of violence avoided.	n/a none	yes 5% USD 2004	Entire sample: CD: $3,481,433, Crime: $423,480, Violence: 736,010. Low-risk group: CD: $-2,059828, Crime: $-1,786,032, Violence: $-9,046,977. High-risk group: CD: $752,103, Crime: $150,738, Violence: $283,542.	Yes, generated 1000 bootstrapped samples n/a
Foster, John M.; et al 2013 USA	Not really, but costs and effects of alternative interventions are provided in table 7 yes	None collected, just intervention costs Reported effect size of comparators, intervention effect reported effect size of 0.03 on middle school math achievement.	none none	no none USD, no price year stated	They report effectiveness/cost of 0.000677. I recalculated at $1477.66. Comparators range from $48.19 for rapid assessment to $8,086,300 for charter schools. Outcomes are reported effect sizes, but outcome for each is unknown and assumed to be different for different studies.	no n/a

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
CEA					
Hollingworth, W.; et al 2012 UK	Reducing smoking in adolescents: cost-effectiveness results from the cluster randomized ASSIST (A Stop Smoking In Schools Trial) Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco	Conduct a cost-effectiveness analysis of a school-based 'peer-led' intervention.	Cluster randomised controlled trial of 59 secondary schools in England and Wales.	12-13 year-old students Secondary schools in England and Wales	The ASSIST programmed trained students to act as peer supporters during informal interactions to encouraged their peers not to smoke.
Kesztyus, D; et al 2013 Germany	Economic evaluation of URMEL-ICE, a school-based overweight prevention programme comprising metabolism, exercise and lifestyle intervention in children European Journal of Health Economics	Measuring the impact of the URMEL-ICE school-based overweight prevention programme on anthropometric measures in primary-school children, computing incremental cost-effectiveness relation?(ICER) and net monetary benefit.	Intervention study with historical control. School-based cluster randomised intervention trail w/o control.	Second grade primary school children Ulm, Germany, for control. Bavarian country of Gunzberg for intervention.	Lecture material integrated into usual curriculum so didn't require additional lessons. Three risk factors address: physical activity, consumption of sweetened beverages, and media use. 28 teaching units over 36 weeks with 6 family homework assignments.
Levaux, HP; et al 2001 USA	Economic evaluation of a 2-dose hepatitis B vaccination regimen for adolescents Pediatrics	Investigate the economic implications of a 2-dose hep B vaccine compared to 3 does for adolescents in 3 settings: public schools, public health clinics, and private sector settings in US.	Model based on primary data collection (CEA).	Adolescents 11-15 Private, public health clinics and public schools across USA	Comparing 2-dose regimen to 3-dose regimen given in three different settings: private, public health clinic and public schools
Noyes, K; et al 2012 USA	Cost-effectiveness of the school-based asthma therapy (SBAT) program Pediatrics	Examine the cost-effectiveness of school-based asthma therapy (SBAT) compared with usual care.	School-based randomised controlled trial with stratification based on smoke exposure (OTHER	Children aged 3-10 Urban school, Rochester City School District, NY	Each child received 1 dose of medication from school nurse once each school day, dose varied depending on severity. In control caregivers encouraged to contact their PCP to discuss asthma symptoms.
Simon, E; et al 2013 Netherlands	An explorative cost-effectiveness analysis of school-based screening for child anxiety using a decision analytic model European Jouranl of Adolescent Psychiatry	Determine CE of 1) one-time school-based screening child focused intervention 2)screening and parent intervention 3)screening and parent or child intervention 4) do nothing for child anxiety.	Model based on real-world data, mainly based on econ evaluation of RCT of this screening programme, followed-up for 2 years (CEA).	8 to 12 The Netherlands, screens take place at primary schools.	Evaluating effectiveness of three types of interventions to a do nothing approach.
Wang, LY; et al 2008 USA	Cost-Effectiveness of a School-Based Obesity Prevention Program Journal of School Health	Evaluate the cost-effectiveness of the Medical College of George FitKid Project, a 3-year, afterschool program designed to prevent obesity among elementary school students	CEA of a randomised, controlled trial of 18 elementary schools in Georgia were randomly assigned. Total of 601 students with retention of 88% and 96% intervention/control in first year. CEA	Third graders Elementary schools in Augusta, GA	2-hour after-school session offered 5 days a week which encouraged youth to make PA a regular part of their schedules. Sessions included 40 mines of academics and snack and 80 minutes of MVPA. Average attendance 49%.
Beets, M. W. 2014 USA	Making healthy eating and physical activity policy practice: The design and overview of a group randomized controlled trial in afterschool programs Contemporary clinical trials	To evaluate the effectiveness of healthy eating and physical activity (HEPA) strategies, which consist of a multi-step, adaptive intervention approach, that addresses price barriers to serving more healthful snacks and professional development training to develop core competencies to promote physical activity.	RCT will be reported to CONSORT guidelines for cluster RCTs. Repeated cross-sectional group randomized controlled trial with delayed treatment group.	Elementary school-aged children (6-12) After school programs (ASP) in the state of California	ASP that serve on a daily basis a fruit or vegetable, eliminate foods and beverages high in added sugar, avoid artificial ingredients and provide at least 30 to 60 minutes of moderate to vigorous physical activity.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
CEA					
Hollingworth, W.; et al 2012 UK	Odds ratio for being a smoker 2 years	Programme cost £32/student, incremental cost/student not smoking at 2 years was £1,500 (95%CI £669-£9947). Students in intervention were less likely to believe they would be a smoker at age 16.	The ASSIST programme reduced smoking among adolescents at a modest cost. Extending intervention to year 8 students would cost approx. £38 million and result in potentially 20,400 fewer adolescent smokers.	Cost-effectiveness analysis Public sector including cost to Local Authority and NHS 1 school year, not discounted but followed up for 2 years	Staff costs, intervention costs and personnel costs.
Kesztyus, D; et al 2013 Germany	Waist circumference 1 year	WC gain was 1.61 and WHtR gain 0.014 significantly less in intervention. Cost €24.09 per child. ICER €11.11 (95CI 8.78:15.02) per cm WC and €18.55(95CI 14.04:26.86) per unit WHtR gain prevented.	New info about cost-effectiveness of structured health promotion embedded in daily routine at primary schools. At WTP of €35 this intervention is cost-effective and this result may help decision makers in implementing programmes to prevent childhood overweight in school settings.	CEA stated societal 1 year	Cost of developing programme, time spent prepping lessons, materials, training. Costs of programme delivery all summed to get a cost per pupil. No health resource use collected.
Levaux, HP; et al 2001 USA	Cost per LY gained	Compliance rose with 2 does which resulted in lower infection rate and greater CE in all settings. In public health clinic, 2 dose dominated 3 dose in LT. ST costs higher for 2 dose, without LT cost offsets of reduced infection.	Improved compliance with 2 doses contribute to a higher probability of adolescents achieving HBV protection, when LT consequences of HBV included, 2 dose is CE in all settings.	Decision analytic model, decision tree for ST societal lifetime	Preparation work for vaccine dose, admin and disposal of vaccine, follow-up of patient who may have missed scheduled dose.
Noyes, K; et al 2012 USA	mean number of symptom free days	$10 per one extra day symptom free. 158 SFD gained per 30-day period per 100 children. Extra $4822 per 100 children per month, net savings $3240.	SBAT was effective and cost-effective in reducing symptoms in urban children with asthma compared with usual care.	CEA Main analysis: Medicaid perspective (payer) also used societal perspective for ICER 1 school year (7-9 months)	4 categories: programmatic costs, health care costs, school attendance fees losses, and parents' productivity losses.
Simon, E; et al 2013 Netherlands	ICER based on ADIS improved child	Strategies 1 and 2 were dominated by 3 and 4, strategy 3 requires additional investment of 107 for each additional ADIS improved child.	Screening followed by child/parent intervention depending on parent anxiety had high increment effects and low costs compared to do nothing approach. Differences between groups small, explorative and first evidence of CE for this screening intervention.	DAM, decision tree, CEA societal 2 years inferred, not directly stated.	Healthcare, direct non-healthcare, indirect and out-of-pocket costs
Wang, LY; et al 2008 USA	Reduction in percent body fat (%BF) 1 year	Intervention cost $174070, $558/student or $956/student who attended >40% of sessions. Usual after-school care costs estimated at $639/student. Students who attend >40% reduced %BF by 0.76% (CI -1.42, -0.09) at an additional cost of $317/student	Students who attended >40% of sessions achieved a sig reduction in %BF at relatively low cost. School-based obesity prevention programs of this type are likely to be a CE use of public funds and warrant careful consideration by policy makers and planners.	CEA states societal 1 year, doesn't give adequate justification for not including the 2 additional years follow-up that is available.	Two types of costs considered: cost of delivering FitKid program after-school and usual cost of after-school care when no intervention is provided.
Beets, M. W. 2014 USA	Children's accelerometer-derived MVPA and time spent sedentary at ASP, HEPA promoting and inhibiting behaviours, types of snacks served and consumed. 3 years, one year baseline 2 years intervention	protocol	protocol	Protocol, stated CEA will be performed. Societal none stated	Costs of snacks and cost of intervention/standard practice.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
CEA						
Hollingworth, W.; et al 2012 UK	Smoking prevention education/standard practice. No health service costs identified.	Staff time including travel time, training materials, venue costs, bus hire, peer support vouchers. OR of being a smoker and smoker prevalence.	none Reduction in smoking prevalence	No, one year programme, but followed up over 2 years none GBP 2008	£1500 per student not smoking. (£32/0.021) CI's calculated using bootstrap sampling of 10,000 iterations.	Sensitivity analysis around privately contracted trainers increased cost £38/student and if employed ASSIST trainers average cost/student falls by £6. none
Kesztyus, D; et al 2013 Germany	Not enough detail, very confusing, poor quality reporting No, control had 'null costs' so missing costs of control	Teacher time, scientific coordinator for training and teacher support. Relative risk for overweight at follow-up and RR for incident WHtR.	none, CEA none	no, one year none 2008 Euro	Cost €24.09 per child. ICER €11.11 (95CI 8.78:15.02) per cm WC and €18.55(95CI 14.04:26.86) per unit WHtR gain prevented. ICER calculation in text does not match mean values reported in Table 2.	Sensitivity analysis varying effectiveness by 10, 20 and 30%, but said it 'shall be tested' as if they haven't done it yet? State all costs were precisely collected during trial so only vary effects at 10, 20 and 30% lower values. no model
Levaux, HP; et al 2001 USA	Yes, 2 and 3-dose regimens yes	Advertising and promo material, equipment in set-up, vaccine/materials, staff time required to administer/ clinical duties Compliance of 3-dose regimens from questionnaires. As no 2-dose data available, compliance derived from 3 doses.	LYs gained	yes 5% 2001 USD CPI	Private sector 2-dose: $964/LY, $1517/infection prevented; public school 2-dose: $1246/LY $1960/infection prevented	One-way and multivariate sensitivity analyses performed, varied from 50% to 150% of base-case values. Varied compliance as well. Probability of vaccine, protection, preventing infection, cost/person/infection prevented/LY.
Noyes, K; et al 2012 USA	yes yes	see costs above SFD	none Cost per SFD	no USD 2009	see results	Bootstrapped and varied unit costs to lower and upper bounds of 95% CI Probability of high anxious or not high anxious.
Simon, E; et al 2013 Netherlands	yes	Resources used for care of the child (anxiety, other psych problems, physical problems) through use of 2 week cost diaries. ADIS (diagnostic interview) 'improved' or 'not improved'		yes 4% 2012 Euro Dutch indexes	€107 per ADIS improved between strategy 3 and 4.	Scenario performed with 1) screening organised during annual visit of school physician 2) optimal participation rates 3) with only direct health care costs (i.e. healthcare perspective). One-way sensitivity Analyses on costs and probabilities increasing/decreasing by 25%.
Wang, LY; et al 2008 USA	Yes, estimated cost of after-school care Only considered two types of costs and barely considered consequences	Personnel, training, transportation and materials to deliver the intervention %BF reduction	none No, stated this was a limitation	No, one year time horizon 2003 USD	see results	Simple sensitivity analysis varying per capita usual after-school care costs in plausible range from $5 to $10.
Beets, M. W. 2014 USA	Briefly as standard practice, no HEPA strategies no	none mentioned Didn't specify from outcomes which would be the main measure of effect they would use in the ICER calculation. Refer to it simply as 'effect.'	None 'Effect' not specified	no USD, no year stated	Refer to calculation of CER, but definition given is an ICER.	none none

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
CEA					
Brooker, Simon; et al 2010 Kenya	Improving educational achievement and anaemia of school children: design of a cluster randomised trial of school-based malaria prevention and enhanced literacy instruction in Kenya Trials	To evaluate the impact of school-based malaria prevention and enhanced literacy instruction on the health and educational achievement of school children in Kenya.	A factorial, cluster randomised trial	Primary school classes 1 to 5 Being implemented in 101 government primary schools on the coast of Kenya	i) Intermittent screening and treatment of malaria in schools by public health workers, ii)training workshops and support for teachers to promote explicit and systematic literacy instruction. Schools randomised to one of four groups receiving either i) the malaria intervention alone, ii) the literacy intervention alone, iii) both combined, or iv) control
Ford, Tamsin; et al 2012 UK	Supporting teachers and children in schools: the effectiveness and cost-effectiveness of the incredible years teacher classroom management programme in primary school children: a cluster randomised controlled trial, with parallel economic and process evaluations Bmc Public Health	To evaluate whether teacher classroom management improves socio-emotional well-being among children as measured by the SDQ and cross-validated with direct observation, parental SDQ and child report on How I Feel About My School where available. Many more, see page 2.	Cluster randomised controlled trial of the Incredible Years teacher classroom management (TCM) course with combined economic and process evaluations.	Children aged 4-9 Primary schools within Devon, Torbay and Plymouth	TCM draws on cognitive social learning theory, about how coercive cycles of interaction between adults and children reinforce unwanted behaviour patterns. It also incorporates strategies for challenging angry, negative, and depressive internal dialogue in adults whilst interacting with children, drawn from cognitive behavioural approaches.
Carabin, H.; et al 2000 Egypt	A population dynamic approach to evaluating the impact of school attendance on the unit cost and effectiveness of school-based schistosomiasis chemotherapy programmes Parasitology	Model the possible costs and effectiveness of reaching non-enrolled children through school-based programmes using empirical data from Egypt.	Population dynamic model	School-aged children Egypt	Four strategies compared: school-based coverage of 85% and school-aged targeted coverage of 25, 50 and 85%. No actual trial, modelled strategies based on previously collected data.
Nishiura, H; et al 2014 Japan	Cost-effective length and timing of school closure during an influenza pandemic depend on the severity Theoretical Biology and Medical Modelling	Optimize the timing and length of school closure during influenza pandemic for cost-effectiveness.	Modelling study with ICER of Yen/LY (OTHER)	Age groups proportional to that of Japan Japanese schools	Modelled effect of different lengths of school closures to 'reactive school closings' closing when many people are infected. Closure was varied from 7, 14 and 21 days.
Muennig, PA; et al 2014 USA	The Cost-effectiveness of New York City's Safe Routes to school program American Journal of Public Health	Evaluate CE of a package of roadway modification in NYC funded under the Safe Routes to School program for both school age and adult users.	Markov model (OTHER)	School age and adults New York City	Federally funded $612M program to build new sidewalks, bicycle lanes, and improve safety at crossings, upgrade signage. Intended to encouraged children to walk/bike to school.
Barrett, Jessica; et al 2015 USA	Cost Effectiveness of an Elementary School Active Physical Education Policy American Journal of Preventive Medicine	The purpose of this study is to estimate the cost effectiveness of a state “active PE” policy implemented nationally requiring that at least 50% of elementary school PE time is spent in moderate to vigorous physical activity (MVPA).	A previously developed cohort model (ACE from Australia) was used to simulate the impact of an active PE policy on physical activity, BMI, and healthcare costs over 10 years for a simulated cohort of the 2015 U.S. population aged 6–11 years. Data were analysed in 2014.	Aged 6-11 years US population of children attending public elementary schools	An "active PE" policy implementing state policy directing state boards of education to include PE curriculum a requirement that 50% of PE time be devoted to MVPA.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
CEA					
Brooker, Simon; et al 2010 Kenya	Educational achievement and anaemia, the hypothesised mediating variables through which education is affected. 24 months	protocol	protocol	Protocol, CEA Assessed from both provider (government) and societal perspective none stated	Assessed using an ingredient approach, based on interviews with individuals involved in delivering the interventions and by consultation of the programme accounting system. The aim is to estimate the cost of scaling-up the interventions in Kenya.
Ford, Tamsin; et al 2012 UK	Total difficulties score from the SDQ completed by class teacher, supplemented by parent SDQs. 30 months, with data collection at baseline, 9, and 18 months	protocol	protocol	Protocol, CEA Broad public sector perspective, including use of all health, education and social care services, plus criminal justice sector resources and criminal activity. Primary economic evaluation will look at T3, 30 month follow-up. Will explore 'longer-term implications.'	A brief questionnaire to parents to capture high cost/high use resource items. The full interview will be used with a random sample of 50 parents to validate the supplement. Educational service use collected from schools. TCM costs will use a micro costing (bottom-up approach.
Carabin, H.; et al 2000 Egypt	Cases of disease prevented Wasn't an intervention study so no follow-up	No. treated, infection and early disease of both bacterium presented in Table 4. The school based strategy which only covered 85% still prevented 77% of cases of disease. Would cost USD $0.06-1.03 to reach non-enrolled children.	Treating non-enrolled children is an important consideration in maximizing the effectiveness of treatment programmes while mainlining a cost-effectiveness comparable to school-based delivery.	Population dynamic model none stated 15 years modelled, 5 for intervention 10 years follow-up	
Nishiura, H; et al 2014 Japan		Not cost-effective. If risk of death three times greater than that of H1N1, the school closure could be regarded as cost-effective.	No fixed timing and duration of school closure that can be recommended as universal guideline. The effectiveness of school closure depends on the transmission dynamics of a particular strain especially the infection fatality risk.	model none stated 1 year, inferred	Annual leave due to needing to stay home with children, and cost to save a single year of life?
Muennig, PA; et al 2014 USA		SRTS associated with net societal benefit of $230M and 2055 QALYs gain in NYC	SRTS reduces injuries and saves money in the long run.	Markov model societal 50 years	Cost of pedestrian injury, transport cost to school, death, program cost
Barrett, Jessica; et al 2015 USA	cohort model simulated over 10 years	: An elementary school active PE policy would increase MVPA per 30-minute PE class by 1.87 minutes (95% uncertainty interval [UI]¼1.23, 2.51) and cost $70.7 million (95% UI¼$51.1, $95.9 million) in the first year to implement nationally. Physical activity gains would cost $0.34 per MET-hour/day (95% UI¼$0.15, $2.15), and BMI could be reduced after 2 years at a cost of $401 per BMI unit (95% UI¼$148, $3,100). From 2015 to 2025, the policy would cost $235 million (95% UI¼ $170 million, $319 million) and reduce healthcare costs by $60.5 million (95% UI¼$7.93 million, $153 million).	: Implementing an active PE policy at the elementary school level could have a small impact on physical activity levels in the population and potentially lead to reductions in BMI and obesity-related healthcare expenditures over 10 years.	CEA Modified societal perspective 10 years	1) training PE teachers; 2) training school principals; 3) replacement of equipment and materials; 4)state PE coordinator time for oversight, implementation and monitoring.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
CEA						
Brooker, Simon; et al 2010 Kenya	Yes, see intervention description not mentioned	none mentioned Improvements in test scores and educational achievement, assessed in terms of differences in standard deviation units.	None See effectiveness	No not stated	ICER will be calculated for each outcome in relation to the status quo, the other health intervention packages tested in the present study, and current interventions.	A comparison with alternative education interventions, especially those conducted in Kenya, will also be undertaken. None
Ford, Tamsin; et al 2012 UK	Briefly as no classroom management not mentioned	Will be collected mainly via Child and adolescent service Use schedule (CA-SUS). Total difficulties score SDQ.	Not a preference measure SDQ	none stated Not stated, GBP assumed	CE will be assessed using the net benefits approach.	Will use non-parametric bootstrapping to explore the probability that each of the treatments is the optimal choice, subject to a range of possible maximum values (ceiling ratio) that a decision-makers might be willing to pay for an additional unit of outcome gained. Didn't mention specifics, just that data from the trial supplemented by data from the literature will be used in decision analytic modelling techniques.
Carabin, H.; et al 2000 Egypt	yes unclear	Unit cost of treatment Reduction in the number of infection or in the number of early disease cases over 15 years.	none Number of disease cases prevented	yes 5% USD no price year stated	None presented, gave unit cost per child of $0.60 for school-based and an additional $1.03 per child treated to reach non-school enrolled children	Performed on unit costs Yes, given in Appendix 2
Nishiura, H; et al 2014 Japan	no not clear	not clear Life years, infections, absolute difference b/t two scenarios.	ICER in Japanese Yen per LY LY	no Japanese Yen	Provided graphically, didn't state ICER, because was never below their threshold.	none stated Age groups, susceptibility of contracting flu, 2ndary transmission, reduction of cases due to closure, risk of death.
Muennig, PA; et al 2014 USA	no not clear	Not described outside of costs listed QALE, questionable methods to obtain QALE.	EQ5D5L given to two senior Pediatric orthopaedic surgeons QALE	Yes, only over 50 years 3% 2013 USD	None reported, supposedly cost saving, QoL .95 too high for injured people.	Found it to be cost saving so didn't run, so only did it on the annual model? A series of 1-way sensitivity along with MC sim including 'plausible boundaries' for values. No. of people at risk, risk ratio of injury, probability of hospitalisation, case fatality ratio, QALE. Costs: total program costs, injury costs, death and transportation costs to school.
Barrett, Jessica; et al 2015 USA	yes, current practice yes	see costs MVPA converted to MET-hours assuming average MET level of 4.5. MET hour increase and BMI unit change	none MET-hour increase per day after 1 year and cost per BMI unit change after 2 years	yes 3% 2014 USD	$0.34 per MET-hour gained (95% UI$0.15 to $2.51) $401 per BMI unit reduced (95% UI $148 to $3,100). Over a 10 year period $1,720 per BMI unit reduced. ($272 to $5,710).	PSA using Monte Carlo simulations in @RISK over 10,000 and 1,000,000 iterations. yes, given in Table 1

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
CUA					
Blakely, Tony;et al 2014 New Zealand	Cost-effectiveness and equity impacts of three HPV vaccination programmes for school-aged girls in New Zealand Vaccine	Estimate the health gains, net-cost and cost-effectiveness of the currently implemented HPV national vaccination programme of vaccination dispersed across schools and primary care, and two alternatives: school-based only, and mandatory school-based vaccination but the opt-out permitted.	Markov macro-simulation model	12-year-old girls and boys New Zealand schools and primary care	1) as currently implemented HPV national vaccination programme for girls only across schools and primary care for 2) school-based only 3) mandatory school-based with active opt out.
Cooper, K.; et al 2012 UK	An economic model of school-based behavioural interventions to prevent sexually transmitted infections International Journal of Technology Assessment in Health Care	Assess the cost-effectiveness of behavioural interventions in school, for the prevention of sexually transmitted infections, in young people.	Economic model	13-15 years UK, school/ community care	Three interventions: teacher led (20 sessions over two years), peer led (three sessions of one hour over one school term), and standard sexual health education.
Frick, K. D.; et al 2004 USA	Modelled cost-effectiveness of the experience corps Baltimore based on a pilot randomized trial Journal of Urban Health-Bulletin of the New York Academy of Medicine	1) Model the cost-effectiveness of the Experience Corps Baltimore using data from a pilot randomized trial, including costs, older adults' health status, and quality of life and cost data from the medical expenditure panel survey and 2) describe the relationship between children experience increased expected lifetime earning through improve educational attainment resulting from exposure to the Experience Corps Baltimore volunteers and the program's cost-effectiveness.	Pilot randomised controlled trial of 6 schools	Older adults and children Baltimore schools	Designed to provide opportunities for older adults to give back to their communities by involving volunteers in high-impact generative activities to provide help for public elementary schools with attendant moderate physical, social, and cognitive engagement of the volunteers. Volunteers provided literacy support, behaviour management, violence prevention, community outreach and library support.
Gerald, JK; et al 2010 USA	Cost-effectiveness of school-based asthma screening in an urban setting Journal of Allergy Clinical Immunology	To conduct a cost-effectiveness analysis of school-based asthma screening strategies.	5 health state Markov model used to evaluating school-based screening in simulated population of urban elementary school children (CUA)	Elementary school-aged children Urban school setting, in Birmingham, Alabama; 95% black, 80% eligible for subsidized lunches	Simulated four mutually excluding screening strategies compared to no screening over 1 year. Four strategies: 1) the Narrow Questionnaire, 2) The Broad Questionnaire, 3)Multi-Stage with Spirometry, and 4) Multi-stage with Exercise testing.
Konig, HH and Barry, JC 2004 Germany	Cost-Utility analysis of orthoptic screening in Kindergarten: a Markov model based on data from Germany Pediatrics	To estimate the long-term cost-effectiveness of a hypothetical screening program for untreated amblyopia in 3-year-old children conducted by orthoptists in all German kindergartens in the year 2000.	CUA with decision tree combined with Markov model.	3-year-old children All German kindergartens	Orthoptic screening for children age 3 in all German kindergartens in the year 2000.
Kowada, A 2012 Japan	Cost effectiveness of Interferon-Gamma Release Assay for School-based tuberculosis screening Molecular Diagnosis & Therapy	To assess the cost effectiveness of school-based tuberculosis (TB) screening using QunatiFEROn TB Gold In-Tube (QFT) versus the tuberculin skin test (TST) and chest x-ray examination (CXR).	Markov models	1st year high school and university students Hypothetical cohort modelled.	QFT vs. TST vs. CXR

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
CUA					
Blakely, Tony;et al 2014 New Zealand	State cost/QALY, but actually Years of life lived in disability Not intervention study so no follow-up	Current programme $18800/QALY; school only $33000/QALY; mandatory $117000/QALY. All interventions generated more QALYs for indigenous and deprived people	A more intensive school-only vaccination programme seems warranted. Reducing in vaccine price will greatly improve CE possibly making mandatory vaccine optimal from a HC perspective	Markov macro-simulation model with 2,000 Monte Carlo iterations Health care system Lifetime, or until participants reach 110.	Vaccine costs $113 , delivery and administration $141 or $126, and cost of enacting new law for third intervention and delivery admin cost of $19 per dose
Cooper, K.; et al 2012 UK	Condom use estimated using meta-analysis of studies identified Not an intervention study so no follow-up. No information of time horizon of model	See summary outcomes and ICER results.	School-based behavioural interventions which provide information and teach young people sexual health skill can bring about improvements in knowledge and increased self-efficacy through these may be limited in terms of impact on sexual behaviour. Some uncertainty around these results.	Cost-utility analysis NHS and PSS Not explicitly stated so unclear if discounting required	Direct costs were treatment of infection and provision of the interventions and taken form published UK sources. All updated using NHS multiplier and reported in Euros. Assume no difference in costs and outcomes of standard education.
Frick, K. D.; et al 2004 USA	Cost/QALY, QALYs not measured, estimated from other sources 2 years	$205,000/QALY older adults, with 12-15 additional students graduation because $49,000/QALY and cost-effective. WTP threshold not defined.	Using conservative modelling assumptions and excluding benefits to teachers, principals and the surrounding community, the Experience Corps Baltimore appears expensive for older adults' health improvements, but requires only small long-term benefits to the target children to make the program cost-effective.	Markov model, with random transitions none stated 2 years	Salaries of supervisory staff, recruitment/training, volunteer stipends, and other operating costs.
Gerald, JK; et al 2010 USA	Cost/QALY, obtained from secondary sources 1 year modelled	Most efficient strategy identified children with previously diagnosed but poorly controlled asthma at a cost of $15000/QALY.	Population-based (school) asthma screening is not cost-effective at $50,000/QALY and has only a 20% chance of being cost-effective at $100,000/QALY. Population-based asthma screening not cost-effective.	Decision tree with Markov model societal 1 year, Markov model has 365 daily cycles	Daily cost included: ASFD, symptom day and exacerbation recover day, medicines, and routine physician visits. Screening costs, health resource use, indirect costs, medication and acute care.
Konig, HH and Barry, JC 2004 Germany	Incremental costs and effects, the ICER Lifetime modelled	Orthoptic screening was €7,397/QALY. Probability of ICER <€25,000/QALY was 84%.	The ICER of orthoptic screening seems to fall within a range that warrants careful consideration by decision-makers. Much of the uncertainty in the results stems from the effect of amblyopia on quality of life.	CUA with decision tree and Markov model Third-party payer perspective Lifetime	Costs of organisation screening exam, fixed costs, variable costs, ophthalmologic exam, and cost of treatment.
Kowada, A 2012 Japan	QALYs gained Lifetime modelled	QFT strategy yielded the greatest benefits at the lowest cost: 16-years-olds $627.89, 29.69835 QALYs; 19-year-olds $646.04, 29.15361 QALYs. TST 16 $943.50, 29.69767 QALYs; 19 $998.62, 29.15288 QALYs. CXR 16 $7,286.24, 29.69767QALYs; 19 $7,305.19, 29.14911 QALYs	The QFT strategy provided the greatest benefits at the lowest cost for school-based TB screening. There appears to be little role for TST or CXR in screening of school populations. Current practices using either TST or CXR screening should be reconsidered on the basis of cost effectiveness.	Markov model Societal Lifetime (up to 80 years)	Direct costs, such as inpatient and outpatient costs and indirect costs arising from loss of productivity. Costs of screening included the labour cost for two physician visits and TST kits. QFT included screening kits, on physician visit and labour cost for lab technicians. CXR material, on physician visit and labour of radiologic technetium, etc.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
CUA						
Blakely, Tony;et al 2014 New Zealand	yes yes	Delivery and administration for each scenario, cost of enacting new immunization law. Health systems cost based on sex and age group and cancer stage of care. Vaccine coverage and subsequent reduction in HPV	Used disability weights Referred to as QALY, but used disability weights applied to non-fatal states to calculate years life lived in disability.	yes 3% base case and 0 and 6% in scenario analyses NZ dollars 2011	School based program vs current program $34,700/'QALY.' For mandatory law vs school-based $122,500/'QALY'	Univariate sensitivity analysis using 2.5 and 97.5 percentile values from the uncertainty distribution for each input parameter showed large variation in ICER. Scenario analyses to change costs, excluding unrelated health care costs, and herd immunity estimates. Sex, age, ethnicity, mortality due to cancer and incidence
Cooper, K.; et al 2012 UK	yes No, did not identify costs and outcomes of the standard sexual health programme. Also appears they assume the same effect for both teacher and peer lead intervention.	For teacher and peer led interventions based on results of two UK trials and valued using UK primary and secondary sources. From published studies and UK reports, condom use	Utility data from studies of groups of patients who had developed complications from STIs. Methods for eliciting preferences not reported. QALY	State that they updated the analysis to reflect changes in discount rates. none stated EUR, no price year stated	€24,268/QALY for teacher led and €96,938 with the peer-led intervention. Probably of less than €36,000=16%	One-way sensitivity analyses undertaken by varying all the model parameters across plausible ranges. A PSA carried out with input values sample from probability distribution using 1,000 Monte Carlo simulations. Presented as CEAC. Yes, detailed in table 4
Frick, K. D.; et al 2004 USA	no not detailed	Volunteer commitment of 15 hours a week (not costed specifically but given a stipend to cover expenses) Self-report health status.	Didn't use one, so used self-reported health status of project older adults' outcomes. QALY (projected from self-report health status)	yes 3% USD 2003	$205,000/QALY for adults. If 12 additional students graduated (0.3%) ICER would be $50,000/QALY and if 15 additional graduate, the program would be cost-saving.	5,000 iterations to obtain a distribution of cost-effectiveness results Health states of older adults
Gerald, JK; et al 2010 USA	In previous paper yes	Screening costs, questionnaire, spirometry, administration, dry visit, diagnostic procedure, medication, MD visit, hospital stay, ED visit QALY, not entirely sure how they got their QALY weights. Symptom days, asthma severity, ED visits, hospitalisations.	Each health state was associated with a 'quality adjusted life day' and used paediatric health outcome measure to calculate QALYs. QALY stated. Reference is given in a table.	No, 1 year time horizon none USD 2006	Not sure how calculated, report Narrow Questionnaire: $151,000/QALY; Broad Questionnaire: $312,000/QALY; in text and NQ: $127,020 and BQ: $251, 850 and MSwSP and MSwET as dominated in table E3.	One way deterministic sensitivity analysis between NQ and status quo repeated with individual cost elements. Sensitivity analysis demonstrated 3 variables accounted for 90% of uncertainty: symptom day preference weight, prevalence, and baseline rate of symptom days. Asthma, diagnosis, controlled, prevalence, and disease severity, see table 1 for model inputs.
Konig, HH and Barry, JC 2004 Germany	Briefly as usual care yes	modelled costs Modelled from estimates from literature.	QALY QALY	Yes 5% EUR 2000	€7,397/QALY	Univariate and Monte Carlo simulation. 90% uncertainty interval of the ICER of €3,452 to €72,637/QALY Screening population, test characteristics, Effectiveness, utilities, costs, mortality, treatment success, discount rate
Kowada, A 2012 Japan	yes yes	Just the costs mentioned were include in a hypothetical cohort. QALYs as identified in the literature.	QALY QALY	yes 3% USD 2009	See Table 4, all ICERs are dominated by the QFT strategy.	One-way and two-way sensitivity analyses, comparing all strategies simultaneously. Each model variable was assigned a distribution based on the values in literature or assumptions. PSA with Monte Carlo simulation also performed. Age, probability of having TB, LTBI, developing TB fromLTBI, mortality rate from TB, all cause mortality.

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
CUA					
Miller, T.; et al 2013 Zimbabwe	Cost-effectiveness of school support for orphan girls to prevent HIV infection in Zimbabwe Prev Sci	Analyse cost per QALY gained in school support as a structural intervention to prevent HIV risk factors among Zimbabwe orphan girls adolescents.	RCT	Adolescent orphan girls grade 6 A rural province in Zimbabwe	Based on Social Development Model, includes school support (fees, uniforms and school supplies) and a helper which is a trained teacher (approx. 1 per 10 participants). Helpers monitored attendance and intervened with absenteeism with access to a small emergency fund.
Moodie, M;et al 2009 Australia	Cost-effectiveness of active transport for primary school children-Walking school bus program International Journal of Behavioural Nutrition and Physical Activity	To assess from a societal perspective the incremental CE of the Walking School Bus (WSB) program for Australian primary school children as an obesity prevention measure.	Modelled effects on BMI and DALYs of he WSB program applied throughout Australia. (CUA)	5 to 7 Australia	Children are accompanied by 2 volunteer adult "conductors" (ratio 1 adult to 8 children) and travel along a set route through a neighbourhood picking up children along the way at designated stops and delivering them to school.
Moodie, M; et al 2013 Australia	The Cost-effectiveness of a successful community-based obesity prevention program: the be active eat well program Pediatric Obesity	Examine the CE of Be Active Eat Well (BAEW), a large, multifaceted, community-based capacity-building demonstration program that promoted healthy eating and PA for Australian children.	CEA/CUA (CUA)	4 to 12 Australia 2003-2006, conducted in rural town of Colac, population 11000	Particular focus on primary school setting. 6 primary schools and 4 preschools participated. Targeted reduced to viewing, consumption of sugar drinks, increased water, fruit and veg and active play after school
Moodie, ML; et al 2010 Australia	The Cost-effectiveness of Australia's Active After-school Communities Program Obesity	Assess from a societal perspective the cost-effectiveness of the Active After-school communities (AASC) program, a key plank of the former Australian Government's obesity prevention program.	Simulation model (CUA). Retrospective policy review, didn't actually have control, but assumed no intervention as a comparator	Prep to grade 6 aged 5-14 Australian schools/ communities from 3-5pm time slot	Funding provided by small grants to participating schools. PA coordinators were appointed to develop and deliver physical activity program specific to the needs of the school over 2-3 session per week for 8 weeks of each of the four school terms per year.
Muennig, P; Woolf, S 2007 USA	Health and Economic Benefits of reducing the number of students per classroom in US primary schools American Journal of Public Health	Estimate the costs associated with reducing class sizes in kindergarten through grade 3 as well as the effects of small class sizes on selected outcomes such as QALYs and future earnings.	Markov model using multiple datasets, but mainly trial data of STAR Project (CUA)	Hypothetical cohort aged 5 to 65 Elementary schools in US	Project STAR was a randomised trial of 12000 students in schools in Tennessee. Randomised to 22-25 student classrooms or 13-17 students. Health economic outcomes not collected, so used regression for estimates of educational attainment and earnings.
Philipsson, A.; et al 2013 Sweden	Cost-utility analysis of a dance intervention for adolescent girls with internalizing problems Cost Eff Resour Alloc	To assess the cost-effectiveness of a dance intervention in addition to usual school health services for adolescent girls with internalizing problems, compared with usual school health services alone.	Prospective, RCT.	Adolescent girls aged 13-18 City in central Sweden, and their schools	Dance twice weekly during 8 months in addition to usual school health services. Participants followed-up 5 times during study.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
CUA					
Miller, T.; et al 2013 Zimbabwe	Cost per unit of improvement in outcome. State cost/QALY but EQ-5D only taken in 2010. None, but lasted 3.33 years	Intervention yielded an estimated US$1472 in societal benefits and .36 QALY gain, costing $6/QALY.	For non-boarders financial benefits exceeded it's costs, boarding was not cost-effective as it did not have any effect on outcome measures relative to girls in treatment group who did not board.	CUA societal 3 years plus one term in the fourth year.	Intervention program costs: supplies, helper fee, boarding charges if boarded.
Moodie, M;et al 2009 Australia	Benefit assessment estimate of health gain using DALY. Increased PA, converted to BMI change and cost-offsets and DALYs over lifetime	Modelled interventions cost $AUD22.8M and resulted in incremental saving of 30 DALYs at a net cost per DALY saved of $AUD0.76M. Evidence base judged 'weak' as no data available documenting an increase in number of children walking due to intervention.	Under current modelling assumptions, the WSB is not an effective or CE measure to reduce childhood obesity. The attribution of some costs to non-obesity objectives (reducing traffic/pollution) is justified for other possible benefits.	CUA/ CEA societal Lifetime, modelled until reached 100 years of age or died.	Unit costs, resource use and assumptions included in additional file 1. Adjusted to reference year using CPI. Cost to participants and families and all sectors involved in delivery of intervention.
Moodie, M; et al 2013 Australia	BMI units saved and DALYs averted over predicted cohort lifetime.	Cost AUD .34M annually, and saved 547 BMI units and 10.2 DALYs. Net cost per DALY saved AUD 29798.	BAEW was affordable and CE, and generated substantial spin-offs in terms of activity beyond funding levels. Elements fundamental to its success and potential CE associated with scaling up require identification.	CUA/CEA societal Lifetime of predicted cohort, takes current cohort 5-19 and follows in five-year groups for remaining life or 100 years.	See table 1, many costs included, very thorough
Moodie, ML; et al 2010 Australia	Cost/DALY saved Lifetime simulated	No. of new children receiving intervention benefit 69,300. One year intervention costs $40.3 M and saved 450 DALYs, cost-offsets $3.7M and cost/DALY $82,000.	Although the program has intuitive appeal, it was not CE under base-case modelling assumption. To improve CE as an obesity prevention measure, a reduction in costs needs to be coupled with increases in no. of participants and PA undertaken.	Simulation model societal Lifetime or age 100	Intervention costs 40.3 million, unit costs and sources, and assumptions in Supplementary table not available. Adjusted to 2001 using relevant consumer price index. Costs to health sector, participants and families, and other sectors involved in the delivery of the intervention.
Muennig, P; Woolf, S 2007 USA	Costs, cost savings and cost per QALY, QALY and future earnings To age 65 simulated	From societal perspective, reducing class size generates $168,000 in cost savings and gain of 1.7 QALYs per graduate. From government perspective costs saving to $15,000/QALY gained.	Reducing class size would be cost saving from a societal perspective. Reducing class size may be more cost-effective than most public health and medical interventions.	Markov model Societal and governmental Up to age 65	Medicare/Medicaid, salary costs, additional schooling costs, higher education costs, include crime costs in sensitivity analyses to make 'conservative.'
Philipsson, A.; et al 2013 Sweden	Cost per QALY 20 months	$25 per dance session. Cost was $670 per participant. Visit to nurse $58, mean number of visits to school nurse decreased by 10.75 in intervention group compared to 6.89. Increased QoL by 0.08 units more than those in the control group.	Intervention with dance twice weekly in addition to usual school health service may be considered cost-effective compared with usual school health services alone, for adolescents with internalizing problems.	CUA societal 20 months	1.) cost of the dance classes, (fraction of costs of dance teacher, rent, equipment and overhead) 2.) cost of using selected healthcare resources, i.e. visits to the school nurses.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
CUA						
Miller, T.; et al 2013 Zimbabwe	No, control not formally described, but there are effects presented in Table 4. EQ-5D not distributed to most married girls because they either dropped out of school or were lost to follow-up.	Education resources used to support the girls. Staying in school, marrying early, and QALY gained from preventing HIV.	Shona-language version of EQ-5D QALY (range 0.985-0.094)	yes 3% USD 2010	Stated as $6/QALY. When I calculate myself using total/pupil cost $1565 given in table two and QALYs given in table 4 (0.807-0.752)= $28454 and Thai scoring (.678-.612)=$23712. When use QALY gain stated in abstract and table 5 (probabilistic estimate .36)=$4347	Sensitivity analysis applying Thai scoring system for EQ-5D (as only other value set available for developing country), range 0.798:-0.454. Also used excel add-in to estimate uncertainty around costs and QALYs through 1 million simulations. Yes, gave uncertainty ranges in Table 1
Moodie, M;et al 2009 Australia	Current practice, 'do nothing' as there's no organised program in place yes	see figure 2 Reduction in BMI, DALYs, increased METS	DALY	yes 3% AUD 2001	Incremental effect reduction of 0.03 BMI units per child. Incremental costs were $22.8M. $0.76M/DALY which exceeds $50,000/DALY	atRisk software conducted Monte Carlo simulations (4000). Optimistic and very optimistic scenarios modelled. See table 2, height weight, BMI, estimated energy expend. METS, etc.
Moodie, M; et al 2013 Australia	Yes, current practice in 12 primary schools across Victoria, where no specific intervention offered. Yes, pathway analysis to id costs	Who did what, to whom, when where and how often. Opportunity cost of time expended. Reduction in BMI, DALYs averted,	DALY	yes 3% AUD 2006	Gross cost per BMI unit saved $399. Net cost per DALY saved of $20,227.	Same as before with atRisk software. Uncertainty distribution not attached to intervention costs since based on detailed evaluation data. Conducted scenario of 50% receiving benefit. See table 2, BMI, cost-offsets, triangular and uniform distributions
Moodie, ML; et al 2010 Australia	Briefly describe current practice as no intervention Can't tell from the description of costs in the article.	Should be reported in supplementary table S1 which is no longer available. Coordinator time, program delivery planning and operation A range of available data used to model BMI change from increase in PA. Change in BMI then converted to DALYs saved.	DALY DALY	yes 3% AUD 2001	AUD$82,000/DALY saved (95%CI $40,000-$165,000)	Monte Carlo simulations to present 95% uncertainty range around costs, benefits and ICER. Univariate sensitivity analysis i) reduction in the ratio of sites per regional co-ordinator, ii) reduction in the number of state level co-ordinators, iii) application of the same wage rate to all site co-ordinators. Iv) combination of scenarios, v) all participants receive full intervention benefit. Height, weight, % of schools interested, no. of children, METs, extra minutes on PA, etc. (Table 2)
Muennig, P; Woolf, S 2007 USA	No, only description was the number of children in the classrooms of the control group. not clear	Additional school and staff costs of small class size, additional schooling achieved and higher education. Additional teaching resource, educational resources, Medicaid. *did not include potential construction to build bigger schools to accommodate smaller class sizes. High school and college graduation, life expectancy, QALY (assumed to be applied to additional life expectancy) came from adult study.	EQ5D taken from MEPS participants age 25-65 QALY	Yes, but only for 12.5 because benefits wouldn't be realised until graduated, unclear 3% not stated	From governmental perspective, intervention was $15,415/QALY (CI $19000-$33000). In all other scenarios, the intervention dominates control. How is this possible? Some grand assumptions.	1-way uncertainty to isolate most influential variables, and Monte Carlo simulation to generate CI around estimates. Classroom size, deprivation(free lunch), high school grads, college attendants, earnings, Medicare/Medicaid, general costs, see Table 2
Philipsson, A.; et al 2013 Sweden	Described as control Just resource use	Dance classes, usual school health services of prevention and care provided by the school nurse. QALY gained as measured by HUI3, translated by a professional translator into Swedish.	Health Utility Index Mark 3 QALY	yes 3% Converted from Swedish krona to USD 2011	US $ 3830/QALY, 95% probability of CE with WTP $50,000/QALY	5,000 bootstrapped ICERS, sensitivity analyses of 50% higher costs and 50% lower effect. ICER $7,660 and $7,180 respectively. no model

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
CUA					
Rein, D. B.; et al 2012 USA	The potential cost-effectiveness of amblyopia screening programs J Pediatric Ophthalmol Strabismus	Estimate incremental cost-effectiveness of amblyopia screening at preschool and kindergarten.	Individual microsimulation model using natural history data of amblyopia	3-100 US general population	Compared costs and benefits of 3 amblyopia screening scenarios: 1) acuity/stereopsis (A/S) screening at kindergarten, 2) A/S screening at preschool and kindergarten, and 3) photo screening at preschool and A/S screening at kindergarten.
Rush, E; et al 2014 New Zealand	Lifetime cost effectiveness of a through-school nutrition and physical programme: Project Energize Obesity Research & Clinical Practice	Project energize aims to improve the overall health and reduce the rate of weight gain of all Waikato primary school children. An existing model used to extrapolated effects of general and Maori child pop of NZ.	Modelling study (CUA)	6 to 8 and 8 to 12 Waikato district is where the trial took place, model assumes NZ pop.	Program to increase physical activity and encourage healthy eating. Multicomponent program delivered to two age groups.
Shepherd, J; et al 2010 UK	The effectiveness and cost-effectiveness of behavioural interventions for the prevention of sexually transmitted infections in young people aged 13-19: a systematic review and economic evaluation Health Technology Assessment	Assess the effectiveness and CE of schools-based skills building behavioural interventions to encourage young people to adapt and maintain safer sexual behaviour and to prevent them from acquiring sexually transmitted infections.	Systematic review of effectiveness and econ evaluation model (CUA)	13-19	Behavioural interventions defined as any activity to encourage your people to adopt sexual behaviours that would protect them from acquiring STIs for which reported sexual behavioural outcome.
te Velde, Saskia J.; et al 2011 Netherlands	Modelling the Long Term Health Outcomes and Cost-Effectiveness of Two Interventions Promoting Fruit and Vegetable Intake among Schoolchildren Economics and Human Biology	To date, future health effects and cost-effectiveness at the longer term have not been estimated for existing school-based fruit and vegetable interventions. The current study aimed to provide an example of how these calculation can be done, by using data of two existing Dutch intervention programs, and to estimate cost-effectiveness for these two interventions.	Retrospective economic evaluation using cluster randomised controlled trial data and epidemiological modelling.	10-12 year-olds Primary schools in the Netherlands.	Detailed descriptions of the two fruit and vegetable interventions published previously, both aimed to improve intake of fruits and vegetables. Pro children provided fruit and veg twice a week, with worksheets and online feedback tool, and parent component. The Schoolgruiten included a free fruit and vegetable scheme and curriculum to increase knowledge and skills related to fruit and veg consumption.
Tengs, TO; et al 2001 USA	The cost-effectiveness of intensive national school-based anti-tobacco education: results from the tobacco policy model Preventive Medicine	Evaluate the CE of enhanced nationwide school-based anti-tobacco education relative to the status quo.	System dynamic simulation modelling study (CUA)	12-13 years US schools nationwide	Intensive school-based tobacco use prevention program given to every 7th and 8th grader in US.
Vijgen, SMC; et al 2008 Netherlands	Cost-effectiveness analyses of health promotion programs: a case study of smoking prevention and cessation among Dutch students Health Education Research	Determine Cost-effectiveness of a Dutch school-based smoking education program.	CUA	adolescents Health promotion in school , Netherlands (model based)	Based on another study of a peer led 45 minute smoking cessation school programme. Social influence group vs control

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
CUA					
Rein, D. B.; et al 2012 USA	Cost per QALY and cost per case avoided Simulated to age 100	No screening most CE with WTP of less than $16,000/QALY. A/S at kindergarten WTP of $17,000 and $21,000. A/S screening at preschool and kindergarten WTP between $22,000 and $75,000. Photo screening WTP over $75,000. All scenarios CE when assuming WTP of $10,500 per case of amblyopia cured.	All three likely to be considered cost-effective relative to may other public health programs. Choice of program depends on budgetary resources and value placed on monocular vision loss prevention.	CUA/CEA Societal perspective excluding cost of informal care and lost productivity from adult visual impairment. 3-100 simulated	Screening costs, German cost estimates of amblyopia treatment by age: $2,102 age 3 and $775 age 10 and older.
Rush, E; et al 2014 New Zealand	Cost per QALY Extrapolated lifetime model	$44.96/child/year to deliver programme. ICER for younger, $30,438, older $24,690. For Maori $28,241 and $22,151 respectively. For middle socioeconomic status schools $23,211, $17,891. Cost effective in a number of scenarios of general, Maori and different age groups.	Project Energize would improve quality and length of life and when compared with other obesity prevention programs previously assessed with this model, it would be relatively CE from the health payer's perspective.	Previously developed model applied funder's/payer's lifetime	Ongoing cost of the intervention estimated from project budget. Healthcare costs of chronic conditions same as in existing model.
Shepherd, J; et al 2010 UK	STIs averted	Few significant differences between interventions and comparators in changes in sexual behaviour outcomes. Some significant differences in knowledge and self-efficacy. Quality of intervention provider influence young people's perceptions. Cost £4.30 and £15/ pupil for teacher and peer led interventions and £20,223 and £80,782/QALY. OR 1.03 not significant.	School-based behavioural intervention can bring about improvements in knowledge, but didn't significantly influence sexual risk-taking behaviour or infection rates.	CUA, systematic review of economic evaluations for prevention of STIs in young people and Bernoulli model for probability of STI infection NHS PSS Lifetime, intervention effects last 1 year	NHS national and local unit costs
te Velde, Saskia J.; et al 2011 Netherlands	Fruit and vegetable intake. Epidemiological modelling estimated the number of DALYs gained over the lifetime of all 10 year olds in the Netherlands. 2 years	Pro Children ICER €5,728/DALY. Schoolgruiten was €10,674/DALY. Probability of being CE 80% and 68% respectively. Pro Children has 70% chance of dominating Schoolgruiten.	Well-designed fruits and vegetable-promoting interventions targeting primary school-age children may be a good investment, but trials with a follow-up of a decade or more are required to enable more rigorous analyses.	CUA Health services perspective lifetime	Costs from the two programmes were used to estimate nationwide implementation costs in the model. Health care costs related to diseases were incorporated into the epidemiological model.
Tengs, TO; et al 2001 USA	Never actually stated None, computer simulation model that relied on secondary data	Over 50 years, CE estimated $4900 and $340000/QALY. Assuming 30% effectiveness that dissipates over 4 years CE is $20000/QALY.	Although not cost saving, a much more intensive school-based anti-tobacco education effort would be an economically efficient investment for the nation.	Simulation modelling study using Marconian system dynamic model, The Tobacco Policy Model. CUA, but no preference measure taken directly. societal 50 years	Used costs from TNT program, estimated average direct medical costs incurred by adults from previous study. To estimate annual cost for each age, gender and smoking status, performed multiple regression model using Hodgson's data.
Vijgen, SMC; et al 2008 Netherlands	Cost per QALY 18 months, mentions insufficient follow-up.	ICER 19900 per QALY. Several assumptions had to be made, lack of effectiveness data on smoking in adolescents.	CE of health promotion programs is lacking. For policy makers, CE is very important because investment now may return gains/savings in the future.	Model based CUA using the Chronic Disease model (CDM) didn't state 100 years, lifetime	Estimated intervention costs, model include health care costs of various smoking related diseases.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
CUA						
Rein, D. B.; et al 2012 USA	Yes, person-level simulation of amblyopia incidence, detection with and without screening at preschool and kindergarten. Estimates from literature	Rates of referral and follow-up care	QALY QALY, decrements assumed by estimates from the literature.	yes 3% 2002 Euros converted to USD and inflated to 2005 using consumer price index medial care component.	For no screening cost was less than $16,000/QALY, screening at kindergarten $17 to $21,000 per QALY. Screening at preschool and kindergarten $22 to $75,000. Photo screening at preschool and kindergarten at $75,000 and greater.	PSA, Sensitivity analysis of A/S screening on rates observed in VIP study. A/S screening sensitivity in preschool and kindergarten, photo screening sensitivity in pre-schoolers and kindergarten
Rush, E; et al 2014 New Zealand	status quo Described elsewhere	Used the same resource use costs as was already built into model. BMI reductions, effect assumed to decay at 1% each year for first 5 years.	QALY QALY	yes 3.5% 2011 NZ$	Younger: $30438/QALY Older: $24690 Maori Y: $28241 Maori O: $22151 Middle SES: $23211, $17891	Varied conditions such as cost of intervention +-10%, discount rate 0-5%, decay of effect which had biggest impact on ICER 5-10% $100K to $500K. Model described elsewhere
Shepherd, J; et al 2010 UK	Yes, teacher 20 sessions, peer led 3 sessions Costs associated with each STI case such as complications, PID and infertility and HRQoL loss.	Medical treatment of STIs, cost of behavioural interventions. Condom use, from systematic review of effectiveness.	HRQoL taken from previous utility studies of patients with condition of interest. QALY, STI cases averted, medical costs saved.	yes 3.5% GBP	£20,223 and £80,782/QALY	Uncertain in clinical effect, HRQoL and resource use. Deterministic and probabilistic sensitivity analysis. One-way deterministic of individual parameters. STI prevalence, single-act transmission probabilistic, condom effectiveness and condom use, number of sexual episodes and number of sexual partners.
te Velde, Saskia J.; et al 2011 Netherlands	yes Not clear what all went into the epidemiological model.	Hospitalization, care, and care by a family physician related to diseases in the epidemiological model. Effects at 2 years for each intervention were input into their model even though they were not statistically significant.	DALY DALY	yes 3% 2003 Euro	Net intervention costs + difference in health care costs/DALYs averted by intervention	PSA, bootstrap of 10,000 iterations, used Ersatz program to simultaneously vary key parameters. Effect in Pro Children and Schoolgruiten and lasting life long, costs of Pro Children and Schoolgruiten, relative risks of diseases, discounting, value of a DALY.
Tengs, TO; et al 2001 USA	'Current average education practice nationwide' Yes, but mostly assumed	Estimated costs included in model Used Quality of Well being scale (QWB) to assess QoL. Based effectiveness on seven sources.	QWB, not sure if preference based QALY	yes 3% 1999 USD using MCPI	Cost/QALY from $24000-$600000. Cost/LY$3.8M and $170M	3 way Monte Carlo sensitivity analysis to assess impact on medical costs, QoL and mortality. Distributions placed on difference costs, QoL, and mortality. Performed many scenarios. Age, gender, smoking status, exposure to nicotine in utero and year, see Table 2.
Vijgen, SMC; et al 2008 Netherlands	control not identified	not identified States effectiveness data is lacking, LYG and QALYs	QALY QALY	yes 4% Euro 2004	1990 Euros /QALY	Sensitivity analysis on key parameters, varied effectiveness in daily and experimental smokers, intervention costs, discount rate and time horizon. Smoking prevalence, incidence mortality and costs of 14 smoking related diseases

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
CUA					
Wang, LY; et al 2001 USA	Cost-effectiveness of a school-based tobacco-use prevention program Arch Pediatr Adolesc Med	Determine he CE of a school-based tobacco-use prevention program.	Using data from the 2 year efficacy study of the Project Toward No Tobacco Use (TNT), conducted decision analysis. Benefits were Lys saved, and QALYs, costs saved. (CUA) Model with base case, worst and best case scenarios using data from previous sources.	12-15 (7-9th grades) Based TNT, school-based in Southern California	10 lessons to counteract social influences and misconceptions that lead to tobacco use, delivered by trained health educators in 8 junior high schools. 2 less booster delivered in 8th grade and followed up o 9th grade
Wang, Li Yan; et al 2011 USA	The Economic Effect of Planet Health on Preventing Bulimia Nervosa Archives of Pediatrics & Adolescent Medicine	To assess the economic effect of the school-based obesity prevention program Planet Health on preventing disordered weight control behaviours and to determine the cost-effectiveness of the intervention in terms of its combined effect on prevention of obesity and disordered weight control behaviours.	Modelling study using previous RCT and economic evaluation evidence.	RCT data on girls 10-14, present study projected up to age 17 Middle schools	The Planet Health program was implemented from 1995 to 1997 and designed to promote healthful nutrition and physical activity among youth. Interdisciplinary, school-based obesity prevention program. Eating disorders positively associated with overweight and obesity in adolescence, there is an interest in integrating prevention of both disorders.
Barber, Sally E.; et al 2013 UK	Pre-schoolers in the playground an outdoor physical activity intervention for children aged 18 months to 4 years old: study protocol for a pilot cluster randomised controlled trial Trials	To undertake a pilot cluster randomised controlled trial of an outdoor playground-based physical activity intervention for parents and their children aged 18 months to 4 years old and to assess the feasibility of conducting a full scale cluster RCT.	Study protocol for cluster RCT and economic evaluation.	18 months to 4 years Bradford, West Yorkshire, UK which includes some of the most deprived areas in UK	Pre-schoolers in the Playground (PiP) comprises a 10-week initiation phase (one school term) followed by a 20-wekk maintenance phase (two school terms).
Quach, J.; et al 2013 Australia	Sleep well - Be well study: Improving school transition by improving child sleep: A translational randomised trial BMJ Open	To determine whether using school-based screening followed by a brief behavioural intervention is cost-effective when delivered by an existing school-based health workforce.	RCT nested in a population-based, cross-sectional survey.	No age specified, just children with sleep problems Primary schools in the southern region of Melbourne, Australia	School nurse arranges a 45 minute session to provide the intervention to the parent at the child's school, covering education about normal sleep requirements and the importance of good sleep hygiene practices, then select acceptable strategies. Two weeks later a follow-up 15 min phone call, and then final 30 face-to-face appointment if needed.
Chestnutt, IG; et al 2012 UK	Protocol for "Seal or Varnish?" (SoV) trial: a randomised controlled trial to measure the relative cost and effectiveness of pit and fissure sealants and fluoride varnish in preventing dental decay Bmc Oral Health	To compare the clinical effectiveness of pit and fissure sealants (PFS) and fluoride varnish (FV) in preventing dental caries in first permanent molars in 6-7 year-olds. Secondary aims: examining the impact of PFS and FV on children and their parents/carers in terms of quality of life/treatment acceptability measures, and implementation in a community setting	Randomised, assessor-blinded, two-arm, parallel group trail	6-7 year old schoolchildren Primary schools in deprived areas of South Wales	Treatment delivered via mobile dental clinic. PFS and FV will be applied by trained dental hygienists. FV will be applied at baseline, 6, 12, 18, 24 and 30 months. PFS will be applied at baseline and re-examined at all follow-ups and reapplied if detached/insufficient.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
CUA					
Wang, LY; et al 2001 USA	Not specifically mentioned 2 years	Cost $16403, prevented estimated 34.9 students from becoming smokers. Savings of $13316/LY and $8482/QALY. TNT was cost savings over a reasonable range of model parameter estimates.	TNT is highly CE compared with other widely accepted prevention interventions. School-based prevention programs of this type warrant careful consideration by policy makers and program planners.	CUA/CEA without preference measures collected. Decision model to model lifetime, based on decision tree up to age 26. societal Lifetime, from estimated established smokers at 26 and over lifetime.	Training, teaching and material costs.
Wang, Li Yan; et al 2011 USA	Intervention costs, medical costs saved, quality-adjusted life years gained, and cost-effectiveness ratio. RCT 2 years, present study modelled to age 17.	An estimated 1 case of bulimia nervosa would have been prevented. As a result, an estimated $33,999 in medical costs and 0.7 QALYs would be saved. At an intervention cost of $46,803, the combined prevention of obesity and disordered weight control behaviours would yield a net savings of $14,238 and a gained of 4.8 QALYs.	Primary prevention programs, such as Planet Health, warrant careful consideration by policy makers and program planners. The findings provide additional argument for integrated prevention of obesity and eating disorders.	Modelling study/CUA societal 10 years	Used medical costs for bulimia nervosa previously reported in the literature. Estimated 10-year cumulative costs based on two studies. Chose cost of CBT and inpatient and outpatient treatment costs. Incorporated 10 year costs of patient returning to treatment. Estimated cumulative costs over 10 years, then calculated average which was used in base-case and sensitivity analyses.
Barber, Sally E.; et al 2013 UK	To assess the feasibility of a full-scale cluster RCT, the pilot will examine recruitment rates, attendance and attrition, see right 52 weeks with follow-ups at 10 and 30 weeks.	protocol	protocol	Protocol, CUA planned none stated Within-trail analysis and if possible modelled costs and benefits over a more appropriate time horizon.	Will explore possibility of using routine databases to capture relevant resource use. Will collect costs of the intervention
Quach, J.; et al 2013 Australia	Child psychosocial functioning at 6 months. 12 months, with 6 months	protocol	protocol	Protocol, two step, CCA, then CEA using PedsQL as main analysis. Secondary analysis presenting combined child and parent QALY CUA. Classified as CUA Health and education and broader societal perspective. 12 months	Costs of intervention and resource use
Chestnutt, IG; et al 2012 UK	The presence or absence of dental caries and caries treatment on all surfaces of all teeth will be recorded using the ICDAS system. 36 months	protocol	protocol	Protocol, type of eval not stated, assumed CUA societal none stated	Costs for each trial participant including number and frequency of service use, travel costs to families, assessment of total cost of PFS and FV including potential costs of treatment avoided.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
CUA						
Wang, LY; et al 2001 USA	no no	Just costs mentioned above Estimated Lys and QALY gains, based on 2 year dataset on TNT.	QALYs assumed from literature. QALY, not directly measured	yes 3% 1990 USD	none	Multivariate sensitivity analysis, varied values of each of 10 key parameters assuming estimates normally distributed. see tables 2-4
Wang, Li Yan; et al 2011 USA	No, only modelled project costs of bulimia. No, no costs or consequences identified for control group that didn't receive the Planet Health intervention.	None collected specifically. Took medical costs of bulimia from the literature. Taken from the RCT in terms of prevented from developing bulimia. Took HRQoL from one study using a preference-weighted 15D.	Preference-weighted 15D, a generic, comprehensive, self-administered instrument for measuring and assessing HRQoL among BN patients. A set of utility preference weights were elicited from the general public. QALY	yes 3% 2010 USD	$2,966/QALY, to be interpreted with caution. This was a modelling study that assumed the effectiveness of the intervention would prevent one person from developing BN. It also assumed cost savings (instead of taking the difference in cost between interventions) and assumed QALY gains from the literature (instead of directly measuring for each intervention).	Sensitivity analyses were performed to test the assumptions made. Univariate and multivariate sensitivity analyses on 5 parameters: percentage of girls with disordered weight control behaviours who had eating disorders, progression probability, medical treatment costs, HRQoL of BN patients, and time to recovery. Multivariate, 10000 Monte Carlo simulation using @RISK. Many, see above. This model is based on many assumptions.
Barber, Sally E.; et al 2013 UK	usual care trial protocol	Number and type of injuries sustained during intervention period and service use reported by parents at each assessment point using previously developed questionnaires. PedsQL for child HRQoL and EQ5D for parental HRQoL	Not specified QALY planned, did not give preference details	none stated GBP	Will be generated	PSA will be conducted Plan for longer term model
Quach, J.; et al 2013 Australia	briefly assumed	Parents will retrospectively recall resource use for child's sleep. Child psychosocial functioning at 6 months, PedsQL	CHU9D, PedsQL, child, EQ-5D parent QALY	yes 5% None stated, GBP assumed	Will present a CUA on the combined parent-child QALYs as a secondary analysis.	Will perform extensive sensitivity analysis none
Chestnutt, IG; et al 2012 UK	Yes, but didn't include a do nothing approach yes	Staff resources, treatment/appointment duration, equipment and materials used. Costs to families collected from parental resource utilisation questionnaire. Will also capture school resource use. HRQoL as measured by the CHU9D being sent home to parents asking children to complete with assistance if necessary.	CHU9D QALY	none stated Not stated, GBP assumed	Did not state if will calculate an ICER	none stated none

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
CUA					
Pearson, Amber; et al 2014 New Zealand	Is expanding HPV vaccination programs to include school-aged boys likely to be value-for-money: a cost-utility analysis in a country with an existing school-girl program BMC Infectious Diseases	The goal of this research was to examine the cost-effectiveness of adding boys to a girls-only program in New Zealand. We modelled the incremental health gain and costs for extending the current girls-only program to boys, intensifying the current girls-only program to achieve 73% coverage, and extension of the intensive program to boys.	: A Markov macro-simulation model, which accounted for herd immunity, was developed for an annual cohort of 12-year-olds in 2011 and included the future health states of: cervical cancer, pre-cancer (CIN I to III), genital warts, and three other HPV-related cancers	cohort of 12 year-old boys New Zealand schools and primary care	Methods followed the Burden of disease epidemiology, equity and cost-effectiveness program protocol. They adapted a previous Markov model on the cost-utility of girls only HPV vaccination to estimate QALYs gain and net health system costs for girls only and girls and boys vaccination.
CBA					
Barnett, Steven 1985 USA	Benefit-Cost Analysis of the Perry Preschool Program and Its Policy Implications Educational Evaluation and Policy Analysis	Benefit-Cost Analysis of the Perry Preschool Program and its long-term effects and examine the analysis as a basis for US public policy decisions regarding early childhood education.	Benefit-cost analysis, of two year program with follow-up at 15 and 19 (CBA)	3 and 4-year-old black children with no discernible physical handicaps Original study conducted in Ypsilanti, MI with children born b/t 1958 and 1962.	Operated from October to May with 3 elements: 1. centre-based for 2.5 hours each morning, evolving from traditional nursery school to cognitively oriented Piagetian approach. 2. Home visiting, teachers visit once a week for 1.5 hours. 3. Group parent meetings.
Belfield, CR; et al 2005 USA	The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using data from the Age-40 follow-up Journal of Human Resources	Perform cost-benefit analysis of the High/Scope Perry Preschool Program using new data on the careers and livelihoods of the participants and control group up to age 40.	123 at risk children in Michigan in 1960. 58 randomly assigned to receive the program and 65 to be in the control. They have been surveyed periodically since the program.	3 and 4-year-old black children with no discernible physical handicaps Ypsilanti, MI	Operated from October to May with 3 elements: 1. centre-based for 2.5 hours each morning, evolving from traditional nursery school to cognitively oriented Piagetian approach. 2. Home visiting, teachers visit once a week for 1.5 hours. 3. Group parent meetings.
Glewwe, Paul; et al 2001 Philippines	Early Childhood Nutrition and Academic Achievement: A Longitudinal Analysis Journal of Public Economics	Investigate the nutrition-learning nexus using a unique longitudinal data set that follows a large sample of Filipino children from birth until the end of their primary education. Also to perform a cost-benefit analysis of return on investment of early childhood nutrition programs.	Statistical and economic analysis of longitudinal data, non-experimental.	Birth to end of primary school Cebu, Philippines 1983-1994	No intervention, analysis focused on the achievement production function, which relates to early nutritional and other academic inputs to a child's scholastic output as measured by achievement test scores.
Reynolds, Arthur J.; et al 2011 USA	Age 26 Cost-Benefit Analysis of the Child-Parent Centre Early Education Program Child Development	Conduct a CBA of the Child Parent Centre (CPC) Education program using outcome data at age 26. 1) Does participation in the CPC continue to demonstrate high economic benefits relative to costs? 2) Do the estimated economic benefits in 2007 USD differ across preschool, school-age, and extended-program participation? 3) Do economic benefits differ by child and family subgroups, including gender, parent education, family risk status, neighbourhood poverty, and length of participation?	Prospective cohort study, quasi experimental design, Title 1 Elementary and Secondary Education Act of 1965 is after Head Start the US's oldest federally funded preschool program. 989 CPC and 550 comparison children followed up to age 26 (approximately 90% of each group). Comparison group matched to intervention on age, eligibility, neighbourhood and family poverty.	Intervention given from age 3 to 9, followed-up to 26. US schools (Chicago area)	Child parent centre early education program (CPC) provide preschool and school-age services up to age 9 for economically disadvantaged children in the Chicago Public schools system. Data for the study was collected from the Chicago Longitudinal Study (CLS) which has follow-up to age 26. Intervention described previously.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
CUA					
Pearson, Amber; et al 2014 New Zealand	QALYs gained simulation over lifetime	At an assumed local willingness-to-pay threshold of US$29,600, vaccination of 12-year-old boys to achieve the current coverage for girls would not be cost-effective, at US$61,400/QALY gained (95% UI $29,700 to $112,000; OECD purchasing power parities) compared to the current girls-only program, with an assumed vaccine cost of US $59 (NZ$113). This was dominated though by the intensified girls-only program; US$17,400/QALY gained (95% UI: dominant	These results suggest that adding boys to the girls-only HPV vaccination program in New Zealand is highly unlikely to be cost-effective. In order for vaccination of males to become cost-effective in New Zealand, vaccine would need to be supplied at very low prices and administration costs would need to be minimised.	CUA Markov macro-simulation model none stated, but may be described elsewhere Stated 12 year-olds over their whole lives.	Used routine, linked admin health data for entire NZ population. Assigned health system costs by sex, and age to healthy state. Added costs for cancer patients, and other disease states from regression and averages respectively. Intervention costs included cost of vaccine (NZ$113) and delivery/admin of $141 in schools and primary care or $126 in schools only.
CBA					
Barnett, Steven 1985 USA	cost/benefit	Significantly better outcomes for intervention in IQ, test scores, graduation, higher education, arrests, employment and welfare. Good evidence it's a good investment for society for one year, relatively good for 2 years.	While study is old, authors argue (in 1985) that structure of school and society hasn't changes enough to threaten generalisability, may be more effective for deprived populations. More evaluations should be planned because stakes too high to rely on Perry School program only.	Benefit-cost analysis Society, 2 groups considered, participants and families and 'taxpayers' child's lifetime	(and benefits) for 7 categories: a)program costs, b)child care, c)elementary and second. education, d) higher education, e)delinquency and crime, f)earnings and employment, and g)welfare.
Belfield, CR; et al 2005 USA	cost benefit ratio Up to age 4	The treatment group obtains significantly higher earnings. For the general public, higher tax revenues, lower criminal justice system expenditures, and lower welfare payments easily outweigh program costs; they repay $12.90 for every $1 invested.	Program gains come mainly from reduced crime by males, and thus subject to uncertainty around those assumptions.	Cost-benefit analysis' stated Societal Up to age 40	Undiscounted USD 2000 program costs estimated as $15,827. Cost information was taken from school district budgets and the program administration unit; both operating costs and capital costs are included.
Glewwe, Paul; et al 2001 Philippines	1983-1994	Heterogeneity in learning endowments, home environment or 'parental tastes' can't fully explain why malnourished children performed relatively poorly in school, positive relationship persists after controlling for such factors.	In simple CBA $1 invested in early childhood nutrition program could potentially return $3 worth of gains in academic achievement.	Simple CBA stated none stated Lifetime assumed as they assumed a working age 16-61 when calculating cost-benefit.	One scenario of benefit assumes a child is given an extra half year of earnings $650. Discounted over 15 years at 3%=$415 5%=$310. Untargeted intervention could achieve an average improvement in heigh0for-aged of 0.6 at $300/child and targeted $600/child.
Reynolds, Arthur J.; et al 2011 USA	CLS data up to age 26.Net present value (all societal costs - all societal benefits). Also expressed as benefit-cost ratio or return on investment.	Those who participated had economic benefits that exceeded costs. Preschool program: $10.83 per dollar invested (18% annual return), school-age program: societal return of $3.97/$1 invested (10% annual return) and extended intervention program (4-6 years) had a return of $8.24 (18%). Males, 1-year preschool participants and children from higher risk families derived greater benefits.	Findings provide strong evidence that sustained programs can contribute to well-being for individuals and society.	CBA stated, but not on an RCT and no valuation undertaken. societal Age 3-26	Preschool average cost per child $5,5597. School-age: $2,010, and extended program $12,304. Did not cost CPC in kindergarten or in comparison (full day), approximated the costs were similar in each.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
CUA						
Pearson, Amber; et al 2014 New Zealand	Yes yes	Health care resource use estimated for each health state based on NZ administrative data. New Zealand data on cancer incidence and survival, and other cause mortality (by sex, age, ethnicity and deprivation). HPV prevalence reduction, disability weights applied for each disease state, and population morbidity estimated from lifetables. Cancer incidence rates predicted from regression of NZ cancer registry data.	QALYs valued using disability weights QALY	yes 3% Used PPPs to convert from USD to NZ dollars 2011	Same level of coverage for boys dominated. Vaccination of boys at currently coverage level for girls $117,500 per QALY, adding boys to an intensified girls only program not cost-effective $247,000/QALY. Intensified girls only program is optimal at $33,000/QoL.	Scenario analyses explored reductions in vaccine cost per dose, and removing her immunity cancer reduction benefits. Monte Carlo simulation with 2,000 draws. All scenario analyses found adding boys was not cost-effective. Coverage achieved in females, herd immunity, HPV burden of disease in males, cost of vaccine and admin, vaccine efficacy and duration of protection, number of diseases.
CBA						
Barnett, Steven 1985 USA	No, probably in other papers, control was assumed to be no intervention control	cost of program, child care Based on assumptions from table 2	none Estimated benefit to society from children up to age 19	yes 5% USD 1981	Positive CB ratio, see table 4. 1 year: $5,525 se $3,200, 2 years $4,987 se $2,936	Varied discount rate: 3, 5 and 7%, reduced benefits by 50% and still exceeds costs
Belfield, CR; et al 2005 USA	Just as the control Important costs of control not identified. But benefits for both groups were identified.	Cost of the program Long-term benefits of the program include increased educational attainment, earnings profiles, tax contributions, lower criminal activity, and welfare payments.	None Monetised in cost-benefit ratio	yes 3% and 7% USD 2000	n/a	Plausible variations on how the net benefits to society vary given the program yields strongly positive benefits to participants.
Glewwe, Paul; et al 2001 Philippines	no alternative n/a	Not collected, not intervention Based on achievement scores and height-for-age	none none	yes 3-5% USD 1994	n/a	no no
Reynolds, Arthur J.; et al 2011 USA	Yes, 550 comparison group that attended full-day kindergarten in in five randomly selected schools. The comparison group were enrolled in 'usual early intervention' available for low-income children. Left out kindergarten costs, but said it was because the focus was on CPC program costs.	Valued as reductions in school remedial services, criminal justice system expenditures, victims of crime expenditures, child welfare and victimization from child abuse and neglect, depression, and substance misuse. These were valued and considered benefits of the program Do not describe valuation, but refer to sticking to 'previous estimation procedures.' In addition to reductions stated in the resource use column, there was also reduced mortality due to smoking and increases in lifetime earnings and tax revenues.	None not CUA. Benefits calculated in dollar terms. Net present value of the program (costs - benefits), also expressed as benefit-cost ratio (benefits/costs) to obtain return per $1 invested.	yes 3% 2007 USD adjusted for inflation	CBA	Monte Carlo simulations of 10,000 iterations. Sensitivity analysis for different model assumptions. Discount rate, earnings estimates, exclusion of intangible crime victim savings, inclusion of smoking benefits.

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
CBA					
Wang, LY; et al 2000 USA	Economic Evaluation of Safer Choices Arch Pediatric Adolescent Medicine	To evaluate cost-effectiveness and cost benefit of Safer Choices school based intervention for high school students.	CEA and CBA of cluster randomised trial with follow-up at 7, 19 and 31 months. Study based on 7 months. (CEA)	High school students, 15-19, 3677 ninth grade students USA, 10 schools in northern California and 10 in southeast Texas randomised to receive intervention or not.	2 year theory-based multicomponent intervention. Primary aim is to reduce number of student engaging in unprotected sex, focusing on school wide change to influence student behaviour.
Eckermann, Simon; et al 2014 Australia	Evaluating return on investment in a school based health promotion and prevention program: The investment multiplier for the Stephanie Alexander Kitchen Garden National Program Social Science & Medicine	To synthesise evidence of attributable impacts on student attitudes, lifestyle and behaviour and the societal (government, school, and wider community) multiplier on initial investment up to and beyond two years triangulated with qualitative evidence to enable informed assessment of expected long-term community impacts and returns from SAKGNP investment.	Return on investment with multiplier assessment to provide key evidence to assess network engagement, ownership and dynamic impacts.	Primary school students aged 8-12 years School/ community in Victoria	The Stephanie Alexander Kitchen Garden National Program (SAKGNP) is a school-based programme designed to promote pleasurable food education in Australian primary schools. Whole school commitment to hiring garden and kitchen specialists to support weekly lesson, linking to official curriculum, involving community volunteers and minimum 2 year commitment.
Heckman, J.; et al 2010 USA	The rate of return to the HighScope Perry Preschool Program Journal of Public Economics	Estimate the rate of return to the HighScope Perry Preschool program, as previous studies ignore the compromises during randomisation and do not report standard errors. The rates of return reported account for those factors.	Rate of return with benefit-to-cost ratios to support their conclusion	Preschool program with follow-up data into 40s Ypsilanti, Michigan	see earlier
Tai, T; Bame, SI 2010 USA	Cost-Benefit Analysis of Childhood Asthma Management through School-based clinic programs Journal of Community Health	Examine the cost-benefit of school-based health clinic programs (SBHC) in managing childhood asthma nationwide for reduction of medical costs of ER, hospital and outpatient physician care and savings in opportunity social costs of lowering absenteeism and work loss and of future earnings due to premature deaths.	CBA using 8 public data sources	5-17 year olds, school-aged children Data sources used to calculate national costs of implementing SBHC using multiple tertiary datasets.	Modelled the implementation of a school-based health clinic in schools to reduce asthma severity and hospital costs.
Hoeflmayr, David and Hanewinkel, Reiner 2006 Germany	Do school-based tobacco prevention programmes pay off? The cost-effectiveness of the 'Smoke-free Class Competition' Public Health Journal of the Royal Institute of Public Health	The objective of this study was to determine the cost-effectiveness of a school-based tobacco prevention programme.	Using data from a previous effectiveness study of the 'Smoke-free Class Competition' (SFC), an economic analysis was conducted to determine the cost-effectiveness of the SFC.	11-14 year old German students German schools	: To take part in the SFC, classes make the decision to be a non-smoking class for 6 months (from autumn to spring). The pupils themselves and their teachers monitor the smoking status of the pupils and report on it regularly. Classes that refrain from smoking can win a number of attractive prizes. In the school year 2001/ 2002, 150,566 German students participated in the SFC, representing approximately 4% of the total target population of 11–14-year-old German students. The effectiveness evaluation is based on 2,142 students who participated in the programme in the school year 1998/1999.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
CBA					
Wang, LY; et al 2000 USA	Reduce number of sexually active high school students and increase condom and contraception use among sexually active students.	Base-case intervention cost $105243, achieving 15% increase in condom use and 11% increase in contraceptive use within 1 year among 345 sexually active students. Cases averted: HIV 0.12, Chlamydia 24.37, gonorrhoea, 2.77, PID 5.86, and pregnancy 18.5. Every dollar invested returns %2.65 in med and social costs savings.	Safer choices is CE and cost saving in most scenarios considered. Program cost data should be routinely collecting evaluations of adolescent prevention programs.	Model, CEA and CBA of Safer Choices Societal, for medical considered public and private sector perspectives 1 year	Intervention, cases of HIV and other SDTs averted, and pregnancy, lost productivity.
Eckermann, Simon; et al 2014 Australia	Return on Australian Government investment on impacts on student lifestyle behaviours, food choices and eating habits surveyed across students and parents. 2 years minimum, some school observed longer	Multiplier impact on total community activity was 5.07 ($226,737/$44,758); 1.60 attributable to school, and 2.47 to wider community activity. All 8 schools observed beyond 2 year continued garden and kitchen classes with an average 17% scaling up and one school fully integrating staff into the curriculum.	Evidence supports the SAKGNP to be a successful health promotion program with high community network impacts and return on investment in practice.	RoI Australian Government 2 years, + some observed after	Government expenditure, kitchen staff expenditure, specialist staff, garden staff and value of volunteer time, school contributions, grant and community contributions.
Heckman, J.; et al 2010 USA	Rate of return, benefit-cost-ratios Use data up to age 40	Rates of return generally fall between 7-10%, substantially lower than previous literature. Returns generally statistically significantly different from 0 for both men and women and above historical return on equity. Benefit-cost ratios support this and range from $7 to $12 for every $1 invested.	Analysis provides a lower-bound on true rate of return to Perry Preschool program, reporting standard errors, and exploring sensitivity to alternate assumptions.	Rate of return Societal Lifetime	Initial program cost per child in undiscounted 2006 USD is $17,759.
Tai, T; Bame, SI 2010 USA	Cost of implementation (based on experiences of 12% of public schools that have these programs) compared to reduction of hospital cost absenteeism and premature death.	see conclusions	Costs: $4.55 Billion compared to savings of $1.69 Billion medical costs. Estimated savings due to absenteeism and premature deaths $23.13 billion. Many assumptions made, recommend future work.	CBA using 8 public data sources, classified as other hospital and societal none stated	School nurse staffing multiplied by asthma prevalence for 5-17 year olds. Cost is based solely on putting nurse into schools = $4.554 billion/year (not included: facility, equipment and support staff) but already provided by school health services. Cost only include programs to run during school hours.
Hoeflmayr, David and Hanewinkel, Reiner 2006 Germany	Number of students prevented from becoming established smokers. 1 year	In the school year 2001/2002, it is estimated that the SFC prevented 3,076 students from becoming established smokers, with net benefits of 5.59 Mio. Euro (direct net benefits) and 15.00 Mio. Euro (total net benefits). The direct benefit/cost ratio was 8.2 and the total benefit/cost ratio was 3.6.	Data suggest that the SFC is a cost-effective school-based intervention.	CBA societal none stated, alluded to lifetime or at least age 28.	The average societal cost saved per smoker avoided applied to number of smokers prevented to obtain benefit valuations, and cost data of the programme were collected and compared against the benefit. Cost data were collected from financial statements of the operating agency, surveys of regional co-ordinators and participating classes (direct and productivity costs). The benefit was the product of the number of students prevented from becoming established smokers, based on a stochastic progression model extending the programme's outcome evaluation, and the (direct and indirect) value per prevented smoker.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
CBA						
Wang, LY; et al 2000 USA	Yes, standard, information based, HIVE prevention curriculum. Consequences of control not identified however was accounted for by cases averted.	Program costs, see table 1 Came from trial which found significant increase in condom and contraceptive use. CE varies depending on: HIV prevalence, STD incidence rates, program costs, medical care costs for HIV and STDs and medical costs of pregnancy.		Not in primary study as was one year, but reported discounting used in other studies where cost data was used for current study. 1994 USD	Intervention cost of $105,243 achieved 15% increase in condom use, 11% increase in contraceptive use. For every dollar invested in the program, $2.65 in total medical and social costs were saved.	Yes, multivariable sensitivity analysis to determine robustness. 6 variables adjusted: probability of HIV and other STD transmission, HIVE prevalence rate, STD incidence rates, condom use per act, contraceptive failure rate, percentage of students using contraceptives and medical cost per case. Prevalence of diseases, and transmission probabilities
Eckermann, Simon; et al 2014 Australia	n/a n/a	None identified, not already mentioned in costs Analysis of the four domains of interest triangulation.	n/a n/a	no none AUD, but no price year stated	n/a	none n/a
Heckman, J.; et al 2010 USA	yes Costs of control not identified, but benefits were.	Program costs Program benefits include education, employment and earnings, criminal activity, tax payments, and use of the welfare system.	None Benefits were monetised.	Yes 3% USD 2006	n/a	Yes, present sensitive analyses of plausible assumptions in each benefit parameter. Alternative assumptions to the costs of crime, bring their rate of return estimates down from those previously published.
Tai, T; Bame, SI 2010 USA	No alternative in this analysis	Cost of school nurse staffing nationwide Reductions in cost come from reductions in: ER, hospital costs, outpatient care due to improved health status, absenteeism and premature death	none stated	Yes for calculating wages lost to premature death 3% USD		
Hoeflmayr, David and Hanewinkel, Reiner 2006 Germany	yes	Direct (17.28 bn euro) and indirect costs (16.6 bn Euro) of smoking in Germany due to forgone labour productivity based on human capital method. Programmed costs collected from the competition schools. Smoking progression model was used to estimate the number of students who were prevented from becoming established smokers	n/a n/a	yes 5%. 0%, 3%, and 10% used in sensitivity analysis Euro, doesn't specifically state price year, alludes to 2001/2002, but then also states other years, unclear.	n/a, The direct benefit/cost ratio was 8.2 and the total benefit/cost ratio was 3.6.	Sensitivity analysis varying discount rate, univariate and multivariate analyses. Monte Carlo analysis in 10,000 cycles performed on model using Crystal Ball 2000 (Decisioneering Inc.) yes, described during sensitivity analysis on a number of parameters.

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
Cost Analysis					
Li, Yanping; et al 2010 China	The nutrition-based comprehensive intervention study on childhood obesity in China (NISCOC): a randomised cluster controlled trial Bmc Public Health	To describe the design of a multi-centred random controlled school-based clinical intervention for childhood obesity in China. Secondary objective is to compare the cost-effectiveness of the comprehensive intervention strategy with two other interventions, one only focuses on nutrition education, the other only focuses on physical activity.	Multi-cantered randomised controlled trial.	Grade 1 to 5, aged 7-13 years 6 centres in Beijing, Shanghai, Chongqing, Shandong province, Heilongjiang province and Guangdong province.	Three interventions included in the present study: nutrition education, physical activity intervention and a comprehensive interventions including both.
Newbury-Birch, D.; et al 2013 UK	A feasibility trial of alcohol screening and brief interventions for risky drinking in young people in a high school setting in the UK: Sips jr-high Alcoholism: Clinical and Experimental Research	To answer the following research questions: 'is it feasible to deliver screening and brief alcohol intervention in schools in England' and 'what are the likely eligibility, consent, participation and retention rates of young people in a UK-relevant trail of brief intervention compared to standard practice?'	Feasibility 3 arm pilot trial with cluster randomisation and level of school and integrated qualitative process evaluation.	Younger adolescents (aged 14-15) in Year 10 High schools across North Tyneside.	Control arm: standard alcohol advice, which will include a leaflet about healthy living. Level one intervention: in addition to control, those who screen positively for alcohol misuse using the alcohol screening questionnaire will take part in a 30 minute personalised session delivered by the Learning Mentor. Level 2: in addition to control and level 1, young people will attend a subsequent one hour session of behaviour change counselling.
Abt, Clark 1966 USA	A COST-EFFECTIVENESS MODEL FOR THE ANALYSIS OF TITLE I ESEA PROJECT PROPOSALS, PART I-VII Technical note for Office of Education, US department of health, education, and welfare	Present design for elementary and secondary education cost-effectiveness model for the analysis of Title 1 ESEA project proposals.	Computer simulation with empirical data being collected for validation. Model is not expected to be predictive, but indicative of the relative cost-effectiveness of alternate Title 1 programs.	School children USA	Title 1 programs are for the disadvantaged, and are directed at increasing learning, through changes in student achievement, attitudes and environmental factors, social behaviours, and community impacts.
Brown, H. S.; et al 2007 USA	The cost-effectiveness of a school-based overweight program International Journal of Behavioural Nutrition and Physical Activity	To assess the cost-utility and cost-benefit of a school-based intervention designed to reduce obesity in adulthood.	Four interventions schools and four matched control schools were randomly selected out of the two largest school districts in El Paso.	Students 8-11 years US schools/primary care (n=896)	The Coordinated Approach to Child Health (CATCH) interventions includes a classroom curriculum, a physical education programme, modifications to the school food service, and family and home-based programmes.
Scherrer, CR; et al 2006 USA	Public Health sealant delivery programs: optimal delivery and the cost of practice acts Public Health Dentistry	Determine optimal combinations of staffing levels and sealant stations for school-based sealant programs.	Discrete event simulation model (OTHER)	Second graders and sixth graders age 8-14 Wisconsin schools. Obtained data from IL, OH, AZ, NM, CO and AL.	A school-based public health intervention to screen children for sealants and provide if needed.
Anderson, R; et al 2014 UK	Cost-effectiveness of classroom-based cognitive behaviour therapy in reducing symptoms of depression in adolescents: a trial-based analysis Journal of Child Psychology and Psychiatry	Determine CE of a universally delivered age-appropriate CBT programme in school classrooms to reduce depression.	Trial-based CEA based on cluster RCT comparing classroom-based CBT with usual school provision. (almost CUA)	12-16 years 8 mixed-sex UK secondary schools, including 3,357 children between the two trial arms	Data collected at baseline, 6 and 12 months.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
Cost Analysis					
Li, Yanping; et al 2010 China	Physical examination, body composition, blood biochemical indices, dietary intake situations, physical activities, physical measurements, obesity related knowledge, cost 1 year (intervention assumed to last that long)	protocol	protocol	Protocol, Alludes to CEA and CBA, but doesn't describe methods. Classified as cost analysis none stated End of intervention assumed	Costs of staff training, intervention materials, teachers and school input and supervising fee.
Newbury-Birch, D.; et al 2013 UK	Total consumption at 12 months using the TLFB-28 within intervention groups. 12 months, with 6 month follow-up	protocol	protocol	Protocol, Does not state, stated will use pilot to design economic evaluation, classified as cost analysis Broad perspective mentioned for entire study 12 months	Will describe the costs of introducing and running the brief intervention and will focus on examining what school resource data should be collected (and how) in terms of ongoing staff and capital costs.
Abt, Clark 1966 USA	Not an intervention study Not an intervention study	No actual results, this was a technical note to aid those decision makers who might use the model.	The model is not intended a research tool rather an evaluation and planning tool.	Simulation model Societal/government, not actually stated None specifically stated. Costs projected for the life of the program.	Indirect and direct costs of the Title 1 programs
Brown, H. S.; et al 2007 USA	No. of overweight cases averted CATCH trial followed-up over three years. Model extrapolated to age 40 to 64 years	CER calculated using general population and Hispanic estimates. See ICER column.	CER for CATCH was $900 well under $30000 threshold and sensitivity analysis reveals the results are robust.	Model based on a single trial and populated with estimates from the literature Societal Model extrapolated to ages 40 to 64 years	Cost of intervention, medical costs, and productivity costs. Medical costs derived from literature on hypertension, hypercholesterolemia, type 2 diabetes, CVD and stroke. Productivity estimated as difference in no. of sick days taken by obese adults compared to non-obese adults.
Scherrer, CR; et al 2006 USA	Optimal quantity of labour and capital to minimise program costs. Simulation of programs for delivering sealant	For general, direct or indirect supervision it is optimal to have only 1 dentist or less. For general supervision, its optimal to have dentist and assistants come on separate days. Sig saving by reducing supervision level.	States could save money by relaxing restriction on the type of personnel who can deliver sealants in public health settings and by productivity gains through proper consideration of staffing. Savings could be used to improve access and reduce disparities.	Model, discrete event simulation societal/Wisconsin 10-year life of sealants	Staffing costs, travel time, equipment. Did not included admin costs.
Anderson, R; et al 2014 UK	Individual self-report on care costs, QALYs (base on EQ5D) symptoms of depression (Short mood and feelings questionnaire) 12 months	Lower QALYs in CBT group -0.05QALY/person (CI: -.09, -.005, p=.03) but 'clinically negligible' difference not found in complete case analysis. Little evidence of any between-arm differences in SMFQ scores or costs per person for CBT vs usual school provision	Universal provision of classroom CBT is unlikely to be either more effective or less costly than usual school provision.	CEA and CUA attempted, classified as cost, but nearly CUA NHS and Social care 1 school year Sept 2009 to July 2010	Cost of delivering programme from detailed project records of staff time and other expenditure: paid time of facilitators, training, ongoing supervision, travel costs, printing, and recruiting schools, room hire and subsistence.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
Cost Analysis						
Li, Yanping; et al 2010 China	Yes, but did not include a do nothing approach. Not enough detail provided	none mentioned Did not state which outcome measure would be the effectiveness measure that would be analysed in the CEA/CBA.	none stated Did not give enough detail	none stated None stated, RMB assumed	Did not state if will calculate an ICER.	Did not state if this would be performed. None
Newbury-Birch, D.; et al 2013 UK	Yes Will use the pilot to identify important costs and consequences.	See costs. Did not state which outcome measure would be the effectiveness measure that would be analysed.	none stated none	none stated None stated, GBP assumed	Did not state if will calculate an ICER	Did not state if this would be performed none
Abt, Clark 1966 USA	No, but the aim of this technical note was not to evaluate two opposing programs, but explain how the model worked. Not an intervention study	Described in the cost simulation part of the model. Report use of churchman-Ackoff approximate measure of value procedure to value outcomes.	n/a n/a	Yes, mentions discounted to present value No rate mentioned USD	Chart 5 shows info needed to determine incremental value per dollar of the Title 1 program being implemented.	none Made up of various sub models: school, instructional process, community, cost, effectiveness and the simulation
Brown, H. S.; et al 2007 USA	Yes, they were matched No, seemed to only identify cost and benefits of the intervention and assume no cost/benefit change from control. Medical costs assumed to accrue in control come from literature.	Intervention costs, medical costs averted (direct costs), labour productivity costs. Used NHIS survey questions on activity limitations to weight health states and estimate QALY. Used an equation in the appendix.	It is unclear how the health states were valued. QALY and monetary benefits	yes 3% USD 2004	Report CERs instead $900 per QALY gained compared to no intervention. When calculating numerator totalled costs of intervention less medical costs due to obesity (didn't take difference in cost b/t groups). Net benefit $68,125	Sensitivity analysis using a triangular distribution for 1000 simulations. In all cases intervention remained cost-effective and net beneficial. 48 in total varied in sensitivity analysis, see table 2 and 4.
Scherrer, CR; et al 2006 USA	yes Costs yes, consequences no	see costs All assumed. Discuss possible sources of effectiveness in discussion, but don't specify its use in the analysis.	costs	yes 3% 2003 USD, assumed	none	Varied sensitivity around dentist's productivity by assuming dentists have higher productivity than hygienists. Mainly just costs of the different strategies.
Anderson, R; et al 2014 UK	Yes, usual school provision Yes, for costs used adapted CSRI	Service use, adapted CSRI, health service visits and hospital inpatient stays EQ-5D, cost per QALY and incremental cost per unit decrease in SMFQ score	EQ-5D QALY	no GBP 2010	Don't report them calculated, but see table 4. My calculations: cost/Daly -£2783, cost/SMFQ £348	none no model

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
Cost Analysis					
Stallard, P; et al 2013 UK	A cluster randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of classroom-based cognitive-behavioural therapy (CBT) in reducing symptoms of depression in high-risk adolescents Health Technology Assessment	This is the same study as Anderson, 2014, but it is the full HTA report.	Trial-based CEA based on cluster RCT comparing classroom-based CBT with usual school provision. (CUA)	12-16 years 8 mixed-sex UK secondary schools, including 3357 children between the two trial arms	Data collected at baseline, 6 and 12 months.
Atherly, Adam; et al 2009 USA	An Economic Evaluation of the School-Based power Breathing Asthma Program Journal of Asthma	Evaluate the cost effectiveness of the 'Power Breathing' programme for asthma among middle and high school students.	Matched quasi-experimental design which matched schools based on grade range, enrolment, income and race/ethnicity which were then randomised to intervention or control group.	Adolescents in grades 6-12 (aimed at ages 11-19) 8 junior high schools and 2 high schools in Kansas and Virginia. N=524	Designed to be implemented in a school-based setting, consists of three 90 minute educational sessions focusing on education about asthma, asthma control strategies and psychosocial concerns. More detailed description on page 596.
Babey, Susan H.; et al 2014 USA	How can schools help youth increase physical activity? An economic analysis comparing school-based programs Preventive Medicine	To assess the following types of school-based opportunities to improve physical activity for youth: after-school programs, before-school programmes, PE classes, extended-day PE, and short physical activity breaks during the school day.	Cost analysis, hypothetical intervention applied with estimated costs and effects populated with data from the literature.	School-aged children US schools	1. After-school program, typically from 3 to 6 PM. 2. Extended school day (40-60 min) with increased time for PE class, mandatory for all students. 3. In-class activity consisting of to 10-min breaks of structured physical activity. 4. Before-school activity program, with volunteer or professional supervision available for 30 minutes before school during regular school days.
Boyle, J; et al 2007 UK	A randomised controlled trial and economic evaluation of direct versus indirect and individual versus group modes of speech and language therapy for children with primary language impairment Health Technology Assessment	To compare language outcomes following direct individual therapy working individually , indirect individual therapy, direct group therapy and indirect group therapy for primary school children with language impairment.	2X2 factorial design trial with control receiving existing levels of support. Includes 'short-run' economic evaluation. (non-economic)	6-11 years School settings in Scotland	Compared 4 different strategies
Brassard, P., et al 2006 Canada	Evaluation of a school-based tuberculosis-screening program and associate investigation targeting recently immigrated children in a low-burden country Pediatrics	Retrospectively evaluate a school-based screening program targeting children at high risk for TB infection in Montreal, Canada, as well as subsequently investigate family and household associates of the schoolchildren with latent TB infection (LTBI), based on adherence to LTBI therapy and cost-benefit analysis.	Retrospective evaluation of school screening for TB in immigrant children	4-18-year-olds Montreal, Canada schools (n=16)	In selected schools, immigrant children were screened using the tuberculin skin test (TST). Those who tested positive was referred for medical evaluation and the family and household associates were also screened.
Curtale, F.; et al 2005 Egypt	Control of human fascioliasis by selective chemotherapy: Design, cost and effect of the first public health, school-based intervention implemented in endemic areas of the Nile Delta, Egypt Transactions of the Royal Society of Tropical Medicine and Hygiene	Assess the efficacy of the programme, and gather operational information including cost	Evaluation of a selective screening programme	School-aged children 10-12 Nile Delta	Unclear what was control and what was intervention, not well described. Assuming intervention was the selective treatment delivered at school, but this comes under heading 2.1 The control programme. Must be the name of the program.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
Cost Analysis					
Stallard, P; et al 2013 UK	Individual self-report on care costs, QALYs (base on EQ5D) symptoms of depression (Short mood and feelings questionnaire)	Lower QALYs in CBT group -0.05QALY/person (CI: -.09, -.005, p=.03) but 'clinically negligible' difference not found in complete case analysis. Little evidence of any between-arm differences in SMFQ scores or costs per person for CBT vs usual school provision.	Universal provision of classroom CBT is unlikely to be either more effective or less costly than usual school provision.	CEA and CUA, classified at other since no report of ICER NHS and Social care 1 school year Sept 2009 to July 2010	Cost of delivering programme from detailed project records of staff time and other expenditure: paid time of facilitators, training, ongoing supervision, travel costs, printing, and recruiting schools, room hire and subsistence.
Atherly, Adam; et al 2009 USA	Cost/symptom free day (gain or loss as compared to baseline) mean comparison and OLS regression with random effects (to correct for clustering) Baseline, post intervention and 3 month follow-up.	Intervention increased symptom free days and control decreased, but not statistically significantly. The decline was statistically significant in intervention. Hypothesised effect of intervention depended on baseline symptoms.	School-based interventions aimed at asthma, properly implemented and administrated are an appropriate use of societal resources.	Stated 'economic evaluation', classified as cost analysis as didn't quite meet definition of CUA/CEA societal 3 months	Direct costs associated with medical service use, emergency department visits, hospitalizations, outpatient care, and prescription drugs as well as cost of devices such as peak flow meters. Indirect costs such as lost productivity (including parent work absence), school absences and waiting times at doctor's office and start up implementation costs.
Babey, Susan H.; et al 2014 USA	Cost/MET-hours-gained per year. Wasn't an intervention study so no follow-up	Costs and 'cost-effectiveness' varied considerable across the four programs. In order of most cost-effective: in-class physical activity breaks, instructor-led physical activity before school, extended school day with mandatory PE and finally on-sited after-school programmes.	Inserting activity breaks into the day is appropriate especially when youth must sit at a desk all day and adults have sedentary jobs. Inserting routine breaks at an early age might translate into physical activity breaks in the workplace.	Stated 'economic analysis', classified as cost analysis none stated None stated explicitly, does say annual costs and outcome is MET hours/year, so can assume 1 year	Mainly program costs or operational costs estimated, costs to family as well, but didn't specify what the costs were made up of other than 'out-of-pocket' expenses to participate in programme.
Boyle, J; et al 2007 UK	Standardised scores on test of expressive and receptive language.	Within trial econ evaluation identified indirect group therapy as the least costly with individual therapy as most costly… effectiveness?		classified as cost analysis Individual child for outcome Within trial, had aims to do longer-term, but effectiveness was not sustained at 12 months so unclear whether any longer term treatment effects would be identified.	Two major components: salary cost associated with each mode of delivery and travel cost associated with the delivery method
Brassard, P., et al 2006 Canada	TST positivity rate Retrospectively reviewed hospital charts from 1998 to 2003	The only predictor of adherence was having more than 2 family members brought in for TB screening (adjusted OR: 2.0; 95%CI: 1.3-3.3). 78% of associates of TST positive children who were positive adhered to therapy. Net benefits from both school-based screening and associate investigation, as stand alone or coordinated.	They demonstrated effectiveness, including cost-effectiveness, of a targeted, school-based screening program in a low-burden country and the extra benefit given by adding associates to such a program.	'Cost-benefit analysis', classified as cost analysis none stated None stated explicitly, states over the 5-year period so assume 1998-2003	Total material and labour costs associated with the school-screening and associate investigations. Cost of treating TB derived from recent CBA of treatment of TB in Montreal.
Curtale, F.; et al 2005 Egypt	Control of human fascioliasis. In terms of prevalence? Intensity? Treatment? Primary Outcome not well described Programme screening from 1998 to 2002	Prevalence in the endemic area was reduced from 5.6% to 1.2%.	The targeted selective approach was appropriate in addressing low prevalence infection, effective in reducing prevalence rates and transmission of the disease, and in the present situation, more cost-effective than mass distribution.	costing none stated none stated	Gives some costs in LEG, but doesn't explain how converted to USD and explicitly state the price year that all costs are reported in.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
Cost Analysis						
Stallard, P; et al 2013 UK	Yes, usual school provision Yes, for costs used adapted CSRI	Service use, adapted CSRI, health service visits and hospital inpatient stays EQ-5D, cost per QALY and incremental cost per unit decrease in SMFQ score	EQ-5D QALY	no 2010 GBP	Don't report them calculated, but see table 4. My calculations: cost/Daly -£2783, cost/SMFQ £348	none no model
Atherly, Adam; et al 2009 USA	Identified as control, no other description except that they 'identified the costs and outcomes of asthma' in both groups. Didn't consider opportunity cost of school time lost due to intervention	Compensation to students and staff for participation, time spent by students, parents, teachers, school nurses, facilitators. Total cost given is $30.37 per student. Symptom free day. Table 1 has effectiveness data, but a cost table that correlates to that resource use is missing.	Mention asthma related quality of life, but don't provide any figures. Symptom free day gained	no none USD no price year stated	Did not calculate correctly. Based calculation on intervention only by multiplying intervention effect on baseline days. Costs are not reported for control (even though said identified costs and outcomes), so could not calculate ICER myself.	none none
Babey, Susan H.; et al 2014 USA	yes Costs identified, consequences less well thought out. Much of the analysis was theoretical with a lot of assumptions.	Operating cost per child per year (from previous study), family costs such as out-of-pocket costs to attend. Did not include potential effect on health care costs. Minutes of moderate to vigorous physical activity, assumed, and MET hours - assumed.	none Metabolic equivalent of a task (MET) hours. This was standardised so the same measure of benefit was applied to each intervention.	no none 2012 USD	Based on all comparators gaining the same effect (i.e. 270 MET hours/year). Incremental not calculated. Cost per MET hour gained calculated for the following: after-school on-site $10.62, after school off-site $11.29, Longer day PE $.65-$.98 (405 vs 270 MET hours gained), In class $0.008-$0.01, Before school volunteer lead, negligible and teacher led $.49.	none none
Boyle, J; et al 2007 UK	yes yes	salary, transportation Change in total CELF-3 total language score	No, outcome was change in CELF-3 score	no GBP	Incremental analyses performed but in non-traditional way	no no
Brassard, P., et al 2006 Canada	no No, seemed to only identify intervention costs.	Testing, wages and materials, appointments, chest radiographs, interpreters TB cases prevented (estimated), therapy adherence. Doesn't describe how benefits were valued.	none none, not CUA	Yes, but only for cases prevented 3% CAD, no year specified	none, 'CBA'	Sensitivity analysis varying cost assumptions and rate of hospitalization and probability of lifetime risk of developing TB. None
Curtale, F.; et al 2005 Egypt	no no	Personnel and material costs of providing intervention. infection intensity	none none, not CUA	no none USD, price year not stated	Cost per child screened $0.50, cost per child treated $2.33	none none

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
Cost Analysis					
Foster, E. Michael 2010 USA	Costs and Effectiveness of the Fast Track Intervention for Antisocial Behaviour Journal of Mental Health Policy and Economics	The Fast Track intervention is a 10-year, multi-component prevention program targeting antisocial behaviour. The intervention identified children at school entry and provided intervention service over a 10-year period. This study examined the intervention's impact on outcomes affecting societal costs using data through late adolescence.	Multi-cohort, muti-site, multi-year randomised control trial of program participants and comparable children and youth in similar schools.	School-aged children grade 1-10 US schools within four sites (Durham, NC; Nashville, TN; Seattle, WA; and rural Pennsylvania)	High risk children identified in schools and those scoring in the top 40% were selected to receive the intervention, 91% agreed (n=3,274). Intervention delivered from grades 1 though 10. Parents offered parent training and home visiting, academic tutoring, and social skill training. Parent and child groups offered 2-hour enrichment program. Group meetings held weekly, biweekly, then monthly each year on. Social emotional learning program PATH adapted in school curriculum.
Gelli, Aulo; et al 2009 Italy	The costs and cost-efficiency of providing food through schools in areas of high food insecurity Food and Nutrition Bulletin	To estimate the programmatic costs and cost-efficiency associated with providing food through schools in food-insecure, developing-country contexts, by analysing global project data from the World Food Programme.	Retrospective cost evaluation in terms of specified unit of effect (i.e. unit of nutrient delivered).	School-aged children Food-insecure developing countries	Providing food through schools from the World Food Programme.
Gelli, Aulo; et al 2011 Italy	New benchmarks for costs and cost-efficiency of school-based feeding programs in food-insecure areas Food and Nutrition Bulletin	Address the need for systematic estimates of the cost of different school feeding modalities, and of the determinants of the considerable cost variation among countries. The project aimed to update current benchmarks for school feeding cost and cost efficiency and understand cost drivers and cost-containment opportunities.	Retrospective cost evaluation in terms of specified unit of effect (i.e. unit of nutrient delivered).	School-aged children Food-insecure developing countries	Providing food through schools from the World Food Programme.
Gesell, SB; et al 2013 USA	Comparative Effectiveness of After-school programs to increase physical activity Journal of Obesity	Comparative effectiveness analysis to evaluate the difference in the amount of physical activity children engaged in when enrolled in a physical activity enhanced after school program based in a community rec centre vs a standard school-based after-school program	Natural experiment with 54 children attending community ASP and 37 attending school-based ASP (cost analysis). An observational prospective cohort and natural experiment in Nashville, TN.	Aged 5-13 Nashville, TN	Both ASP followed similar formats from 3-6PM after school, time for snack, homework and play and did not focus on a single activity. Intervention set in community rec and had staff led games. School ASP set in cafeteria had arts and crafts and playing in playground.
Guay, M.; et al 2003 Canada	Effectiveness and cost comparison of two strategies for hepatitis B vaccination of schoolchildren Canadian Journal of Public Health	To compare the effectiveness and costs of school-based and clinic base hepatitis B vaccine programmes.	Cost-effectiveness analysis of multi-centre trial in 2 community clinics and 55 schools, authors state quasi-experimental. Reclassified by CRD as cost consequence as no summary measure of benefit.	8-9 (grade 4 in Canada) School, community care, Monteregie, Canada	Natural experiment where one community clinic replaced school-based vaccine programme with vaccine offered in community clinics after school hours.
Joseph, C. L.; et al 2007 USA	A web-based, tailored asthma management program for urban African-American high school students American journal of respiratory and critical care medicine	Develop and evaluate a multimedia, web-based asthma management program to specifically target urban high school students. The program uses 'tailoring' in conjunction with theory-based models to alter behaviour through individualized health messages based on the user's beliefs, attitudes and personal barriers to change.	RCT, Students reporting asthma symptoms randomised to tailored web programme versus generic asthma websites. N=314	9th through 11th graders, average age 15 Six Detroit public high schools	Web-based programme that focuses on three core behaviours: controller medication adherence, rescue inhaler availability, and smoking cessation/reduction. Four education computer session that uses normative and positive feedback. Core behaviour status determined at session 1 and reassessed during sessions 2-4.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
Cost Analysis					
Foster, E. Michael 2010 USA	Impact on societal cost 10 years	The intervention lacked both the breadth and depth of effects on costly outcomes to demonstrate cost-effectiveness or even effectiveness	The most intensive psychosocial intervention ever fielded did not produce meaningful and consistent effects on costly outcomes. The lack of effects through high school suggest that the intervention will not become cost-effective as participants progress through adulthood.	costing study payer perspective 10 years	Didn't actually cost all resource use. Asked about health service use, criminal activity, school service use, substance abuse and financial assistance and tested for difference between groups using regressions or survival models. Prior analysis was the intervention cost $58,000/child, and did report mental health services costs $6,249 vs $4,905 control vs intervention and general health $6,572 vs $5,466, but not enough to offset cost of program.
Gelli, Aulo; et al 2009 Italy	Yearly expenditure per child None, assumed WFP will continue	Average yearly expenditure US$21.59 per child ranging from $11 to $52. Fortified biscuit most CE in terms of micronutrient delivery, onsite meals best for total calories. Transportation and logistics main cost drivers.	Choice of program objectives will dictate type of food. Fortified biscuits can provide substantial nutritional inputs at a fraction of the cost of school meals, making an appealing option.	costing study WFP assumed, none stated none stated	From WFP Standard Project Reports: expenditures, beneficiaries, and food distribution. From WFP Country Offices' estimated year expenditure by beneficiary.
Gelli, Aulo; et al 2011 Italy	Standardized yearly average school feeding cost per child None, assumed WFP will continue	Average school feeding cost per child non including school-evaluation costs was US$48. Cost was lowest for biscuit programs at US$23 and highest for take-home ration programmes reaching families US$75. Combination on-site and take home $61.	Both costs and effects should be considered carefully when designing school feeding interventions. The average costs of school feeding estimated are higher than those found in earlier studies but fall within range of earlier studies.	costing study WFP assumed, none stated none stated	The standardised costs =cost per beneficiary project expenditure, number of on-site feeding days, standard parameter for ration kcals per modality, planned ration kcals, theoretical food tonnage and actual food tonnage delivered.
Gesell, SB; et al 2013 USA	Physical activity using ActiGraph GTiM accelerometers. 12 weeks	At baseline, 43% of the multi-ethnic sample was overweight/obese, the mean age was 7.9 years. Cost analysis suggested that children attending tradition school-based ASPs, at an average cost of $17.67 per day would need an additional daily investment of $1.59 per child for 12 weeks to increase their moderate-vigorous physical activity.	A low-cost, alternative after-school program featuring adult-led physical activities in a community recreation centre was associated with increased physical activity compared to standard-of-care school-based-after-school programs.	States CEA, but no ICER, costing study societal inferred 12 weeks	Just implementation costs for 12 week study period = $1184 per child ($19.25 daily per child) vs $1087 per child ($17.67 daily per child) for school based ASP.
Guay, M.; et al 2003 Canada	Percentage of school cohort vaccinated None, vaccine	With community clinics vaccine coverage fell to 73%, compared to over 90% in schools. Societal costs were $63 in community clinic and $40 in school.	Results demonstrate the advantage of a school versus community based immunization programme.	Authors did not derive a summary measure of health benefit, therefore CCA societal 1 year	Only costs that varied between programmes counted (i.e. cost of vaccine not included). Measured costs: labour costs in clinic, running expenses of clinic, labour costs in the school, other costs, and cost incurred by the parents on taking their children to the clinic.
Joseph, C. L.; et al 2007 USA	Number of symptom days in the last 2 weeks. 12 months	12 months treatment students report fewer symptom days 0.5(0.4-0.8), p=0.003; symptom-nights 0.4(0.2-0.8), p=0.009; school days missed 0.3(0.1-0.7), p=0.006; restricted activity days 0.5(0.3-0.8), p=0.02; and hospitalisations 0.2(0.2-0.9), p=0.01.	Cost estimates were $6.66 per participating treatment group student. A web-based tailored approach to changing negative asthma management behaviours is economical, feasible and effective in improving asthma outcomes in a traditionally hard-to-reach population.	costing study none stated 12 months	Referral coordinator labour costs $6.66 ($8.05 per treatment student referred and $11.73/student contacted.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
Cost Analysis						
Foster, E. Michael 2010 USA	No, just described as control group. No	Yes, see costs, collected important resource use but didn't report costs associated with each group in a table. Some figures reported in text i.e. health care use, but not all. It is not clear what the overall main effectiveness measure was. It can be inferred effect would have measured by a decrease in societal resources (i.e. cost savings as compared to the control).	none none	no none USD, no price year stated	n/a	no n/a
Gelli, Aulo; et al 2009 Italy	No, an evaluation of programmes already implemented There isn't really an appropriate control	see costs Nutrient/kcal delivered	none none	no none USD, no price year stated	Not ICER, but reported cost/nutrient delivered. Standardised beneficiary $21.59 ($4.51-$96.81); cost/100kcal $3.60; cost/mg iron $2.33; cost/100mg vitamin A $8.73; and cost/100mg iodine $813.32	no no
Gelli, Aulo; et al 2011 Italy	No, an evaluation of programmes already implemented No appropriate control	see costs Food quantity, calories, proteins, and micronutrient content delivered were used to assess the cost-efficiency.	none none	no none USD, no price year stated	No ICER, but reported cost/child. School feeding project cost/child $29, Standardized school feeding cost/child $48, standardised range $15-$213.	no no
Gesell, SB; et al 2013 USA	no no	Implementation costs, facility costs, staff costs Physical activity, body fat percentage, fitness	none	no none USD 2010	No ICER, costing study.	none none
Guay, M.; et al 2003 Canada	yes Yes, table two reports costs to clinics, schools, parents and a unit cost/student vaccinated.	see costs Community vaccine rate 73%, school 90-95%.	none none	No none CAD 1997-8, not stated directly	n/a, but mean costs and effects available to calculate. Most expensive school strategy dominates community strategy.	no none
Joseph, C. L.; et al 2007 USA	yes Does not report costs clearly for each alternative in a table.	Referral contact, average 31 minutes per student, hospitalisations, ED visits Symptom days, asthma severity, ED visits, hospitalisations	Mention Disease specific QoL questionnaire developed by Juniper et al, 1994. Not clear, report QOL domains but not mention of scale.	no none USD, no price year stated	none stated	none none

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
Cost Analysis					
Meng; et al 2013 China	BMI One academic year	The comprehensive intervention was the only combination to lower BMI significantly as compared to its control. BMI and BAZ increment was 0.65 kg/m2(SE 0.09) and 0.01(SE0.11) which was significantly lower than the control (0.82+-0.09 for BMI, 0.10+-0.11 for BAZ).	The school-based integrated obesity intervention program was cost-effective for children in urban China.	CEA, but only CERs reported, classified as cost analysis State social perspective, assume meant societal. 1 academic year	Costs collected by retrospective interview of project coordinator in each centre. Collected a) intervention costs; b) evaluation costs; and c)the development cost on structuring the program before the intervention.
Salisbury, C.; et al 2002 UK	Attendance for asthma review, symptom control, and quality of life Six months	See summary outcomes for primary outcome results. Those at school clinics had greater knowledge of asthma (+0.38, 95% CI= 0.19 to 0.56), most positive attitudes (+0.21, 95%CI = 0.05 to 0.36), and better inhaler technique (P<0.001).	The schools asthma clinic increased uptake of asthma reviews. There were improvements in various process measures,, but not in clinical outcomes.	Not a formal economic evaluation, but could potentially conduct one if the right information is given, cost analysis NHS none mentioned	Economic outcomes were total costs of different models of asthma care, taking into account consultations in school, practice or hospital, and the costs of drug treatment.
Shemilt, I.; et al 2004 UK	Not specified, key outcomes defined by balance sheet approach from broad cost-benefit perspective.	No significant impact on outcomes, conduct was only variable to show significant decrease due to funding but may be by chance. Some signs in unexpected direction.	Little quantitative evidence that the clubs had impacted on health, education or social outcomes. Levels of funding were not significant determinant of observed outcomes in either type of school (primary/secondary).	Cost analysis Not specified, alluded to multiple levels 1 year	Funds associated with setting up and maintaining breakfast clubs
Wang, Li Yan; et al 2003 USA	Cases of adulthood overweight prevented and QALYs saved. 2 years	Base-case assumptions, intervention cost $33,677 or $14 per student per year. The program would prevent 1.9% of the female students from becoming overweight adults resulting in 4.1 estimated QALYs saved. $4305 per QALY saved and net saving to society of $7,313.	The Planet Health program is cost-effective and cost-saving as implemented. School-based prevention programs of this type are likely to be cost-effective uses of public funds and warrant careful consideration by policy makers and program planners.	CEA stated, classified as cost analysis societal 25 years, from age 40 to 65.	Intervention costs estimated from retrospective cost analysis of the program. Estimated medical costs saved by intervention as costs averted per case of adulthood overweight prevented. Finally costs of lost productivity averted per case of adult overweight prevented. All projected costs, based on a lot of assumptions.
Young, T. L.; et al 2003 USA	Satisfaction questionnaires completed by providers, nurses, children and parents. 2 years	Provider, nurse, child and parent satisfaction were high. Average family savings per encounter were 3.4 hours of work time ($43) and $177 in emergency department or $54 in physician costs. Including travel saving for families ranged $101 to $224 per encounter.	Telehealth technology was effective in delivering Pediatric acute care to children. It was an acceptable alterative to traditional health care delivery. The POTS-based technology helps to make this cost-effective alternative for improving access to primary and psychiatric health care for underserved children.	Not an economic evaluation. Societal assumed 2 years	Cost of technology and training, school nurse, and consultant time. This was compared to the average cost of a visit to the ED or paediatrician. Potential savings to parents included time away from work and travel avoided by having their child's condition diagnosed and treated in the school telehealth program.

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
Cost Analysis					
Meng; et al 2013 China	The Costs and Cost-Effectiveness of a School-Based Comprehensive Intervention Study on Childhood Obesity in China Plos One	To evaluate the costs and cost-effectiveness of a comprehensive intervention program for childhood obesity.	A multi-centre randomised controlled trial conducted in six large cites during 2009-2010.	Primary school students aged 6-13 (n=8301) Schools in Beijing and 5 other cities.	Three interventions were compared: Nutrition education, PA, and comprehensive intervention which was a combination of nutrition and PA.
Salisbury, C.; et al 2002 UK	A randomised controlled trial of clinics in secondary schools for adolescents with asthma British Journal of General Practice	To compare a nurse-led clinic in schools versus car in general practice for adolescents with asthma.	RCT in four schools with parallel observational study in two schools. From the detail provided, does not appear to be a formal economic evaluation.	Adolescents 12-14 Six comprehensive schools in the North Bristol NHS Trust	In the randomised trial, pupils were invited to attend asthma review at a nurse-led clinic either in school, or in general practice. The parallel observation study compared pupils invited to practice care within and outside the randomised trial.
Shemilt, I.; et al 2004 UK	A national evaluation of school breakfast clubs: where does economic fit in? Child: Care, Health and Development	Describe the economics of UK school breakfast clubs, estimate costs and investigate relationship between costs and outcomes.	Cluster RCT aim (didn't happen) with postal survey of one year follow-up, secondary econ analysis (other)	Primary and secondary school English schools serving deprived areas	Improve children's nutrition and education by providing breakfast to children who might otherwise start the day without eating.
Wang, Li Yan; et al 2003 USA	Economic Analysis of a School-Based Obesity Prevention Program Obesity Research	To assess the cost-effectiveness and cost-benefit of Planet Health, a school-based intervention designed to reduce obesity in youth of middle-school age children.	CEA stated of RCT of Planet Health	RCT girls 10-14, model projections women 40-65 Middle schools	Planet Health is a school-based intervention designed to reduce obesity in youth of middle-school age. An interdisciplinary curriculum was infused into four major subject areas in into physical education focusing on television viewing, decreasing consumption of high-fat foods, increasing fruit and vegetable intake and moderate and vigorous physical activity.
Young, T. L.; et al 2003 USA	Effectiveness of school-based telehealth care in urban and rural elementary schools Pediatrics	To evaluate the quality and cost effectiveness of health care provided in urban and rural elementary school-based telehealth centres, using plain old telephone system (POTS) technology.	A prospective study using and exploratory design. Used a convenience sample of students who had parental consent.	Students aged 6 to 12 1 urban and 2 rural elementary schools, one rural school dropped out, so just 2.	Each school was staffed with a full-time school nurse and part-time mental health therapist. The consultant clinical site was staffed by paediatricians and Pediatric nurse practitioners. POTS uses regular telephone lines as this is simpler and less expensive to operate. Each school had two lines and a fax and POTS transmitter sending ears, nose, and throat endoscope to a clinical site.

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
Cost Analysis						
Meng; et al 2013 China	Yes, but each alternative seemed to have a separate control in each area. Detailed costs were described, but inclusion of evaluation costs was questionable.	Labour, training, materials, travel and accommodation. BMI, BMI z-score reduction and cases of overweight and obesity prevented, estimated based on observed BMI reduction in intervention and applied to control.	none None, see effectiveness data	no none USD 2010	Calculated CERs and effects was difference between before and after BMI, not a comparison between interventions. Can calculate yourself from tables 2 and 4: Control vs Nutrition = 52.8/(.74-.72)= $2640/BMI reduction point Nutrition vs PA =(52.3-52.8)/(.76-.74)= -$25. PA vs Combi PA dominates in all cities, more effective and less costly. Not what the authors conclude.	no none
Salisbury, C.; et al 2002 UK	Yes, the control was described as normal care. Does not give enough detail, e.g. 'resource use were collected from clinic records, general practice records, and follow-up questionnaires'	See left. Tables do give details of resource use. Practice costs include doctor consultations, practice nurse consultations, and drug costs. School asthma clinic costs include nurse salaries, admin salaries, postage and consumables. Hospital costs include admissions, outpatient, and A&E visits. Proportion of patients who had a review for asthma in six months, and HRQoL as measured by the Paediatric Quality of Life questionnaire, standardised UK version, PAQLQ.	No mention of conversion of utility scores to QALY. None.	no n/a GBP no year stated.	None stated, mean costs given: School £56.63, GP £38.11 difference =£18.52. Difference in means for Quality of life -0.06. Negative ICER result, no CE plane given or bootstrap performed.	None n/a
Shemilt, I.; et al 2004 UK						
Wang, Li Yan; et al 2003 USA	no no	Not identified in retrospective cost analysis except cost described. CER = (C-NA)/NQ and net benefit =NA+NB - C, had to estimate C, N, Q, and B. Individual estimates of QALYs and productivity loss costs not available in the literature. All effectiveness data estimated from literature/their own methods.	QALY estimated by author QALY	yes 3% 1996 USD	CER = $4,305, not an ICER	Sensitivity analysis varying 10 parameters, one-way and all together. See Table 1 See table 1, most parameters estimates from literature, or author's calculations. Difficult to tell how heterogeneous the estimates from literature are to the current study's population of interest.
Young, T. L.; et al 2003 USA	No alternative included in observation, rather alternative forms of medicine i.e. visit to ED/paediatrician described when making cost savings assumptions. No	See costs, no additional resource use collected. No effectiveness data collected except relating to satisfaction. This was an observation study so findings were descriptive.	none none	no none USD, no year stated	none	none none

Authors Year of Publication Country of Origin	Title Journal	Aims/Objectives	Study Design	Age Setting	Intervention Description
Non-economic evaluations					
Kristensen, Alyson H.; et al 2014 USA	Reducing Childhood Obesity through US Federal Policy A Microsimulation Analysis American Journal of Preventive Medicine	To estimate the impact of three federal policies to reduce childhood obesity prevalence in 2032 after 20 years of implementation.	Markov microsimulation model of policy impacts on diet/physical activity and BMI	School-aged population, 6-18 years USA national policy, schools	Three federal policies simulated: 1) after school physical activity programs, 2) sugar sweetened beverage excise tax, and 3) ban on fast food television advertising targeting children.
Bruzzese, Jean-Marie; et al 2006 USA	Using School Staff to Establish a Preventive Network of Care to Improve Elementary School Students' Control of Asthma Journal of School Health	Test the efficacy of a preventive care network for children with asthma which included 1) school nurses coordinating relationships between families, primary care providers (PCP) and school personnel and 2) school personnel and PCPs receiving training regarding asthma management.	Randomised controlled design of 44 schools from the 5 boroughs of NYC.	Kindergarten to grade 5. Age approx. 5-11. US schools/primary care (n=591)	School health team consisting of full-time school nurse, school physician (2 days/month), public health assistant (2-3 days/week), teacher/administrator and parent. Participated in 3 day training workshop then worked with parents to develop an asthma management plan.
Massoni, Sebastien; et al 2012 France	How to Improve Pupils' Literacy? A Cost-Effectiveness Analysis of a French Educational Project Economics of Education Review	Evaluate the Action Lecture program and use the estimation of impact on academic achievement to conduct a cost-effectiveness analysis and take a reduction of the class size program as a benchmark.	Non-randomised intervention and control study. Schools had to apply, and as not many did all who applied were admitted. Control schools have to be sourced separately and agree to participate. Use a difference in difference method.	Paris nursery and primary schools, six schools, over 400 participants.	Unconventional intervention in which pupils don't have any courses for 2 weeks, but work together on a specific topic with different activities (reading, research, museum visit, writing, etc.). The goal is to develop the takes for reading and discovery and increase motivation to attend school.
Wade, T. J.; et al 2008 USA	Improvements in health-related quality of life among school-based health centre users in elementary and middle school Ambulatory Pediatrics : the official journal of the Ambulatory Pediatric Association	To examine the role of school-based health centres (SBHCs) on changes in student health-related quality of life over a 3-year period among elementary and middle school students.	Three-year longitudinal prospective study.	Kindergarten to grade 8. Approximately age 5 to 14. Four intervention elementary schools and four comparison schools	SBHC in elementary schools. Most research on SBHC is for adolescents and reports improved student health. The research for SBHC in elementary schools improves access and reduces emergency department use, but it is unclear whether improvements in health translate to younger students.

Authors Year of Publication Country of Origin	Primary Outcome Follow-up	Results (OR, RR, risk ratio, CI, p-value, mean diff)	Conclusions	Type of evaluation Viewpoint/perspective Time horizon	Costs
Non-economic evaluations					
Kristensen, Alyson H.; et al 2014 USA	BMI and changes in percentage of overweight or obese youth 20 years	Afterschool physical activity programs would reduce obesity most among 6-12 year-olds (1.8 percentage points) and advertising ban would reduce obesity the least (0.9 percentage points). Sugar sweetened beverages excise tax would reduce obesity most 13-18 years (2.4 percentage points).	All three policies would reduce childhood obesity prevalence by 2032, however a national $0.01/ounce SSB excise tax is the best option.	Markov microsimulation model, on effects alone none stated 20 years	Not actually modelled, only policy impacts were modelled on decreasing childhood obesity prevalence
Bruzzese, Jean-Marie; et al 2006 USA	No. of days with symptoms 2 years	Few improvements in health outcomes achieved, no impact on use of urgent health care services, school attendance, or caregiver's QoL.	1) school staff has limited time for asthma care; partnerships may be required with medical centres, 2) community clinicians need better links to school health services, and 3) families need better connections to school health services and community clinicians.	Non-economic but did include QoL and resource use. none stated none stated	
Massoni, Sebastien; et al 2012 France	Because didn't evaluate any achievements had to add them in with questionnaires before and after. Primary outcome: impact of project. 5 months	The Action lecture had a positive impact on academic results and reading attitudes. The authors state the programme is cost-effective but then never actually give any costs.	Impact of Action Lecture were positive and more efficient than a class size reduction programme.	Authors state CEA, no costs actually reported, non-economic. none stated None stated, 2 months assumed	*Notes cont. Makes a good point in discussion how education isn't standardised like medicine, few economic evaluations and no common indicator of benefit such as QALY.
Wade, T. J.; et al 2008 USA	HRQoL as measured annually by the PedsQL 4.0 self-report, and parent proxy-reported scores. 3 years	Adjusting for school and individual-level covariates, there was a significant improvement in student-reported HRQoL over the 3 years compared to the comparison group. Other significant predictors included age, gender, health insurance, and household income. There were no differences across groups from parent proxy reports of HRQoL.	The SBHC model of health care delivery improves student-reported HRQoL among younger, elementary, and middle school children. It appears to influence children with impeded access to care who can benefit most.	Not an economic evaluation, but picked up to use of HRQoL measure. This is a multivariate regression analysis using prospective data. not stated none stated	No attempt to measure costs

Authors Year of Publication Country of Origin	Alternatives described? Important costs and consequences of each alternative identified?	Resource use Effectiveness data	Type of preference measure Measure of benefit (QALY, DALY)	Use of discounting? Rate used Currency	ICER	Analysis of uncertainty Model Parameters
Non-economic evaluations						
Kristensen, Alyson H.; et al 2014 USA	yes No, only consequences	none Estimated average effect size from PubMed search from 2000-2012 and modified as needed due to varied nature of evidence.	none none	no none USD but only used as a modelled strategy excise tax	none	none Age, race, BMI, physical activity, diet, program specific parameters in table 2
Bruzzese, Jean-Marie; et al 2006 USA		Nurse time recorded, number of urgent visits to clinician, ED visits, and hospitalizations collected but not costed. Pediatric Asthma Caregiver's Quality of Life Questionnaire (PACQLQ)				
Massoni, Sebastien; et al 2012 France	Yes, but control was limited as the evaluation was described as being quite intrusive no	none Points in reading scores	none none	no none none	Stated as 208.5 points per teaching position for class size reduction and 280 points per teaching position for the Action lecture programme.	no none
Wade, T. J.; et al 2008 USA	Comparison school group that didn't have a SBHC. No	not identified HRQoL as measured by PedsQL 4.0 self-report and parent proxy report	HRQoL not transformed into QALYs none	no none none	none	Conducted two sets of analyses, with and without income included in the regressions. Also conducted bivariate analysis with users and nonusers of the SBHCs. None

Appendix 4: CHEERS Checklist

	CHEERS Item No.	1	2	3	4	5	6	7	8	9	10	11a	11b	12	13a	13b	14	15	16	17	18	19	20a	20b	21	22	23	24
Study No.	**Author, Year**																											
1	Abt, C; 1966	cost analysis		√		√																				√		
2	Anderson, R; 2014	√, cost analysis	√	√		√	√	√	√		√						√					√				√	√	√
3	Ansell, J.; 2002	√		√		√		√			√	n/a		n/a		n/a	√	n/a	n/a								√	
4	Atherly, A; 2009	√, cost analysis		√		√	√	√		n/a	√		n/a	n/a		n/a		n/a	n/a					n/a		√		√
5	Babey, S; 2014	cost analysis	√	√		√		√			√	n/a	√	n/a	n/a	n/a		n/a	n/a			√	n/a	n/a		√	√	√
6	Barber, S; 2013	√, protocol	√	√		√		√	√		√																√	√
7	Barnett, S; 1985	√, CBA	√	√	√		√	√	√	√	√						√									√	√	
8	Barrett, J; 2015	√	√	√		√	√	√	√	√	√	n/a	√	n/a	n/a		√				√		n/a	√		√	√	√
9	Beets, M; 2014	√, protocol		√		√	√	√																			√	
10	Belfield, CR; 2005	√, CBA		√		√	√		√	√	√			n/a			√				√					√		
11	Bertrand, E; 2011	√	√	√		√	√	√	√	√	√	n/a	√	n/a	n/a	√		√			√	√	n/a		n/a	√		
12	Blakely, T; 2014	√	√	√	√	√	√	√		√	√	n/a	√		n/a	√	√	√	√		√	√	n/a	√	√	√	√	√
13	Boyle, J; 2007	√, cost analysis	√	√		√																						
14	Brassard, P; 2006	√, cost analysis	√	√		√				√	√		n/a	n/a		n/a		n/a	n/a					n/a	n/a		√	
15	Brooker, S; 2010	√, protocol	√	√	√	√	√	√			√																√	√
16	Brown, H; 2007	√, cost analysis	√	√	√	√	√	√	√	√	√						√				√				√	√		
17	Bruzzese, J; 2006	non-economic		√										n/a												√		
18	Carabin, H.; 2000	√		√		√		√	√	√	√	n/a		n/a	n/a			√			√		n/a	√		√	√	
19	Chestnutt, I; 2012	√, protocol	√	√		√	√	√			√																√	√
20	Cooper, K.; 2012	√	√	√	√	√	√	√			√	n/a			n/a				√		√		n/a	√	n/a	√	√	√
21	Crowley, D; 2014	√		√		√		√		√	√	n/a		n/a	n/a			√					n/a		n/a	√	√	√
22	Curtale, F.; 2005	cost analysis		√		√							n/a	n/a		n/a		n/a	n/a					n/a	n/a			√
23	Eckermann, S; 2014	other	√	√		√	√		√			n/a	n/a	n/a	n/a	n/a		n/a	n/a	√		n/a	n/a	n/a	n/a	√	√	
24	Ford, T; 2012	√, protocol	√	√		√	√	√	√		√			n/a													√	√

	CHEERS Item No.	1	2	3	4	5	6	7	8	9	10	11a	11b	12	13a	13b	14	15	16	17	18	19	20a	20b	21	22	23	24
Study No.	**Author, Year**																											
25	Foster, E; 2006	√	√	√	√	√	√	√	√	√	√	√	n/a	n/a		n/a	√	n/a	n/a				√	n/a	√	√	√	√
26	Foster, E; 2010	√, cost analysis	√	√	√	√	√	√	√				n/a	n/a	√	n/a		n/a	n/a	√	n/a		n/a	n/a	√	√	√	
27	Foster, J; 2013	√		√		√						√		n/a		n/a		n/a	n/a	√		√		n/a		√		
28	Frick, K; 2004	√		√		√			√	√	√					√	√		√	√		√			√	√	√	
29	Gelli, A; 2009	cost analysis	√	√		√					√	√	n/a	n/a	√	n/a		n/a	n/a		√			n/a	√	√		
30	Gelli, A; 2011	cost analysis	√	√		√					√	√	n/a	n/a	√	n/a		n/a	n/a		√			n/a	√	√		
31	Gerald, J; 2010	√	√	√	√	√	√	√	√	√	√	n/a	√		n/a	√	√	√			√		n/a			√	√	√
32	Gesell, S; 2013	√, cost analysis	√	√		√					√						√									√	√	√
33	Glewwe, P; 2001	√, CBA		√		√				√	√	n/a	n/a	n/a	n/a	n/a	√	n/a	n/a	√	n/a	n/a	n/a	n/a	n/a	√	√	
34	Guay, M.; 2003	cost analysis	√	√		√	√	√			√	√	n/a	n/a	√	n/a	√	n/a	n/a			√		n/a	n/a		√	
35	Heckman, J; 2010	√, CBA		√		√	√		√	√	√			n/a			√		√	√	√			√		√	√	
36	Hoeflmayr, D; 2006	√, CBA	√	√		√	√	√		√	√	n/a		n/a	n/a							n/a	n/a		n/a	√	√	√
37	Hollingworth, W.; 2012	√	√	√		√	√	√	√	√	√	√	n/a	n/a	√	n/a	√	n/a	n/a	√	n/a	√	√	n/a	n/a	√	√	√
38	Joseph, C; 2007	cost analysis	√	√		√		√				√	n/a		√	n/a		n/a	n/a					n/a	n/a	√	√	√
39	Kesztyues, D; 2013	√	√	√		√	√		√	√	√	√	n/a	n/a		n/a	√	n/a	n/a	√				n/a	n/a	√	√	
40	Konig, H; 2004	√	√	√	√	√	√	√	√	√	√	n/a	√		n/a	√	√	√	√		√		n/a	√	√	√	√	
41	Kowada, A; 2012	√	√	√		√	√	√	√	√	√	n/a	√		n/a		√	√				√	n/a			√	√	√
42	Levaux, H; 2001	√	√	√	√	√	√	√	√	√	√		√			√	√		√		√	√		√		√	√	
43	Li, Y; 2010	√, protocol	√	√		√		√																			√	√
44	Liping, M; 2013	√, cost analysis	√	√		√	√	√			√		n/a	n/a	√	n/a	√	n/a	n/a					n/a	n/a	√	√	√
45	Miller, T.; 2013	√		√		√	√	√	√	√	√	√	n/a	√	√	n/a	√	n/a	n/a		√	√		n/a	n/a	√	√	√
46	Moodie, M; 2009	√	√	√	√	√	√	√	√	√	√		√			√	√		√	√	√	√		√		√	√	√
47	Moodie, M; 2013	√	√	√	√	√	√	√	√	√	√		√			√	√		√	√	√	√		√		√	√	√
48	Moodie, M; 2010	√	√	√		√	√	√	√	√	√		n/a	n/a			√				√	√			n/a	√	√	√
49	Muenning, P; 2007	√	√	√		√	√	√		√	√		√			√			√		√	√		√		√	√	

	CHEERS Item No.	1	2	3	4	5	6	7	8	9	10	11a	11b	12	13a	13b	14	15	16	17	18	19	20a	20b	21	22	23	24
Study No.	**Author, Year**																											
50	Muenning, P; 2014	√	√	√	√	√	√	√		√	√		√	√			√		√							√	√	
51	Newbury-Birch, D.;	√, protocol		√				√	√																		√	
53	Noyes, K; 2012	√	√	√	√	√	√	√	√		√		√		√		√		√			√				√	√	√
54	Pearson, A; 2014	√	√	√		√		√		√	√	n/a		n/a	n/a	√	√	√				√	n/a		√	√	√	√
55	Philipsson, A.; 2013	√	√	√		√	√	√	√	√	√		n/a	n/a		n/a	√	n/a	n/a			√	√	n/a	n/a	√	√	√
56	Quach, J.; 2013	√, protocol	√	√		√	√	√		√	√						√										√	√
57	Rein, D; 2012	√	√	√	√	√	√	√	√	√	√	n/a		n/a	n/a	√	√		√		√	√	n/a	√		√	√	
58	Reynolds, A; 2011	√, CBA	√	√	√	√	√	√	√	√	√	√	n/a				√	n/a	n/a		√	√		√	√	√		
59	Rush, E; 2014	√		√	√	√	√	√	√	√	√		√			√	√				√			√	√	√	√	
60	Salisbury, C.; 2002	cost analysis	√	√		√	√	√			V	√														√	√	
61	Scherrer, C; 2006	cost analysis	√	√		√	√	√	√	√						√	√	√								√		
62	Shemilt, I; 2004	cost analysis	√	√		√																				√	√	
63	Shepherd, J; 2010	√		√	√	√	√	√	√	√	√		√			√	√	√	√					√		√	√	√
64	Simon, E; 2013	√		√	√	√	√	√		√	√		√			√	√	√	√		√	√		√		√		√
65	Stallard, P; 2013	√, cost analysis	√	√		√	√	√	√		√						√					√				√	√	√
66	Tai, T; 2010	√, CBA		√		√		√		√		n/a	√		n/a	√	√		√				n/a			√		
67	te Velde, S; 2011	CBA		√		√	√	√	√	√	√	n/a		√	n/a		√		√		√		n/a	√		√	√	√
68	Tengs, T; 2001	√	√	√	√	√	√	√	√	√	√		√			√	√		√		√	√		√		√		
69	Vijge, S; 2008	√		√		√		√	√	√	√						√	√	√					√		√		√
70	Wade, T; 2008	non-economic	√	√							√	√	n/a	n/a		n/a								n/a		√	√	
71	Wang, L; 2000	√	√	√		√	√	√	√		√		√			√	√		√					√		√	√	
72	Wang, L; 2001	√	√	√	√	√	√	√	√	√							√	√	√							√		
73	Wang, L; 2003	√, cost analysis	√	√		√	√		√	√	√	n/a			n/a			√	√				n/a	√		√	√	
74	Wang, L; 2008	√	√	√	√	√	√	√	√		√		√		√		√					√				√	√	
75	Wang, L; 2011	CBA	√	√		√	√	√	√	√	√	n/a		√	n/a	√	√		√				n/a	√		√	√	√
76	Young, T; 2003	√, cost analysis	√	√		√																				√	√	

Appendix 5: PRISMA Checklist

Section/topic	#	Checklist item	Reported in section #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	3
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	na
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3.1
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3.2.1
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	na
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	3.2.1
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Appendix 1
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	3.2.2, 3.2.3
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	3.2.4

Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	3.2.4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	3.2.3, 3.2.4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	na
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	3.2.5

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	3.2.2, 3.2.3
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	3.2.5
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	3.3
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Appendix 3
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Appendix 4
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	a) Appendix 3, b) na
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	na
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Appendix 4

Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	na
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	3.4.1
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	3.4.2
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	3.5
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	na

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

Appendix 6: CHEERS checklist—Items to include when reporting economic evaluations of health interventions

Section/item	Item No	Recommendation	Reported in section number
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.	✓ 4
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.	N/A see abstract
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study.	✓ 4.1
		Present the study question and its relevance for health policy or practice decisions.	✓4.3.2
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	✓4.4.1
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	✓4.3.1
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	✓4.4.1
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	✓4.3.1, 4.3.3
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	✓4.4.1
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	✓4.4.1
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.	✓ 4.3.3.1
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.	✓4.3.3, 4.4.3
	11b	*Synthesis-based estimates:* Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	N/A

Section/item	Item No	Recommendation	Reported in section number
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	✓ 4.2.1.2
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.	✓4.4.2
	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.	N/A
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.	✓ 4.4.2.1
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	N/A
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	N/A
Analytical methods	17	Describe all analytical 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 (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	✓ 4.4.4
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.	✓ 5.3-5.6
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.	✓ 5.6
Characterising uncertainty	20a	*Single study-based economic evaluation:* Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	✓ 5.6, 5.7
	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.	N/A
Characterising 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.	N/A
Discussion			

Section/item	Item No	Recommendation	Reported in section number
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	✓ 5.7, 5.8
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 non-monetary sources of support.	✓ 4.3
Conflicts of interest	24	Describe any potential for conflict of interest of 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.	✓ 4.3

Appendix 7: Stata syntax for scoring SDQ replicated from SDQinfo.org website

SDQ: Generating scores in STATA

The scoring algorithm is based on the 25 variables plus impact items for each questionnaire. The algorithm expects to find these variables with specific names: the first letter of each variable name is 'p' for the parent SDQ, 's' for the self-report SDQ and 't' for the teacher SDQ. After this first letter, the variable names are as follows:

consid = Item 1 : considerate
restles = Item 2 : restless
somatic = Item 3 : somatic symptoms
shares = Item 4 : shares readily
tantrum = Item 5 : tempers
loner = Item 6 : solitary
obeys = Item 7 : obedient
worries = Item 8 : worries
caring = Item 9 : helpful if someone hurt
fidgety = Item 10 : fidgety
friend = Item 11 : has good friend
fights = Item 12 : fights or bullies
unhappy = Item 13 : unhappy
popular = Item 14 : generally liked
distrac = Item 15 : easily distracted
clingy = Item 16 : nervous in new situations
kind = Item 17 : kind to younger children
lies = Item 18 : lies or cheats *[for the SDQ for 2-4 year olds, replace 'lies' with 'argues']*
bullied = Item 19 : picked on or bullied
helpout = Item 20 : often volunteers
reflect = Item 21 : thinks before acting
steals = Item 22 : steals *[for the SDQ for 2-4 year olds, replace 'steals' with 'spite']*
oldbest = Item 23 : better with adults than with children
afraid = Item 24 : many fears
attends = Item 25 : good attention
ebddiff = Impact question: oveall difficulties in at least one area
distres = Impact question: upset or distressed
imphome = Impact question: interferes with home life
impfrie = Impact question: interferes with friendships
impclas = Impact question: interferes with learning
impleis = Impact question: interferes with leisure

For each of these items, if the first response category (not true, no, not at all) has been selected, this is coded as zero, the next response category (somewhat true, yes-minor, just a little) is coded as one and so on.

For each informant, the algorithm generates six scores. The first letter of each derived variable is 'p' for parent-based scores, 's' for self-report-based scores and 't' for teacher-based scores. After this first letter, the names of the scores are as follows:

emotion = emotional symptoms
conduct = conduct problems
hyper = hyperactivity/inattention
peer = peer problems
prosoc = prosocial
ebdtot = total difficulties
impact = impact

```
*** Recoding variables and then scoring the parent SDQ scores

recode pobeys (0=2) (1=1) (2=0) (else=.), gen(qobeys)
recode preflect (0=2) (1=1) (2=0) (else=.) , gen(qreflect)
recode pattends (0=2) (1=1) (2=0) (else=.), gen(qattends)
recode pfriend (0=2) (1=1) (2=0) (else=.), gen(qfriend)
recode ppopular (0=2) (1=1) (2=0) (else=.), gen(qpopular)

recode pdistres (0=0) (1=0) (2=1) (3=2) (.=0), gen(qqdistres)
recode pimphome (0=0) (1=0) (2=1) (3=2) (.=0), gen(qqimphome)
recode pimpfrie (0=0) (1=0) (2=1) (3=2) (.=0), gen(qqimpfrie)
recode pimpclas (0=0) (1=0) (2=1) (3=2) (.=0), gen(qqimpclas)
recode pimpleis (0=0) (1=0) (2=1) (3=2) (.=0), gen(qqimpleis)

egen nemotion=robs(psomatic pworries punhappy pclingy pafraid)
egen pemotion=rmean(psomatic pworries punhappy pclingy pafraid) if
nemotion>2
replace pemotion=round(pemotion*5)

egen nconduct=robs(ptantrum qobeys pfights plies psteals)
egen pconduct=rmean(ptantrum qobeys pfights plies psteals) if nconduct>2
replace pconduct=round(pconduct*5)

egen nhyper=robs(prestles pfidgety pdistrac qreflect qattends)
egen phyper=rmean(prestles pfidgety pdistrac qreflect qattends) if
nhyper>2
replace phyper=round(phyper*5)

egen npeer=robs(ploner qfriend qpopular pbullied poldbest)
egen ppeer=rmean(ploner qfriend qpopular pbullied poldbest) if npeer>2
replace ppeer=round(ppeer*5)

egen nprosoc=robs(pconsid pshares pcaring pkind phelpout)
egen pprosoc=rmean(pconsid pshares pcaring pkind phelpout) if nprosoc>2
replace pprosoc=round(pprosoc*5)

egen nimpact=robs(pdistres pimphome pimpfrie pimpclas pimpleis)
gen pimpact=qqdistres+qqimphome+qqimpfrie+qqimpclas+qqimpleis if
(nimpact!=0)
replace pimpact=0 if pebddiff==0

drop qobeys qreflect qattends qfriend qpopular qqdistres qqimphome
qqimpfrie qqimpclas qqimpleis nemotion nconduct nhyper npeer nprosoc
nimpact

gen pebdtot=pemotion+pconduct+phyper+ppeer

*** Recoding variables and then scoring the child SDQ scores
```

```
recode sobeys (0=2) (1=1) (2=0) (else=.), gen(robeys)
recode sreflect (0=2) (1=1) (2=0) (else=.) , gen(rreflect)
recode sattends (0=2) (1=1) (2=0) (else=.), gen(rattends)
recode sfriend (0=2) (1=1) (2=0) (else=.), gen(rfriend)
recode spopular (0=2) (1=1) (2=0) (else=.), gen(rpopular)

recode sdistres (0=0) (1=0) (2=1) (3=2) (.=0), gen(rrdistres)
recode simphome (0=0) (1=0) (2=1) (3=2) (.=0), gen(rrimphome)
recode simpfrie (0=0) (1=0) (2=1) (3=2) (.=0), gen(rrimpfrie)
recode simpclas (0=0) (1=0) (2=1) (3=2) (.=0), gen(rrimpclas)
recode simpleis (0=0) (1=0) (2=1) (3=2) (.=0), gen(rrimpleis)

egen nemotion=robs(ssomatic sworries sunhappy sclingy safraid)
egen semotion=rmean(ssomatic sworries sunhappy sclingy safraid) if
nemotion>2
replace semotion=round(semotion*5)

egen nconduct=robs(stantrum robeys sfights slies ssteals)
egen sconduct=rmean(stantrum robeys sfights slies ssteals) if nconduct>2
replace sconduct=round(sconduct*5)

egen nhyper=robs(srestles sfidgety sdistrac rreflect rattends)
egen shyper=rmean(srestles sfidgety sdistrac rreflect rattends) if
nhyper>2
replace shyper=round(shyper*5)

egen npeer=robs(sloner rfriend rpopular sbullied soldbest)
egen speer=rmean(sloner rfriend rpopular sbullied soldbest) if npeer>2
replace speer=round(speer*5)

egen nprosoc=robs(sconsid sshares scaring skind shelpout)
egen sprosoc=rmean(sconsid sshares scaring skind shelpout) if nprosoc>2
replace sprosoc=round(sprosoc*5)

egen nimpact=robs(sdistres simphome simpfrie simpclas simpleis)
gen simpact=rrdistres+rrimphome+rrimpfrie+rrimpclas+rrimpleis if
(nimpact!=0)
replace simpact=0 if sebddiff==0

drop robeys rreflect rattends rfriend rpopular rrdistres rrimphome
rrimpfrie rrimpclas rrimpleis nemotion nconduct nhyper npeer nprosoc
nimpact

gen sebdtot=semotion+sconduct+shyper+speer

*** Recoding variables and then scoring the teacher SDQ scores

recode tobeys (0=2) (1=1) (2=0) (else=.), gen(uobeys)
recode treflect (0=2) (1=1) (2=0) (else=.) , gen(ureflect)
recode tattends (0=2) (1=1) (2=0) (else=.), gen(uattends)
recode tfriend (0=2) (1=1) (2=0) (else=.), gen(ufriend)
recode tpopular (0=2) (1=1) (2=0) (else=.), gen(upopular)

recode tdistres (0=0) (1=0) (2=1) (3=2) (.=0), gen(uudistres)
recode timpfrie (0=0) (1=0) (2=1) (3=2) (.=0), gen(uuimpfrie)
recode timpclas (0=0) (1=0) (2=1) (3=2) (.=0), gen(uuimpclas)

egen nemotion=robs(tsomatic tworries tunhappy tclingy tafraid)
egen temotion=rmean(tsomatic tworries tunhappy tclingy tafraid) if
nemotion>2
replace temotion=round(temotion*5)

egen nconduct=robs(ttantrum uobeys tfights tlies tsteals)
```

```
egen tconduct=rmean(ttantrum uobeys tfights tlies tsteals) if nconduct>2
replace tconduct=round(tconduct*5)

egen nhyper=robs(trestles tfidgety tdistrac ureflect uattends)
egen thyper=rmean(trestles tfidgety tdistrac ureflect uattends) if
nhyper>2
replace thyper=round(thyper*5)

egen npeer=robs(tloner ufriend upopular tbullied toldbest)
egen tpeer=rmean(tloner ufriend upopular tbullied toldbest) if npeer>2
replace tpeer=round(tpeer*5)

egen nprosoc=robs(tconsid tshares tcaring tkind thelpout)
egen tprosoc=rmean(tconsid tshares tcaring tkind thelpout) if nprosoc>2
replace tprosoc=round(tprosoc*5)

egen nimpact=robs(tdistres timpfrie timpclas)
gen timpact=uudistres+uuimpfrie+uuimpclas if (nimpact!=0)
replace timpact=0 if tebddiff==0

drop uobeys ureflect uattends ufriend upopular uudistres uuimpfrie
uuimpclas nemotion nconduct nhyper npeer nprosoc nimpact

gen tebdtot=temotion+tconduct+thyper+tpeer
```

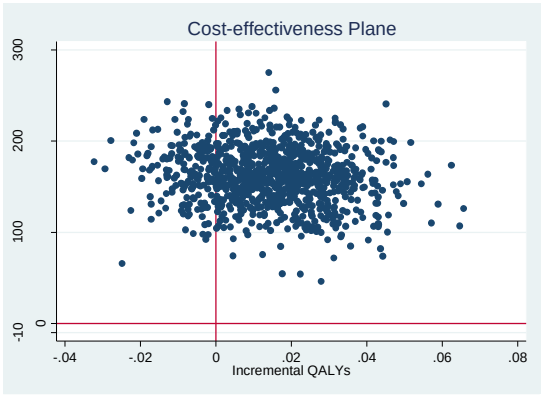
The predictive algorithm was converted into STATA syntax by Anna Goodman, Richard Rowe and Ye Gan

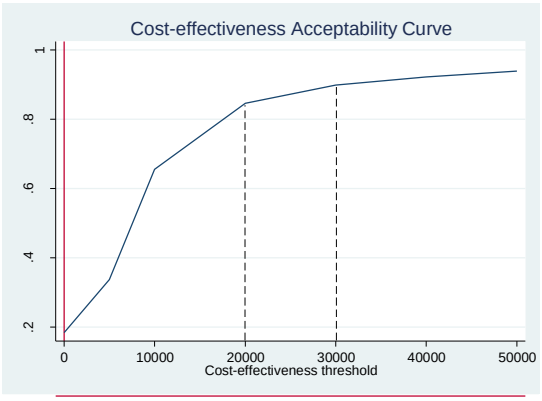
---

Last modified : 8/01/10

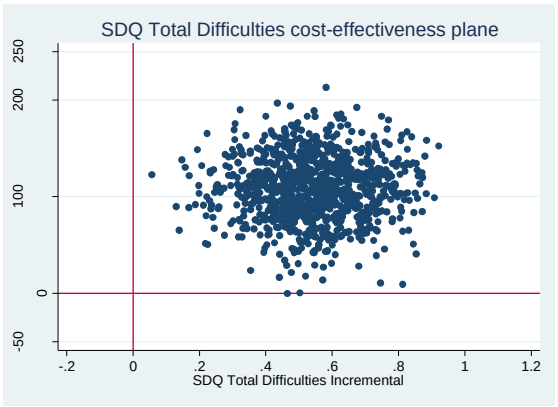
Appendix 8: Cost-effectiveness planes and CEACs for all sensitivity analyses

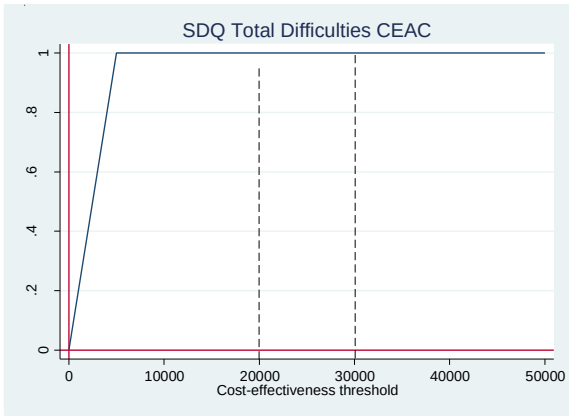
SA0: Base-case Adolescent Values



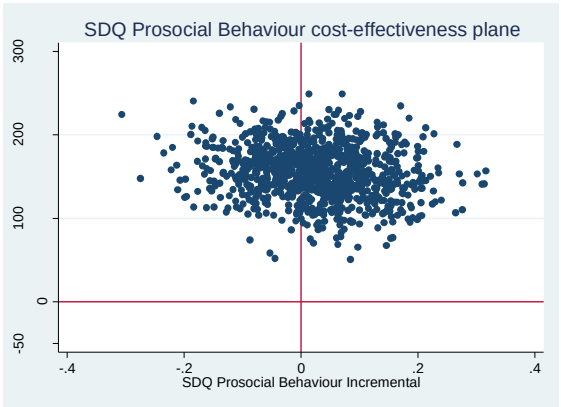


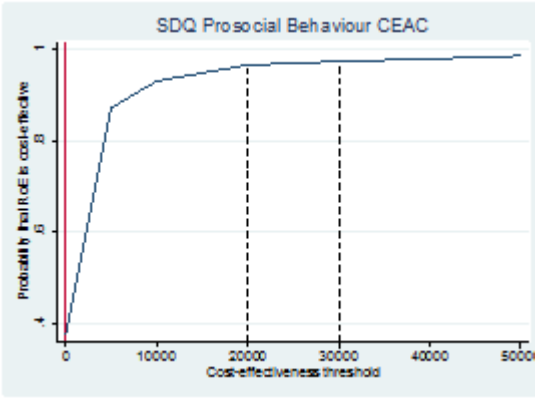
SA1: SDQ Total Difficulties (CEA)





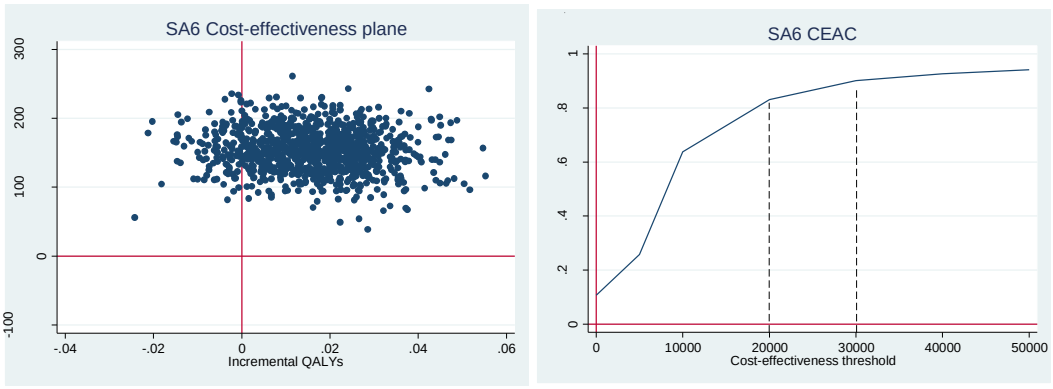
SA2: SDQ Prosocial Behaviour (CEA)





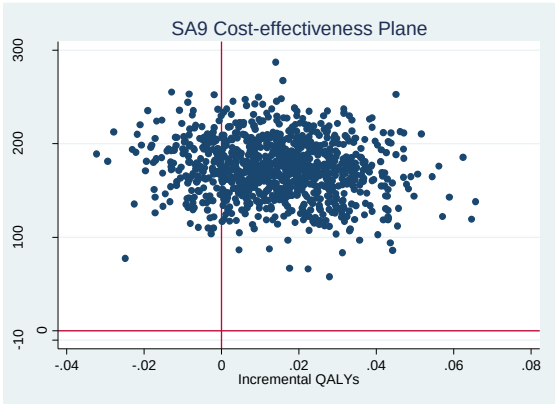
SA3: CHU9D mapped from SDQ (Equation 10)

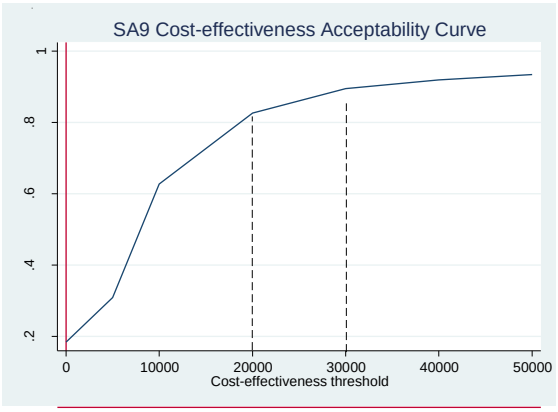
SA4: CHU9D estimated from original tariff (UK adults)



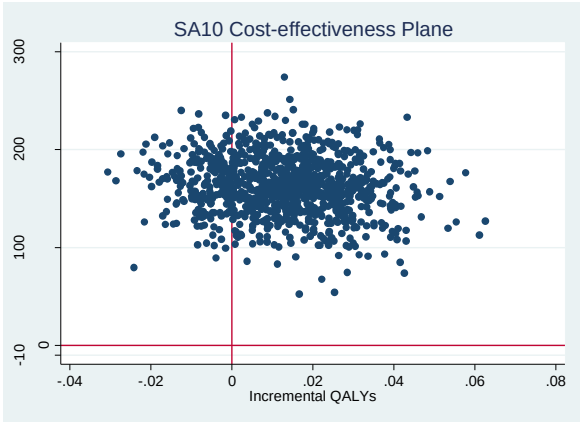
SA5: Training and material costs not annuitised

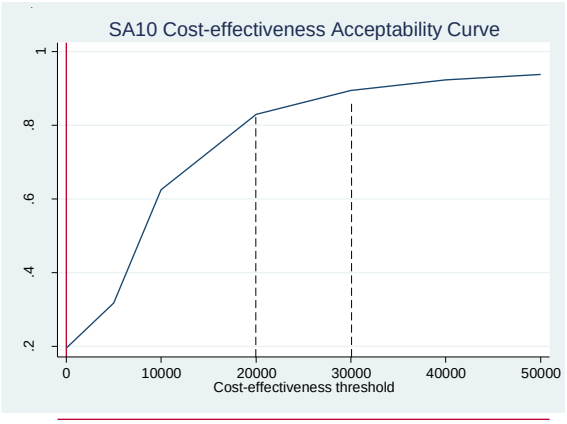
SA6: Training and material costs annuitised over 3 years



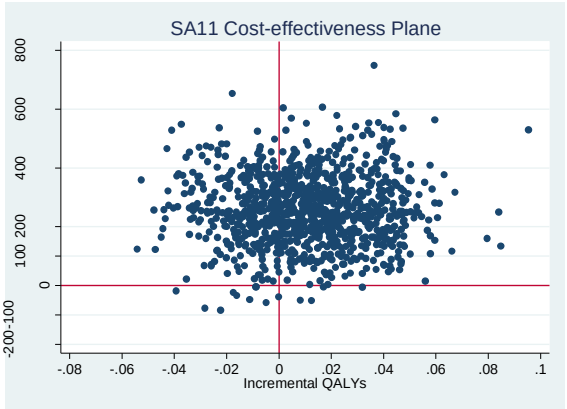


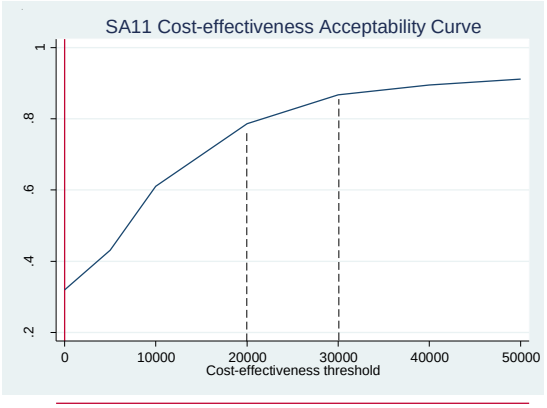
SA7: Use of more traditional 3.5% discount rate for costs and outcomes



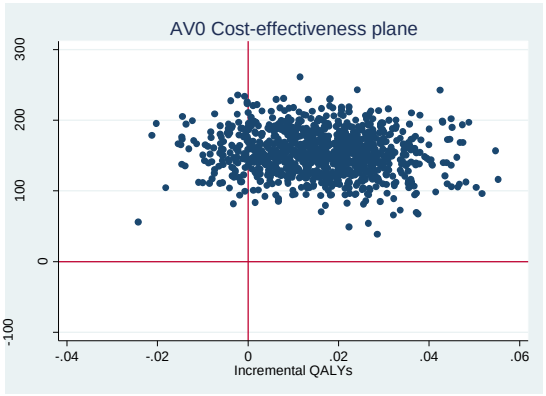


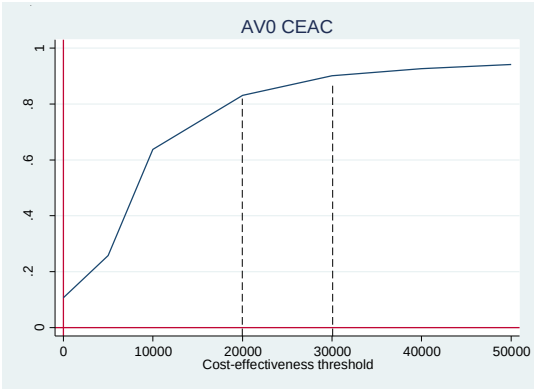
SA8: Available-case analysis assuming MCAR



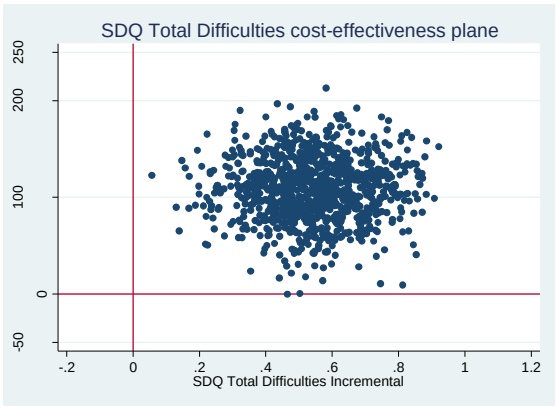


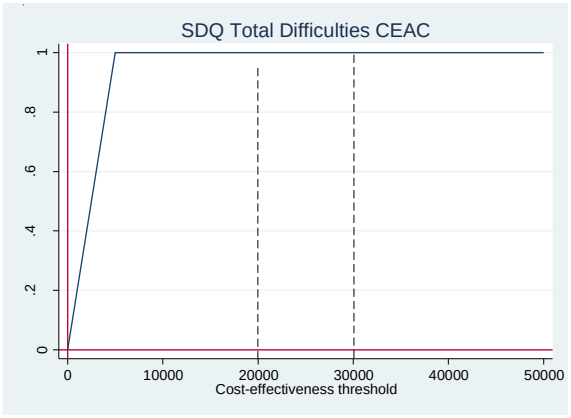
AV0: CHU9D estimated from original tariff (UK adults)



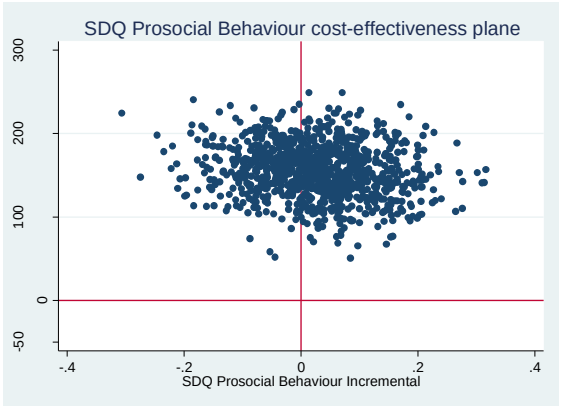


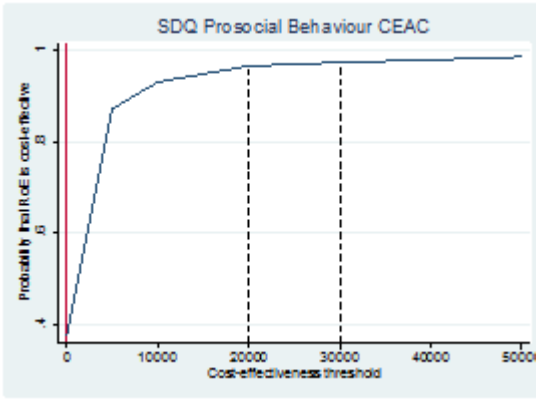
AV1: SDQ Total Difficulties (CEA)



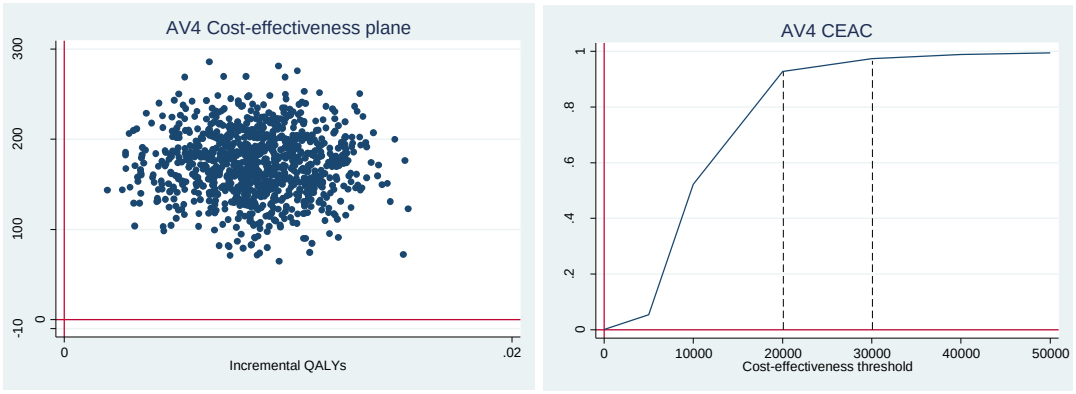


AV2: SDQ Prosocial Behaviour (CEA)

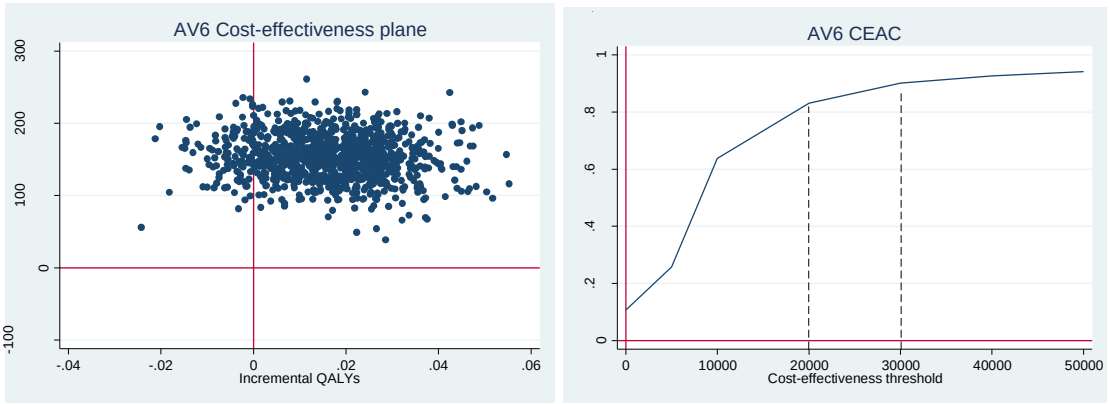




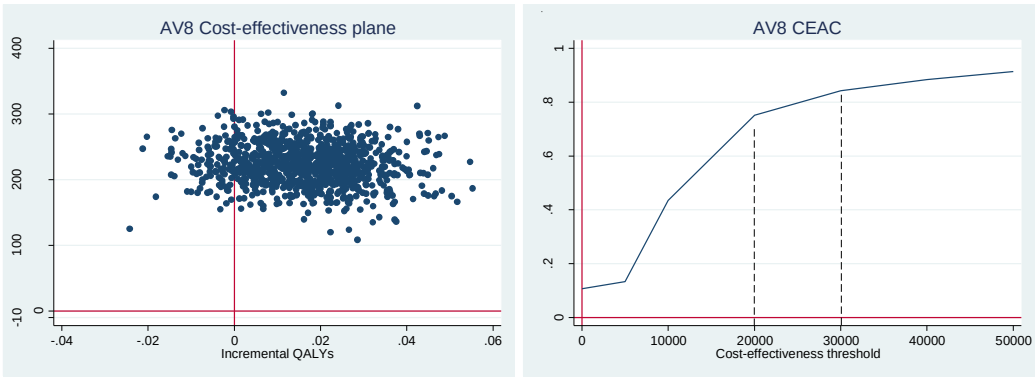
AV3: CHU9D mapped from SDQ (Equation 10)



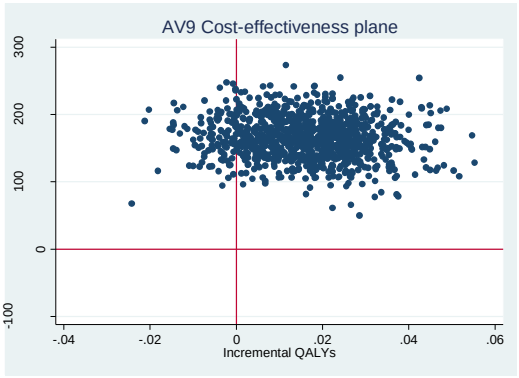
AV4: CHU9D estimated from alternative tariff (Australian adolescents)

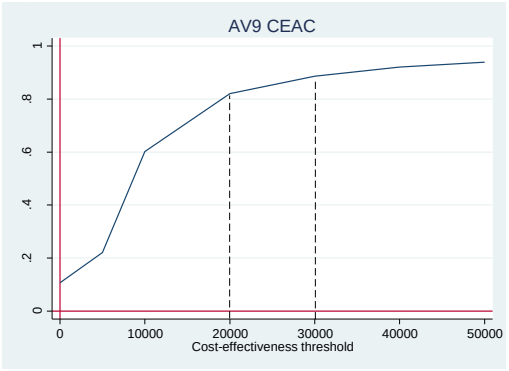


AV5: Training and material costs not annuitised

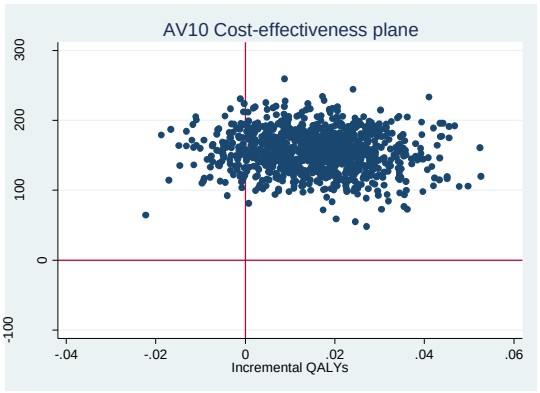


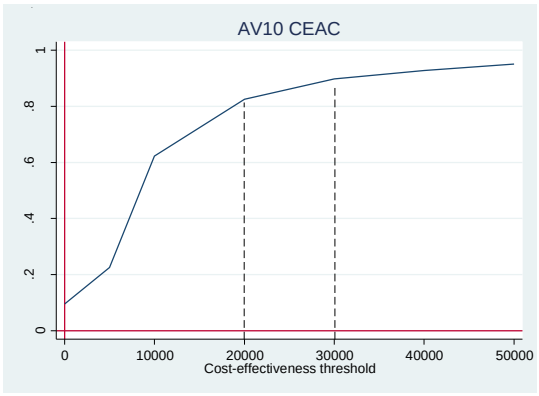
AV6: Training and material costs annuitised over 3 years



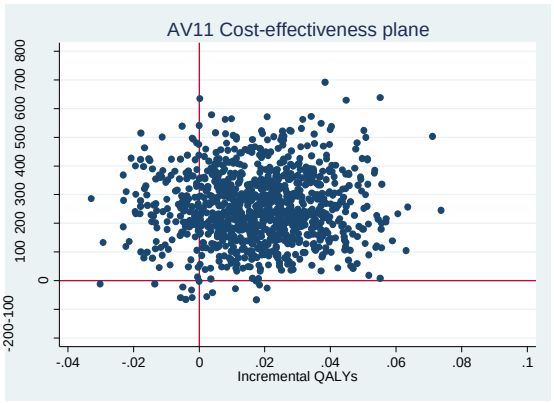


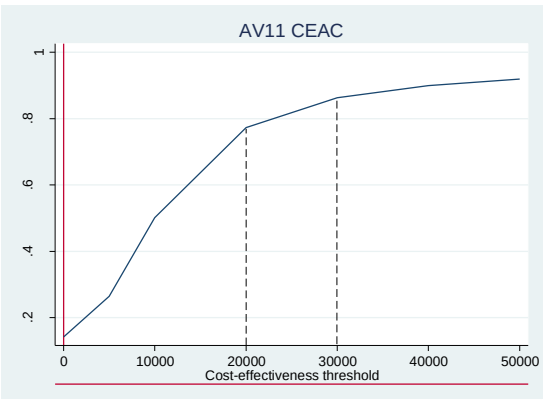
AV7: Use of more traditional 3.5% discount rate for costs and outcomes





AV8: Available-case analysis assuming MCAR





List of References

1. World Health O, Srinivasa Murthy R. The World health report 2001 : mental health, new understanding, new hope. Geneva: World Health Organization, 2001.
2. NICE. Social and emotional wellbeing for children and young people. https://www.nice.org.uk/advice/lgb12/chapter/Introduction2013.
3. Petrides KV, Frederickson N, Furnham A. The role of trait emotional intelligence in academic performance and deviant behavior at school. *Personality and Individual Differences* 2004; **36** (2): 277-293.
4. Ciarrochi J, Deane FP, Anderson S. Emotional intelligence moderates the relationship between stress and mental health. *Personality and Individual Differences* 2002; **32** (2): 197-209.
5. NICE Guidlines PH12. (2008). Social and emotional wellbeing in primary education. 2015-05-04. URL:http://www.nice.org.uk/guidance/PH12/chapter/introduction. Accessed: 2015-05-04. (Archived by WebCite® at http://www.webcitation.org/6YH50hWjm). Public Health Guidance.
6. Weissberg RP, Greenberg MT. School and community competence-enhancement and prevention programs. In: Damon W, ed. Handbook of Child Psychology. Vol. 4. 5th edn. New York: Wiley, 1998.
7. Guerra NG, Bradshaw CP. Linking the prevention of problem behaviors and positive youth development: core competencies for positive youth development and risk prevention. *New Directions for Child Adolescent Development* 2008; **2008** (122): 1-17.
8. Power C, Manor O, Fox J. Health and Class: The Early Years. London: Chapman and Hall, 1991.
9. Caspi A, Moffitt TE, Newman DL, Silva PA. Behavioral observations at age 3 years predict adult psychiatric disorders. Longitudinal evidence from a birth cohort. *Archives of General Psychiatry* 1996; **53** (11): 1033-9.
10. Adi Y, Killoran A, Janmohamed K, Stewart-Brown S. Systematic review of the effectiveness of interventions to promote mental wellbeing in children in primary education. Report 1: Universal approaches which do not focus of violence or bullying: University of Warwick2007.
11. Paradis AD, Koenen KC, Fitzmaurice GM, Buka SL. Impact of persistent and adolescent-limited antisocial behaviour on adult health outcomes. *Journal of Epidemiology and Community Health* 2016; **70** (10): 1004-1010.
12. Petrou S, Johnson S, Wolke D, Hollis C, Kochhar P, Marlow N. Economic costs and preference-based health-related quality of life outcomes associated with childhood psychiatric disorders. *The British Journal of Psychiatry* 2010; **197** (5): 395-404.
13. Knapp M, McDaid D, Parsonage M. Mental health promotion and mental illness prevention: the economic case. London: Personal Social Services Research Unit, London School of Economics and Political Science2011.
14. Ashdown DM, Bernard ME. Can Explicit Instruction in Social and Emotional Learning Skills Benefit the Social-Emotional Development, Well-being, and Academic Achievement of Young Children? *Early Childhood Education Journal* 2012; **39** (6): 397-405.
15. Coelho VA, Marchante M, Sousa V. "Positive Attitude": A multilevel model analysis of the effectiveness of a Social and Emotional Learning Program for Portuguese middle school students. (1095-9254 (Electronic)).

16. Conduct Problems Prevention Research G. The Effects of a Multiyear Universal Social–Emotional Learning Program: The Role of Student and School Characteristics. *Journal of consulting and clinical psychology* 2010; **78** (2): 156-168.
17. Bond L, Patton G, Glover S, et al. The Gatehouse Project: can a multilevel school intervention affect emotional wellbeing and health risk behaviours? *Journal of Epidemiology & Community Health* 2004; **58** (12): 997-1003.
18. Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, Schellinger KB. The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child development* 2011; **82** (1): 405-32.
19. Morse TCaAG. Financial sustainability
of schools: Department for Education2016.
20. Local highways maintenance: economic costs and benefits tool. In: Transport Df, ed2015.
21. Sullivan L. Introduction to the Population Health Intervention Research Initiative for Canada. *Canadian journal of public health = Revue canadienne de sante publique* 2009; **100** (1): Suppl I5-7.
22. Masters R, Anwar E, Collins B, Cookson R, Capewell S. Return on investment of public health interventions: a systematic review. *J Epidemiol Community Health* 2017.
23. NICE. Judging whether public health interventions offer value for money2013.
24. Reynolds AJ, Temple JA, Robertson DL, Mann EA. Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *JAMA* 2001; **285** (18): 2339-46.
25. Barnett WS. Long-term cognitive and academic effects of early childhood education on children in poverty. *Preventive Medicine* 1998; **27** (2): 204-7.
26. Goodman A, Joshi H, Nasim B, Tyler C. Social and emotional skills in childhood and their long-term effects on adult life: University College London, Institute of Education2015.
27. Heckman JJ, Moon SH, Pinto R, Savelyev PA, Yavitz A. The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics* 2010; **94** (1-2): 114-128.
28. Belfield CR, Nores M, Barnett S, Schweinhart L. The High/Scope Perry Preschool Program. *Journal of Human Resources* 2006; **41** (1): 162-190.
29. Hummel-Rossi B, Ashdown J. The state of cost-benefit and cost-effectiveness analyses in education. *Review of Educational Research* 2002; **72** (1): 1-30.
30. NICE. How NICE measures value for money in relation to public health interventions2013.
31. Schultz TW. Investment in Human Capital. *The American Economic Review* 1961; **51** (1): 1-17.
32. Becker GS. Human Capital: A theoretical and empirical analysis, with special reference to education: National Bureau of Economic Research, 1964.
33. Mincer J. Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2. 1974.
34. Woessmann L. The economic case for education. *Education Economics* 2016; **24** (1): 3-32.
35. Remme M, Martinez-Alvarez M, Vassall A. Cost-Effectiveness Thresholds in Global Health: Taking a Multisectoral Perspective. (1524-4733 (Electronic)).
36. Weatherly H, Drummond M, Claxton K, et al. Methods for assessing the cost-effectiveness of public health interventions: Key challenges and recommendations. *Health Policy* 2009; **93** (2–3): 85-92.
37. Ungar WJ. Economic Evaluation in Child Health. Oxford: Oxford University Press, 2010.

38. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Second Edition edn. Oxford: Oxford University Press, 2017.
39. NICE: Guide to the methods of technology appraisal. June 2008. *National Institute for Health and Clinical Excellence*.
40. Santos RG, Chartier MJ, Whalen JC, Chateau D, Boyd L. Effectiveness of school-based violence prevention for children and youth: a research report. *Healthcare quarterly (Toronto, Ont)* 2011; **14 Spec No**: 80-91.
41. Schonert-Reichl, K. Smith V, Zaidman-Zait A. Effectiveness of the Roots of Empathy Program in Fostering the Social-Emotional Development of Primary Grade Children2002.
42. Smith V. Roots of Empathy Whole Schools Evaluation Report: Examining Variability of Program Implementation of the Roots of Empathy. Alberta, Canada: University of Alberta2008.
43. Schonert-Reichl KA, Smith V, Zaidman-Zait A, Hertzman C. Promoting Children's Prosocial Behaviors in School: Impact of the "Roots of Empathy" Program on the Social and Emotional Competence of School-Aged Children. *School Mental Health* 2012; **4** (1): 1-21.
44. MacDonald A, McLafferty M, Bell P, et al. Evaluation of the Roots of Empathy programme by North Lanarkshire Psychological Service. North Lankarkshire Psychological Service2013.
45. Wrigley J, Makara K, Elliot D. Evaluation of Roots of Empathy in Scotland 2014-2015, Final Report for Action for Children (unpublished)2015.
46. Kendall G, Schonert-Reichl K, Smith V, et al. The Evaluation of Roots of Empathy in Western Australian Schools 2005. Western Australia: Telethon Institute for Child Health Research2006.
47. Greenberg MT. School-based prevention: current status and future challenges. *Effective Education* 2010; **2** (1): 27-52.
48. Greenberg MT, Weissberg RP, O'Brien MU, et al. Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *Am Psychol* 2003; **58** (6-7): 466-74.
49. Browne G, Gafni A Fau - Roberts J, Roberts J Fau - Byrne C, Byrne C Fau - Majumdar B, Majumdar B. Effective/efficient mental health programs for school-age children: a synthesis of reviews. 2004 (0277-9536 (Print)).
50. Payton J, Weissberg R, Durlak J, et al. The positive impact of social and emotional learning for kindergarten to eight-grade students: Findings from three scientific reviews. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning2008.
51. Sutton P, Love J, Bell J, et al. The emotional well-being of young people: are view of the literature. Aberdeen: Robert Gordon University2005.
52. Wilson SJ, Lipsey MW. School-Based Interventions for Aggressive and Disruptive Behavior: Update of a Meta-Analysis. *American journal of preventive medicine* 2007; **33** (2 Suppl): S130-S143.
53. Gordon M. Roots of Empathy: Responsive parenting, caring societies. *Keio Journal of Medicine* 2003; **52** (4): 236-243.
54. Miller PA, Eisenberg N. The relation of empathy to aggressive and externalizing/antisocial behavior. *Psychological bulletin* 1988; **103** (3): 324-44.
55. Connolly P, Miller S, Kee F, et al. A cluster randomised controlled trial evaluation and cost-effectiveness analysis of the Roots of Empathy schools-based programme for improving social and emotional wellbeing outcomes among 8-9 year olds in Northern Ireland - *in submission to NIHR PHR*2016.

56. Hayward M. Mentalization-based Treatment for Borderline Personality Disorder: A Practical Guide. *The Psychiatrist* 2008; **32** (5): 200-200.
57. Craig P, Dieppe P, Macintyre S, Michie S: Developing and evaluating complex interventions: new guidance. 2014-06-30. URL:http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/. Accessed: 2014-06-30. (Archived by WebCite® at http://www.webcitation.org/6QiiYPnA5).
58. Shiell A, Hawe P, Gold L. Complex interventions or complex systems? Implications for health economic evaluation. *BMJ* 2008; **336** (7656): 1281-3.
59. Evans R, Murphy S, Scourfield J. Implementation of a School-Based Social and Emotional Learning Intervention: Understanding Diffusion Processes Within Complex Systems. *Prevention Science* 2015; **16** (5): 754-764.
60. Husereau D, Jacobs P, Manns B, Hoomans T, Marshall D, R; T. Econonomic Evaluation of Complex Health System Interventions: Institute of Health Economics2014.
61. Payne K, McAllister M, Davies LM. VALUING THE ECONOMIC BENEFITS OF COMPLEX INTERVENTIONS: WHEN MAXIMISING HEALTH IS NOT SUFFICIENT. *Health Econ* 2013; **22** (3): 258-271.
62. Datta J, Petticrew M. Challenges to evaluating complex interventions: a content analysis of published papers. *BMC Public Health* 2013; **13** (1): 1-18.
63. Byford S, Sefton T. Economic Evaluation of Complex Health and Social Care Interventions. *National Institute Economic Review* 2003; **186** (1): 98-108.
64. NICE. Guide to the methods of technology appraisal 2013. 2016-08-12. URL:https://www.nice.org.uk/process/pmg9/chapter/1-foreword. Accessed: 2016-08-12. (Archived by WebCite® at http://www.webcitation.org/6ji0JYTLa): National Institute for Health and Care Excellence04 April 2013.
65. OECD (2017), Life expectancy at birth (indicator). doi: 10.1787/27e0fc9d-en (Accessed on 15 August 2017)
66. OECD. Fiscal Sustainability of Health Systems: OECD Publishing.
67. Appleby J. Spending on health and social care over the next 50 years. 2013.
68. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Fourth edn. Oxford: Oxford University Press, 2015.
69. Mankiw G. Principles of economics. Seventh edn. Stamford, CT: Cengage Learning, 2015.
70. Coulanges F. Oikonomia: An inquiry into beginnings of economic thought and language. *Kyklos* 1958; **11** (1): 29-57.
71. Mushkin SJ. Toward a definition of health economics. *Public Health Reports* 1958; **73** (9): 785-794.
72. McIntosh E, Clarke P, Frew E, Louviere J. Applied Methods of Cost-Benefit Analysis in Health Care. Oxford: Oxford University Press, 2010.
73. Grossman M. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy* 1972; **80** (2): 223-255.
74. Sabik LM, Lie RK. Priority setting in health care: lessons from the experiences of eight countries. *International Journal for equity in health* 2008; **7** (1): 4.
75. NICE. Who We Are. 2017-02-17. URL:https://www.nice.org.uk/about/who-we-are. Accessed: 2017-02-17. (Archived by WebCite® at http://www.webcitation.org/6oLLnaSBE)
76. HTAGlossary.net. Also available at: http://htaglossary.net/Health+Technology+Assessment+%28HTA%29&highlight=health%20technology.

77. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the Economic Evaluation of Health Care Programmes. Third edn. Oxford: Oxford University Press, 2005.
78. Hjelmgren J, Berggren F, Andersson F. Health economic guidelines--similarities, differences and some implications. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2001; **4** (3): 225-50.
79. NICE. Methods for the development of NICE public health guidance (third edition). 2014-07-02.
URL:http://http://www.nice.org.uk/article/PMG4/chapter/1%20Introduction. Accessed: 2014-07-02. (Archived by WebCite®http://www.webcitation.org/6QlR7U5Py: National Institute for Health and Clinical Excellence2012.
80. Clarke, Philip. *Health economic evaluation - a brief history.* Seminar given at the University of Glasgow, 21st September 2016.
81. Calkins GN. Some Results of Sanitary Legislation in England Since 1875. *Publications of the American Statistical Association* 1891; **2** (14): 297-303.
82. Chapin CV. THe relative values of public health procedures. *Journal of the American Medical Association* 1917; **LXIX** (2): 90-95.
83. Rebelo L. The Origins and the Evolution of Health Economics: a discipline by itself? Led by economists, practitioners or politics?: Católica Porto Business School, Universidade Católica Portuguesa2007.
84. Arrow KJ. Uncertainty and the welfare economics of medical care. *The American economic review* 1963: 941-973.
85. Klarman HE. Economics of Health. New York: Columbia University Press, 1965.
86. Klarman HE, John O'S F, Rosenthal GD. Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease. *Medical Care* 1968; **6** (1): 48-54.
87. Eldessouki R, Smith MD, Health Care System Information Sharing: A Step Toward Better Health Globally. Value Health Regional Issues 2012; 1:118-129
88. Fox-Rushby J, Cairns J. Economic Evaluation. England: Open University Press, 2005.
89. McIntosh E, Barlow J, Davis H, Stewart-Brown S. Economic evaluation of an intensive home visiting programme for vulnerable families: a cost-effectiveness analysis of a public health intervention. *Journal Of Public Health (Oxford, England)* 2009; **31** (3): 423-433.
90. Black WC. The CE Plane. *Medical Decision Making* 1990; **10** (3): 212-214.
91. Glick H, Doshi J, Sonnad S, Polsky D. Economic Evaluation in Clinical Trials. Oxford: Oxford University Press, 2007.
92. O'Brien BJ, Briggs AH. Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. *Statistical methods in medical research* 2002; **11** (6): 455-68.
93. EuroQol-a new facility for the measurement of health-related quality of life. . *Health Policy* 1990; **16** (3): 199-208.
94. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions. Health Utilities Index. *Pharmacoeconomics* 1995; **7** (6): 503-20.
95. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002; **21** (2): 271-92.
96. Stevens K. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics* 2012; **30** (8): 729-747.
97. Taylor TR. Understanding the choices that patients make. *The Journal of the American Board of Family Practice* 2000; **13** (2): 124-33.
98. Torrance GW, Thomas WH, Sackett DL. A Utility Maximization Model for Evaluation of Health Care Programs. *Health Services Research* 1972; **7** (2): 118-133.

99. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making* 1989; **9** (2): 142-9.
100. Nord E. An alternative to QALYs: the saved young life equivalent (SAVE). *Bmj* 1992; **305** (6858): 875-7.
101. EuroQol Group. 2014-07-28. URL:http://www.euroqol.org/eq-5d-products/eq-5d-y.html. Accessed: 2014-07-28. (Archived by WebCite® at http://www.webcitation.org/6RP3CtTXI).
102. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *BMJ* 1990; **300** (6719): 230-5.
103. Coast J, Smith RD, Lorgelly P. Welfarism, extra-welfarism and capability: The spread of ideas in health economics. *Social Science & Medicine* 2008; **67** (7): 1190-1198.
104. NIHR today. Also available at: http://www.nihr.ac.uk/about-us/nihr-today/ (accessed April 2017).
105. NIHR. Public Health Research Programme Full Proposal Guidance. Also available at: http://www.nihr.ac.uk/funding-and-support/documents/current-funding-opportunities/phr/CURRENT_PHR_SAF_Full_Guidance_June-2016.pdf (accessed April 2017).
106. Robinson R. Cost-benefit analysis. *British Medical Journal* 1993; **307** (6909): 924-926.
107. van Mastrigt GA, Paulus AT, Aarts MJ, Evers SM, Alayli-Goebbels AF. A qualitative study on the views of experts regarding the incorporation of non-health outcomes into the economic evaluations of public health interventions. *BMC Public Health* 2015; **15**: 954.
108. Folland, Goodman, Stano. The Economics of Health and Health Care: Pearson New International Edition. Seventh Edition edn. Essex: Pearson Education Limited, 2014.
109. Ryan M, Gerard K, Amaya-Amaya M. Using Discrete Choice Experiments to Value Health and Health Care. Vol. 2: Springer, 2008.
110. Byford S, McDaid D, Sefton T. Because it's worth it: a practical guide to conductiong economic evaluation in the social welfare field. York: Joseph Rowntree Foundations, 2003.
111. Briggs AH, O'Brien BJ. The death of cost-minimization analysis? *Health Econ* 2001; **10** (2): 179-84.
112. Drummond M, Stoddart G, Torrance G. Methods for the Economic Evaluation of Health Care Programmes. New York: Oxford University Press, 1987.
113. Dakin H, Wordsworth S. Cost-minimisation analysis versus cost-effectiveness analysis, revisited. *Health Econ* 2013; **22** (1): 22-34.
114. NICE. Smoking: brief interventions and referrals. PH1. March 2006. Also available at: https://www.nice.org.uk/guidance/ph1.
115. OECD. Fiscal Sustainability of Health Systems: Bridging Health and Finance Perspectives. Paris: OECD publishing2015.
116. Vallejo-Torres L, Garcia-Lorenzo B, Castilla I, et al. On the Estimation of the Cost-Effectiveness Threshold: Why, What, How? *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2016; **19** (5): 558-66.
117. NICE. Guide to the Methods of Technology Appraisal. April 2004. Also available at: http://webarchive.nationalarchives.gov.uk/20080611223138/http://www.nice.org.uk/niceMedia/pdf/TAP_Methods.pdf.
118. Claxton K, Martin S, Soares M, et al. Methods for the Estimation of the NICE Cost Effectiveness Threshold: Final Report. CHE Research Paper 81. University of York: CHE Center for Health Economics2013.

119. Appleby J, Devlin N, Parkin D. NICE's cost effectiveness threshold. *BMJ : British Medical Journal* 2007; **335** (7616): 358-359.
120. House of Commons Health Committee HC 27-I. National Institute for Health and Clinical Excellence: first report of session 2007-08. Vol. 12008.
121. Barnsley P, Towse A, Karlsberg Schaffer S, Sussex J. Critique of CHE Research Paper 81: Methods for the estimation of the NICE cost effectiveness threshold. Occasional Paper 13/01. London: Office of Health Economics2013.
122. Dillon AS. Carrying NICE over the threshold. 2015. Also available at: https://www.nice.org.uk/news/blog/carrying-nice-over-the-threshold (accessed 04/04/2017).
123. Hutubessy R, Chisholm D, Edejer TT. Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. *Cost effectiveness and resource allocation : C/E* 2003; **1** (1): 8.
124. Marseille E, Larson B, Kazi DS, Kahn JG, Rosen S. Thresholds for the cost-effectiveness of interventions: alternative approaches. *Bull World Health Organ* 2015; **93** (2): 118-24.
125. Henry D. Economic Analysis as an Aid to Subsidisation Decisions. *PharmacoEconomics* 1992; **1** (1): 54-67.
126. Guidelines for the economic evaluation of pharmaceuticals: Canada. November 1994. Also available at: https://www.cadth.ca/media/pdf/peg_e.pdf.
127. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *Jama* 1996; **276** (15): 1253-8.
128. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. *JAMA* 2016; **316** (10): 1093-1103.
129. Patient Protection and Affordable Care Act, 42 USC §18001 et seq (2010).
130. Neumann PJ, Weinstein MC. Legislating against Use of Cost-Effectiveness Information. *New England Journal of Medicine* 2010; **363** (16): 1495-1497.
131. World Health Organization. The case for investing public health: a public health summary report for EPHO 82014.
132. Department of Health
Departmental Report 2008. 2008. Also available at: http://webarchive.nationalarchives.gov.uk/20101126054708/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/AnnualReports/DH_084908.
133. Ramsey S, Willke R, Briggs A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value in Health* 2005; **8** (5): 521-33.
134. Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. *BMJ* 2011; **342**.
135. 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.
136. Briggs A, Claxton K, Sculpher M. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press, 2006.
137. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Med Decis Making* 2012; **32** (5): 733-43.
138. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making* 2012; **32** (5): 678-89.

139. Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research: some lessons from recent UK experience. *Pharmacoeconomics* 2006; **24** (11): 1055-68.
140. Kindig D, Stoddart G. What Is Population Health? *American journal of public health* 2003; **93** (3): 380-383.
141. Kindig DA. Understanding Population Health Terminology. *Milbank Quarterly* 2007; **85** (1): 139-161.
142. Orton LC, Lloyd-Williams F, Taylor-Robinson DC, Moonan M, O'Flaherty M, Capewell S. Prioritising public health: a qualitative study of decision making to reduce health inequalities. *BMC Public Health* 2011; **11**: 821.
143. The Life Scientific. Sir Michael Marmot. BBC Radio 4
144. Cecchini M, Sassi F, Lauer JA, Lee YY, Guajardo-Barron V, Chisholm D. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. *Lancet* 2010; **376** (9754): 1775-1784.
145. 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**: 1001.
146. Chapel J. CDC Coffee Break: Economic Evaluation: Alternatives to ROI to show societal benefits2016.
147. Wyke S, Hunt K, Gray CM, et al. Public Health Research. Football Fans in Training (FFIT): a randomised controlled trial of a gender-sensitised weight loss and healthy living programme for men - end of study report. Southampton (UK): NIHR Journals Library Copyright (c) Queen's Printer and Controller of HMSO 2015. This work was produced by Wyke et al. under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK., 2015.
148. McIntosh E, Lawson K, Donaldson C, et al. Report on a workshop on methods for the economic evaluation of population health interventions: conceptual and practical challenges. Occasional Paper no 23: MRC/CSO Social and Public Health Sciences Unit2012.
149. Supporting public service transformation: cost benefit analysis guidance for local partnerships: Crown copyright2014.
150. NICE. Developing NICE guidelines: the manual. NICE, 2014. Also available at: https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview.
151. NICE. The social care guidance manual.Updated July 2016.
152. Porta M, Greenland S, Hernan M, Silva IdS, Last J, eds. A Dictionary of Epidemiology. Sixth edn. New York, NY: Oxford University Press2014.
153. Meyer BD. Natural and quasi-experiments in economics. *Journal of business & economic statistics* 1995; **13** (2): 151-161.
154. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003; **7** (27): iii-x, 1-173.
155. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. *BMJ* 2016; **352**.
156. Devlin N, Sussex J. Incorporationg multiple criteria in HTA: methods and processes. London: OHE Research: Office of Health Economics2011.

157. Thokala P, Duenas A. Multiple Criteria Decision Analysis for Health Technology Assessment. *Value in Health* 2012; **15** (8): 1172-1181.
158. Belton V, Stewart TJ. Multiple Criteria Decision Analysis: An integrated Approach: Kluwer Academic Publishers, 2002.
159. Marsden G, Towse A, Henshall C. How can HTA meet the needs of health system and government decision makers?: Office of Health Economics2017.
160. Devlin N, Ijzerman MJ, Marsh K, Thokala P. MCDA for Health Care Decision Making - An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force - DRAFT for REVIEW only: ISPOR 2015.
161. Marsh K, M IJ, Thokala P, et al. Multiple Criteria Decision Analysis for Health Care Decision Making--Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2016; **19** (2): 125-37.
162. McDaid D, Sassi F, Merkur S, eds. Promoting Health, Preventing Disease: the economic case. England: Open University Press2015.
163. Rose G. Strategy of prevention: lessons from cardiovascular disease. *BMJ* 1981; **282**: 1847-1851.
164. Jousilahti P, Laatikainen T, Salomaa V, Pietila A, Vartiainen E, Puska P. 40-Year CHD Mortality Trends and the Role of Risk Factors in Mortality Decline: The North Karelia Project Experience. *Global heart* 2016; **11** (2): 207-12.
165. Mackenbach JP, Lingsma HF, van Ravesteyn NT, Kamphuis CB. The population and high-risk approaches to prevention: quantitative estimates of their contribution to population health in the Netherlands, 1970-2010. *European journal of public health* 2013; **23** (6): 909-15.
166. Mathes T, Antoine S-L, Prengel P, Bühn S, Polus S, Pieper D. HEALTH TECHNOLOGY ASSESSMENT OF PUBLIC HEALTH INTERVENTIONS: A SYNTHESIS OF METHODOLOGICAL GUIDANCE. *International Journal of Technology Assessment in Health Care* 2017: 1-12.
167. Curtis, Lori J. Economics, Health Economics, Evaluation, and Public Health. 2014-11-18. URL:http://www.nccph.ca/docs/Curtis_HE_PH_En.pdf. Accessed: 2014-11-18. (Archived by WebCite® at http://www.webcitation.org/6UBMrmkYW)
168. Chokshi DA, Farley TA. The Cost-Effectiveness of Environmental Approaches to Disease Prevention. *New England Journal of Medicine* 2012; **367** (4): 295-297.
169. Schwappach DB. The economic evaluation of prevention – Let's talk about values and the case of discounting. *Int J Public Health* 2007; **52** (6): 335-336.
170. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)—Explanation and Elaboration: A Report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value in Health* 2013; **16** (2): 231-250.
171. Cookson R. QALYs and the capability approach. *Health Econ* 2005; **14** (8): 817-29.
172. The ideas and influence of Alan Williams: Be reasonable - Do it my way! In: Mason Anne R, Towse A, eds. Oxford: Radcliffe Publishing2008.
173. Brazier J, Tsuchiya A. Improving Cross-Sector Comparisons: Going Beyond the Health-Related QALY. *Appl Health Econ Health Policy* 2015; **13** (6): 557-65.
174. Wildman J, McMeekin P, Grieve E, Briggs A. Economic evaluation of integrated new technologies for health and social care: Suggestions for policy makers, users and evaluators. *Soc Sci Med* 2016; **169**: 141-148.
175. Petrou S, Gray R. Methodological challenges posed by economic evaluations of early childhood intervention programmes. *Applied Health Economics and Health Policy* 2005; **4** (3): 175-181.

176. Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ. Discounting and decision making in the economic evaluation of health-care technologies. *Health Econ* 2011; **20** (1): 2-15.
177. Claxton K, Sculpher M, Culyer AJ. Mark versus Luke? Appropriate methods for the evaluation of public health interventions. CHE Research Paper 31. Centre for Health Economics: University of York2007.
178. Smith RD, Yago M, Millar M, Coast J. Assessing the macroeconomic impact of a healthcare problem: the application of computable general equilibrium analysis to antimicrobial resistance. *J Health Econ* 2005; **24** (6): 1055-75.
179. NICE. Appraising life-extending, end of life treatments. 2014-11-20. URL:http://www.nice.org.uk/guidance/gid-tag387/resources/appraising-life-extending-end-of-life-treatments-paper2. Accessed: 2014-11-20. (Archived by WebCite® at http://www.webcitation.org/6UEAnXcHl)
180. COOKSON R, DRUMMOND M, WEATHERLY H. Explicit incorporation of equity considerations into economic evaluation of public health interventions. *Health Economics, Policy and Law* 2009; **4** (02): 231-245.
181. Townsend P, Davidson N, Whitehead M. Inequalities in Health: the Black Report and the Health Divide. Second edn. London: The Penguin Group, 1992.
182. Owen L, Morgan A, Fischer A, Ellis S, Hoy A, Kelly MP. The cost-effectiveness of public health interventions. *J Public Health (Oxf)* 2012; **34** (1): 37-45.
183. Levin H. Waiting for Godot: cost-effectiveness analysis in education. *New directions for evaluation* 2001; **2001** (90): 55-68.
184. Levin HM. Cost-effectiveness Analysis. International Encyclopedia of Economics of Education. 2nd edn. Oxford: Pergamon, 1995.
185. Gilead T. Education and the Rationale of Cost–Benefit Analysis. *British Journal of Educational Studies* 2014; **62** (4): 373-391.
186. Merkur S, Sassi F, McDaid D. Promoting health, preventing disease: is there an economic case? : World Health Organization Policy Summary 62013.
187. U.S. Department of Education. National Center for Education Statistics. *Measuring Resources in Education: From Accounting to the Resource Cost Model Approach.* Working Paper No. 1999-16, by Jay G. Chambers. Project Officer, William J. Fowler, Jr. Washington, D.C.: 1999.
188. Sinha I, Jones L, Smyth RL, Williamson PR. A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children. *PLoS medicine* 2008; **5** (4): e96.
189. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare. In: Dissemination CfRa, ed. University of York2009.
190. Popay J, Roberts H, Sowden A, et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme2006.
191. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP journal club* 1995; **123** (3): A12-3.
192. Scottish Intercollegiate Guidelines Network (SIGN) Search Filters. Edinburgh: 2015. Also available at: http://www.sign.ac.uk/methodology/filters.html#.
193. Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized trials. *Cochrane Database Syst Rev* 2007 (2): Mr000001.
194. Walker DG, Wilson RF, Sharma R, et al. Best Practices for Conducting Econoic Evaluations in Health Care: A Systematic Review of Quality Assessment Tools. Rockville, MD: Agency for Healthcare Research and Quality (US)2012.

195. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996; **313** (7052): 275-83.
196. Chiou CF, Hay JW, Wallace JF, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care* 2003; **41** (1): 32-44.
197. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care* 2005; **21** (2): 240-5.
198. International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Also available at: https://www.ispor.org/Health-Economic-Evaluation-Publication-CHEERS-Guidelines.asp.
199. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMJ : British Medical Journal* 2013; **346**.
200. The EQUATOR Network: Enhancing the QUAlity and Transperency Of health Research Also available at: https://www.equator-network.org/reporting-guidelines/cheers/.
201. Cochrane Handbood for Systematic Reviews of Interventions. In: Higgins JP, Green S, eds2011.
202. Dixon-Woods M, Agarwal S, Young B, Jones D, Sutton A. Integrative approaches to qualitative and quantitative evidence. *London: Health Development Agency* 2004; **181**.
203. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Bmj* 2009; **339**: b2535.
204. Bertrand E, Mallis M, Bui NM, Reinharz D. Cost-effectiveness simulation of a universal publicly funded sealants application program. *Journal of Public Health Dentistry* 2011; **71** (1): 38-45.
205. Crowley DM, Jones DE, Coffman DL, Greenberg MT. Can we build an efficient response to the prescription drug abuse epidemic? Assessing the cost effectiveness of universal prevention in the PROSPER trial. *Preventive Medicine* 2014; **62**: 71-77.
206. Foster EM, Jones D. Can a costly intervention be cost-effective?: An analysis of violence prevention. Archives of general psychiatry. Vol. 63, pp. 1284-91, 2006.
207. Foster JM, Toma EF, Troske SP. Does Teacher Professional Development Improve Math and Science Outcomes and Is It Cost Effective? *Journal of Education Finance* 2013; **38** (3): 255-275.
208. Levaux HP, Schonfeld WH, Pellissier JM, Cassidy WM, Sheriff SK, Fitzsimon C. Economic evaluation of a 2-dose hepatitis B vaccination regimen for adolescents. *Pediatrics* 2001; **108** (2): 317-325.
209. Noyes K, Bajorska A, Fisher S, Sauer J, Fagnano M, Halterman JS. Cost-effectiveness of the school-based asthma therapy (SBAT) program. *Pediatrics* 2013; **131** (3): e709-e717.
210. Wang LY, Gutin B, Barbeau P, et al. Cost-effectiveness of a school-based obesity prevention program. *Journal of School Health* 2008; **78** (12): 619-624.
211. Barrett JL, Gortmaker SL, Long MW, et al. Cost effectiveness of an elementary school active physical education policy. *American journal of preventive medicine* 2015; **49** (1): 148-159.
212. Beets MW. Making healthy eating and physical activity policy practice: The design and overview of a group randomized controlled trial in afterschool programs. Contemporary clinical trials. Vol. 38, pp. 291-303, 2014.
213. Ford T, Edwards V, Sharkey S, et al. Supporting teachers and children in schools: the effectiveness and cost-effectiveness of the Incredible Years teacher classroom

management programme in primary school children: a cluster randomised controlled trial, with parallel economic and process evaluations. *BMC Public Health* 2012; **12**: 719-719.
214. Simon E, Dirksen CD, Bögels SM. An explorative cost-effectiveness analysis of school-based screening for child anxiety using a decision analytic model. *European Child & Adolescent Psychiatry* 2013; **22** (10): 619-630.
215. Kesztyues D, Schreiber A, Wirt T, 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.
216. Hollingworth W, Cohen D, Hawkins J, et al. Reducing smoking in adolescents: cost-effectiveness results from the cluster randomized ASSIST (A Stop Smoking In Schools Trial). Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco. Vol. 14, pp. 161-8, 2012.
217. Brooker S, Okello G, Njagi K, et al. Improving educational achievement and anaemia of school children: design of a cluster randomised trial of school-based malaria prevention and enhanced literacy instruction in Kenya. *Trials [Electronic Resource]* 2010; **11**: 93.
218. Carabin H, Chan MS, Guyatt HL. A population dynamic approach to evaluating the impact of school attendance on the unit cost and effectiveness of school-based schistosomiasis chemotherapy programmes. *Parasitology* 2000; **121**: 171-183.
219. Nishiura H, Ejima K, Mizumoto K, et al. Cost-effective length and timing of school closure during an influenza pandemic depend on the severity. Vol. 11. Theoretical Biology and Medical Modelling, pp. 5, 2014.
220. Cooper K, Shepherd J, Picot J, et al. An economic model of school-based behavioral interventions to prevent sexually transmitted infections. Vol. 28. International Journal of Technology Assessment in Health Care, pp. 407-414, 2012.
221. Konig HH, Barry JC. Cost-utility analysis of orthoptic screening in kindergarten: a Markov model based on data from Germany. *Pediatrics* 2004; **113** (2): e95-108.
222. Philipsson A, Duberg A, Moller M, Hagberg L. Cost-utility analysis of a dance intervention for adolescent girls with internalizing problems. *Cost effectiveness and resource allocation : C/E* 2013; **11** (1): 4.
223. Shepherd J, Kavanagh J, Picot J, et al. The effectiveness and cost-effectiveness of behavioural interventions for the prevention of sexually transmitted infections in young people aged 13-19: a systematic review and economic evaluation. *Health Technology Assessment* 2010; **14** (7): 1-+.
224. te Velde SJ, Lennert Veerman J, Tak NI, Bosmans JE, Klepp K-I, Brug J. Modeling the Long Term Health Outcomes and Cost-Effectiveness of Two Interventions Promoting Fruit and Vegetable Intake among Schoolchildren. *Economics and Human Biology* 2011; **9** (1): 14-22.
225. Vijgen SMC, van Baal PHM, Hoogenveen RT, de Wit GA, Feenstra TL. Cost-effectiveness analyses of health promotion programs: A case study of smoking prevention and cessation among Dutch students. *Health Education Research* 2008; **23** (2): 310-318.
226. Barber SE, Jackson C, Akhtar S, et al. "Pre-schoolers in the playground" an outdoor physical activity intervention for children aged 18 months to 4 years old: study protocol for a pilot cluster randomised controlled trial. *Trials [Electronic Resource]* 2013; **14**: 326.
227. Chestnutt IG, Chadwick BL, Hutchings S, et al. Protocol for "Seal or Varnish?" (SoV) trial: a randomised controlled trial to measure the relative cost and effectiveness of pit and fissure sealants and fluoride varnish in preventing dental decay. *Bmc Oral Health* 2012; **12**.

228. Frick KD, Carlson MC, Glass TA, et al. Modeled cost-effectiveness of the Experience Corps Baltimore based on a pilot randomized trial. *Journal of Urban Health* 2004; **81** (1): 106-17.
229. Gerald JK, Grad R, Bailey WC, Gerald LB. Cost-effectiveness of school-based asthma screening in an urban setting. *J Allergy Clin Immunol* 2010; **125** (3): 643-50, 650 e1-650 e12.
230. Muennig P, Woolf SH. Health and economic benefits of reducing the number of students per classroom in US primary schools. *American journal of public health* 2007; **97** (11): 2020-2027.
231. Muennig PA, Epstein M, Li G, DiMaggio C. The cost-effectiveness of New York City's Safe Routes to School program. Vol. 104. American Journal of Public Health, pp. 1294-1299, 2014.
232. Rein DB, Wittenborn JS, Zhang X, Song M, Saaddine JB. The potential cost-effectiveness of amblyopia screening programs. *Journal of pediatric ophthalmology and strabismus* 2012; **49** (3): 146-55; quiz 145, 156.
233. Tengs TO, Osgood ND, Chen LL. The cost-effectiveness of intensive national school-based anti-tobacco education: results from the tobacco policy model. *Prev Med* 2001; **33** (6): 558-70.
234. Wang LY, Crossett LS, Lowry R, Sussman S, Dent CW. Cost-effectiveness of a school-based tobacco-use prevention program. *Arch Pediatr Adolesc Med* 2001; **155** (9): 1043-50.
235. 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.
236. Blakely T, Kvizhinadze G, Karvonen T, Pearson AL, Smith M, Wilson N. Cost-effectiveness and equity impacts of three HPV vaccination programmes for school-aged girls in New Zealand. *Vaccine* 2014; **32** (22): 2645-56.
237. Moodie M, Haby M, Galvin L, Swinburn B, Carter R. Cost-effectiveness of active transport for primary school children — Walking School Bus program. *The International Journal of Behavioral Nutrition and Physical Activity* 2009; **6**.
238. Moodie ML, Herbert JK, de Silva-Sanigorski AM, 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.
239. Moodie ML, Carter RC, Swinburn BA, Haby MM. The cost-effectiveness of Australia's Active After-School Communities program. *Obesity* 2010; **18** (8): 1585-92.
240. Rush E, Obolonkin V, McLennan S, et al. Lifetime cost effectiveness of a through-school nutrition and physical programme: Project Energize. Vol. 8. Obesity Research and Clinical Practice, pp. e115-e200, 2014.
241. Quach J, Gold L, Arnup S, Sia KL, Wake M, Hiscock H. Sleep well - Be well study: Improving school transition by improving child sleep: A translational randomised trial. BMJ Open. Vol. 32013.
242. Pearson AL, Kvizhinadze G, Wilson N, Smith M, Canfell K, Blakely T. Is expanding HPV vaccination programs to include school-aged boys likely to be value-for-money: a cost-utility analysis in a country with an existing school-girl program. *BMC Infectious Diseases* 2014; **14** (1): 351.
243. Miller T, Hallfors D, Cho H, Luseno W, Waehrer G. Cost-effectiveness of school support for orphan girls to prevent HIV infection in Zimbabwe. *Prevention Science* 2013; **14** (5): 503-512.
244. Kowada A. Cost effectiveness of interferon-gamma release assay for school-based tuberculosis screening. *Molecular diagnosis & therapy* 2012; **16** (3): 181-90.

245. König H-H, Barry J-C, Leidl R, Zrenner E. Economic evaluation of orthoptic screening: results of a field study in 121 German kindergartens. *Investigative Ophthalmology & Visual Science* 2002; **43** (10): 3209-3215.
246. Barnett WS. Benefit-Cost Analysis of the Perry Preschool Program and Its Policy Implications. *Educational Evaluation and Policy Analysis* 1985; **7** (4): 333-342.
247. Reynolds AJ, Temple JA, White BAB, Ou S-R, Robertson DL. Age 26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program. *Child development* 2011; **82** (1): 379-404.
248. Wang LY, Davis M, Robin L, Collins J, Coyle K, Baumler E. Economic evaluation of Safer Choices: a school-based human immunodeficiency virus, other sexually transmitted diseases, and pregnancy prevention program. *Arch Pediatr Adolesc Med* 2000; **154** (10): 1017-24.
249. Tai T, Bame SI. Cost-benefit analysis of childhood asthma management through school-based clinic programs. *J Community Health* 2011; **36** (2): 253-60.
250. Hoeflmayr D, Hanewinkel R. Do school-based tobacco prevention programmes pay off? The cost-effectiveness of the 'Smoke-free Class Competition'. *Public Health* 2008; **122** (1): 34-41.
251. Glewwe P, Jacoby HG, King EM. Early Childhood Nutrition and Academic Achievement: A Longitudinal Analysis. *Journal of Public Economics* 2001; **81** (3): 345-368.
252. Eckermann S, Dawber J, Yeatman H, Quinsey K, Morris D. Evaluating return on investment in a school based health promotion and prevention program: The investment multiplier for the Stephanie Alexander Kitchen Garden National Program. *Social Science & Medicine* 2014; **114**: 103-112.
253. Barnett MA, Thompson S. The role of perspective taking and empathy in children's Machiavellianism, prosocial behavior, and motive for helping. *The Journal of Genetic Psychology: Research and Theory on Human Development* 1985; **146** (3): 295-305.
254. Abt CC, National Center for Educational Statistics WDC, Abt Associates ICMA. A COST-EFFECTIVENESS MODEL FOR THE ANALYSIS OF TITLE I ESEA PROJECT PROPOSALS, PART I-VII1966.
255. Brown HS, 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**: 47.
256. Scherrer CR, Griffin PM, Swann JL. Public health sealant delivery programs: optimal delivery and the cost of practice acts. *Medical Decision Making* 2007; **27** (6): 762-71.
257. Atherly A, Nurmagambetov T, Williams S, Griffith M. An economic evaluation of the school-based "power breathing" asthma program. *The Journal Of Asthma: Official Journal Of The Association For The Care Of Asthma* 2009; **46** (6): 596-599.
258. Babey SH, Wu S, Cohen D. How can schools help youth increase physical activity? An economic analysis comparing school-based programs. *Preventive Medicine* 2014; **69**: S55-S60.
259. Brassard P, Steensma C, Cadieux L, Lands LC. Evaluation of a school-based tuberculosis-screening program and associate investigation targeting recently immigrated children in a low-burden country. *Pediatrics* 2006; **117** (2): e148-56.
260. Foster EM. Costs and effectiveness of the fast track intervention for antisocial behavior. The journal of mental health policy and economics. Vol. 13, pp. 101-19, 2010.
261. Gesell SB, Sommer EC, Lambert EW, et al. Comparative effectiveness of after-school programs to increase physical activity. *J Obes* 2013; **2013**: 576821.

262. Guay M, Clouatre AM, Blackburn M, et al. Effectiveness and cost comparison of two strategies for hepatitis B vaccination of schoolchildren. *Canadian Journal of Public Health* 2003; **94** (1): 64-67.
263. Joseph CL, Peterson E, Havstad S, et al. A web-based, tailored asthma management program for urban African-American high school students. *American Journal of Respiratory & Critical Care Medicine* 2007; **175** (9): 888-95.
264. Wang LY, Yang Q, Lowry R, Wechsler H. Economic Analysis of a School-Based Obesity Prevention Program. *Obesity Research* 2003; **11** (11): 1313-1324.
265. Young TL, Ireson C. Effectiveness of school-based telehealth care in urban and rural elementary schools. *Pediatrics* 2003; **112** (5): 1088-94.
266. Newbury-Birch D, O'Neil S, Gilvarry E, et al. A feasability trial of alcohol screening and brief interventions for risky drinking in young people in a high school setting in the UK: Sips jr-high. Alcoholism: Clinical and Experimental Research. Vol. 37, pp. 147a, 2013.
267. Anderson R, Ukoumunne OC, Sayal K, et al. Cost-effectiveness of classroom-based cognitive behaviour therapy in reducing symptoms of depression in adolescents: a trial-based analysis. *J Child Psychol Psychiatry* 2014; **55** (12): 1390-7.
268. Stallard P, Phillips R, Montgomery A, et al. A cluster randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of classroom-based cognitive-behavioural therapy (CBT) in reducing symptoms of depression in high-risk adolescents. *Health Technology Assessment* 2013; **17** (47).
269. Boyle J, McCartney E, Forbes J, O'Hare A. A randomised controlled trial and economic evaluation of direct versus indirect and individual versus group modes of speech and language therapy for children with primary language impairment. *Health Technol Assess* 2007; **11** (25): iii-iv, xi-xii, 1-139.
270. Gelli A, Al-Shaiba N, Espejo F. The costs and cost-efficiency of providing food through schools in areas of high food insecurity. *Food and Nutrition Bulletin* 2009; **30** (1): 68-76.
271. Gelli A, Cavallero A, Minervini L, Mirabile M, Molinas L, de la Mothe MR. New benchmarks for costs and cost-efficiency of school-based feeding programs in food-insecure areas. *Food and Nutrition Bulletin* 2011; **32** (4): 324-332.
272. Salisbury C, Francis C, Rogers C, et al. A randomised controlled trial of clinics in secondary schools for adolescents with asthma. *British Journal of General Practice* 2002; **52**: 988-996.
273. Shemilt I, Mugford M, Moffatt P, et al. A national evaluation of school breakfast clubs: Where does economics fit in? *Child: Care, Health and Development* 2004; **30** (5): 429-437.
274. Ansell J, Guyatt HL. Comparative cost-effectiveness of diagnostic tests for urinary schistosomiasis and the implications for school health programmes. *Annals of Tropical Medicine and Parasitology* 2002; **96** (2): 145-153.
275. Curtale F, Hassanein YA, Savioli L. Control of human fascioliasis by selective chemotherapy: Design, cost and effect of the first public health, school-based intervention implemented in endemic areas of the Nile Delta, Egypt. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2005; **99** (8): 599-609.
276. Li Y, Hu X, Zhang Q, et al. The nutrition-based comprehensive intervention study on childhood obesity in China (NISCOC): a randomised cluster controlled trial. *Bmc Public Health* 2010; **10**.
277. Meng L, Xu H, Liu A, et al. The Costs and Cost-Effectiveness of a School-Based Comprehensive Intervention Study on Childhood Obesity in China. *PloS one* 2013; **8** (10).

278. Meng L, Xu H, Liu A, et al. The costs and cost-effectiveness of a school-based comprehensive intervention study on childhood obesity in China. Vol. 8. PLOS ONE, pp. e77971, 2013.
279. Joseph CL, Peterson E, Havstad S, et al. A web-based, tailored asthma management program for urban African-American high school students. American journal of respiratory and critical care medicine. Vol. 175, pp. 888-95, 2007.
280. Foster EM. Costs and effectiveness of the fast track intervention for antisocial behavior. *Journal of Mental Health Policy and Economics* 2010; **13** (3): 101-119.
281. Bruzzese JM, Evans D, Wiesemann S, et al. Using school staff to establish a preventive network of care to improve elementary school students' control of asthma. *Journal of School Health* 2006; **76** (6): 307-12.
282. Wade TJ, Mansour ME, Line K, Huentelman T, Keller KN. Improvements in health-related quality of life among school-based health center users in elementary and middle school. *Ambulatory Pediatrics* 2008; **8** (4): 241-9.
283. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; **340**.
284. Barber SE, Akhtar S, Jackson C, al. E. Preschoolers in the Playground: a pilot cluster randomised controlled trial of a physical activity intervention for children aged 18 months to 4 years. *NIHR Journals Library* 2015; **Public Health Research No. 3.5**.
285. Beets MW, Weaver RG, Turner-McGrievy G, et al. Making Healthy Eating Policy Practice: A Group Randomized Controlled Trial on Changes in Snack Quality, Costs, and Consumption in After-School Programs. *Am J Health Promot* 2016; **30** (7): 521-31.
286. Giles EL, Scott S, Coulton S, et al. Development of a multicentre randomised controlled trial of screening and brief alcohol intervention to prevent risky drinking in young people in a high-school setting (SIPS JR-HIGH). *The Lancet* 2015; **38** (6): S37.
287. Brooker S, Halliday KE. Impact of Malaria Control and Enhanced Literacy Instruction on Educational Outcomes among Kenyan School Children: A Multi-Sectoral, Prospective, Randomized Evaluation. 3ie Series Report2014.
288. Ratcliffe J, Stevens K, Flynn T, Brazier J, Sawyer MG. Whose values in health? An empirical comparison of the application of adolescent and adult values for the CHU-9D and AQOL-6D in the Australian adolescent general population. *Value In Health* 2012; **15** (5): 730-736.
289. Adlard N, Kinghorn P, Frew E. Is the UK NICE “Reference Case” Influencing the Practice of Pediatric Quality-Adjusted Life-Year Measurement within Economic Evaluations? *Value in Health* 2014; **17** (4): 454-461.
290. Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. *Health Technol Assess* 2001; **5** (4): 1-157.
291. Longworth L, Rowen D. NICE DSU Technical Support Document 10: the use of mapping methods to estimate health state utility values: Decision Support Unit2011.
292. World Health Organization. Childhood overweight and obesity. 2017. Also available at: http://www.who.int/dietphysicalactivity/childhood/en/ (accessed 16/02/2017).
293. Kwan SYL, Petersen PE, Pine CM, Borutta A. Health-promoting schools: an opportunity for oral health promotion. *Bulletin of the World Health Organization* 2005; **83**: 677-685.
294. Banke-Thomas AO, Madaj B, Charles A, van den Broek N. Social Return on Investment (SROI) methodology to account for value for money of public health interventions: a systematic review. *BMC Public Health* 2015; **15**: 582.
295. Burdge RJ, Vanclay F. SOCIAL IMPACT ASSESSMENT: A CONTRIBUTION TO THE STATE OF THE ART SERIES. *Impact Assessment* 1996; **14** (1): 59-86.

296. World Health Organization. Health Impact Assessment (HIA). 2003. Also available at: http://www.who.int/hia/about/en/.
297. Heckman JJ, Masterov D. The Productivity Argument for Investing in Young Children. IZA Discussion Paper No 2725 Available at SSRN: http://ssrncom/abstract=9821172007.
298. Belli PC, Bustreo F, Preker A. Investing in children's health: what are the economic benefits? *Bull World Health Organ* 2005; **83** (10): 777-84.
299. Heckman JJ. Schools, Skills, and Synapses. *Econ Inq* 2008; **46** (3): 289.
300. Janssens A, Rogers M, Thompson Coon J, et al. A systematic review of generic multidimensional patient-reported outcome measures for children, part II: evaluation of psychometric performance of English-language versions in a general population. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2015; **18** (2): 334-45.
301. Ungar WJ. Challenges in health state valuation in paediatric economic evaluation: are QALYs contraindicated? *Pharmacoeconomics* 2011; **29** (8): 641-52.
302. Apajasalo M, Sintonen H, Holmberg C, et al. Quality of life in early adolescence: a sixteen-dimensional health-related measure (16D). *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1996; **5** (2): 205-11.
303. Moodie M, Richardson J, Rankin B, Iezzi A, Sinha K. Predicting time trade-off health state valuations of adolescents in four Pacific countries using the Assessment of Quality-of-Life (AQoL-6D) instrument. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2010; **13** (8): 1014-27.
304. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2010; **19** (6): 875-86.
305. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care* 1996; **34** (7): 702-22.
306. Saigal S, Rosenbaum P, Stoskopf B, et al. Development, reliability and validity of a new measure of overall health for pre-school children. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2005; **14** (1): 243-57.
307. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res* 2010; **19** (6): 875-886.
308. Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the pediatric quality of life inventory. *Medical Care* 1999; **37** (2): 126-39.
309. 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.
310. Development of a preference-based index for the Pediatric Quality of Life Inventory (PedsQL) for use in economic evaluation. Also available at: https://www.findaphd.com/search/projectdetails.aspx?PJID=40915.
311. Ratcliffe J, Flynn T, Terlich F, Stevens K, Brazier J, Sawyer M. Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. *Pharmacoeconomics* 2012; **30** (8): 713-727.
312. Crump RT, Beverung LM, Lau R, Sieracki R, Nicholson M. Reliability, Validity, and Feasibility of Direct Elicitation of Children's Preferences for Health States. *Medical Decision Making* 2016; **37** (3): 314-326.

313. Stevens K. Valuation of the Child Health Utility 9D Index. *PharmacoEconomics* 2012; **30** (8): 729-47.
314. 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 : an international journal of quality of life aspects of treatment, care and rehabilitation* 2012; **21** (4): 717-25.
315. 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.
316. The development of a paediatric health related quality of life measure for use in economic evaluation: The Child Health Utility 9D (CHU 9D). https://www.shef.ac.uk/scharr/sections/heds/mvh/paediatric. Accessed 15.03.172017.
317. The CHILDSPLA project. http://childspla.lshtm.ac.uk/about-the-project/ Accessed 15.03.17
318. Nuffield Foundation (2012) *Social trends and mental health: introducing the main findings* London: Nuffield Foundation.
319. Collishaw S, Maughan B, Natarajan L, Pickles A. Trends in adolescent emotional problems in England: a comparison of two national cohorts twenty years apart. *J Child Psychol Psychiatry* 2010; **51** (8): 885-94.
320. Collishaw S, Goodman R, Pickles A, Maughan B. Modelling the contribution of changes in family life to time trends in adolescent conduct problems. *Soc Sci Med* 2007; **65** (12): 2576-87.
321. Macleod KL, Crowther R, Stewart-Brown S. The health of children and young people. In: Gillam S, Yates J, Badrinath P, eds. Essential Publich Health. Second edn: Cambridge University Press2012.
322. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry* 1997; **38** (5): 581-6.
323. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry* 2001; **40** (11): 1337-45.
324. Becker A, Woerner W, Hasselhorn M, Banaschewski T, Rothenberger A. Validation of the parent and teacher SDQ in a clinical sample. *European Child & Adolescent Psychiatry* 2004; **13**: ii11-ii16.
325. Roy BV, Veenstra M, Clench-Aas J. Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry* 2008; **49** (12): 1304-1312.
326. Information for researchers and professionals about the Strengths & Difficulties Questionnaire. 2015-05-04. URL:http://www.sdqinfo.com/. Accessed: 2015-05-04. (Archived by WebCite® at http://www.webcitation.org/6YH56saPy).
327. Achenback T. Manual for the Child Behavior Checklist/4-18 and 1991 Profile. Burlington, VT: University of Vermont Department of Psychiatry1991.
328. Rutter M. A children's behaviour questionnaire for completion by teachers: preliminary findings. *Journal of Child Psychology and Psychiatry* 1967; **8** (1): 1-11.
329. Furber G, Segal L, Leach M, Cocks J. Mapping scores from the Strengths and Difficulties Questionnaire (SDQ) to preference-based utility values. *Qual Life Res* 2014; **23** (2): 403-411.
330. Rousseau P. Empathy. *American Journal of Hospice and Palliative Medicine* 2008; **25** (4): 261-262.

331. Bohart AC, Greenberg LS. Empathy: Where are we and where do we go from here? Empathy reconsidered: New directions in psychotherapy. Washington, DC, US: American Psychological Association, 1997.
332. Elliott R, Bohart AC, Watson JC, Greenberg LS. Empathy. *Psychotherapy* 2011; **48** (1): 43-49.
333. Concise Medical Dictionary. 8th edn. online: Oxford University Press2014.
334. Northern Ireland Statistics and Research Agency. *Northern Ireland Multiple Deprivation Measure 2010*. Belfast: NISRA
335. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in medicine* 2011; **30** (4): 377-99.
336. Eldridge S, Kerry S. A Practical Guide to Cluster Randomised Trials in Health Services Research. West Sussex: Jon Wiley & Sons, Ltd, 2012.
337. Gomes M, Grieve R, Nixon R, Edmunds WJ. Statistical methods for cost-effectiveness analyses that use data from cluster randomized trials: a systematic review and checklist for critical appraisal. *Med Decis Making* 2012; **32** (1): 209-20.
338. Curtis L. Unit Costs of Health and Social Care 2014. Personal Social Services Research Unit: University of Kent2014.
339. Royal College of Nursing. NHS Agenda for Change pay scales - 2011/2012. 2016-05-13. URL:https://www2.rcn.org.uk/__data/assets/pdf_file/0005/372992/004106.pdf. Accessed: 2016-05-13. (Archived by WebCite® at http://www.webcitation.org/6hTLC4Ghy)
340. OECD (2015), Purchasing power parities (PPP) (indicator). doi: 10.1787/1290ee5a-en (Accessed on 02 September 2015).
341. Hale J, Cohen D, Ludbrook A, Phillips C, Duffy M, Parry-Langdon N. Moving from evaluation into economic evaluation: a health economics manual for programmes to improve health and well-being. Online: UK Health Promotion and Health Econmics Forum
342. NHS reference costs 2013 to 2014: Policy paper. In: Health Do, ed2014.
343. Office for National Statistics. Statistical bulletin: 2011 Annual Survey of Hours and Earnings (SOC 2000). 2014-07-01. URL:http://www.ons.gov.uk/ons/rel/ashe/annual-survey-of-hours-and-earnings/ashe-results-2011/ashe-statistical-bulletin-2011.html. Accessed: 2014-07-01. (Archived by WebCite® at http://www.webcitation.org/6Qjo0o4yG).
344. British National Formulary. 2014-07-24. URL:http://www.bnf.org/bnf/index.htm. Accessed: 2014-07-24. (Archived by WebCite® at http://www.webcitation.org/6RIurjC3U).
345. Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. *Qualitative health research* 2010; **20** (3): 340-51.
346. Faria R, Gomes M, Epstein D, White IR. A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. *PharmacoEconomics* 2014; **32** (12): 1157-1170.
347. Little RJA, Rubin DB. Statistical Analysis with Missing Data. Second edn. Hoboken, New Jersey: Wiley & Sons, Inc.2002.
348. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing... presumed at random: cost-analysis of incomplete data. *Health Econ* 2003; **12** (5): 377-92.
349. Rubin DB. Inference and missing data. *Biometrika* 1976; **63** (3): 581-592.
350. Díaz-Ordaz K, Kenward MG, Grieve R. Handling missing values in cost effectiveness analyses that use data from cluster randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2014; **177** (2): 457-474.
351. Burton A, Billingham LJ, Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. *Clin Trials* 2007; **4** (2): 154-61.

352. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons, 1987.
353. Schafer JL. Multiple imputation: a primer. *Statistical methods in medical research* 1999; **8** (1): 3-15.
354. Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. *Journal of health services research & policy* 2004; **9** (4): 197-204.
355. 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.
356. Briggs AH, O'Brien BJ, Blackhouse G. Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annual Review of Public Health* 2002; **23** (1): 377-401.
357. Van Hout B, Al M, Gilad S, Gordon G, Rutten F. Costs, effects and C/E-ratios alongside a clinical trial. *Health Econ* 1994; **3**: 309 - 319.
358. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess* 1999; **3** (2): 1-134.
359. Bachmann MO, Fairall L, Clark A, Mugford M. Methods for analyzing cost effectiveness data from cluster randomized trials. *Cost effectiveness and resource allocation : C/E* 2007; **5**: 12.
360. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *American Journal of Epidemiology* 1981; **114** (6): 906-914.
361. Gomes M, Ng ES, Grieve R, Nixon R, Carpenter J, Thompson SG. Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. *Med Decis Making* 2012; **32** (2): 350-61.
362. 2014-based National Population projections at mid-years by age last birthday in five year age groups for all individual years 2014-2039, and additional years 2041, 2046, 2051, 2056, 2061 and 2064 by the Office for National Statistics published 29 October 2015.
363. Robertson W, Fleming J, Kamal A, 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 Technol Assess* 2017; **21** (1): 1-180.
364. Julious SA, Horspool MJ, Davis S, et al. PLEASANT: Preventing and Lessening Exacerbations of Asthma in School-age children Associated with a New Term - a cluster randomised controlled trial and economic evaluation. *Health Technol Assess* 2016; **20** (93): 1-154.
365. EEF Guidance on Cost Evaluation: Education Endowment Foundation2015.
366. Culyer AJ. Cost-effectiveness thresholds in health care: a bookshelf guide to their meaning and use. *Health Economics, Policy and Law* 2016; **11** (4): 415-432.
367. NICE. Supporting investment in public health: Review of methods for assessing cost effectiveness, cost impact and return on investment2011.
368. Minnis H. Conversation with Helen Minnis 14th October.2016.
369. Edwards RT, Jones C, Berry V, et al. Incredible Years parenting programme: cost-effectiveness and implementation. *Journal of Children's Services* 2016; **11** (1): 54-72.
370. Siegel JE. Cost-effectiveness analysis and nursing research--is there a fit? *Image--the journal of nursing scholarship* 1998; **30** (3): 221-2.
371. Feinstein L, Sabates R, Anderson T, Sorhaindo A, Hammond C. What are the effects of education on health? Copenhagen: OECD Centre for Educational Research and Innovation2006.

372. Deighton J, Croudace T, Fonagy P, Brown J, Patalay P, Wolpert M. Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: a review of child self-report measures. *Child and adolescent psychiatry and mental health* 2014; **8**: 14.
373. White J, Connelly G, Thompson L, Wilson P. Assessing wellbeing at school entry using the Strengths and Difficulties Questionnaire: professional perspectives. *Educational Research* 2013; **55** (1): 87-98.
374. Booker CL, Skew AJ, Sacker A, Kelly YJ. Well-Being in Adolescence—An Association With Health-Related Behaviors: Findings From Understanding Society, the UK Household Longitudinal Study. *Journal of Early Adolescence* 2014; **34** (4): 518-538.
375. Ohl M, Mitchell K, Cassidy T, Fox P. The Pyramid Club Primary School-Based Intervention: Evaluating the Impact on Children's Social-Emotional Health. *Child and Adolescent Mental Health* 2008; **13** (3): 115-121.
376. Ewing DL, Monsen JJ, Kwoka M. Behavioural and emotional well-being of children following non-directive play with school staff. *Educational Psychology in Practice* 2014; **30** (2): 192-203.
377. Wolpert M, Cheng H, Deighton J. Measurement Issues: Review of four patient reported outcome measures: SDQ, RCADS, C/ ORS and GBO - their strengths and limitations for clinical use and service evaluation... Strengths and Difficulties Questionnaire... Revised Child Anxiety and Depression Scale... (Child) Outcomes Rating Scale... Goals Based Outcomes. *Child & Adolescent Mental Health* 2015; **20** (1): 63-70.
378. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry* 2000; **177**: 534-539.
379. Earnshaw J, Lewis G. NICE Guide to the Methods of Technology Appraisal: Pharmaceutical Industry Perspective. *PharmacoEconomics* 2008; **26** (9): 725.
380. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *The European Journal of Health Economics* 2010; **11** (2): 215-25.
381. University of London. UCL Institute of Education. Centre for Longitudinal Studies, Millennium Cohort Study. Colchester, Essex: UK Data Archive [distributor].
382. Skovgaard AM, Olsen EM, Houmann T, et al. The Copenhagen County child cohort: design of a longitudinal study of child mental health. *Scandinavian Journal of Public Health* 2005; **33** (3): 197-202.
383. Australian Mental Health Outcomes and Classification Network. 2015-10-29. URL:http://www.amhocn.org/. Accessed: 2015-10-29. (Archived by WebCite® at http://www.webcitation.org/6cdxT5FqW).
384. Child Outcomes Research Consortium. 2015-10-29. URL:http://www.corc.uk.net/. Accessed: 2015-10-29. (Archived by WebCite® at http://www.webcitation.org/6cdxaH5q9).
385. Boyer NR, Miller S, Connolly P, McIntosh E. Paving the way for the use of the SDQ in economic evaluations of school-based population health interventions: an empirical analysis of the external validity of SDQ mapping algorithms to the CHU9D in an educational setting. *Qual Life Res* 2016; **25** (4): 913-923.
386. Scoring the Strengths & Difficulties Questionnaire for ages 4-172015-05-04. URL:http://www.sdqinfo.org/py/sdqinfo/b3.py?language=Englishqz(UK). Accessed: 2015-05-04. (Archived by WebCite® at http://www.webcitation.org/6YH5WiCL5).
387. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *American journal of epidemiology* 2010; **171** (5): 624-32.

388. 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.
389. Xu F, Chen G, Stevens K, 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-e89222.
390. Canaway AG, Frew EJ. 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. *Qual Life Res* 2013; **22** (1): 173-183.
391. Petrou S, Rivero-Arias O, Dakin H, et al. The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. *Pharmacoeconomics* 2015; **33** (10): 993-1011.
392. Madan J, Khan KA, Petrou S, Lamb SE. Can Mapping Algorithms Based on Raw Scores Overestimate QALYs Gained by Treatment? A Comparison of Mappings Between the Roland-Morris Disability Questionnaire and the EQ-5D-3L Based on Raw and Differenced Score Data. *Pharmacoeconomics* 2017; **35** (5): 549-559.
393. The Green Book: Appraisal and Evaluation in Central Government. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf. London: HM Treasury2016.
394. Flodmark CE, Marcus C, Britton M. Interventions to prevent obesity in children and adolescents: a systematic literature review. *International journal of obesity (2005)* 2006; **30** (4): 579-89.
395. van Sluijs EM, McMinn AM, Griffin SJ. Effectiveness of interventions to promote physical activity in children and adolescents: systematic review of controlled trials. *Bmj* 2007; **335**: 703.
396. Bélanger E, Rodríguez C, Groleau D. Shared decision-making in palliative care: A systematic mixed studies review using narrative synthesis. *Palliative Medicine* 2011; **25** (3): 242-261.
397. Noyes J, Edwards RT. EQ-5D for the Assessment of Health-Related Quality of Life and Resource Allocation in Children: A Systematic Methodological Review. *Value in Health* 2011; **14** (8): 1117-1129.
398. Petrou S, Murray L, Cooper P, Davidson LL. The accuracy of self-reported healthcare resource utilization in health economic studies. *Int J Technol Assess Health Care* 2002; **18** (3): 705-10.
399. Achana F, Petrou S, Khan K, Gaye A, Modi N. A methodological framework for assessing agreement between cost-effectiveness outcomes estimated using alternative sources of data on treatment costs and effects for trial-based economic evaluations. *The European journal of health economics : HEPAC : health economics in prevention and care* 2017.
400. Thorn JC, Coast J, Cohen D, et al. Resource-use measurement based on patient recall: issues and challenges for economic evaluation. *Appl Health Econ Health Policy* 2013; **11** (3): 155-61.
401. Franklin M, Davis S, Horspool M, Kua WS, Julious S. Economic Evaluations Alongside Efficient Study Designs Using Large Observational Datasets: the PLEASANT Trial Case Study. *PharmacoEconomics* 2017; **35** (5): 561-573.
402. Schweinhart L, Montie J, Xiang Z, Barnett S, Belfield C, Nores M. Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40. Ypsilanti, Michigan: High/Scope Educational Research Foundation, pp. 194-215, 2005.

403. Schweinhart L, Weikart DP. Young Children Grow Up: the effects of the Perry Preschool Program on Yourths Through Age 15. Ypsilanti, Michiagan: High/Scope Educational Research Foundation, 1980.
404. Chen G, Stevens K Fau - Rowen D, Rowen D Fau - Ratcliffe J, Ratcliffe J. From KIDSCREEN-10 to CHU9D: creating a unique mapping algorithm for application in economic evaluation. (1477-7525 (Electronic)).
405. Mpundu-Kaambwa C, Chen G, Russo R, Stevens K, Petersen KD, Ratcliffe J. Mapping CHU9D Utility Scores from the PedsQLTM 4.0 SF-15. (1179-2027 (Electronic)).
406. Wood H, Arber M, Glanville JM. SYSTEMATIC REVIEWS OF ECONOMIC EVALUATIONS: HOW EXTENSIVE ARE THEIR SEARCHES? *Int J Technol Assess Health Care* 2017: 1-7.
407. McNeal Jr RB, Hansen WB. An examination of strategies for gaining convergent validity in natural experiments: DARE as an illustrative case study. *Evaluation Review* 1995; **19** (2): 141-158.
408. West SL, O'Neal KK. Project D.A.R.E. Outcome Effectiveness Revisited. *American journal of public health* 2004; **94** (6): 1027-1029.