An investigation of economic evaluation methods for public health interventions: meeting the needs of public health decision-makers

A thesis presented for the degree of

Doctor of Philosophy at Newcastle University

(Faculty of Medical Sciences)

Sarah Ruth Hill

BA, MSc (University of York)

September 2019

Abstract

Market failure typically arises from the provision of public goods in private markets due to their being non-excludable and non-rivalry in consumption. Consequently, public goods will generally be under-provided in private markets and government intervention is commonplace to help ensure efficient provision. Programmes for prevention in health can be considered public goods. Prevention largely falls within remit of public health which, since 2013, has been the responsibility of local authorities (LAs) in England.

Due to limited resources to fund public health activities in England, the prioritisation of resources is paramount. Economic evaluation can assist in guiding resource allocation decisions. Recommendations for economic evaluation methods to appraise public health interventions are less clear than for health technologies. Identifying relevant methodologies for public health appraisal is important to address the complexity of public health programmes and the LA setting of public health decisions in England. This study aimed to engage public health decision-makers (PHDMs) to identify the most beneficial economic evaluation tool(s) to meet their needs.

This research focused on a sub-sect of public health: interventions related to alcohol consumption. A systematic review was conducted to identify economic evaluations of interventions to reduce alcohol misuse since 2006. The majority of evaluations identified were cost-utility analyses (CUAs). Limited consideration of methodological challenges specific to public health was found.

A qualitative interview study with PHDMs in North-East England was then conducted which identified limited use and knowledge of health economic tools amongst the PHDMs. A desire for the incorporation of broader outcomes in evaluations to incorporate the local decision-making context was established.

Building on the outcomes from the review and qualitative study, a cost-benefit analysis (CBA) and social return on investment (SROI) were conducted to evaluate a brief alcohol intervention in schools. In order to conduct the CBA, a contingent valuation (CV) study was carried out to obtain a monetised measure of benefit. Despite debate in the literature on the suitability of the CV method to elicit true economic values for public goods, particularly for goods offering limited private consumption value, CV has been used in the field of environmental economics for decades to ascertain non-use values for goods. The method

has also been used to value healthcare goods. The outcomes of the CV survey additionally provided information on the mechanics of the decisions made by the public via examination of predictors of willingness-to-pay for the intervention. It can also help elucidate reasons given in support of the intervention (or lack of), potentially relevant to PHDMs. Such examination help bridge the gap between the fields of health economics and psychology.

Both the CBA and SROI identified positive societal benefit from the intervention. These results were shown to PHDMs at a workshop, alongside results from a previously conducted CUA and a cost-consequence analysis (CCA) of the same intervention. The workshop elicited PHDM attendees' preferences for the practical use of methodologies. SROI was most preferred, yet concern remained over its use and interpretation by PHDMs. A "one-tool-fits-all" approach was doubted as being appropriate by some attendees. Combining CCA alongside another evaluation was proposed favourably by PHDMs. Providing locally relevant analysis via additional sensitivity analyses was also regarded as highly beneficial by attendees.

Acknowledgements

This thesis would not have been possible without the help and support of many individuals. I would like to take this opportunity to acknowledge my gratitude to them.

Firstly, I would like to wholeheartedly thank my supervisors: Professor Luke Vale, Professor David Hunter, Dr Yemi Oluboyede, and Dr Emily Henderson. Their support, guidance, and honest feedback have made this work what it is. I have been extremely fortunate to have had a fantastic supervisory team, who have allowed me to grow as an independent researcher, yet who have always been available to provide advice when needed.

I am also grateful to Fuse for providing me with the opportunity to undertake this PhD. Without their financial support this PhD would not have been possible. Access to the Fuse network of like-minded researchers and individuals has been invaluable throughout this research. Furthermore, I would like to thank the SIPS Jr HIGH project team and Viviana Albani for allowing me access to their data.

I would also like to express gratitude to my colleagues in the health economics group at the Institute of Health & Society. To those who have given up their time to discuss or read my work, I am eternally grateful.

I would like to extend my thanks to the public health decision-makers who voluntarily gave up their time to participate in the qualitative interviews and workshop conducted for this research. Additionally, I must acknowledge colleagues who have assisted me at various stages of this research: Wanwuri Akor and Jo Gray.

I have been fortunate enough to have embarked on my PhD journey with several other students; Emma, Beth, Tass, Charlotte, Jess, Rashmi, and all the fellow PhD colleagues with whom I have shared the “PhD room” over the past four years. I will be forever appreciative of the solidarity you have provided throughout the challenging stages of the PhD.

I owe a great debt of gratitude to my parents, Peter and Valerie, who have provided a roof over my head for the past four years and have been an unlimited source of love and support. Despite the perhaps perceived endlessness of this endeavour, you have suffered my presence graciously, and have always ensured that I have never gone hungry after late nights in the office. To my sister, Sue, thank you for instilling the importance of proper grammar in me from such a young age. To my youngest brother, Dan, for our lunch-time coffees which

provided much needed escapes from the computer. Finally, to my oldest sibling, Andrew, for at least feigning interest in my research during your visits home.

Thanks must also go to my oldest friends, who have been my greatest cheerleaders and who have been so understanding when the thesis took over social engagements such as weddings and hen parties.

Last, but not least, I am truly grateful to James for your love, kindness and patience. Despite the physical distance between us, you have always been considerate of my work and allowed me to be indulgent with my time without complaint.

Table of Contents

List of Tables

List of Figures

List of Boxes

Published work from this thesis

Published protocol

This is the protocol for work presented in Chapter Four

Hill SR, Oluboyede Y, Vale L, Hunter D, Henderson E. *Economic evaluation and priority-setting in public health: protocol for a review of the methods used to evaluate interventions for the prevention and reduction of excessive alcohol consumption*. PROSPERO: International prospective register of systematic reviews; 2016. Report No.: CRD42016039063

Published manuscript

This is a shortened version of the review presented in Chapter Four

Hill SR, Vale L, Hunter D, Henderson E, Oluboyede Y. Economic evaluations of alcohol prevention interventions: Is the evidence sufficient? A review of methodological challenges. *Health Policy*. 2017;121(12):1249-62

Abbreviations

A&E	Accident and Emergency
A-SAQ	Adolescent Single Alcohol Question
ASBI	Alcohol screening and brief intervention
AUDIT	Alcohol Use Disorders Identification Test
BCR	Benefit-cost ratio
CADTH	Canadian Agency for Drugs and Technologies in Health
CBA	Cost-benefit analysis
CCA	Cost-consequence analysis
CCG	Clinical Commissioning Group
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CI	Confidence interval
CPI	Consumer price index
CRD	Centre for Research and Dissemination
CUA	Cost-utility analysis
CV	Contingent valuation
DALY	Disability Adjusted Life Year
DCE	Discrete choice experiment
DsPH	Directors of Public Health
EU	European Union
GLM	Generalised Linear Model
GP	General practitioner
HTA	Health technology assessment
HUI	Health Utilities Index
HYE	Healthy Year Equivalent
ICER	Incremental cost-effectiveness ratio
LA	Local authority
MCDA	Multi Criteria Decision Analysis
NDTMS	National drug treatment monitoring system
NHS	National Health Service
NHS EED	NHS Economic Evaluation Database
NICE	National Institute for Health and Care Excellence
NSB	Net societal benefit
OLS	Ordinary least squares
PBMA	Programme budgeting and marginal analysis
PHDM	Public health decision maker
PHE	Public Health England
PSS	Personal social services
QALY	Quality adjusted life year

RCT	Randomised controlled trial
ROI	Return on investment
SAPM	Sheffield Alcohol Policy Model
SD	Standard deviation
SE	Standard error
SIPS	Screening and Intervention Programme for Sensible Drinking
SROI	Social return on investment
UK	United Kingdom
US	United States
USA	The United States of America
WEMWBS	Warwick-Edinburgh mental wellbeing score
WHO	World Health Organisation
WTA	Willingness to accept
WTP	Willingness to pay

PART I. Introduction, Theory, Literature review and Qualitative study

Chapter 1. Introduction

The overall aim of this doctoral research project is to identify the most beneficial health economic evaluation tool(s) to meet the needs of public health decision-makers (PHDMs). Whilst the relevance of various economic evaluation tools to aid decision-making can be debated by health economists on the methods' academic and theoretical merits, these arguments may pose less relevance to everyday decision-making. It is equally important that the end-users of economic evaluation evidence are able to use the information in a manner that serves the purposes of those making resource allocation decisions. Therefore, this thesis seeks to explore the decision-making needs of PHDMs to add to the debate from the perspective of the end-users of the information.

This chapter begins by presenting a background of the developments in public health in England over the past decade. Changes to the context of contemporary public health decision-making have brought to the fore the complexities surrounding the health economic appraisal of public health programmes. These complexities are outlined in section 1.1 and insights from current literature on potential approaches to address them are presented. Section 1.2 describes the motivation for this research, drawing on existing guidance for the appraisal of public health interventions and outlines remaining areas for enquiry. The research questions posed for the thesis are presented in section 1.3. This is followed by a brief introduction to the case study examined for the empirical work in section 1.4. The final section outlines the structure of the thesis.

1.1 Public health in England

The focus of public health in the United Kingdom (UK) has shifted several times since its inception in the nineteenth century; moving from improving environmental conditions to tackling infectious disease and then to embedding the welfare state in post-war Britain. Twenty-first century public health has evolved yet again, moving away from a focus on communicable disease towards tackling non-communicable disease by concentrating on the behavioural, environmental and socioeconomic factors that influence health (2). Public health was defined by Sir Donald Acheson as *"…the science and art of preventing disease, prolonging life and promoting health through the organised efforts of society." (3).*

Interest in the wider, socioeconomic determinants of health was partly the catalyst for the shift in responsibilities of public health provision in England. These responsibilities were

returned to local authorities (LAs) in 2013 having resided within the NHS since 1974. The reform was intended to enable the wider public health needs of local communities to be best met by integrating public health with the broader LA functions, such as provision of social care, housing, environmental health, leisure and transport services (4). Although the new responsibilities were welcomed by local councils (5), agents unaccustomed to public health decision-making, such as locally elected politicians, need to be supported in their roles. Consequently, new approaches to evaluation and prioritisation of public health (dis)investments may be required in the LA setting.

Resources available to fund public health activities are, however, limited. This is pertinent given the continual UK government cuts to the ring-fenced public health grant provided to LAs in England. In 2016, an announcement was made stating that the grant would be reduced by 3.9% on average in real-terms per annum between 2016 and 2020 (6). As of the 2019/20 public health grant, this amounted to a cash reduction of £250 million to English LAs since 2016 (7). Whilst LAs are able to invest in public health initiatives using other LA funds, the cuts to the ring-fenced public health grant represent only part of a larger suite of austerity measures. Local government grants have been subjected to substantial cuts over the past decade (8). Additionally, funding diverted towards public health initiatives using non-public health grant finance has an opportunity cost for other LA responsibilities. Funds spent on public health programmes cannot also be used to finance initiatives in other sectors which LAs are accountable for such as housing, transport infrastructure, and environmental protection. Although, there may be overlap between public health initiatives and those within other LA sectors, for example, providing cycle-lanes may straddle both public health and transport infrastructure improvements. This presents an additional challenge to public health agents and puts greater pressure on decision-makers to efficiently allocate scarce resources. An approach that can aid decisions over resource allocation is economic evaluation.

1.1.1 *Public health appraisal and prioritisation*

Economic evaluation is defined as *"the comparative analysis of alternative courses of action in terms of both their costs and consequences"* (9, p.4) and is ingrained in the analysis of health technologies in the UK as a means to assist the efficient allocation of resources (10). Limited resources to fund healthcare and public health initiatives require that decisions have to be made over which programmes and services are financed and supported and which are

not. Health technology assessment (HTA) in the UK has traditionally aimed to maximise health subject to an exogenous budget constraint via the application of a "reference case" approach to the appraisal of health technologies (11). This approach, outlined by the National Institute for Health and Care Excellence (NICE), uses formal economic evaluation in the form of cost-effectiveness analysis (CEA) in which technologies are appraised based on their associated costs and health outcomes, measured as quality adjusted life years (QALYs). The QALY measure combines quantity and quality of life into a single outcome (9). The sub-set of economic evaluation in which QALYs constitute the outcome measure is typically referred to as cost-utility analysis (CUA) (see Chapter Three for further discussion of these frameworks).

Additional factors besides cost-effectiveness also have to be considered when appraising any intervention but arguably especially public health programmes, such as equity and social justice concerns (12-14), and burden of disease (15). Furthermore, within the context of LAs, public health constitutes a wider remit than providing health and care services; public health concerns improving the well-being of the population. It encompasses securing economic prosperity, educational achievement, and occupational prospects to name but a few (16). Therefore, aiming for health maximisation may be inappropriate given the additional concerns of local PHDMs and the need to additionally capture intervention impacts on non-health outcomes (17). Consequently, the reference case approach for HTA proposed by NICE with its focus on health (as measured by QALYs) (11) is unlikely to be appropriate for the appraisal of public health interventions (18). The relevance of the current CUA framework to public health appraisal has also been questioned by scholars in the field (19-23). Academics have thus been exploring ways in which the CEA framework may be adapted for public health interventions in light of these additional considerations. Advances such as adjusting the way in which the QALY is currently derived in order to incorporate wider outcomes (24), developing outcome measures which capture capability wellbeing rather than health (25), and developing methods of incorporating health equity concerns into CEA (26, 27) have all been suggested, yet these approaches remain absent from routine practice (28).

Beyond the CEA framework, alternative approaches have also been considered. The UK treasury has recommended frameworks such as cost-benefit analysis (CBA) for the evaluation of public sector policies and projects for over a decade in order to capture a broad range of impacts such as economic, environmental, and social (29, 30). NICE also

broadened its recommended approach to evaluating public health programmes in the third edition of the methods guidance for public health appraisal (31). In the updated edition, greater emphasis is placed on the use of methods such as CBA and cost-consequence analysis (CCA) for the evaluation of public health interventions in light of the transfer of responsibility for public health commissioning to LAs. However, CEA and CUA remain essential to guidance, with CBA and CCA forming secondary analyses.

Evaluative frameworks from beyond the standard health economic sector, such as return on investment (ROI) and its broader counterpart, social return on investment (SROI) (32, 33), have also gained popularity for public health priority-setting. A desire to promote investment in prevention via public health initiatives has resulted in the examination of the ROI of public health programmes by academics (34) and Public Health England (PHE) (35). However, caution has been raised regarding the potential risks to resource allocation of basing decisions on ROI outcomes insofar as evidence of cost-saving (i.e. a positive ROI) may become a necessary condition on which public health programmes are to be implemented (36).

Prioritisation tools such as Programme Budgeting Marginal Analysis (PBMA) (37-39) and Multi Criteria Decision Analysis (MCDA) (15, 40) have also garnered interest for assisting public health decision-making in recent years. MCDA has particularly been considered due to its ability to balance evidence on the cost-effectiveness of public health programmes with other evidence relevant to LA decision-making (41).

1.1.2 *The complexity of public health interventions*

The public health system is multi-sectoral in that interventions in other sectors, such as housing or the environment, can affect the health of the population, or conversely, interventions designed to improve health can impact other sectors as a consequence in the change of the public's behaviours and actions (42). Public health interventions are thus often described as complex (43) due to the inclusion of multiple components, which act independently and inter-dependently (44). For example, Petticrew (45) describes "urban regeneration" programmes as an example of complex interventions due to the integration of several components from educational to housing interventions, each contributing to the overall programme.

Furthermore, public health interventions, which may not be considered complex in themselves, may operate within complex systems such as primary care, hospitals, or schools (46). These contexts are considered complex as it is often difficult to distinguish intervention impact from other influencing factors within the system.

This complexity presents additional issues for the evaluation of public health programmes compared to health technologies. The challenges associated with evaluating complex public health programmes have been considered by academics for over a decade (20, 46-49). Yet, there remains limited guidance for public health appraisal. A review of guidance for the economic evaluation of public health interventions published in 2013 (17) identified a need for further guidance in the field of public health given the many challenges that face those attempting economic evaluations. Edwards *et al.* (17) suggested that those undertaking evaluations of public health interventions should think beyond the traditional health economist's tools and consider health effects as part of a wider set of relevant outcome measures, such as health inequalities, social and environmental impacts (17). Since the review by Edwards *et al.* was published (17), NICE updated their manual for producing guidance to amalgamate guidance for health technologies, public health and social care interventions. Consequently, the reference case was amended to include guidance on the economic evaluation of public health interventions (10). The recommendations with respect to economic evaluation frameworks, however, remain largely unchanged from those stated in the earlier version of guidance (31).

1.2 Motivation for research

Despite the exploration of new approaches to evaluate public health interventions over the past decade, evaluation using the standard CUA framework (9) has remained the most commonly drawn on evidence for NICE public health guidance (21). Perhaps more importantly, studies have shown that local policy-making in healthcare often overlooks economic evaluation evidence during prioritisation (50-52). There are several barriers to the use of economic evaluations in the decision-making process ranging from the technical to the political. The transfer of public health responsibilities to LAs has brought political factors to the forefront of priority setting. Context and values are now more influential than ever before in public health priority-setting (53), particularly given the involvement of locally elected council members. Their requirements to satisfy stakeholders, including their constituents, and their own professional interests, such as future re-election (54), contribute

to the reported limited role played by economic evaluation evidence in local public health decision-making. Furthermore, experience-driven knowledge is often privileged over academic evidence when council members address local policy interests (54).

It is unclear what criteria are used for public health decision-making, particularly at the LA level, and what role economic evaluation plays in those decisions. Consequently, research is needed to investigate how economic evaluation evidence can be best generated and relayed to PHDMs in order to aid decision-making and prioritisation of the limited public health funding. Prior to the move back to LAs, Phillips *et al.* (19) explored the role of economics in public health prioritisation. CBA and CUA were identified as preferred to CEA or CCA in Phillips *et al.'s* study (19), however, the research was limited in terms of the very small sample of policy-makers included (n=8). Consequently, the authors recommended further research in conjunction with policy-makers in order to build on their findings. Furthermore, the applicability of the study findings to the current state-of-play of public health decision-making in a LA context may have to be considered, given that different agents now fill the roles of PHDMs.

1.3 Thesis aim and research questions

Since the onset of the doctoral research reported in this thesis there have been some further developments in the field of economic evaluation and prioritisation in public health (55-58). However, the question of the most appropriate economic evaluation framework for aiding public health prioritisation decisions remains unresolved. Therefore, this research aims to add to the literature and shed light on this issue.

In order to achieve this aim, three research questions are posed:

1. With respect to current economic evaluation and priority-setting tools:
    a. What evidence is currently available and which methods are used by the health economic research community to evaluate public health interventions?
    b. Does the quality of evidence produced meet recommendations for health economic evaluations of public health interventions from the available guidance?
2. With respect to the use of health economic evidence by PHDMs:
    a. To what extent is health economic evidence used by PHDMs to aid decision-making?

b. To what extent do PHDMs have sufficient knowledge of health economic tools to appropriately use the available evidence?
c. What barriers do PHDMs perceive exist to the use of health economic evidence as it is currently produced?

3. Is a particular method of economic evaluation, or combination of methods, most beneficial to PHDMs for their decision-making needs?

The systematic review reported in Chapter Four addresses research question one. The qualitative interview study with PHDMs reported in Chapter Five addresses question two. Finally, the findings form a workshop with PHDMs reported in Chapter Ten addresses question three.

1.4 Introduction to the empirical case study

The empirical work conducted for this doctoral research project is based on a case study of a public health intervention. Details of the case study are discussed in greater detail in Chapter Six with a brief overview presented here. The intervention chosen was a school-based alcohol screening and brief intervention (ASBI) programme for students in Year 10 (ages 14-15), which aimed to prevent misuse of alcohol in young people. The intervention was evaluated in a UK based, randomised controlled trial (RCT), SIPS Jr HIGH. Details of the intervention and the trial findings are reported in Giles *et al.,* 2019 (59). This case study provided an example of a complex public health issue, due to the potentially negative impacts of alcohol on young people ranging from health impacts (60) to anti-social behavioural problems (61) and educational attainment (62). Consequently, this case study was relevant to exploring the benefits of various economic evaluation frameworks for evaluating complex public health interventions.

1.5 Thesis structure

This thesis is comprised of three parts. Part One introduces the thesis, outlines the economic theory underpinning economic evaluation and reviews the relevant literature. It concludes with a qualitative interview study, which informs the choice of economic evaluation methods that are the focus of the second part of the thesis. Part Two outlines the empirical work conducted to prepare economic evaluation evidence to present to PHDMs in a workshop. Part Three discusses the findings from a workshop that was held to elicit feedback on alternative methods of economic evaluation which could be used to appraise

public health interventions. The thesis concludes with a discussion of all the empirical work conducted. A synopsis of each chapter is presented below.

Part One

The subsequent chapters in Part One are outlined as follows:

Chapter Two outlines the economic theory underpinning the methodologies presented in this thesis. This consists of an overview of consumer choice theory and the welfarist and extra-welfarist theoretical paradigms. The implications of each of these schools of thought for contemporary economic evaluation and addressing questions of efficiency in resource allocation are presented.

Chapter Three outlines each economic evaluation and priority-setting methodology considered in this thesis. Four of the standard methods of economic evaluation used to appraise health technologies are outlined (CEA, CUA, CCA and CBA) in addition to SROI. Priority-setting frameworks such as MCDA and PBMA are also outlined. The explanation of CBA also includes a brief discussion of the various approaches to monetising benefits and outlines approaches that can be used to elicit stated preferences, such as contingent valuation (CV) and discrete choice experiments (DCEs).

Chapter Four describes the conduct of a systematic review which explores the use of the methods discussed in Chapter Three in the published literature. The review focuses on the use of tools to appraise interventions to prevent alcohol misuse, in accordance with the case study used as the basis for the empirical work conducted in this thesis.

Chapter Five investigates the use of economic evaluation evidence to aid priority-setting by PHDMs. A broad range of PHDMs are interviewed for the study, ranging from Directors of Public Health (DsPH) to locally elected council members with a health brief. Particular attention is paid to barriers and enablers to the use of health economic evidence and the current state of PHDMs' knowledge of health economic concepts. The chapter explores whether local PHDMs have any strong views on the merits of particular economic evaluation and priority-setting tools. The findings from the chapter direct the remainder of the empirical work in Part Two by focussing attention on four methods of economic evaluation: CUA, CCA, CBA and SROI.

Part Two

Chapter Six describes the development of a contingent valuation (CV) survey which was used to elicit a value of willingness-to-pay (WTP) in order to conduct a CBA. The SIPS Jr HIGH case study is also described in detail.

Chapter Seven outlines the methods of recruitment and analysis used to examine the information gathered from the CV survey. The chapter also discusses the results of the CV study described in Chapter Six. This chapter estimates a mean value of WTP which is used in the CBA conducted in Chapter Eight.

Chapter Eight reports the methods and results of both a CBA and SROI analysis appraising the ASBI programme examined by the SIPS Jr HIGH trial. The results from both of these evaluations fed into the workshop with PHDMs in order to provide examples of how each method can be used to evaluate a public health intervention. Short reports of each method of economic evaluation were produced for the workshop with PHDMs in order to provide examples of the sort of information that could be produced using each approach.

Chapter Nine outlines the adaptation of the novel CBA and SROI evaluations, and a CUA and CCA that had been previously conducted for the within-trial health economic analysis of the SIPS Jr HIGH study, into a short report form. Each report outlined the methods, results and sensitivity analyses that were conducted for each of the four evaluations.

Part Three

Chapter Ten provides details of the workshop and the process of obtaining feedback on the four methods of economic evaluation presented for review. The chapter also outlines the findings with respect to PHDMs' opinions of the merits and demerits of each method for aiding local public health decision-making. The implications of the findings for conducting and presenting economic evaluations of public health interventions are also discussed.

Chapter Eleven summarises the thesis findings with respect to each of the research questions posed in section 1.3. The implications of the research for practice are outlined alongside areas where a need for further research has been identified. The chapter also discusses the limitations of the overall doctoral research project and ends with some concluding remarks.

Chapter 2. Economic theory underpinning economic evaluation

This chapter outlines the theoretical basis of economic evaluation and its variants which are used in this thesis. The chapter begins with an explanation of consumer choice theory and its assumptions. This leads to a description of the archetypal economic problem of utility maximisation. Section 2.2 moves towards applying the principles outlined at a societal level within the economic framework of welfare economics. Following this, the decision-rules for optimising social welfare determined by welfare economic theory are discussed. An alternative framework, extra-welfarism, is introduced in section 2.3, and section 2.4 outlines the problem of market failure in healthcare and public health provision. The final section draws out the implications of each school of thought for methods of economic evaluation.

2.1 Consumer choice theory

Microeconomic theory provides a framework for decision-making in relation to consumption for a rational consumer. Consumer choice theory is built on the central tenet of optimisation: of the numerous combinations (bundles) of goods[1] that an individual can consume, he/she will choose the bundle that he/she most prefers. In economics, the term utility is used to denote a 'subjective sensation associated with the consumption of goods (e.g. satisfaction, satiation of need etc.)' (63, p.16). Thus, the bundle of goods which provides the consumer with the greatest utility is that which he/she will most prefer.

2.1.1 *Preference axioms*

In order for consumer choice theory to be upheld, several axioms of preferences are assumed (63).

1. Complete preferences

It is assumed that preferences for bundles of goods form a closed loop such that for any two bundles of goods (A and B) one of the following statements is always true:

i. A is preferred to B
ii. B is preferred to A
iii. A is indifferent to B

[1] The term *goods* is used in this sense to denote anything that an individual may wish to consume, including both good and services

2. Transitivity

Preferences are transitive so that of three bundles of goods, A, B and C, if A is preferred to B and B is preferred to C then A will also be preferred to C. Thus, preferences regarding goods are consistent.

3. Reflexivity

Reflexivity ensures that any bundle of goods is indifferent to itself.

4. Continuity

If bundle A is preferred to bundle B and a third bundle, C, is sufficiently close to B to be comparable then A will also be preferred to C.

5. Non-satiation

This assumption implies that consumers will never be satiated with goods; more is always preferred to less. If two bundles, A and B, consist of varying quantities of two goods, x and y, the bundle which has a greater quantity of good x and no less of good y will always be preferred. The most preferred bundle will, thus, lie on a higher indifference curve[2] than the least preferred bundle, since for the same quantity of good y, the preferred bundle contains a greater quantity of good x. This axiom holds for all goods which are not considered "bad" by the consumer, i.e. things that one might prefer to have less of.

6. Diminishing rates of marginal substitution

Taking a bundle containing two goods, x and y, as the quantity of x is increased, the quantity of y foregone to retain indifference within the bundle gets less. The rate of substitution of good y for x, therefore, diminishes as the quantity of good x within the bundle increases. The implication of this axiom is that indifference curves are convex to the origin and that a consumer never prefers to forego all of one good in favour of another.

2.1.2 *Utility maximisation*

Consumers choose between bundles of goods such that they maximise their own utility. Typically, however, utility maximisation is constrained by a budget. Therefore, the problem becomes maximising utility subject to this budget constraint. For a hypothetical world where

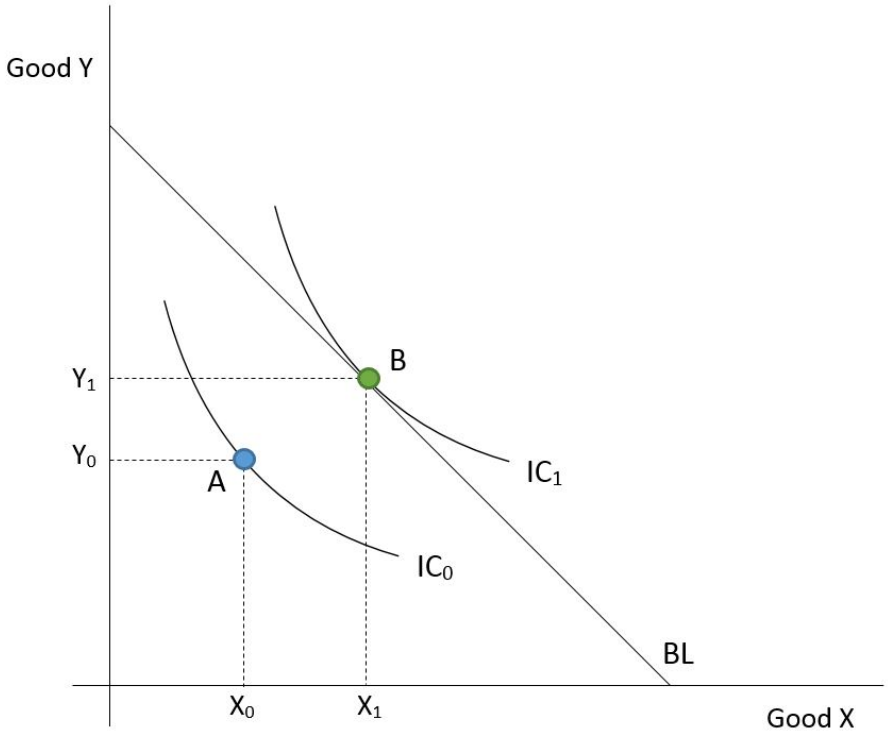
[2] An indifference curve is a line on a graph of two goods that represents different combinations of consumption of each good for which a consumer is indifferent between, which generate a given level of utility.

only two competing goods exist, X and Y, this can be described algebraically. Equation 2.1 sets this out for a rational consumer, where U_i denotes the utility function of consumer *i*, income is represented by *M* and P_x and P_y denoted the price of goods X and Y, respectively (63).

$$Max\ (U_i) = f(X,Y)\ subject\ to\ M \geq P_x X + P_y Y \qquad (2.1)$$

Maximising utility subject to a budget constraint can also be demonstrated graphically, taking account of each of the axioms of preferences described in section 2.1.1. Figure 2.1 illustrates utility maximisation in the hypothetical, two-good world where consumer *i's* income imposes a budget constraint shown by BL. All points along BL represent levels of consumption of goods X and Y which exhaust consumer *i's* income. Assuming one cannot be in debt, points above BL are not feasible, whilst any point below BL is within the consumer's budget.

Figure 2.1 Utility maximisation subject to a budget constraint



An indifference curve is a line which displays all combinations of goods in which a consumer is indifferent between, i.e. a rational consumer derives the same level of utility from any combination of goods along the curve (64). Figure 2.1 shows two indifference curves: IC_0 and IC_1. Goods bundle A on indifference curve IC_0 is operating within consumer *i's* budget since IC_0 is below BL. Bundle B corresponds to an increase in consumption of both goods X and Y

compared to A (X_0 to X_1, Y_0 to Y_1), thus, increasing consumer *i's* utility from consumption of both goods. The increase in utility is represented by a higher indifference curve, IC_1. Utility is maximised at bundle B because it is the point of tangency between BL and IC_1; any movement along the indifference curve IC_1will result in an increase in one good at the expense of the other but as it is above BL it cannot be afforded (64).

2.2 Welfare economics

The theory of consumer choice described in the previous section is concerned with one individual. However, when concerned with the allocation of public resources, such as healthcare, the scope is much broader towards that of society. Welfare economics can be defined as the analysis of social desirability of states of the world, or allocations of resources, in terms of the utility gained from such allocations by individuals in the society (65, p.205) and is the theoretical basis of economic evaluation.

The normative analysis of healthcare is based on four tenets of neo-classical welfare economics (66, 67): utility maximisation, individual sovereignty, consequentialism, and welfarism. Utility maximisation, as described in section 2.1.2, is achieved when a rational consumer chooses between bundles of goods to maximise his/her utility. Individual sovereignty implies that individuals can best appraise their own welfare and, consequently, this rejects paternalistic judgements about individual welfare. Consequentialism, a normative ethical stance which states that actions should be judged based solely on their consequences, is here applied to economic actions. Thus, the appropriateness of economic policies or actions should be based on their outcomes, not the process involved in producing the resultant effects. Finally, welfarism extends the consequentialist stance in relation to individual utility; therefore, actions, such as resource allocation, should be appraised on the utility generated to individuals affected by the action.

2.2.1 *Pareto optimality*

The classical school of welfare economics considered utilities to be cardinally measurable (i.e. that consumption bundles can be given a measurable value) and comparable across members of a society (66). Within this school of thought, societal welfare could be optimised by maximising the utilities of all members of society. The assumption that utilities possessed cardinal properties was relaxed with the introduction of neo-classical welfare economics and ordinal utility theory (i.e. that consumption bundles can be ranked by preference but

without the need to attach a measurable value). The solution to the problem of optimising societal resource allocation was, thus, replaced with Pareto optimality (66).

Pareto optimality is achieved via the consideration of Pareto improvement. A "weak" Pareto improvement is said to have occurred if an economic change, such as a reallocation of resources, increases the utility of society as a whole. A "strong" Pareto improvement constitutes a reallocation of resources which makes at least one person better off without making any one person worse off. The strength of each improvement reflects the value judgement required to agree that the course of action is beneficial to society. Thus, the "weak" improvement, in which all individuals benefit from some action, is less objectionable than the "strong" improvement, in which only one individual is required to benefit (65). An allocation of resources is, therefore, Pareto optimal at the point in which any further allocation that could improve one individual's utility could only be made at the expense of another individual's utility. In other words, no further movement can be made that is a strong Pareto improvement.

Several different distributions of resources could simultaneously be considered Pareto optimal, yet the Pareto optimality criterion fails to provide guidance on which distribution of resources should be applied (assuming allocations are mutually exclusive). The value judgements made when adhering to Pareto principles do not include judgments on equitable distributions of resources but are concerned solely with the relative size of societal gains and ensuring no person is left worse-off (68). Distributional concerns, nevertheless, are often held in regard by society, as demonstrated by government instituted income redistribution schemes such as progressive taxation and benefit payments (64).

Furthermore, few policy decisions in reality adhere to Pareto conditions, necessitating a further evolution of Pareto principles if they are to be applicable to real-world policy decisions (65, 68). The notion of potential Pareto improvement was developed in order to allow individuals who lose out from some economic action to be compensated by those who gain to ensure there is no net loss from an action (68). In a simplified world constituting two people, a policy would be potentially Pareto improving if the utility gain to Person A is larger than the loss of utility to Person B. Person A could, therefore, compensate person B the exact amount of utility lost from policy implementation (so that Person B's utility level remains unchanged) whilst Person A maintains a utility gain compared to their pre-policy utility level.

Ensuring compensation is adequately distributed from winners to losers of policy action is, in reality, difficult. To this end, two economists, John Hicks and Nicholas Kaldor, proposed variants of the notion of potential Pareto improvement, commonly referred to as the Kaldor-Hicks criterion. The Kaldor-Hicks criterion allows Pareto conditions to be operationalised in a more realistic economic context whereby no actual compensation needs to be paid for an action to be considered potentially Pareto improving (69, 70). Pareto principles are considered to be maintained in policy action, so long as utility losses could be hypothetically compensated from the utility gains.

Given how ubiquitous money is, discussions surrounding the use of potential Pareto improvement to guide policy decisions have commonly centred on the ability to compensate individuals in monetary terms rather than directly compensating utility. This can be reasoned from the notion that utility cannot be directly observed or compared, therefore, money is used as a proxy for utility since it can be used to purchase goods, which contribute to improvements in utility via the consumption of those goods (65). However, this relies on an implicit assumption that a unit of currency contains equal social value amongst winners and losers (68).

Since money is used as a proxy for measuring welfare change (i.e. the change in utility levels) from policy action, two concepts are drawn upon to measure changes in utility using money: compensating variation and equivalent variation. Both concepts can be considered from the perspective of some economic action producing either gains or losses to an individual (see Table 2.1). Compensating variation requires income changes made from an ex-post perspective relative to policy action, in order to return an individual to their original utility level. Equivalent variation requires income changes made from an ex-ante perspective relative to policy action, in order to equate an individual's original utility level to the new level that would be obtained following policy action. Both of these concepts are further illustrated graphically in Figures 2.2 and 2.3.

Table 2.1 Compensating and equivalent variation for gains and losses from policy action

	For policy action which generates gain in utility for an individual	**For policy action which generates a loss in utility for an individual**
Compensating variation	The amount of income taken away from a person following policy action to return that person to their pre-policy level of utility	The amount of income required to be given to a person following policy action to return that person to their pre-policy level of utility
Equivalent variation	The amount of income required to be given to a person prior to policy implementation to raise their pre-policy utility to the same level as would be realised if the policy is implemented.	The amount of money required to be taken away from a person prior to policy implementation that would lower their pre-policy utility to the same level that would be realised if the policy is implemented.

Table adapted from Table 9.1 in Morris *et al.* (2012)(65)

2.2.2 *Compensating variation*

The concept of compensating variation following a policy action which generates a gain for an individual is graphically represented in Figure 2.2. In Figure 2.2, two indifference curves, IC_0 and IC_1 demonstrate levels of utility in relation to a combination of provision of a public good (e.g. healthcare) and expenditure on private goods (measured in monetary units). IC_1, which is north-east of IC_0, represents a higher utility level. Suppose the individual is initially on IC_0 and spends Y_0 on private goods whilst enjoying X_0 of a public good (point A). If some policy action increases the quantity of the public good provided from X_0 to X_1 and the individual maintains his/her current consumption of expenditure on private goods (Y_0), he/she would move to point B on indifference curve IC_1. In order to return the individual to his/her original level of utility (IC_0) whilst enjoying X_1 consumption of the public good, his/her income would need to be reduced by the amount BC, which would reduce his/her expenditure on private goods. Thus, BC represents the compensating variation, which is to say the amount of income taken away to compensate for the increase in the public good.

Figure 2.2 Compensating variation

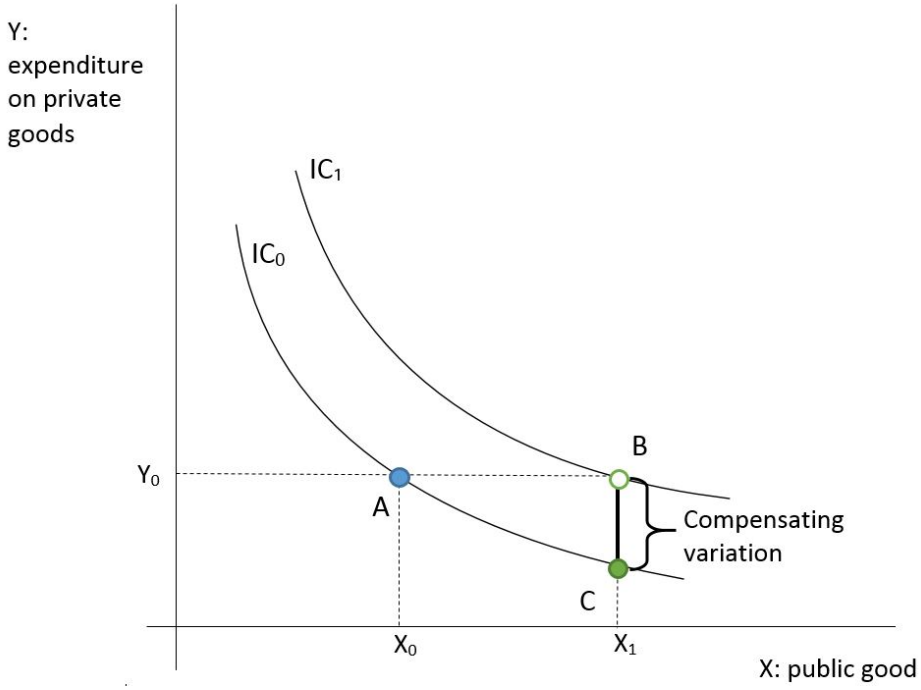


Figure adapted from Bateman *et al.,* 2002 (1)

The link between compensating variation and willingness-to-pay (WTP) is illustrated in Figure 2.2 since the amount BC can be viewed as the amount that the consumer is willing to pay (i.e. expenditure which could otherwise have been used to consume private goods) in order to increase the provision of the public good to X_1.

2.2.3 *Equivalent variation*

Equivalent variation can be demonstrated using the same individual's preferences as the compensating variation example above. The same indifference curves as Figure 2.2 are depicted in Figure 2.3 and, as before, the individual starts at point A spending Y_0 on private goods whilst enjoying X_0 of a public good. In a scenario in which a policy could be implemented that would result in a gain for the individual with respect to his/her consumption of the public good, equivalent variation represents the income that would need to be given to the individual so that his/her expenditure on private goods would be equally preferred to the increase in the public good from X_0 to X_1. As in Figure 2.2, an increase in public good provision to X_1 would place the individual at point B on indifference curve IC_1 if he/she were to maintain his current expenditure on private goods (Y_0). At the higher level of utility represented by IC_1, the individual could spend Y_1 on private goods for

the pre-policy provision of the public good, X_0 (point D). Since points B and D are both on IC_1, the amount of expenditure DA on private goods is equally preferred to the increase in public good provision of the amount AB. Equivalent variation is, therefore, depicted by DA, which is to say the amount of income that would need to be given to the individual in order to raise his/her original utility to an equivalent level to that which would be obtained following the policy implementation.

Figure 2.3 Equivalent variation

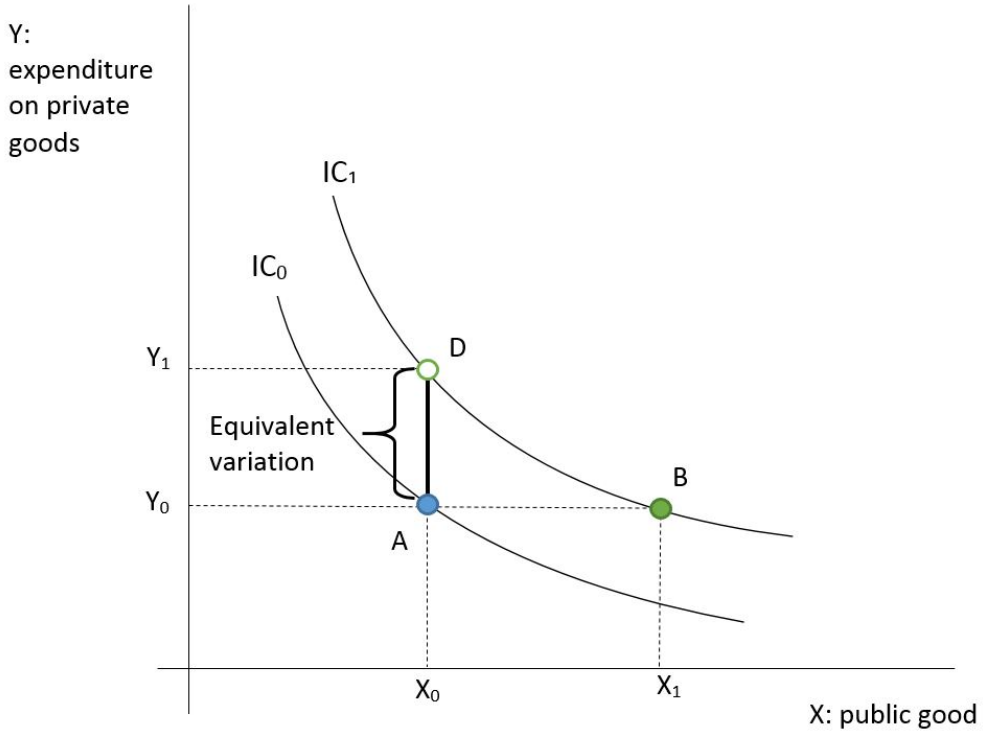


Figure adapted from Bateman *et al.,* 2002 (1)

Equivalent variation can be linked to the notion of willingness-to-accept (WTA), since DA is the amount the consumer is willing to accept (i.e. extra expenditure for the consumption of private goods) in order to forego the increase in the provision of the public good to X_1 and retain the same increased level of utility.

In situations where markets are available for goods, market forces achieve equilibrium between the supply and demand for goods via the mechanism of prices. This theory was first set out by Adam Smith in 1776 in his book *An Inquiry into the Nature and Causes of the Wealth of Nations,* and is a central tenet to the idea that market forces will achieve an efficient allocation of resources, i.e. neither under nor over-provision of goods (55). In a perfectly competitive market, neither producers nor consumers of goods believe that their action will influence the price of goods (63). Under the *First Theorem of Welfare Economics,*

market equilibrium will be Pareto optimal providing that all goods which constitute to individual utility exist within a competitive market (for detailed proof of this, please refer to Gravelle & Rees., 2004, Chapter 13 (63)). However, where market forces are weak, or no market exists, eliciting measures of WTP or WTA can be used to value non-market goods (1). For example, if a new policy is introduced which will generate gains for some members of society and losses for others, the notions of compensating and equivalent variation can be applied to value the welfare changes resulting from the policy. Illustrating this using compensating variation (see Figure 2.2), individuals who would gain from the policy can be asked their WTP for the policy to be implemented, taking into account their expected increase in utility level as a result of policy implementation. Whereas individuals who would experience a loss in utility from the policy could be asked their WTA the policy implementation, considering the compensation required to maintain their original utility level.

2.3 The rise of extra-welfarism

Welfare economic theory has its critics when it is applied to evaluating resource allocation related to healthcare. Much of the criticism lies with the assumptions required for welfare economic theory to hold (see the four tenets described in section 2.2). For example, the tenet of individual sovereignty has been rejected on the argument that individuals are not necessarily the best judge of their own welfare with regards to healthcare due to insufficient information (71).

Perhaps most prominently, however, is a rejection of the fourth tenet, welfarism. An argument against welfarism follows that measuring social welfare on the basis of individual utility is an overly narrow approach to considering social welfare (67, 72). Rather, criticism of the welfarist approach in economics was noted as far back as 1979 by economist and philosopher, Amartya Sen (73), chiefly arguing that it is overly narrow in its consideration of only welfare related to the consumption of goods and services, excluding non-utility information. In agreement with Sen, Tony Culyer (74) proposed that such non-utility aspects are wholly relevant to judgements over healthcare resources, for instance, "*non-goods characteristics of individuals (like whether they are happy, out of pain, etc.)"* (p.36). Due to this extension to the welfarism axiom, the framework proposed by Culyer has been termed *extra-welfarist;* following this new framework, within the healthcare sector, quantifiable measures of health benefit should be considered most important (74, 75).

However, in the context of health and healthcare, the limited utility criticism can be countered on the basis that demand for health has been presented as a derived demand via its impact on individual utility (i.e. a rational individual will gain greater utility from experiencing good health compared to poor health) (76). Following this argument, utility is a function of both consumption (e.g. of healthcare goods and services) and health as a goal itself.

Nevertheless, a question has been raised over the critical normative assumptions required to use Paretian welfare economic theory to measure social welfare when WTP (or WTA) is used to value welfare changes. These consist of the assumptions of constant marginal returns to income across individuals in society (65, 68, 77) and equal weight to welfare changes across society (65). An ethical judgment is required to address these issues and in the absence of established judgement they often remain ignored (65, 77).

Thus, extra-welfarism integrates two new concepts for evaluation within the realm of health services: need (rather than demand) for healthcare and health (as opposed to utility) outcomes. In principle, extra-welfarism need not be viewed only in the confines of health but as a framework that incorporates any relevant information for decisions on societal welfare beyond utility. However, in practice, it has placed an almost exclusive focus on health outcomes (66).

2.4 Market failure in the provision of healthcare and public health programmes

The market for healthcare is not a competitive market. Certain properties of healthcare generate market failure, such as demand for healthcare as a derived demand for health (mentioned previously); it produces externalities (e.g. vaccinated individuals positively impact the rest of society via herd immunity); there is asymmetry of information between providers and healthcare recipients (doctors know far more about disease and treatment than patients); and uncertainty exists regarding the need for healthcare and its effectiveness (66).

Furthermore, market failure is also present with respect to public health programmes, which may extend beyond healthcare, for instance improvements in transport infrastructure or air quality. Goods such as clean air and cycle lanes are considered public goods because they are non-rivalry, in other words consumption by one individual does not prevent or reduce its consumption others (63). Additionally, public goods are often non-excludable, meaning that

their consumption by any individual cannot be prevented, which can lead to "free-riders" enjoying the benefits of a public health programme without contributing to its provision (63). Public health prevention activities could also be considered public goods since many of such programmes naturally exhibit the characteristics of public goods, i.e. they are non-rivalry and are non-excludable. For example: water fluoridation programmes to prevent tooth decay and taxes on sugar in drinks to reduce sugar consumption as an obesity prevention measure. The consumption of neither fluoride in water nor drinks subject to additional tax precludes consumption by others (i.e. they are non-rivalry) and no individual can be prevented from accessing either fluoridated water or tax-imposed sugary drinks (i.e. they are non-exludable).

The market failure present in healthcare and public health goods markets justifies government intervention for their provision. Left to market forces, market failure would lead to an inefficient allocation of healthcare and public health service resources (63). Therefore, governments can take over the role of provision to attempt to improve the efficiency of production of healthcare and public health goods (63). However, given finite resources available to governments, decisions are required on the quantity and diversity of healthcare and public health goods to provide. Economic evaluation can assist in making these decisions.

2.5 The implication of welfarism and extra-welfarism for economic evaluation

Due to market failure in healthcare and public health, the valuation of goods in this field cannot be left to market pricing. Welfare economics laid the foundation for economic evaluation via its introduction of WTP to value welfare change in circumstances where market prices cannot be used. The Paretian approach to economic evaluation, i.e. via CBA, is therefore grounded in welfarism in which measures of benefit derived from the provision of healthcare and public health goods can be elicited from society via methods such as CV (see Chapter Three for further discussion of CV). Even in instances of preventive action in which an individual may not feel he/she derives any personal benefit (e.g. a sugar tax for individuals who do not routinely consume sugar-containing beverages), altruistic behaviour towards prevention can be observed from those willing to pay for such goods (78).

Under the CBA framework, all costs and benefits are valued and compared in order to identify potential Pareto improvements in the provision of services. This approach is concerned with efficiency; do the benefits to society outweigh the costs? Distributional

concerns, such as whether a potential Pareto improvement is equitable, are not addressed by CBA; such concerns require social value judgements which do not relate to efficiency. This framework for economic evaluation is typically associated with addressing questions of overall allocative efficiency, i.e. whether something is worth doing, irrespective of any budget constraint. Allocative efficiency concerns the production and allocation of resources in order to produce optimal levels of output, distributed in a way to meet consumers' values (55, 66).

Supporters of CBA advocate the method for its theoretical foundations in welfare economics and its ability to determine whether an action is warranted on the basis of societal preferences (1, 79). However, CBA is used far less extensively for the evaluation of health-related interventions, particularly since the introduction of CEA in health economics. The measurement and valuation of all societal costs and benefits in order to identify potential Pareto improvements is not simple. Therefore, partial CBAs, in which some aspects of costs or benefits are not valued, may still provide information relevant to decision-making in the field of healthcare (80). CEA, in which costs are valued but benefits are not, can be viewed as a partial CBA (80). Whilst this constrained version of CBA may be unable to address issues of overall allocative efficiency, it could still contribute to examining production allocative efficiency and technical efficiency (80). Under production allocative efficiency, inputs are valued but outputs remain in physical units (80) whilst technical efficiency is concerned with organising the production of goods so that inputs are minimised for the production of a given output (66). Healthcare is typically provided within the constraint of finite resources, particularly in England where government funding for healthcare is constrained by a fixed budget. Therefore, in the context of aiming to improve technical efficiency, CEA may be sufficient (77).

Following the development of extra-welfarism and the argument that welfare changes resulting from healthcare should be evaluated with respect to changes in health (rather than overall utility) (74), a new form of economic evaluation gained precedence for the evaluation of healthcare goods (and, later, also public health programmes). This framework was CUA and is considered to be an extension of CEA, however, it can also be considered a "non-welfarist", simplified form of CBA since both its inputs and outputs are valued (80). Whilst the outputs of a CUA are not valued commensurate to costs, as is the case in CBA, attaching a relevant monetary value to the output of a CUA allows for a direct comparison between

costs and benefits. Thus, CUA could theoretically assess overall allocative efficiency to judge whether an action is worthwhile, irrespective of a budget constraint, provided that a monetary value estimated from individuals' willingness to pay can be identified for the CUA output. However, in practice, CUA is typically used to examine technical efficiency (81) via the objective of maximising health from available resources (82).

CUA most commonly values health benefits using a tool such as the QALY (9). A QALY is a composite measure of mortality and morbidity, combining estimated length of life with an estimate of health-related quality of life (the QALY is discussed further in Chapter Three). This measure maximises a notion of health rather than utility, in keeping with the extra-welfarist framework for the allocation of healthcare resources.

Chapter One introduced alternative tools that may play a role in evaluating public health interventions, which have not traditionally been used to evaluate healthcare technologies, such as SROI and ROI and prioritisation tools such as MCDA and PBMA (55). These approaches to evaluation have been introduced as practical tools to aid decisions regarding complex interventions, rather than by virtue of any theoretical superiority, or even equivalence, to economic evaluation methods conventionally applied in HTA.

These tools are yet to be broadly discussed with respect to the welfarist/extra-welfarist debate, however, SROI is said to have its roots in CBA (83, p.128). Nicholls (83) outlines similarities between CBA and SROI, such as their objective of valuing change resulting from action related to wellbeing and their use of money as a proxy for the value of change. However, the "rigour" of SROI analyses need not be as substantial as that of CBA, provided that the analysis is accountable and *"good enough for the decision it is being used to inform*" (83, p.130). Furthermore, the use of WTP values for non-market goods is recommended for SROI (84), which would place it within the welfarist camp (17, 55). Nevertheless, whether elements of SROI are considered to align with the welfarist paradigm, SROI remains described as less theoretical than CBA (33); its development drew on various schools of thought, including sustainable accounting and financial accounting, in addition to CBA (83).

The following chapter (Chapter Three) will formally introduce each of the evaluative approaches relevant to this doctoral project.

Chapter 3. Economic evaluation methodology

Chapter Two provided an outline of the economic theory underpinning methods of economic evaluation. The grounding in welfarism and extra-welfarism of contemporary economic evaluation was discussed. In addition, the contrast between CBA and CEA/CUA in relation to how each addresses questions of efficiency was outlined. Whilst distinct from economic evaluation methodologies, priority-setting tools are also potentially relevant for assisting in public health decision-making. This chapter introduces each of the economic evaluation and priority-setting methods discussed in this thesis in greater detail.

The following methods are described: CEA and CUA, CCA, CBA (including an outline of methods to value benefit, which are relevant to the empirical work reported later in Chapter Seven), SROI, PBMA, and MDCA.

3.1 Cost-effectiveness analysis and cost-utility analysis

CEA is a form of economic evaluation which uses non-monetary measures of benefit in the form of natural units, for example: life years gained for a treatment for cardiovascular disease, kilograms lost for an obesity-reduction intervention, or units of alcohol consumed for a minimum alcohol pricing intervention. A variation of CEA that has gained popularity in recent years is CUA. CUA is recommended as the primary method of analysis for healthcare interventions by several institutions such as NICE in England (10) and by key commentators internationally (85, 86). CUA measures benefit using a measure of weighted health gain, which comprises both a mortality and morbidity component. The term “utility” in CUA refers to individual or societal preferences for health outcomes (9); within the CUA framework, utility differs from the economic sense of the term introduced in Chapter Two which underpins welfare economic theory.

Outcomes for a CEA come directly from clinical or epidemiological evidence, whilst CUAs use a valued measure of health improvement. The health improvement measure incorporates a health-related quality of life score for the state of health achieved by an intervention which is then used to weight the length of life that is affected by the intervention’s outcome (9). The most commonly used generic measure for CUA is the QALY (87), however, alternative measures have also been developed such as the Disability Adjusted Life Year (DALY) or the Healthy Years Equivalent (HYE) (88, 89). However, claims made by Mehrez & Gafni (88) that the HYE is theoretically superior to the QALY have been widely criticised (90) and its

superiority to the QALY challenged (91). As such, the HYE has not been considered feasible in practice (92) although, the DALY is in use as the preferred outcome measure for CUA by the World Health Organisation (93).

There are several tools available to measure the health-related quality of life component of a QALY, such as the EQ-5D (94) (favoured by NICE (10)), the SF-6D (95), or the Health Utilities Index (96). These tools measure generic quality of life, which are not specific to a particular disease area. As such, generic QALYs are generated. However, in certain cases a generic measure may be insufficient to capture health improvement and a condition specific measure of health-related quality of life may be more relevant (97). Condition specific measures have been developed to generate QALYs for several health conditions, such as oncology (98), epilepsy (99), and multiple sclerosis (100).

The use of QALYs in CUA allows the cost-effectiveness of interventions for varying health conditions to be compared, whereas CEA restricts comparisons to those between similar interventions in which the same outcome measure can be obtained. It is this ability to make broad comparisons that has promoted CUA to become the preferred method of analysis for HTA in several countries (101).

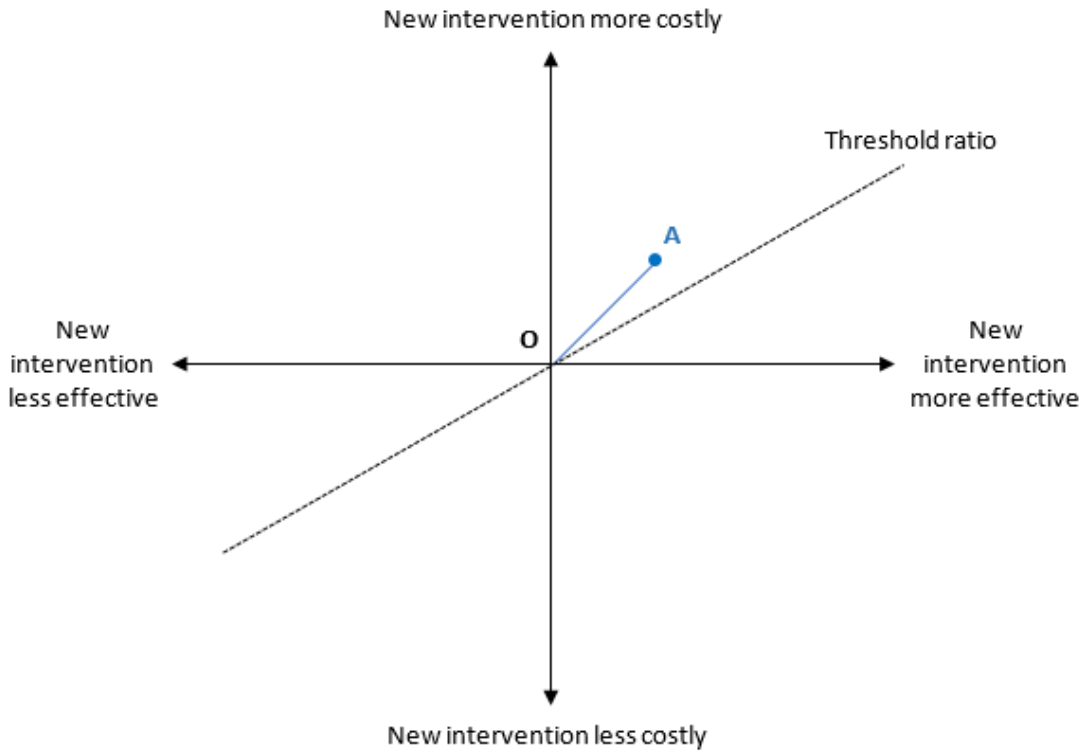
3.1.1 *Decision rules in CEA and CUA*

Both CEA and CUA combine health outcomes with the costs of action to assess the worthiness of some intervention. Where an intervention generates additional health benefit for an additional cost (compared to some alternative course of action), the results of both methods of analysis are typically reported in the same manner: an incremental cost-effectiveness ratio (ICER). An ICER represents the additional cost required to achieve a one-unit improvement in health when comparing the outcomes of one course of action against an alternative. An ICER is calculated once dominant alternatives (i.e. those which are both less costly and more effective than the alternative courses of action) and dominated (i.e. those that are both more costly and less effective) have been removed from the available courses of action for comparison. The ICER calculation for two alternatives is illustrated in Equation 3.1, where C_B and C_A represent the total costs for alternatives B and A, respectively and where E_B and E_A represent the total effects associated with each alternative. In the ICER calculation $C_B > C_A$ because the comparison is always made for the most expensive alternative compared to the next most expensive.

$$ICER = \frac{\Delta C}{\Delta E} = \frac{C_B - C_A}{E_B - E_A} \quad (3.1)$$

The results from a CUA or CEA can be displayed graphically on a cost-effectiveness plane (Figure 3.1) (9) in which incremental effectiveness is displayed along the x axis and incremental cost along the y axis. The point estimate of incremental cost and effectiveness of the intervention is represented by point A and the origin of the plane represents the comparator. The ICER of the new intervention compared to the comparator is, therefore, the slope of the line OA.

Figure 3.1 Representation of an ICER on a cost-effectiveness plane



A decision rule is then applied to interpret whether the findings justify using resources in a particular way. Decisions are based on subjective judgements of the value of the outcome, or alternatively, a relative comparison can be made between ICERs of substitutable interventions which have used the same outcome measure during the evaluation process. In CUA, relative comparisons can also be used, however, relative comparison in both CEA and CUA relies on the assumption that there are constant returns to scale with respect to cost-effectiveness (102). In other words, an assumption is made that scaling provision of the intervention to meet societal need does not reduce cost-effectiveness and that the ICER remains constant.

Deciding whether an action represents an acceptable use of resources within a constrained budget requires a threshold value of a cost per outcome. When addressing questions of technical efficiency (introduced in Chapter Two), the threshold value should represent some notion of opportunity cost, given that any resources used for one programme cannot be used for another (103). The cost-effectiveness threshold is depicted as the dotted line crossing through the origin of the cost-effectiveness plane in Figure 3.1. The slope of this line represents the threshold ratio to which acceptable cost-effectiveness is compared. In England, an acceptable cost for the gain of one QALY is typically considered to be between £20,000 and £30,000 for healthcare interventions (10), however, alternative values have been deliberated in extenuating circumstances, for instance: life-extending treatment for terminal diseases at the end of life (up to £50,000 per QALY) or for very rare diseases (between £100,000 and £300,000 per QALY) (104). The region of cost-effectiveness on a cost-effectiveness plane is the area below the threshold line. In Figure 3.1, point A is above the threshold line and, thus, outside of the area of cost-effectiveness.

Whilst CUA has gained general acceptance for the evaluation of healthcare interventions, its use for public health appraisal is not accepted so consensually. There has been a call for greater consideration of CBA or for the consideration of alternative evaluative approaches such as CCA (82) and SROI (17, 32). Additionally, academics seeking to capture a broader measure of well-being, as opposed to health, have considered the capability approach as a potential alternative to traditional CUA using QALYs as the outcome measure (25, 105). The capabilities approach was theorised by Amartya Sen in the 1980s (106) and academics have since attempted to operationalise Sen's ideas into tools able to capture and measure capability, i.e. what an individual can do rather than what he/she actually does. Tools such as the ICECAP-A (ICEpop CAPability measure for Adults) (107) and ICECAP-O (ICEpop CAPability measure for Older people) (108) or the OCAP-18 (Operationalising the CAPabilities approach using a questionnaire of 18 capabilities-specific items) (109) and OxCAP-MH (an instrument specifically designed to operationalise the capabilities approach in mental health) (110) have been developed to capture capability well-being. However, developing these tools for use in economic evaluations remains to be established (109, 111). Due to the limited practical use of the capabilities approach within the economic evaluation paradigm currently, it is not discussed in great detail in this chapter or included as an alternative evaluative approach in the empirical work contained in this thesis.

3.2 Cost-consequence analysis

CCA is a method of setting out all the relevant costs and outcomes of an intervention in a clear format. Unlike other methods of economic evaluation, CCA does not aggregate the costs and outcomes into a single value representing cost-effectiveness. For this reason, there is no real decision rule in CCA. The purpose of the method is to provide a broad array of information to decision-makers on the impacts of an intervention (impacts can include any effect of relevance, both health-related and non-health-related). However, subjective judgement over the relative importance of the outcomes and costs presented is required on behalf of the decision-maker due to the lack of aggregation.

CCA has been termed a "balance sheet" approach by some commentators (112) due to the disaggregated nature of the method in which costs and benefits are presented. According to McIntosh (112), the "balance sheet" approach is a form of CBA since the same guidance should be followed with respect to the identification of all costs and benefits incurred by an intervention. However, CCA differs in that the valuation of benefits stage of a CBA is not conducted.

CCA can be a useful tool to comprehensively introduce relevant outcomes and costs and, therefore, be approachable to a range of different decision-makers who may have diverse perspectives (113). However, decisions made solely using CCA can lack transparency and could be open to criticism for "cherry-picking" results to fit an agenda (114) that may not be in society's best interests (115).

3.3 Cost-benefit analysis

Chapter Two introduced CBA as emerging from welfare economic theory, for which the focal concern is that of societal welfare. A well-constructed CBA, therefore, considers all costs and benefits, on whomsoever they fall, associated with a course of action (116). In this method of economic evaluation, both benefits and costs are measured in commensurate units, typically monetary units, to enable direct comparison. This enables a simple decision rule such that an action can be considered worthwhile if benefits outweigh costs. The decision-rule follows from the Hicks-Kaldor criterion which states that societal gainers could theoretically compensate losers and society would remain better off; thus, so long as aggregate benefits are greater than aggregate costs, welfare should be improved on a

societal level. Equation 3.2 displays the calculation of net societal benefit (NSB) for programme i.

$$\boldsymbol{NSB_i = \sum_{t=1}^{n} \frac{b_i(t) - c_i(t)}{(1+r)^{t-1}}} \quad \textbf{(3.2)}$$

When a programme is evaluated for longer than one year, the present value of the NSB is calculated where $b_i(t)$ is the discounted monetary benefit for programme i in year t, $c_i(t)$ is the monetary costs of programme i in year t, $1/(1+r)^{t-1}$ is the discount factor for interest rate r, and n is the time-horizon of the programme. The formula for NSB is simplified if the time-horizon of a programme is 12 months or less since no discounting is required. The NSB calculation for programme i, thus, becomes Equation 3.3.

$$\boldsymbol{NSB_i = \sum b_i - c_i} \quad \textbf{(3.3)}$$

CBA aims to identify whether NSB > 0, i.e. the net benefit to society is greater than the costs. The outcome of a CBA can also be presented as a benefit-cost ratio (BCR) in which total benefits are divided by total costs. Equation 3.4 displays the CBA decision rule using a BCR.

$$\boldsymbol{\frac{b_i}{c_i} > 1} \quad \textbf{(3.4)}$$

Both the NSB and BCR can be used to appraise whether a programme is valuable to society, however, Keating & Keating (117) stress caution when using the BCR since maximising the difference between cost and benefits (NSB) is not necessarily equitable to maximising the BCR (117). It is for this reason they recommend using BCRs to eliminate projects that would never be worthwhile implementing (from an allocative efficiency standpoint) (i.e. BCR<1). However, when deciding between multiple projects, the NSB would be the better indicator of maximising benefit.

3.3.1 *Valuing benefit in CBA*

The use of monetised benefits in CBA necessitates monetary values to be placed on outcomes for which no market value exists. In health economic CBAs this includes health outcomes. At the methodology's initial foray into health economics in the 1960s and 1970s, the human capital approach was used to capture benefit (118). This approach considers health impacts in terms of reduced healthcare costs and increased productivity costs (or prevention of reduced productivity). However, it has many limitations such as discriminating

against those outside the workforce (81), ignoring quality of life (119), and a lack of theoretical grounding in welfare economics (120).

Another method that has been used in evaluations of healthcare interventions is a financial value of a QALY. Similar to the human capital approach, monetised QALYs are not grounded in welfare economic theory, nonetheless, evaluations using this method are considered CBAs because costs and benefits are measured in commensurate units. However, it could be argued that this approach is not valuing all benefits associated with an action, particularly a public health intervention for which benefits may extend beyond those captured by the QALY measure.

The welfare economic theoretical framework implies that benefits should be measured using a value of WTP or WTA, derived from the notions of compensating variation and equivalent variation outlined in Chapter Two. WTP and WTA can be measured using two approaches: revealed preferences or stated preferences. The revealed preference approach elicits WTP or WTA via observed behaviour such as choosing occupations associated with a high risk of injury for a stated salary (119). However, it is often difficult to ascertain an appropriate value using revealed preference techniques for non-marketed goods. In many developed countries, such as the UK, healthcare is not provided in the market place, therefore, observing individual preferences for healthcare services using purchase observation is difficult.

An alternative method for ascertaining WTP or WTA uses the stated preference approach. This technique elicits values through the use of hypothetical scenarios in which individuals state their WTP for some defined benefit-inducing action (or their WTA compensation to forego the benefit generated from some action). Theoretically, the value of WTP or WTA should be near equivalent in the absence of income effects, however, studies examining this have found this may not necessarily be the case (121). A variety of explanations for the disparity between WTP and WTA have been proposed from degrees of substitutability between non-market and market goods (122) to the impact of “loss aversion” (i.e. weighing losses greater than gains) (123). Consequently, it is typically recommended to elicit WTP over WTA where possible (124).

3.3.2 *Contingent valuation*

Stated preference valuations are generally established using survey methods such as the contingent valuation method (1). Environmental and transport economics initiated the use of CV to measure WTP (125-127). CV methodology creates hypothetical scenarios within which individuals are requested to provide the maximum monetary value that they would be "willing-to-pay" for the action described in the scenario, contingent on a suitable market existing. The amount an individual states that he/she is willing to pay indicates strength of preference for the good being valued.

Although initially introduced via the environmental and transport economic fields, CV has been used to value healthcare since the 1970s (128) and has gained popularity for valuing a range of healthcare goods and services such as healthcare management, medical treatments, surgery, and pharmaceuticals (129). However, several methodological issues surround the CV method regarding its hypothetical survey design, the vehicle of (hypothetical) payment employed in the surveys, the various methods of eliciting WTP within the CV framework, and whether CV elicits true economic values from respondents. These criticisms of CV are discussed briefly below, however, for full discussions of these issues please refer to Bateman *et al.,* 2002 (1), Mitchell *et al.,* 2013 (124), and McIntosh *et al.,* 2010 (130).

Hypothetical design

Critics of the CV method have argued that the hypothetical nature of CV surveys is unduly removed from real-world behaviour to elicit reliable WTP preferences. Testing this argument has, however, produced mixed results with some studies eliciting different WTP outcomes using revealed and hypothetical approaches e.g. Carson *et al.,* 1996 (131), whilst more recent studies have indicated that valuations elicited via either approach may be similar (132) particularly if responses are adjusted for certainty (133).

Elicitation method

The CV method is *"not an unique methodology"* (134, p.103) due to the numerous ways in which WTP can be elicited, each having unique methodological advantages and disadvantages. Originally, either a bidding game or open-ended question format was employed (135). Bidding-games follow an auction style arrangement in which several monetary amounts are sequentially presented to respondents, either increasing or

decreasing in value, based on the response to the previous value. This would continue until the respondent's maximum WTP was established, or a predetermined range of values had been exhausted. On the other hand, open-ended questions simply ask a respondent for their maximum WTP value.

Criticism of the open-ended format lies with the perception that outright questions on WTP with no reference can be difficult for respondents to answer (1), thus leading to large proportions of non-responses, zero valuations, and outliers (124). The bidding-game format can equally generate biased responses due to the potential influence on final valuations caused by the choice of starting bid value (136).

Alternative elicitation techniques were developed in response to the issues with the bidding-game and open-ended formats. The dichotomous choice method, also referred to as "take-it-or-leave-it" because a single value is offered and respondents either agree to pay that value or refuse, gained popularity as it most closely represents how individuals are accustomed to making payment decisions. The method is, however, insensitive to an individual's maximum WTP. Hanemann (137) suggested a double-bounded variation which aimed to improve statistical efficiency by offering a second payment value in response to the respondent's first response. Whilst dichotomous choice approaches are less burdensome for respondents compared to a bidding-game exercise, the limited information provided on the respondent's actual maximum WTP necessitates large sample sizes in order to obtain a precise estimate of average, maximum WTP (1).

A further elicitation approach, the payment card method, was developed by Mitchell & Carson (138) as an alternative to open-ended and bidding-game formats and quickly gained popularity amongst CV researchers to elicit health benefits (139, 140). The payment card approach encourages respondents to choose their maximum WTP from a list of possible values. The approach is able to provide more sensitive detail on a respondent's maximum WTP than the bidding-game method, whilst being less burdensome for both respondents and researchers. Nevertheless, this approach is also subject to biases related to the range of values presented (1). In order to minimise range bias from the payment card format, an adaptation of the approach was developed, the random card sort, in which payment cards are presented in a random order to respondents rather than as a list. Mixed reports have been observed using this method (139), however, studies have shown it to pass theoretical

validity tests (141) and it has been argued to be preferable compared with the standard payment card format (142).

Nevertheless, there remains debate over the most appropriate method of elicitation to use in CV studies. Research on the variation in WTP responses using different elicitation methods revealed that the method chosen can have a significant impact on the resultant WTP (143, 144), therefore, careful consideration of which method to use must be given.

Payment vehicle

The payment vehicle refers to the way in which the hypothetical payment would be taken; this can be framed in numerous ways such as additional taxation, increased insurance premiums, increases to utility bills, or a one-off (voluntary) payment. Research has demonstrated that the choice of payment vehicle can impact WTP values (145-147).

Whether a payment is framed as either a voluntary or mandatory (collective) contribution has been argued to have an impact on WTP values (148). Voluntary payments may inflate WTP responses if respondents wish to reflect an interest in having the option to purchase the good/service at a later date. Conversely, collective payments (e.g. mandatory tax payment for all payers) may extract respondents' true WTP in the knowledge that their response is consequential (149).

The most appropriate vehicle will be dependent on the context of the CV question, considering both the sample population and the item under scrutiny (150). For example, a healthcare good being valued by a general public sample who are accustomed to receiving healthcare, which has been funded via taxes, may be more familiar with a contribution towards taxation, whereas individuals who are accustomed to an insurance-based healthcare system may respond better to a payment in the form of an increased insurance premium.

Elicitation of true economic value

A key premise of the use of CV to measure benefit for use in a CBA is that the WTP value derived reflects the value of the change in welfare to the respondent. This premise is based on the economic notions of compensating and equivalent variation discussed in Chapter Two (see Sections 2.2.2 and 2.2.3). When valuing public goods, however, this premise has been called into question, particularly in the arena of valuing environmental goods for which an

individual may receive no private consumption benefit (for example, preservation of a national park that an individual never plans to visit). In such examples, the value elicited from a CV reflects existence value, in which a good represents value to an individual merely due to its existence, regardless of any intention or ability to use it (151, 152).

Kahneman *et al.* (78) presented a psychological study of WTP and argue that respondents follow a "contribution model" in their decision on the amount they are WTP, rather than a "purchase model" assumed by the economic theory underpinning CV which represents the economic value at which an individual is truly indifferent between a state of the world in which a good is provided or not provided (78, pp.310-11). The upshot of this argument is that the outcome of a CV measuring the value of a public good does not truly reflect the economic value of the good to society but reflects an expression of attitude that the good should be provided. This conclusion is drawn from a comparison with charitable contributions in which individuals are willing to contribute some amount to support a cause, with the expectation that some action will occur regardless of their WTP (78, p.311).

The aforementioned study by Kahneman *et al.* (78) specifically discussed public environmental goods, rather than public health goods, and by the authors' admission was insufficient to permit inferences about the CV method (p.314). However, their study suggests that WTP values elicited for public goods should be considered carefully and care should be taken to demonstrate that derived WTP represents economic value rather than just attitudes. Nevertheless, expressions of attitude towards public programmes obtained from a CV study, whether revealing true economic value or just some form of contribution value in support of action, may still be important in order to guide the allocation of public resources. Particularly in the arena of public goods where decisions are made for the wellbeing of society, aligning resource allocation with public attitudes may be relevant. To investigate further the support of the public, the mechanism of WTP decisions can be explored by examining the characterisitcs of the public which predict positive WTP, or if WTP is assumed to represent attitude rather than economic value, which groups of society value the provision of the public good in question.

3.3.3 *Discrete choice experiments*

An alternative method for eliciting WTP is a discrete choice experiment (DCE). DCEs have gained popularity amongst health economists since the 1990s (153-155). DCEs have a

foundation in choice behaviour theory, specifically random utility theory (see Louviere *et al.,* 2010 (156) for more detail on random utility theory). DCE surveys present hypothetical scenarios which provide information on a selection of attributes of the item under valuation. Each scenario offers two or more options in which the levels of the attributes are varied; each option will consist of identical attributes, however, the level associated with each attribute will be varied. In order to ascertain the value of WTP from DCEs, one of the attributes must be price or an appropriate approximation for price. From respondents' responses to several choice sets, researchers can indirectly estimate values of WTP via statistical modelling of responses (157, 158). For further detail on DCEs please see Ryan *et al.,* 2007 (153) and Johnson *et al.,* 2013 (159).

DCEs have been cited with the ability to overcome some of the biases attributed to the CV method (160) and are potentially more informative than CV surveys because relevant information is gained on respondents' preferences for various attributes of a good or service, rather than a preference for a composite good. Additionally, the indirect valuation may be preferable in healthcare contexts given a general aversion to paying for healthcare directly (153).

However, there is a trade-off between achieving greater information and complexity. DCEs can be cognitively burdensome for respondents; it has been shown that informational fatigue can occur during DCEs, which can lead to irrational responses (160). As such, CV remains a valid tool for measuring WTP, particularly where an overall value for intervention is sufficient and information on varied attributes is not necessary (153).

3.4 Social return on investment

SROI has been described as *"a framework for measuring and accounting for [a] broader concept of value"* (84, p. 8). This approach to economic evaluation has generated traction as a holistic method of assessing value for money (32). The scope of SROI is broader than the other forms of economic evaluation reported in this chapter, with the exception of CBA; an SROI can include costs and effects from improvements in wellbeing to social and environmental impacts (84). The broad scope of outcomes that can be included in an SROI makes the approach particularly relevant to public policy.

The decision rule used in SROI is similar to that used in CBA because all costs and outcomes are monetised. The results of an SROI are presented as a ratio of investment to impact (see

Equation 3.5). A ratio value greater than one indicates an activity that is socially beneficial for the given level of investment.

$$SROI = \frac{Net\ present\ value\ of\ impact}{Net\ present\ value\ of\ investment} \quad \textbf{(3.5)}$$

The SROI ratio can be closely compared with the BCR (Equation 3.4), however, there is a significant difference between the two calculations with regard to the contents of the numerator and denominator. In a BCR, the denominator includes cost-savings (e.g. via reductions in healthcare service use) in addition to costs associated with implementing the programme. In the SROI ratio, however, the investment (i.e. the denominator) only includes the cost of implementing the programme. Cost-savings are included in the value of the programme's impact (i.e. the numerator), alongside any other monetarised benefits that the SROI analysis may have identified. In the BCR, however, the numerator only includes monetised benefits.

SROI encourages the involvement of stakeholders during the evaluative process to aid the identification of relevant impacts and as a source of information on key aspects of impacts necessary for evaluation, such as the duration of an outcome and the likely outcome in the absence of intervention (32, 161). Stakeholders can also be used to identify relevant sources of value for intervention impacts. Whilst the use of stakeholders is not unique to SROI, within *The Seven Principles of Social Value* (83), stakeholder engagement is implicit in the SROI analysis process. Researchers conducting CUAs and CBAs may engage with stakeholders for certain elements of analysis, e.g. conceptualising models (162) or forming attributes for a DCE study (163), however, stakeholder engagement is not implicit in these evaluations, rather it has emerged as good-practice.

Value is attributed to outcomes in SROI using financial proxies. For some outcomes, market prices for a suitable proxy may be available, for example, cost-savings due to reduced visits to a General Practitioner (GP) can use the average cost of attendance to a GP as a proxy value. For non-market outcomes, revealed and stated preference techniques can be used to elicit values in a similar manner to CBAs (84). Monetised QALYs have also been suggested as an appropriate measure of health outcome by PHE (164).

An SROI evaluation consists of a six-stage process beginning with identifying the scope of the evaluation and identifying relevant stakeholders and closing with the final reporting of results. The six stages are as follows:

1. Establishing scope and identifying key stakeholders
2. Mapping outcomes
3. Evidencing outcomes and applying values
4. Establishing impact
5. Calculating the SROI
6. Reporting and embedding results

The six stages of an SROI represent a process that any economic evaluation researcher would find familiar. The key difference between SROI and other modes of economic evaluation is in the rigor necessary for the analysis. Whilst a CUA or CBA would typically obtain estimates of effect from an RCT or, in the case of a public health initiative, perhaps a natural experiment (22). However, impact can be estimated using less scientifically regarded sources for an SROI, such as asking beneficiaries what the counterfactual would have been had they not been recipients of an initiative (83).

An in-depth exploration of each of the six steps of an SROI analysis is beyond the scope of this chapter, however, Nicholls *et al.* (84) provide an extensive guide explaining the practical requirements for each of these stages alongside an in-depth explanation behind the rationale of each step. Nonetheless, the fourth stage, in which the methods of data analysis are initiated, will be outlined briefly here.

In order to avoid over-estimation in SROI analyses, four factors are considered in the estimation of impact: (i) deadweight (DW), (ii) displacement (DI), (iii) attribution (A), and (iv) drop-off (DO). Firstly, deadweight is calculated to represent the counterfactual of the intervention, i.e. the expected outcome in the absence of any action. Secondly, displacement refers to the movement of outcomes from one area to another as a result of the action, e.g. moving crime from one area to another as a result of an intervention at a specific location (33). Thirdly, the level of outcome that can be reasonably attributed to the intervention is assessed. Finally, drop-off accounts for the reduction of intervention impacts over time (84, 165). The total impact of an action is, therefore, a function of each of the above four factors, as depicted in Equation 3.6.

$$\textit{Impact value} = (\textit{Fiscal impact} + \textit{social impact}) - DW - DI - A - DO \qquad \textbf{(3.6)}$$

3.5 Priority-setting tools

An alternative approach to resource allocation decisions, particularly within the confines of a budget, is the use of priority-setting tools. Whilst economic evaluations can inform decision-makers about the relative efficiency of competing programmes, priority-setting tools can be used on a larger scale. Tools such as programme budgeting and marginal analysis (PBMA) and multi-criteria decision analysis (MCDA) can examine a whole host of programmes to which a decision-maker may wish to allocate the resources available within a given budget. Priority-setting tools can be beneficial for decision-makers by providing a rational framework in which decisions to invest (and disinvest) can be made with legitimacy (166). PBMA and MCDA tools have received particular interest within the research community since the turn of the century (15, 37, 166-168). Both approaches are described below.

3.5.1 *Programme budgeting and marginal analysis*

PBMA is a framework that can assist the allocation of resources in order to meet the needs and priorities of either a local or national agenda. The process can be considered in two parts: *programme budgeting* in which current and past budget allocations are appraised to provide a baseline for future allocations, and *marginal analysis* in which the marginal impacts on costs and outcomes as a result of a proposed investment or disinvestment are examined (39). The concept of opportunity cost plays a prominent role in PBMA since it is necessary to understand the impact on the benefits foregone as a result of any change in resource allocation (167).

Conducting a PBMA is an inclusive and transparent process involving a multi-disciplinary panel consisting of a broad range of individuals, for example, clinicians, local government representatives, third sector stakeholders, etc. A first stage is to divide the budget into broad programmes covering different areas of care (e.g. smoking cessation, treatment for alcohol misuse, sexual health, etc. in the case of an entire public health budget). Information can then be compiled on the resources and costs of all programmes under investigation, which can be combined with the level of activity within each programme (e.g. number of service users) in order to identify the programme budget (166). The multi-disciplinary panel then works collaboratively to propose potential changes to the current arrangement of resources identified during the programme budgeting stage in terms of areas of reduction and areas for expansion. These proposed actions are then scored and ranked based on pre-determined

criteria (38). The scoring and ranking are supported by examination of relevant evidence (which could include economic evaluation evidence, such as CUA, CBA, SROI, etc.) providing information from which each criterion can be considered.

The marginal analysis is then operationalised using the ranked list of actions in order to adjust resource allocation from those areas ranked lowest to those ranked highest. This process should generate a formulation that best meets the desired objectives of the local or national agenda. Alternatively, a less formulaic approach is to compare each option that was shortlisted for expansion against the areas for reduction. An explicit valuing of each option of growth against the shortlisted areas for reduction would need to be undertaken until no further reallocation of resources is estimated to produce further marginal gains.

3.5.2 *Multi-criteria decision analysis*

MCDA is another technique which introduces rationality and transparency into decision-making in order to avoid adhoc processes which may lead to resource allocations which fail to maximise social welfare (15). MCDA can be particularly pertinent for public health decision-making in local government settings since both health economic and non-health economic outcomes can be considered (40). This allows a broad range of outcomes to be compared simultaneously in order to address the varied priorities of public health decision-makers.

MCDA consists of four key steps: (i) identifying alternatives to be appraised, (ii) agreeing criteria with which to appraise alternatives, (iii) examining the alternatives against the criteria to produce criteria scores, and (iv) weighting the criteria scores to reflect the relative importance of each criteria to the overall objective and ultimately produce a holistic assessment of each alternative (40, 169). The criteria considered can be as broad as necessary; some examples to illustrate the breadth of possible criteria are: clinical effectiveness, cost-effectiveness, ease of implementation, burden of disease, equity, impact on familial relationships, crime, environmental damage, certainty of outcomes, and quality of evidence (15, 40, 170). The criteria here demonstrate that MCDA need not necessarily be viewed as a substitute to economic evaluation since evidence of intervention value is often a key criterion; however, MCDA can expand the decision space to include elements beyond evidence on cost-effectiveness to place value alongside supplementary priorities such as distributive justice, social cohesion, and reducing health inequalities.

3.6 Summary

This chapter has outlined a variety of economic evaluation and priority-setting tools available to researchers evaluating public health interventions. CEA, CUA, CBA, and CCA have been used, or at least acknowledged, within the health economic field as alternative tools for the appraisal of healthcare technologies for several decades (9). CUA using QALYs as the outcome measure has become increasingly accepted since the 1990s (82) and has become a favoured evaluative approach for HTA (101) and public health appraisal (10). The complex challenges associated with valuing public health initiatives, both methodologically and contextually (introduced in Chapter One), have generated exploration and consideration of alternative approaches to appraising public health programmes. Therefore, tools such as SROI and formal approaches to prioritisation (e.g. via the use of MCDA or PBMA frameworks) have been introduced to the health economist's toolkit in recent years. The following chapter examines whether, and if so how, the tools discussed in this chapter have been adopted by public health economists over the past decade.

Chapter 4. A systematic review of public health interventions to prevent alcohol misuse

Chapter Four reports an investigation of the existing economic evaluation evidence for public health interventions to prevent alcohol misuse. A version of this chapter has been published (114) which reports on studies published between 2006 and 2016. This chapter, however, also includes the results of an updated search (in March 2019) to account for studies published since 2016.

This chapter addresses the first research question of the thesis outlined in Chapter One. The systematic review reported in this chapter explores the economic evaluation and priority-setting tools evidence currently available for public health interventions. Furthermore, the quality of the available evidence is also examined in relation to the inclusion of methodological elements that have been identified previously as relevant to the appraisal of public health interventions (48). The first section of this chapter outlines the rationale for conducting the review and section 4.2 outlines the review objectives. The methods used to conduct the review are reported in section 4.3 and section 4.4 reports the review findings. The final section discusses the study results with regard to the review objectives and considers the implications for further empirical work.

4.1 Review rationale

The previous chapter outlined methods of economic evaluation available to assess the relative value of public health interventions. As discussed in Chapter One, public health interventions can be complex, interacting with sectors beyond healthcare, and often impact individuals who are not the direct recipients of the intervention. These unique characteristics may necessitate alternative approaches to evaluation compared to those adopted to evaluate healthcare technologies (17, 171).

A previous review of economic evaluations from the public health field, which were published up to 2005 (48), identified the predominance of CEA and CUA studies (63%) and the remaining literature consisted of CCA studies (27%); no CBA studies were identified. Weatherly and colleagues (48) regarded four methodological challenges as being inherent when evaluating public health interventions: (i) attribution of effects, (ii) measuring and valuing outcomes, (iii) identifying intersectoral costs and consequences, and (iv) incorporating equity considerations. The authors concluded that:

> *"The existing empirical literature is very disappointing, offering few insights on how to respond to these challenges. This severely limits the usefulness of economic evaluation in this field."* (48, p.92)

Whilst comprehensive, and inclusive of salient methodological considerations regarding the economic evaluation of public health interventions, the review has limitations. Firstly, only one source of evidence was used to identify literature (the NHS Economic Evaluations Database (NHS EED)). Although the NHS EED is a relevant database for economic evaluations, it does not include evidence of SROI, ROI, or priority-setting methods which Chapter Three identified as potentially relevant for evaluating public health interventions. Secondly, the included literature is limited to the year 2005, the same year in which the academic discussion on evaluating public health interventions was initiated in fervour by Michael Kelly and colleagues (171). Although a review examining guidance for the economic evaluation of public health interventions (17) identified some published literature prior to 2005, the majority of relevant guidance was published post-2005.

The timeline of evidence identified by Edwards and colleagues (17) indicates that many of the methodological issues raised regarding the evaluation of public health interventions may not have been considered by the study authors identified in the review by Weatherly *et al.* (48). Therefore, in order to examine whether the academic arguments posed for a shift from the economic evaluation paradigm set up for healthcare technologies towards more holistic evaluative approaches for public health have been supported in practice, an updated literature review is required.

4.2 Review objectives

There has been significant growth in published health economic evaluations over the past 50 years as the field of health economics has developed; a bibliometric review of health economic publications demonstrates the near-exponential cumulative increase in articles since the 1960s (172). Due to this trend in the growth of economic evaluation publications, it was expected that the number of potentially relevant articles spanning the past decade over the entirety of public health would be beyond the scope of this thesis to examine. As such, the focus of the review covered in this chapter was narrowed to one area within public health: prevention of alcohol misuse.

Prevention of alcohol misuse was considered pertinent due to its importance in English public health priorities (173). Additionally, alcohol misuse is associated with significant negative-externalities to individuals and society beyond those immediately affected and falling on non-health-specific sectors (e.g. via anti-social behaviour (174), traffic accidents (174), familial relations (175) etc.). Consequently, interventions to prevent alcohol misuse typically exhibit characteristics suited to a broader perspective of evaluation. Given the recent interest, discussed in Chapter One, in capturing the value of health and non-health outcomes from public health interventions, prevention of alcohol misuse was thus perceived as a relevant focus for review.

This review aims to explore whether the limited guidance for the economic evaluation of public interventions, discussed in section 4.1 and Chapter One, has led to the adoption of alternative economic evaluation frameworks in lieu of the relative status-quo of CUA favoured for HTAs in England (10). Additionally, it will consider whether Weatherly *et al.'s* (48) conclusions regarding their disappointment in the literature's response to methodological challenges is still pertinent. Hence, this study has two objectives which address the first overarching research question of the thesis (see Chapter One, section 1.3):

1. To identify methods of economic evaluation and priority-setting used to evaluate interventions to prevent or reduce alcohol misuse
2. To examine the quality of studies in relation to the extent at which authors address the public health specific methodological challenges identified by Weatherly *et al.* (48)

4.3 Review methods

A protocol for the review was registered in May 2016 with PROSPERO (176).

4.3.1 *Eligibility criteria*

Economic evaluations were defined for this review as the comparative analysis of alternatives with respect to their associated costs and consequences, including, but not confined to, health consequences. Priority-setting methods were defined as a systematic method of deciding where investments (and disinvestments) should be made to best meet the needs of communities.

Studies were eligible for inclusion if they met the following criteria:

- Economic evaluations or methods of priority-setting with a focus on preventing alcohol misuse or reducing excessive alcohol consumption
- The study population was non-treatment seeking and not currently engaged in treatment for alcohol dependency
- Published in English

Studies were ineligible for inclusion if they met the following conditions:

- Evaluations of pharmacotherapies as these would fall within HTA, rather than public health evaluation, which often uses far more prescriptive methodology, with the dominant method being CUA (11)
- Evaluations of treatments for alcohol dependency, e.g. detoxification or rehabilitation, as these would not be considered preventive. An exception was made for treatments part of a preventive regime, such as screening and brief intervention for non-treatment seeking individuals
- Evaluations of interventions to prevent harm or injury caused as a result of alcohol consumption, such as traffic accidents resulting from drink-driving, unless the primary objective of the study was reducing alcohol consumption
- The study focussed on a narrow population by virtue of a clinical condition e.g. HIV positive individuals, for whom the interventions may be tailored towards improving specific clinical conditions
- Evaluations of interventions to improve general health unless alcohol consumption was the primary focus and alcohol-related outcomes were reported independently

4.3.2 *Search strategy (January 2006-May 2016)*

A literature search was undertaken using NHS EED and Scopus to identify studies published between January 2006 and May 2016. As discussed in section 4.1, NHS EED is a useful resource to identify economic evaluations, however, alternative methods are unlikely to be identified; therefore, an additional search was conducted in Scopus to capture additional SROI and priority-setting studies. The use of multiple databases to identify economic evaluation studies in systematic reviews is recommended in order to reduce database bias, i.e. to reduce the likelihood that a relevant record will be missed (177).

The NHS EED database ceased to be updated from 31st December 2014, therefore, a further search was conducted using Medline, Embase, psychINFO and Cinahl to capture studies published between January 2015 and May 2016 (full search strategies for each database can be viewed in Appendix A). A hand search of relevant health economics and economics journals was conducted alongside reference and citation searches of included items. Journals chosen for hand-searching were identified from an initial scoping search for items reporting priority-setting methods for any area of public health.

Grey literature sources, in the form of public health/health economic conference abstracts, OpenGrey, governmental departments' websites, voluntary organisations' websites, and dissertation and thesis abstracts via ProQuest, were also searched for additional records.

The main search strategy used to identify records from NHS EED was developed with assistance from an information specialist (Shannon Robalino). Research has shown that the inclusion of an information specialist or librarian in systematic reviews is associated with higher quality search strategies compared to reviews conducted without specialist assistance (178). Keywords were identified from an initial scoping search in PubMed. Key words used for the NHS EED search are listed in Box 4.1.

Box 4.1 Key terms used in literature searches

economics	ROI
health Economics	return on investment
economic Evaluation	intoxicate
healthcare Cost	beer
costs	wine
cost analysis	drinking behaviour
value for money	alcoholic beverages
budget	binge drinking
MCDA	alcohol drinking
PBMA	alcoholism
option appraisal	drinking behaviour
multi criteria decision analysis	alcohol use disorder
program budgeting marginal analysis	alcohol abuse
priority setting method	alcohol beverages
social return on investment	alcohol addiction
SROI	alcohol consumption

4.3.3 *Search strategy (May 2016-March 2019)*

To ensure the literature reported in this chapter is as up to date as possible, a supplementary search was conducted in March 2019 using the same search terms described in section 4.3.2. An additional search was conducted in PubMed for studies published by first authors of the items identified in the initial review. It was hypothesised that those authors may have published recently in the same field. An internet search of key search terms and study authors from the previous review was also conducted to identify potentially relevant grey literature.

4.3.4 *Data collection*

Results from each search were imported into an Endnote library and duplicates removed prior to screening titles and abstracts for inclusion. Two researchers, Yemi Oluboyede (YO) and I (SH), independently reviewed all titles and abstracts against the inclusion and exclusion criteria listed in section 4.3.1*,* as recommended by the Cochrane Handbook for conducting systematic reviews (179). Any divergence in opinion regarding inclusion was discussed and agreed on without the need for third-party involvement. I reviewed the full-texts of shortlisted studies from the title and abstract screen in order to assess eligibility. YO verified the studies selected for inclusion, checking for accuracy and eligibility (180). Results of the screening process are described in section 4.4.1.

The review conducted by Weatherly *et al.* (48) discussed earlier in this chapter was also published in greater depth as a report for the Public Health Research Consortium (181). The report included full details of the data extraction form developed for the review. Since the review reported here was examining similar details to those examined by Weatherly and colleagues (48), the data extraction items reported by Drummond *et al.* (181) provided a basis for data extraction of the review reported in this chapter. Items relevant to this study specifically, such as methods of priority-setting adopted, enhanced the data extraction.

The finalised data for extraction included: intervention and comparators, type of study, population and setting, follow-up length, time-horizon for analysis, discounting, perspective, method of economic evaluation or priority-setting, extrapolation of data, reported justification of economic evaluation method, strengths and weaknesses of methods used, outcomes measured, costs included by sector, whether productivity changes were

accounted for, equity considerations, any reported implications for policy from results, and how to use results. The data extraction form can be viewed in Appendix C.

The data extraction process was simplified for the updated review since newly identified items would not be used to inform further empirical studies in this thesis. For example, detail regarding reported implications for policy and how to use the results from the evaluation, would not be utilised and, therefore, were excluded. The simplified data extraction form can be viewed in Appendix C.

I completed the data extraction and a second reviewer validated the extraction of all included studies to ensure the initial data extraction was accurate and complete (180).

4.3.5 *Quality assessment*

Quality of study reporting was conducted based on the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist (182). Several checklists are available for the assessment of economic evaluation studies (183-185), however, the CHEERS checklist has become comparatively popular since its development in 2013 (186). It was developed to optimise economic evaluation reporting in published studies by consolidating the multitude of checklists and guidelines available into a unique reporting guide.

The CHEERS checklist was designed to critique the quality of reporting rather than the quality of the study, however, the two are intimately linked. Due to the purpose of the checklist, various criteria specific to the writing of abstracts, titles, and discussion are included but were not considered relevant to this review, therefore, they were excluded from the quality assessment. Weatherly *et al.* (48) identified that CBAs were often mis-reported in the literature and that the reporting of a societal perspective was not always consistent due to differences in the interpretation of societal perspectives by commentators from the UK (187) and the United States (US) (85). Therefore, two additional criteria were added to examine elements occasionally reported inconsistently:

1. Accuracy of economic evaluation method reported
2. Accuracy of perspective reported

The full list of checklist items is reported in Box 4.2.

Box 4.2 List of items for study quality assessment, based on CHEERS checklist

1.	Form of economic evaluation (or priority setting) clearly reported
1a.	Reported form of economic evaluation is accurate
2.	Target population and subgroups are reported
3.	Setting of evaluation is reported
4.	Study perspective is reported
4a.	Reported perspective is accurate
5.	Comparator interventions reported
6.	Time horizon for evaluation is reported
7.	Discount rate is reported *
8.	Relevance of outcomes measures is reported
9.	Measurement of effectiveness described
10.	Methods of valuation of preference-based outcomes is described *
11.	Methods of estimating resource use are described
12.	Methods of valuing resources in terms of unit
13.	Details of currency and price adjustments for inflation or currency conversion are given
14.	Description and justification of decision-analytic model (or other type of model) are provided *
15.	Assumptions related to model are described and explained *
16.	Analytic methods to support evaluation are reported (e.g. statistical analysis to address skewed data, missing data, extrapolation etc.)
17.	Values and ranges of each component of cost and outcome are reported
18.	Incremental analysis is reported (mean values of costs and outcomes and mean differences provided)
19.	Uncertainty characterised via sensitivity analysis on key parameters
20.	Heterogeneity characterised via discussion of results

*Refers to items that may not be applicable to all studies

4.3.6 *Synthesis of data*

When neither meta-analysis nor qualitative analysis are appropriate methods to synthesise data, the Cochrane Handbook (188) recommends the use of narrative synthesis. Thus, a narrative synthesis was conducted to examine the identified studies based on the methodological challenges reported in section 4.1. Critics of narrative syntheses have exhibited concern over the introduction of bias due to the potential to focus attention on a select few studies (189). Examining each study according to the methodological challenges, identified a priori, minimised the introduction of such bias.

4.4 Results

4.4.1 *Literature search results*

The initial search (January 2006-May 2016) identified 771 records; after deduplication 619 titles and abstracts were screened for eligibility. The search update (May 2016-March 2019) identified 1560 records, which was reduced to 1521 after duplicates were removed. Despite the second search covering a shorter time-period, over twice the number of records were identified. The reduced number of records in the initial search is suspected due to the use of the NHS EED, which is a database of economic evaluations collated via weekly searches of several literature databases (Medline, EMBASE, PsychINFO, CINAHL and PubMed). The use of NHS EED dramatically reduces the number of irrelevant records identified from search terms used to detect economic evaluations. Since the NHS EED has not been updated since December 2014, both the March 2019 search and the final year of the May 2016 search conducted equivalent literature searches of the databases used to identify records for the NHS EED (see section 4.3.2). The final year of the May 2016 search contributed to a significant proportion of the total 771 records identified since only 127 items were retrieved from NHS EED for the period January 2006-December 2014. Consequently, the inability to use the NHS EED for the almost three-year period of the review update resulted in a comparatively large number of records being retrieved from the four databases listed in section 4.3.2.

The process of screening titles and abstracts selected 45 records considered to be potentially eligible for inclusion from the May 2016 review. However, further examination of the full-text was necessary in order to confidently include studies in the review. The process was repeated for the March 2019 search, in which 1492 records were excluded and the full-texts of 29 were examined.

The final screening process of examining the selected full-texts against the eligibility criteria reported in section 4.3.1 shortlisted 23 records for inclusion in the 2016 review and six records for the 2019 search update. One of the records identified in the 2019 review (190) was a publication of a report included in the 2016 review (191) and provided no additional information, therefore, was excluded to prevent duplication of evidence.

In addition to those records retrieved from the database searches, the additional hand-search of relevant journals and search of grey literature for the 2016 review identified five

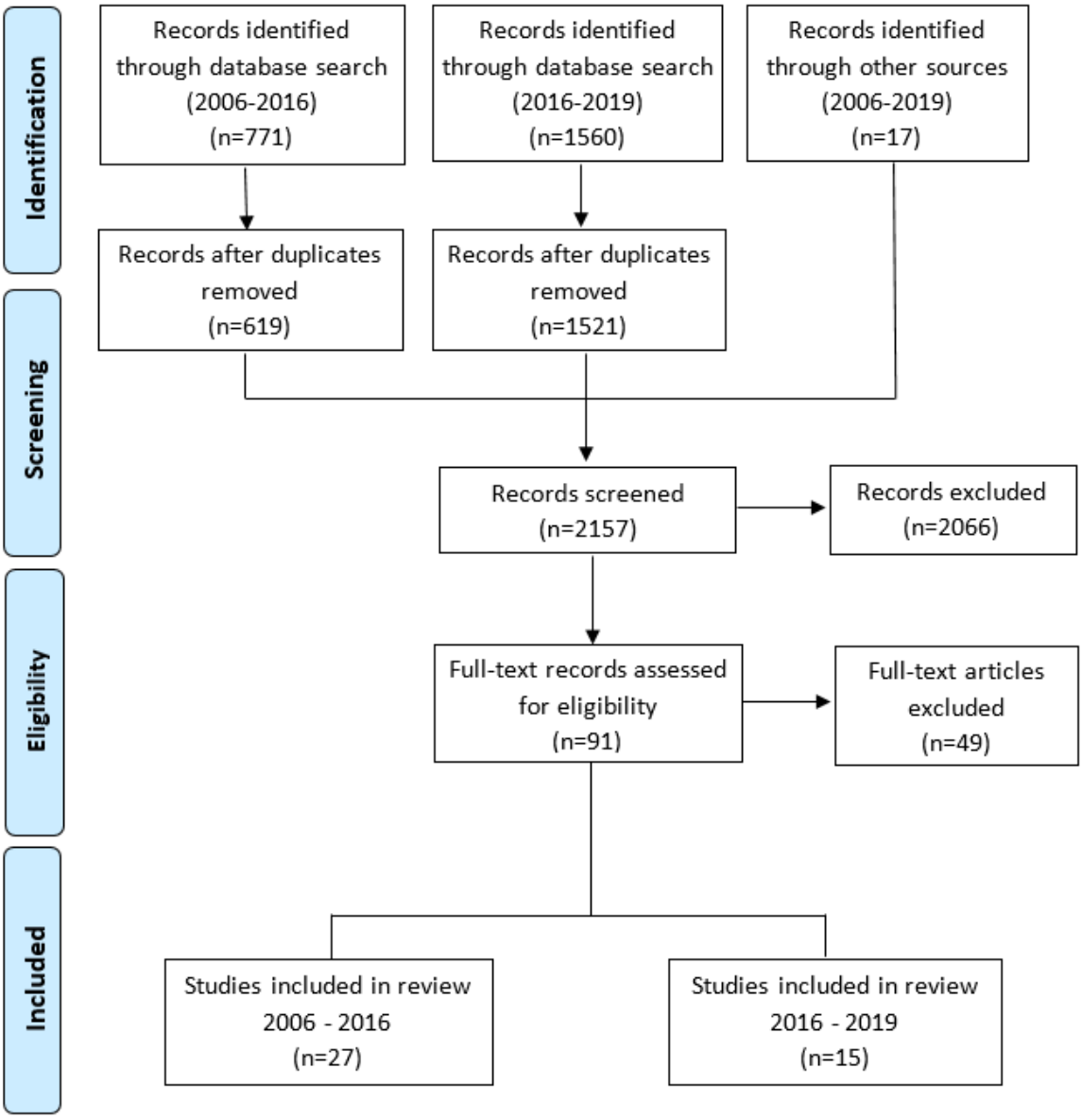
potentially relevant studies, of which full-texts were examined. Of those five, two were considered eligible. A further two eligible studies were discovered by examining the reference lists of the other included records. A total of 27 studies were included in the 2016 review from the combined peer-reviewed and grey literature sources.

The updated search in 2019 included an exploration of records published by authors of studies included in the initial 2016 review; this identified three additional records eligible for inclusion.

The grey literature search of authors also identified a number of reports produced by the Sheffield Alcohol Research Group (192). These reports applied the Sheffield Alcohol Policy Model (SAPM) to a range of alcohol policies and in several country settings. Reports listed for Canada and Northern Ireland could not be accessed, therefore, assessment of those studies was not possible. Due to advances in the SAPM over time, multiple versions of the same study were reported in some cases; in these instances, only the most recent report was included to avoid repetition. Consequently, seven records were considered eligible from the grey literature search. A total of 15 records were included in the 2019 review from both the peer-reviews and grey literature sources.

A PRISMA flow diagram that describes the search processes for both reviews is presented in Figure 4.1.

Figure 4.1 PRISMA flow diagram of the literature search process



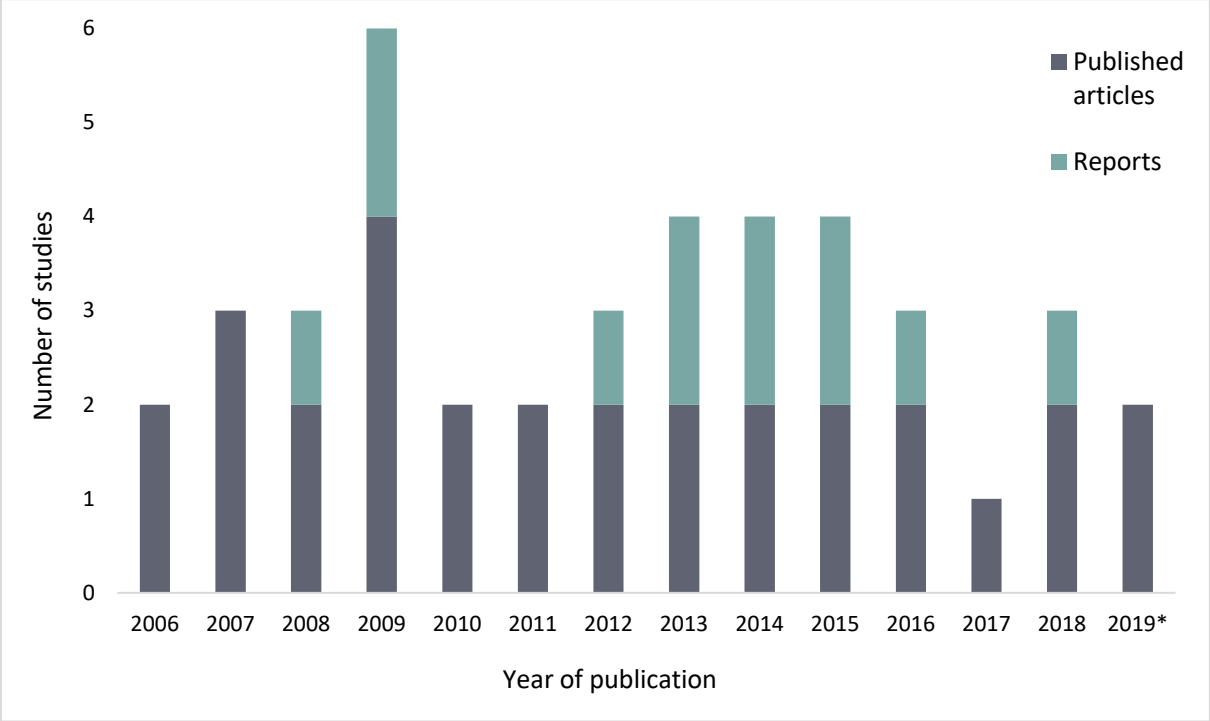
From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

4.4.2 *Overview of studies*

The included studies were published between the years 2006 and 2019; Figure 4.2 demonstrates the distribution of publications across this time period. Relatively few studies were identified between 2017 and 2019, which may be due to a time lag between article submission and publication, given that the literature search was conducted at the beginning of the year in 2019. The lighter shaded bars in Figure 4.2 demonstrate reports identified from searches of grey literature.

Figure 4.2 Distribution of items by publication year



*2019 only includes studies published January - March inclusive

Studies evaluated interventions in the following countries[3]: UK (including studies focussing on England, Scotland, Wales or Northern Ireland individually) (n=12), USA (n=9), Australia (n=5), Netherlands (n=5), Denmark (n=2), Canada (n=2), and the following countries each featured once: Estonia, Italy, Sweden, New Zealand, Czech Republic, Germany, Republic of Ireland, Thailand, Poland, Europe (all EU countries). One study (193) was conducted in a "Western context" with no country specified.

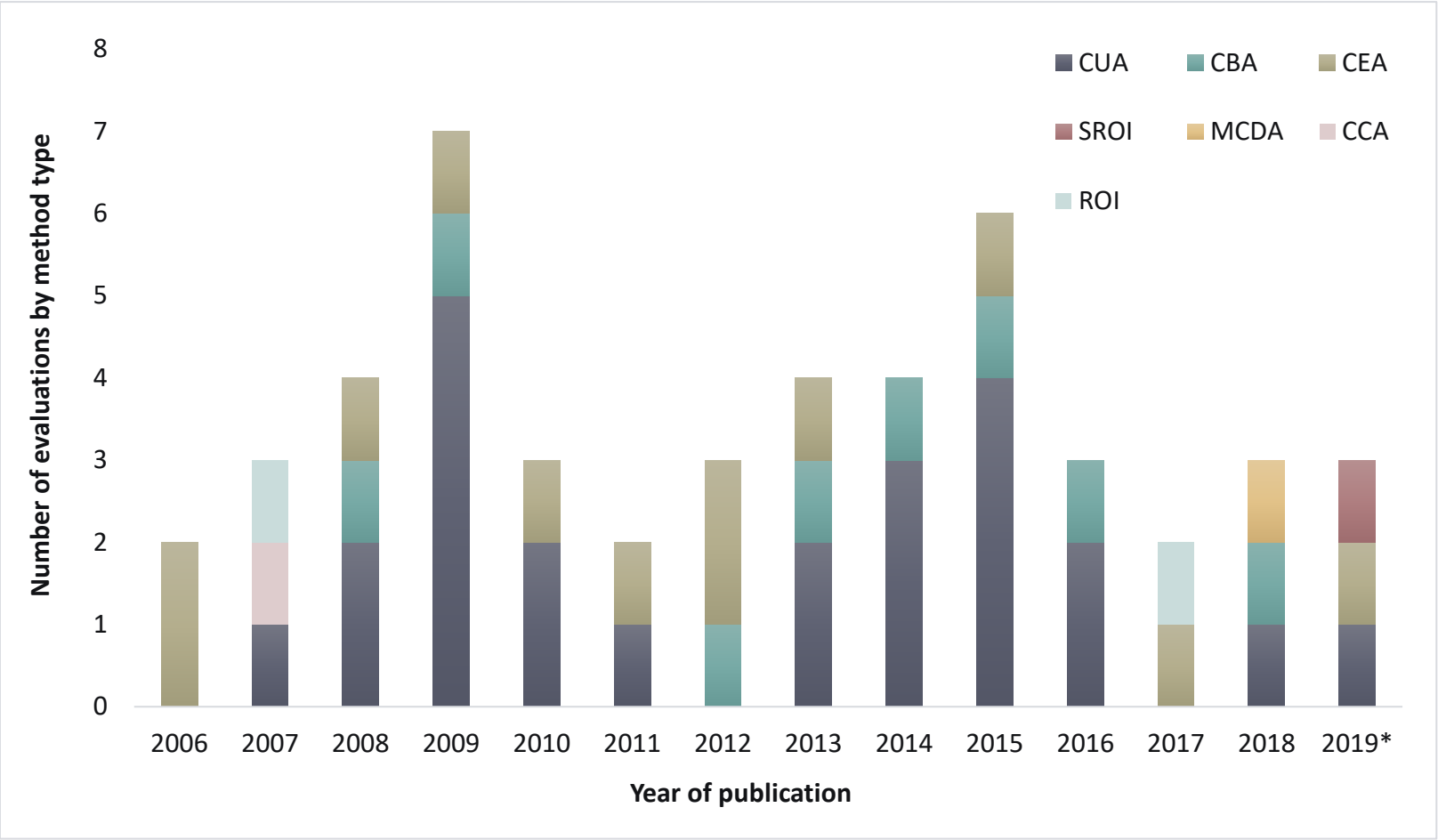
The distribution of economic evaluation and priority-setting methods by publication year is shown in Figure 4.3 to demonstrate their distribution of use over time. The modal methods of economic evaluation used were CUA (n=24) followed by CEA (n=12) and CBA (n=8). Few instances of ROI (n=2), CCA (n=1), SROI (n=1) and MCDA (n=1) were identified. The two examples of MCDA and SROI were published in the year prior to the 2019 review, no studies using a priority-setting technique or SROI were identified in the 2016 review. The ROI evaluations conducted in 2007 and 2017 considered only financial intervention costs and associated financial savings from intervention implementation and included neither a broad series of societal costs and benefits integral to an SROI, nor a monetary valuation of health

[3] The total number of items reported by country is greater than the number of reports included in the review due to several studies featuring multiple countries for comparative purposes.

consequences necessitated for a welfare economic grounded CBA (81). All CBA cases featured in this review valued health gains via a monetarisation of QALYs, no examples were found of benefits valued via approaches consistent with welfare economic theory (i.e. stated or revealed preferences) or human capital valuations (see Chapter Three for detail on CBA benefit valuation methods).

The modal intervention evaluated was ASBI, which was assessed in 16 (38%) studies, followed by tax increases on alcohol (n=9, 21%) and minimum unit pricing (n=7, 17%). Figure 4.4 displays the interventions evaluated and their associated frequencies.

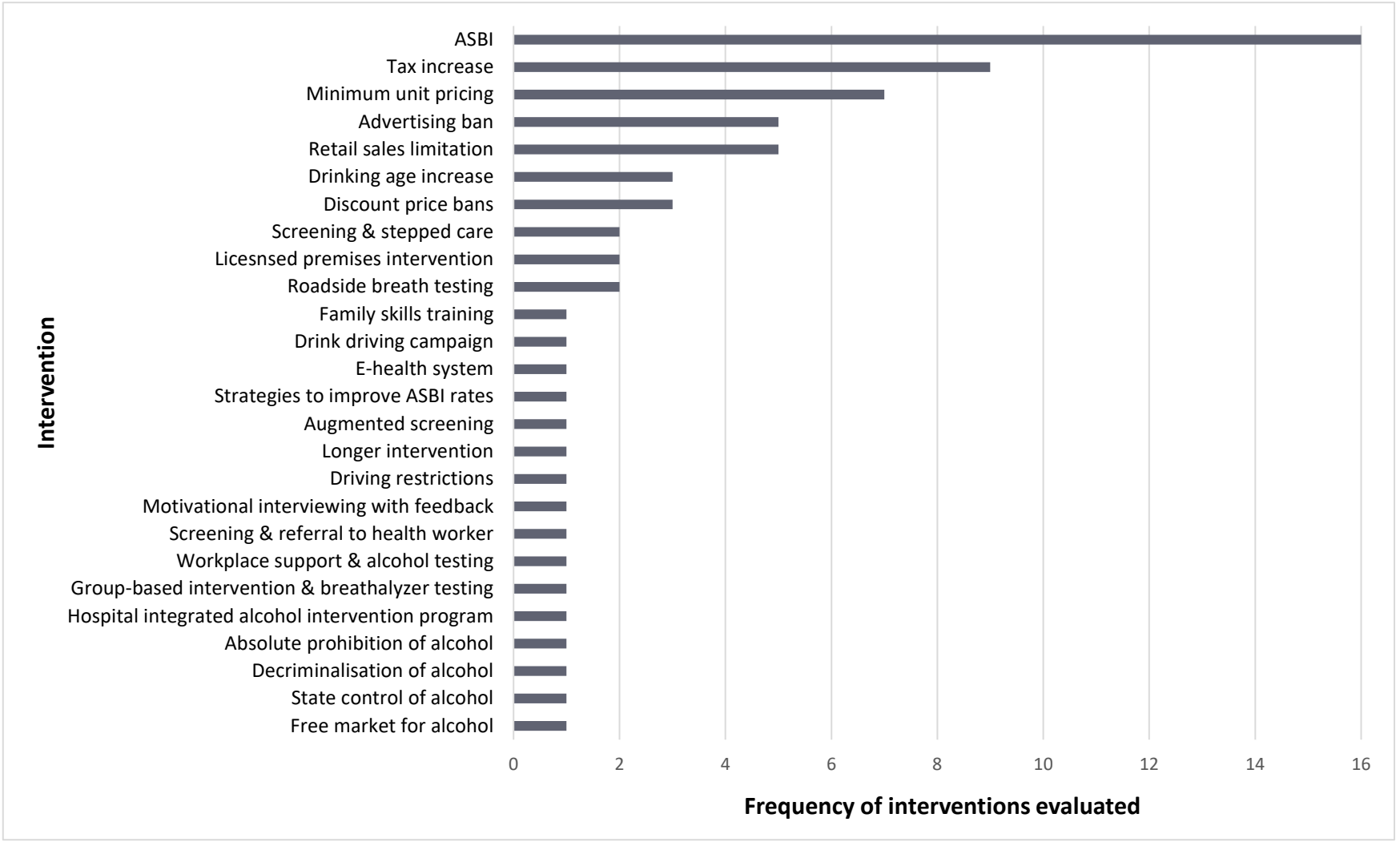
Figure 4.3 Methods of economic evaluation of priority-setting used by year



*2019 only includes studies published January - March inclusive

The cumulative quantity of economic evaluations exceeds the total quantity of included records since several studies reported multiple evaluations

Figure 4.4 Frequency of interventions evaluated



*The cumulative frequency of interventions exceeds the total quantity of included records since several studies evaluated multiple interventions

4.4.3 *Quality of studies*

Table B.1 and Table B.2 in Appendix B report the results from the quality assessment based on the CHEERS checklist. Scores are presented in the table as the proportion of eligible items reported in each study. The items contributing to each score have not been weighted. No recommendations are made for scoring studies using the CHEERS checklist; therefore, no formal attempt has been made to do so. The proportionate scores are merely illustrative of the range of items included in each study and cannot necessarily be used to compare quality relative to other studies.

All studies reported over 50% of the items in the checklist, with the lowest proportion being 57% (see Table B.1 and Table B.2 in Appendix B). No studies included all eligible checklist items.

Notably, none of the records included a preference-based outcomes valuation, preferring to use readily available tariffs for outcome measures; although, one study (194) reported using utilities for alcohol-related health states, which had been previously elicited by colleagues of the authors.

Two criteria were added to the checklist to interrogate accuracy in reporting of economic evaluation type and study perspective. Ten studies did not clearly define the form of economic evaluation undertaken (i.e. CEA, CUA, etc.) (Table B.1 and Table B.2 in Appendix B). For each study, the accurate form of analysis was deduced via examination of the outcome measures and methods of aggregating costs and outcomes. For example, a study reporting an aggregated outcome as an incremental cost per QALY would be considered a CUA. Of those studies that reported the form of analysis, three reported the method inaccurately (195-197). Miller *et al.* (196) and Li *et al.* (197) each describe conducting a CBA, however, an examination of the measures of benefit indicated that the evaluations would be better described as financial ROIs (see section 4.4.2). Mansdotter *et al.* (195) report conducting a CEA, however, there is no final aggregation of costs and outcomes, therefore, it was considered a CCA.

Thirteen of the studies (191, 194, 198-208) reported undertaking a CEA despite measuring health consequences in either QALYs or Disability Adjusted Life Years (DALYs). Whilst not entirely inaccurate, since CUA is a specific form of CEA, precise reporting would distinguish these studies from those which use natural units as an outcome measure.

Nine of the studies did not report the perspective of analysis (193, 209-216). Of the studies that reported an analytic perspective, one reported the perspective inaccurately (195); despite the authors claiming to follow a societal perspective it would be better interpreted as a payer perspective since only costs to the employer implementing the intervention were considered. The societal perspective is recommended by both the first and second Washington panels on cost-effectiveness in health and medicine (85, 86); however, none of the studies reporting to adhere to a societal perspective included all costs regardless of who incurs them, as stipulated in the first panel's definition of this perspective (85). Neither did any of those studies provide any justification for legitimately excluding costs that would be considered necessary to represent a societal perspective; a list of those costs can be found in a published summary of the second panel's recommendations (217). Equally, a handful of studies (200-202, 204, 218) claiming to have followed a healthcare sector perspective included a broader range of costs than recommended (217), such as the inclusion of governmental costs to implement changes in alcohol tax rate.

Productivity costs (defined as work productivity lost from illness) were excluded from two studies reporting a societal perspective (219, 220). Explicitly accounting for productivity is a topic of debate with some experts arguing for this approach (187, 221) whilst others arguing that productivity is implicitly accounted for in the generation of QALYs (11, 85). Whilst it is beyond the scope of this thesis to argue for a side in the productivity costs debate, what is apparent is a need for clear and consistent guidance on this issue if evaluations are to be comparable.

4.4.4 *Attributing effects of public health interventions*

Intervention effects may be observed directly from trials of the intervention, from natural experiments, or before-and-after studies, although results from the latter should be considered cautiously (222). Alternatively, models can simulate intervention effects by drawing on data from secondary sources. Across the literature from both reviews nine randomised, and one non-randomised, controlled trials featured and one before-and-after study (Table 4.1). The majority of the trials (n=9) were identified in the 2016 review, whilst the peer-reviewed literature identified in the 2019 update constituted predominantly modelling studies. The only RCT identified in the 2019 review (223) used intervention effects obtained from the trial to populate the SAPM model in order to ascertain estimates of cost-effectiveness. Only one of eight RCTs identified in the 2016 review followed a similar

approach; Neighbors *et al.* (219), used trial outcomes to develop a decision-analytic model to generate cost-effectiveness estimates.

Thirty of the records identified between 2006 and 2019 were purely modelling studies (Table 4.1). A range of modelling approaches were used: decision-analytic models, Markov microsimulation models, statistical and mathematical models, disease models such as the Chronic Disease Model (see Appendix A in van Baal *et al.,* 2008 (224)), the ALCMOD alcohol model (225), the SAPM (226), and a "meta-model" approach (227). Data to populate the models were taken from epidemiological literature, results from previous trials, longitudinal studies, national surveys and databases, and meta-analyses in order to estimate future costs and outcomes.

Cobiac *et al.* (200) did not expand on the data sources used to populate their model, therefore, the quality of their data sources could not be examined. The remaining modelling studies reported utilising appropriate data with reference to the population of interest, for example seeking sources from the same country and age group. However, a handful reported the use of data from other countries where relevant data was unavailable for the location of study (194, 198, 201, 206, 212, 227-229). Where data from different contexts were used, two studies reported some form of adjustment to improve relevance to the country of study (212, 229).

Table 4.1 Overview of studies included in the review and key details

Author	Year	Country of study	Title	Method of analysis[a]	Study type	Time Horizon	Perspective	Sectors of costs included in analysis
Angus *et al.* (a)	2014	Italy	Cost-effectiveness of a programme of screening and brief interventions for alcohol in primary care in Italy	CUA	Model	30 years	Healthcare sector	Healthcare
Angus *et al.* (b)	2014	Republic of Ireland	Model-based appraisal of minimum unit pricing for alcohol in the Republic of Ireland	CBA	Model	20 years	Unspecified	Healthcare Criminal justice
Angus *et al.*	2015	England	Modelling the impact of Minimum Unit Price and Identification and Brief Advice policies using the Sheffield Alcohol Policy Model Version 3	CUA & CBA	Model	20 years	Societal & Healthcare sector	Healthcare Criminal justice
Angus *et al.*	2016	Europe	Estimating the cost-effectiveness of brief interventions for heavy drinking in primary healthcare across Europe	CUA	Model	Unspecified	Healthcare sector	Healthcare
Angus *et al.*	2018	Wales	Model-based appraisal of the comparative impact of Minimum Unit Pricing and taxation policies in Wales: Final report	CBA	Model	20 years	Unspecified	Healthcare Criminal Justice
Angus *et al.*	2019	England, Poland and Netherlands	Cost-effectiveness of strategies to improve delivery of brief interventions for heavy drinking in primary care: results from the ODHIN trial	CEA & CUA	Model alongside RCT	30 years	Healthcare sector	Healthcare

Table 4.1 cont. Overview of studies included in the review and key details

Author	Year	Country of study	Title	Method of analysis[a]	Study type	Time Horizon	Perspective	Sectors of costs included in analysis
Barbosa *et al.*	2015	United States	The cost-effectiveness of alcohol screening, brief intervention and referral to treatment in emergency and outpatient medical settings	CUA & CEA	Non-randomised study & Model	6 months	Provider & Societal	Healthcare Criminal Justice Other (automobile accident costs)
Barrett *et al.*	2006	United Kingdom	Cost-effectiveness of screening and referral to an alcohol health worker in alcohol misusing patients attending an A&E department	CEA	RCT	12 months	Societal	Healthcare Social care Criminal Justice Voluntary sector
Brennan *et al.*	2009	England	Independent review of the effects of alcohol pricing and promotion: Part B	CBA	Model	10 years	Unspecified	Healthcare Criminal justice Employment
Byrnes *et al.*	2010	Australia	Cost-effectiveness of volumetric alcohol taxation in Australia	CUA	Model	Unspecified	Healthcare sector	Healthcare Government
Cobiac *et al.*	2009	Australia	Cost-effectiveness of interventions to prevent alcohol-related disease and injury in Australia	CUA	Model	Lifetime	Healthcare sector	Healthcare Patient out-of-pocket Government
Cobiac *et al.*	2018	New Zealand	Cost-effectiveness of raising alcohol excise taxes to reduce the injury burden of road traffic crashes	CUA	Model	Lifetime	Healthcare sector & Societal	Healthcare Criminal Justice Government Other (automobile accident costs)
Cowell *et al.*	2012	United States	Cost-effectiveness analysis of motivational interviewing with feedback to reduce drinking among a sample of college students	CEA	RCT	3 months	Provider	Education

Table 4.1 cont. Overview of studies included in the review and key details

Author	Year	Country of study	Title	Method of analysis[a]	Study type	Time Horizon	Perspective	Sectors of costs included in analysis
Crawford *et al.*	2015	United Kingdom	The clinical and cost-effectiveness of brief advice for excessive alcohol consumption among people attending sexual health clinics: a randomised controlled trial	CUA	RCT	6 months	Healthcare sector (NHS & PSS)	Healthcare Social care
De Wit *et al.*	2016	Netherlands	Social cost-benefit analysis of regulatory policies to reduce alcohol use in The Netherlands	CBA	Model	50 years	Societal	Healthcare Law enforcement Criminal Justice Education Government
Drummond *et al.*	2009	United Kingdom	Effectiveness and cost-effectiveness of a stepped care intervention for alcohol use disorders in primary care: pilot study	CUA	RCT	6 months	Unspecified	Healthcare Criminal Justice Social care Other (accident costs)
Havard *et al.*	2012	Australia	Randomized Controlled Trial of Mailed Personalized Feedback for Problem Drinkers in the Emergency Department: The Short-Term Impact	CEA	RCT	6 weeks	Provider	Healthcare
Holm *et al.* (a)	2014	Denmark	Cost-effectiveness of changes in alcohol taxation in Denmark: A Modelling study	CUA	Model	Lifetime	Healthcare sector	Healthcare Law enforcement Government
Holm *et al.* (b)	2014	Denmark	Cost-Effectiveness of Preventive Interventions to Reduce Alcohol Consumption in Denmark	CUA	Model	Lifetime	Healthcare sector	Healthcare Law enforcement Government

Table 4.1 cont. Overview of studies included in the review and key details

Author	Year	Country of study	Title	Method of analysis[a]	Study type	Time Horizon	Perspective	Sectors of costs included in analysis
Ingels *et al.*	2013	United States	Cost-effectiveness of the strong African American families-teen program: 1-year follow-up	CEA	RCT	12 months	Societal	Social care Voluntary sector Out of pocket costs
Kapoor *et al.*	2009	United States	Cost-effectiveness of screening for unhealthy alcohol use with % carbohydrate deficient transferrin: results from a literature-based decision analytic computer Model	CUA	Model	Lifetime	Societal	Healthcare Out of pocket costs
Lai *et al.*	2007	Estonia	Costs, health effects and cost-effectiveness of alcohol and tobacco control strategies in Estonia	CUA	Model	100 years	Societal	Healthcare Law enforcement Out of pocket costs Government
Li *et al.*	2017	USA	Economic Analyses of an Alcohol Misconduct Prevention Program in a Military Setting	CEA & ROI	Before-and-after study	Unspecified	Employer	Healthcare Private (University staff costs) Other (Airforce staff costs)
Mansdotter *et al.*	2007	Sweden	A cost-effectiveness analysis of alcohol prevention targeting licensed premises	CCA & CA	Non-randomised study	5 years	Payer[b]	Law enforcement Private
Meng *et al.*	2012	Scotland	Model-based appraisal of alcohol minimum pricing and off-licensed trade discount bans in Scotland using the Sheffield alcohol policy Model (V2)	CBA	Model	10 years	Unspecified	Healthcare Criminal justice Employment

Table 4.1 cont. Overview of studies included in the review and key details

Author	Year	Country of study	Title	Method of analysis[a]	Study type	Time Horizon	Perspective	Sectors of costs included in analysis
Meng *et al.*	2013	England	Modelled income group-specific impacts of alcohol minimum unit pricing in England 2014/15: Policy appraisals using new developments to the Sheffield Alcohol Policy Model (v2.5)	CBA	Model	10 years	Unspecified	Healthcare Criminal justice
Miller *et al.*	2007	United States	Effectiveness and benefit-cost of peer-based workplace substance abuse prevention coupled with random testing	ROI	Model	Lifetime	Employer & limited societal	Healthcare Private
Miller & Hendrie	2008	United States	Substance Abuse Prevention Dollars and Cents: A Cost-Benefit Analysis	CBA	Model	Lifetime	Societal & State	Healthcare Education Criminal Justice Law enforcement Other (property damage expenses)
Navarro *et al.*	2011	Australia	The potential cost-effectiveness of general practitioner delivered brief intervention for alcohol misuse: evidence from rural Australia	CEA	Model	12 months	Unspecified	Healthcare
Neighbors *et al.*	2010	United States	Cost-effectiveness of a motivational intervention for alcohol-involved youth in a hospital emergency department	CUA & CEA	Model alongside RCT	6 months	Provider & Societal	Healthcare Criminal Justice Other (automobile accident costs)

Table 4.1 cont. Overview of studies included in the review and key details

Author	Year	Country of study	Title	Method of analysis[a]	Study type	Time Horizon	Perspective	Sectors of costs included in analysis
Purshouse *et al.*	2009	England	Modelling to assess the effectiveness and cost-effectiveness of public health related strategies and interventions to reduce alcohol attributable harm in England using the Sheffield Alcohol Policy Model version 2.0	CUA	Model	30 years	Healthcare sector (NHS & PSS) & Public sector	Healthcare Criminal justice Employment
Purshouse *et al.*	2013	United Kingdom	Modelling the cost-effectiveness of alcohol screening and brief interventions in primary care in England	CUA	Model	30 years	Healthcare sector (NHS & PSS)	Healthcare Social care Other (automobile accident costs)
Rogeberg *et al.*	2018	No specific country, "Western context"	A new approach to formulating and appraising drug policy: A multi-criterion decision analysis applied to alcohol and cannabis regulation	MCDA	MCDA	Unspecified	Unspecified	Healthcare Government Law enforcement
Sassi *et al.*	2015	Canada, Czech Republic and Germany	Health and economic impacts of key alcohol policy options	CUA	Model	40 years	Unspecified	Healthcare Government
Shanahan *et al.*	2006	Australia	Modelling the costs and outcomes of changing rates of screening for alcohol misuse by GPs in the Australian context	CEA	Model	Unspecified	Government	Healthcare Government

Table 4.1 cont. Overview of studies included in the review and key details

Author	Year	Country of study	Title	Method of analysis[a]	Study type	Time Horizon	Perspective	Sectors of costs included in analysis
Smit *et al.*	2011	Netherlands	Modelling the cost-effectiveness of healthcare systems for alcohol use disorders: how implementation of eHealth interventions improves cost-effectiveness	CUA	Model	12 months	Healthcare sector	Healthcare
Solberg *et al.*	2008	United States	Primary care intervention to reduce alcohol misuse: ranking its health impact and cost effectiveness	CUA	Model	Lifetime	Healthcare sector & Societal	Healthcare Criminal Justice Social care Out of pocket costs Other (automobile accident costs)
Tanaree *et al.*	2019	Thailand	Integrated treatment program for alcohol related problems in community hospitals, Songkhla province of Thailand: A social return on investment analysis	SROI	Mixed-methods SROI	5 years	Societal	Healthcare Out of pocket costs
Tariq *et al.*	2009	Netherlands	Cost-effectiveness of an opportunistic screening programme and brief intervention for excessive alcohol use in primary care	CUA & CEA	Model	100 years	Healthcare sector	Healthcare
Van den Berg *et al.*	2008	Netherlands	The cost-effectiveness of increasing alcohol taxes: a modelling study	CUA & CEA	Model	100 years	Healthcare sector	Healthcare

Table 4.1 cont. Overview of studies included in the review and key details

Author	Year	Country of study	Title	Method of analysis[a]	Study type	Time Horizon	Perspective	Sectors of costs included in analysis
Watson *et al.*	2013	United Kingdom	AESOPS: a randomised controlled trial of the clinical effectiveness and cost-effectiveness of opportunistic screening and stepped care interventions for older hazardous alcohol users in primary care	CUA	RCT	12 months	Healthcare sector (NHS & PSS)	Healthcare Social care
Zur & Zaric	2016	Canada	A microsimulation cost–utility analysis of alcohol screening and brief intervention to reduce heavy alcohol consumption in Canada	CUA	Model	Lifetime	Health payer	Healthcare

[a] Analysis method interpreted by SH, study authors' stated methods sometimes differed from those stated here

[b] Perspective reported here is interpreted by SH based on costs included, study authors reported adhering to a societal perspective

4.4.5 *Extrapolating long-run outcomes*

Nine of the studies (193, 195-197, 218, 219, 227, 230, 231) failed to clearly report the time horizon used in the analysis, however, an estimate could be inferred for four of those (195, 196, 227, 230) (Table 4.1). The majority of the studies using modelling techniques (n=24, 75%) (Table 4.1) employed a time-horizon of 10 years or longer, 16 of those extrapolated intervention impacts beyond 30 years to ensure the long-term effects of the interventions were captured.

The trial-based studies typically conducted analyses for follow-up periods of 12 months or less (Table 4.1) (195, 232). Exceptions to this were Ingels *et al.* (232) whose analyses extended to 18 months and Mansdotter *et al.* (195) whose analysis extended to five years, although, as mentioned earlier in this section, the precise time-horizon considered was unclear for the latter study.

Tanaree *et al.* (233) reported a five-year time horizon with the justification that after this period drinking problems tend to relapse; it is assumed that repeat intervention may be required at this point, although this was not explicitly discussed by the study authors.

4.4.6 *Outcome measures*

Within identical economic evaluation formats there remains inconsistency amongst the outcome measures used. A range of natural units related to alcohol was used in the CEA studies (e.g. per unit reduction in alcohol consumption, or per one-day decrease in heavy drinking days).

Within the CUA framework, there are far fewer options for relevant outcome measures, however, some studies used QALYs (favoured by NICE) whilst others used DALYs (recommended by the WHO). Even within the QALY paradigm, all QALY measures may not be entirely comparable if the quality of life element has been derived from a different measure. Within the included studies, QALYs were generated using a range of methods: the EQ-5D (e.g. Crawford *et al.* 2015 (234)), SF-12 (209), Health and Activities Limitation Index (HALex) (219), the Health Utilities Index (HUI) (229) and by condition-specific utility values (194) (see section 3.1).

However, different countries and organisations recommend different tools for the measurement of quality of life when calculating QALYs; for example, NICE in England

recommends the EQ-5D (10) whilst the Canadian Agency for Drugs and Technologies in Health (CADTH) recommends the use of any generic classification system, e.g. EQ-5D, SF-6D or HUI (235). Therefore, the use of the HUI by Zur & Zaric (229) would be considered appropriate for the setting of their evaluation (Canada), even though it may not be considered ideal in England.

The CBA, ROI and SROI studies reported outcomes in monetary units, as is standard for these analyses. As highlighted earlier in this chapter (section 4.4.2), the identified CBAs across the two reviews used a monetised value of a QALY to determine the worth of health-related benefits. Values ranged from €45,000 (approximately £38,300 in 2014 prices[4], e.g. Angus *et al.,* 2014 (216)) to £60,000 (in 2014 prices) (e.g. Angus *et al.,* 2015 (207)). Miller & Hendrie (236) did not report how the benefit value was calculated for their CBA, however, QALY data were reported thus an assumption that QALYs were monetised was made on this basis. The SROI study (233) provided the only example of revealed preference methods being used to value outcomes, whereby the market value of an item similar to the outcome of interest is assigned to approximate the worth the non-marketed outcome.

4.4.7 *Intersectoral costs and consequences*

Authors considered costs from 10 unique sectors plus additional costs related to automobile accidents, Airforce staff costs and property damage costs (Table 4.1). Approximately one third of the studies across the 2006-2019 period (n=13, 31%) explicitly refer to taking a societal perspective (Table 4.1). Nine studies did not specify the perspective used, however, examining the costs reported in those studies, a broad or public sector perspective could be assumed for six (209, 212-216). Health service perspectives were followed by 17 studies (40%) and nine studies reported employer, government, or payer perspectives (Table 4.1). Several studies (196, 199, 206-208, 219, 220, 236) reported analyses according to multiple perspectives, to provide flexibility in the use of their findings depending on the audience. The reported MCDA (193) does not adhere to an analytical perspective, however, the authors state the importance of the MCDA criteria acknowledging perspectives of *"users, their surroundings and broader society"* (193, p.147). The criteria included: the financial cost

[4] Conversion from 2013 Euros to 2014 pounds sterling facilitated by the *CEMG - EPPI-Centre Cost Converter* available at https://eppi.ioe.ac.uk/costconversion/default.aspx

of implementing and enforcing policies, generating state revenue, and reducing public costs indirectly related to the policy (e.g. spillover on health policy budgets).

None of the studies from the 2016 review included any non-health related consequences, although, three of the studies from the 2019 review (193, 233, 237) considered non-health outcomes. Those evaluated by Tanaree and colleagues (233) ranged from the interventional effect on recipients' wellbeing to the impact on family members and reductions in alcohol-related traffic accidents. De Wit and colleagues (237) examined benefits to alcohol consumers, producers and retailers, the tax authority, and healthcare whilst Rogeberg *et al.* (193) assessed the merits of various policies according to their impact on health, social aspects, the political environment, crime, and public wellbeing.

Additionally, six of the SAPM reviews (207, 208, 212-215) accounted for the health-related impact of injuries attributed to alcohol use in their calculation of the costs of crime. These studies used published estimates of QALY losses associated with a range of crime types (e.g. assault, rape, etc.) to which a financial value of a "crime QALY" was applied (238). Whilst use of these crime QALYs accounts for the indirect impacts of alcohol use, compared to the direct effects on alcohol users, they remain health-related outcomes.

4.4.8 *Examining population heterogeneity and inferences for equity*

Equity was rarely discussed in any of the identified studies; however, some authors report a limited examination of the effects of population heterogeneity. An inference may be drawn from some of the stratified analyses with regards to the greatest value gain via sub-group targeting of interventions.

Several studies reported results stratified by either gender, age, or drinking status. The findings from Angus *et al.* (198) and Neighbors *et al.* (219) indicate that value to society could be improved by targeting interventions at males, due to significantly different cost-effectiveness estimates for subgroups of men and women. Conversely Zur & Zaric (229) found little difference in cost-effectiveness between males and females. Tanaree *et al.* (233) estimated the most favourable SROI for high-risk, non-dependent drinkers compared to those with dependency issues, whilst the SAMP reports (208, 212-215) generally reported the greatest reductions in alcohol consumption, and associated health-gains, in harmful drinkers relative to moderate or hazardous consumers. Kapoor *et al.* (194) examined cost-utility outcomes by age-cohort and demonstrated significantly reduced cost-effectiveness in

a 75-year-old cohort. This finding is unsurprising given the calculation of QALYs includes improvements in length of life, which naturally disadvantages older individuals, ceteris parabis, and is a common argument in the QALY paradigm debate (239, 240).

Three additional records (195, 204, 211) reported some model parameters stratified according to population characteristics, however, the heterogeneity was not reflected in the overall analysis of cost-effectiveness.

The only inclusion of equity considerations was found in four of the SAPM reports, which included an analysis of intervention outcomes according to socioeconomic status or income level (207, 213, 215, 216). The analyses concluded that the interventions examined (primarily minimum unit pricing or other tax increases on alcohol) reduced health inequalities, mechanised by the greatest concentration of effect falling on the most deprived groups.

4.5 Discussion

This review aimed to address the first overarching research question of the thesis by examining the methods of economic evaluation and priority-setting used in the current literature base and critiquing the quality of the identified studies. The review succeeded in this aim by exploring whether the literature dating back to 2006 indicates an adoption of alternative economic evaluation techniques by the public health economic research field with a particular focus on interventions to prevent alcohol misuse. Furthermore, the quality of the current literature regarding public health specific methodological challenges was critiqued. The following section of this chapter discusses the study results with regard to each of these two study objectives. Furthermore, implications for further empirical work will be discussed based on the findings of the 2016 review, reflecting only the data available at the time the empirical work was conducted.

Although this review focused exclusively on alcohol-related interventions, it is reasonable to expect that the findings would also be applicable to other areas within the jurisdiction of public health, since Weatherly and colleagues' previous review (48) examining public health broadly did not suggest a unique quality differential for evaluations of alcohol-related interventions.

4.5.1 *Methods of economic evaluation*

Across all literature identified in this review (January 2006 – March 2019), CUA remains the most prevalent evaluative method. However, Figure 4.3 depicts a relative decline in the proportion of CUA studies since 2015 and more regularity in the use of CBAs from 2012. The final two years and three months' (2017- March 2019, inclusive) worth of published literature revealed the greatest diversity in methods used (Figure 4.3); the only examples of MCDA and SROI identified in this literature review were published during this time. The distribution of economic evaluation methods over the time-period reviewed indicates some movement towards the adoption of alternative economic evaluation techniques. Whilst over-interpretation of the limited evidence found in this study must be avoided, given the relative domination of CUA and CEA in the first decade of review (2006-2016), the introduction of alternative techniques suggests a shift in research precedence may be occurring. However, it is worth noting that neither the SROI nor MCDA studies identified were conducted in the UK. It is, therefore, unlikely that the adoption of these methods was influenced by the guidance documents introduced in Chapter One, specifically the most recent NICE guidance for public health evaluations (10), as posited in section 4.2.

Although an example of a study using priority-setting methods was identified in the 2019 review (193), they remain considerably rare in the published literature in this field. Methods such as PBMA and MCDA can be advantageous to decision-makers by assisting resource allocation decisions using a systematic and transparent approach. These approaches allow for the consideration of multiple criteria relevant to decision-makers' specific needs and priorities, which is particularly pertinent to a PHDM whose interests likely extend beyond purely health maximisation. The ability of these tools to incorporate the particular needs of decision-makers undertaking the prioritisation exercise is one of their merits and may explain the paucity of these studies in the published literature. Due to the individual nature of priority-setting exercises, they may be used in practice within local government or other agencies but not reported in the academic literature. Therefore, the limited number of studies identified in this review is expected to be an underestimate of the number of priority-setting exercises being conducted in this field.

4.5.2 *Methodological development*

The second objective of this review was to reflect on the identified literature with regards to the four methodological challenges for evaluating public health interventions, originally reported by Weatherly and colleagues (48). By way of reminder, these challenges were: (i) attribution of effects, (ii) measuring and valuing outcomes, (iii) identifying intersectoral costs and consequences, and (iv) incorporating equity considerations. This section will consider whether the current literature demonstrates any methodological improvement in these four areas, following the "disappointment" expressed by Weatherly *et al.* (48) at the state of insight into the challenges from the empirical base they identified.

Attributing effects

The proportion of RCT studies identified over the entirety of this review (n=9, 22%) is substantially less than that identified by Weatherly *et al.* (48) (38%); however, extrapolating long-run outcomes from trial data appears to be a consistent challenge, with only one of the studies (223) able to extrapolate outcomes to a significant time horizon (30 years). Angus and colleagues (55) used trial outcomes to populate a version of the SAPM, supplementing with country-specific baseline data from the literature. The uncertainty imposed by any assumptions made in order to complete the model was assessed in a series of deterministic sensitivity analyses. This study was first published online in September 2018, and, as one of the most recent studies identified in the updated review, may indicate the beginning of a change in the way in which trials of public health interventions are evaluated.

The remaining study distribution in Weatherly *et al.'s* review (48) was divided evenly between non-randomised and review studies (31% for each) and the authors reported the use of modelling in the majority of these. The proportion of studies using modelling techniques in this current review is greater (75%) than that identified by Weatherly *et al.* (48), who stressed the need for further research into methods of evidence synthesis to identify all relevant data for modelling. One novel approach to synthesising evidence for modelling outcomes in different settings was the "meta-model" developed by Angus *et al.* (227), identified in the 2019 review update. This technique allowed cost-effectiveness outcomes for universal screening and brief intervention to be estimated for all countries in Europe, based on data available from a small sample of five countries, via the identification of key parameters believed to impact on estimates of cost-effectiveness. This approach

could prove beneficial for public health decision-making in the UK if tailored to provide outcomes for local communities based on data identified in smaller, localised samples.

Measuring and valuing outcomes

None of the studies reviewed by Weatherly *et al.* (48) considered outcomes beyond health or valued outcomes using a direct measure of WTP. A similar conclusion would be made for the studies identified in the 2016 search of this review; however, some recent examples of more holistic evaluations were identified in the 2019 update. Absent use of WTP to directly value outcomes, however, remains constant throughout the entirety of this current review. Two economic evaluations (233, 237) considered a range of outcomes including, and going beyond, health in an attempt to capture the full impact to society and intervention recipients. Whilst health outcomes examined by De Wit *et al.* (237) were monetised in the same way as the other CBAs identified by this review, using a financial value of QALYs, Tanaree and colleagues (233) made use of revealed preference techniques to assign values to impacts without market values (see Chapter Three for details on revealed preference methods). The use of proxy valuation can introduce uncertainty into results; however, it enables the inclusion of outcomes that would otherwise be overlooked.

The previous review (48) argued for more extensive use of WTP methods to value outcomes or at a minimum, sector-specific generic outcomes such as crime QALYs and education QALYs. Some evidence of improvement was identified in this current review on that latter point with the inclusion of "crime QALYs" in six of the SAPM reports (207, 208, 212-215). The "crime QALYs" were monetised using a value of £81,000 per QALY, which was calculated from the value of a statistical life estimated by Carthy *et al.* (241) using a combination of contingent valuation and standard gamble techniques. The UK based CBA studies that monetised QALYs (207, 212-214), also reported using a value, recommended by the Department of Health (242), which is based on estimates obtained using stated preference techniques.

Earlier in this section it was stated that none of the identified studies used WTP to *directly* value outcomes, which is an important distinction when the financial values used to monetise QALYs are examined. WTP techniques were in actuality incorporated into the CBAs, via the indirect route of providing financial values of QALYs. Two points are worth making here. Firstly, the use of WTP in these instances does not nullify the argument made earlier that current CBAs using monetised QALYs are preventing the full realisation of the

potential of this methodology, since the use of QALYs is still restricting the outcomes to consider only health outcomes. Secondly, the use of CV methods, or other techniques to obtain WTP estimates such as DCEs, should not be lacking in the CBA literature due to an objection to the mode of valuation, since this method has been accepted as an indirect valuation of health outcomes. It is possible that a generic, monetised value of a QALY is used rather than eliciting a bespoke WTP valuation due to the resources required to undertake a stated preference study, a point raised by the Department of Health (242).

Alternatively, Weatherly *et al.* (48) suggested the use of MCDA techniques to value a broad range of costs and outcomes, none of which were identified in their review of the public health literature. One example of a relevant MCDA was identified in the 2019 update (193) to explore different policy options available for the regulation of drugs and alcohol. Thus, some progress in the measuring and valuation of public health outcomes does seem apparent, although they appear only in their infancy being published in the final two years of the review update.

Intersectoral costs and consequences

Weatherly and colleagues (48) reported negligible evidence of intersectoral consideration, as did this current review. Utilising CCA to apportion impact on other sectors was recommended by the previous review authors (48), yet no comprehensive use of this technique was identified in this current review. The general equilibrium approach of simultaneously considering the consequences of interventions across an array of different sectors proposed by Weatherly *et al.* (48) to incorporate intersectoral consequences, does not appear to have been used by any of the study authors identified in this literature review.

The range of costs considered is determined by the perspective of analysis chosen, the broadest of which being societal. The Washington panel on cost-effectiveness in health and medicine proposes the use of the societal perspective in economic evaluations (85, 217) and NICE (10) recommends either a public sector or local government perspective for evaluations of public health interventions. A similar proportion of studies reported to adhere to a societal perspective in both this review and that undertaken by Weatherly *et al.* (48) (31% in both), however, broad perspectives could be inferred from a number of studies that did not report the perspective adhered to (see section 4.4.7).

A lack of consensus is evident in this review amongst the approaches taken for societal perspectives. As highlighted in section 4.4.3, several studies presented a narrow interpretation of this perspective and discrepancy exists over the inclusion of productivity costs. There are, however, time and resource constraints which may prevent a comprehensive collection of data required to truly reflect a societal perspective (243) and some costs and benefits may be difficult to capture, such as productivity costs or benefits to family members or carers who do not directly benefit from an intervention.

Equity

Incorporating equity in economic evaluations was rarely considered in the studied literature from both this review and that by Weatherly *et al.* (48), despite being a globally recognised area of need (244). Only four studies, identified in the 2019 grey-literature search, mentioned potential impacts on inequalities resulting from intervention implementation (207, 213, 215, 216).

Methods of incorporating equity are still in relative infancy in the health economics literature, therefore, it is unsurprising that this element is largely missing from the evaluations identified in this review. Research into methods to incorporate equity considerations into economic evaluations via CEA has been published recently (26, 27) and an investigation into the equity impacts of public health interventions has also been recently published (245). It is beyond the scope of this study to elaborate at length the methods that are available within economic evaluation to incorporate equity considerations. However, it is evident that the studies identified in this review fall short in this area.

4.5.3 *Limitations of the review*

Whilst all reasonable effort was made to ensure rigour in this review, it is not without some limitations. Grey literature was examined in the initial 2016 review, however, the identification of unpublished reports dating back to 2009 during the 2019 update suggests the original review may not have captured all sources of unpublished data. However, it is likely that the majority of the readily available literature for this area has now been captured following the updated searches.

The scope of this review was limited to interventions that directly aim to reduce or prevent the misuse of alcohol and did not include interventions to prevent harm as a result of consuming alcohol. Areas such as transport economics and sexual health would likely include

economic evaluations relevant to this broader scope of alcohol-induced harm and would provide a worthwhile area for exploration in future research. Additionally, the exclusion of non-English studies due to limited resources for translation may have restricted the studies examined in the review.

Limitations related to the choice of a narrative review should also be recognised. This method of synthesis was chosen due to the heterogeneity of interventions and methods allowed in included studies, however, the approach has been criticised for its potential lack of transparency and introduction of bias via the focus on a select few studies (189). Therefore, it is possible that this choice of analysis introduced bias into the review, however, proactive attempts to minimise bias were put in place by structuring the review around elements of economic evaluation methodology identified as pertinent to public health in the previous review by Weatherly *et al.* (48). Each study was examined for relevance to each methodological challenge to ensure equal representation of all the literature during the review.

4.5.4 *Implications for further empirical work*

The findings from this review informed both the qualitative interview study (Chapter Five) and later economic evaluations (Chapters Six to Eight). At the time, only literature identified by the initial review in May 2016 was available, therefore, the following implications draw only upon the data obtained in the early review. Four points of further investigation were identified and are listed below.

This review has been unable to identify a clear trend in the adoption of alternative economic evaluation methods. Figure 4.3 demonstrates the maintenance of the CUA prominence until 2016. A slight proportional decline in studies reporting a CUA is evident from 2015 and from 2012 onwards CBAs were published with more regularity. Nevertheless, there is insufficient evidence to determine whether this is the beginning of a genuine movement away from the precedence of CUA, a temporal anomaly, or just natural variation. If, as some contemporary literature suggests (17, 171), CUA potentially lacks relevant scope for current public health decision-making, the continued reliance on this technique by health economists is concerning for the applied use of evidence. Interest in CUA and alternative techniques should, therefore, be explored with PHDMs during the qualitative interview study reported in Chapter Five.

No evidence of priority-setting techniques was found in the 2016 review. As discussed in section 4.5.1, this does not preclude their use in the field but merely suggests projects may remain unpublished. The use of priority-setting techniques within local public health contexts should, thus, also be an area of exploration during the qualitative interviews with PHDMs.

Equity considerations remain a rarity in the literature. Some evidence of sub-group analysis was found; however, none of the literature from the 2016 review explicitly discussed equity implications. The importance of this to PHDMs should be examined and any sub-group analysis relevant to decision-making identified.

In a similar fashion to the lack of equity considerations, intersectoral costs and consequences were also not prevalent in the retrieved literature. Whether the incorporation of non-health outcomes into health economic evaluations in public health is important to decision-makers should be established. If so, this will require some radical change to the current economic evaluation paradigm.

4.5.5 *Validation of empirical work using the review update*

Decisions were made for the further empirical work in this thesis on the basis of the initial May 2016 review, however, the updated review is able to validate to some extent these choices. Recent literature has brought into focus the use of more holistic methods able to incorporate broader consequences such as SROI, CBA, and MDCA. Interest in these methodologies justifies exploring these methods in later empirical work (elaborated on in Part Two of this thesis).

4.5.6 *Remaining gaps in the evidence base*

Taking into account all literature identified between 2006 and 2019, there remain gaps in the evidence base for economic evaluations of preventive alcohol interventions. Monetary valuation of outcomes directly using WTP estimates is still lacking amongst studies conducting a CBA. The reliance on financial values of QALYs to value non-market goods such as health may put evaluations at risk of excluding potentially relevant value and failing to extract the full potential from the CBA method. Additionally, the use of SROI and priority-setting methods is minimal. The context of the only identified SROI is fairly narrow, reporting an evaluation of a hospital-based intervention, which required the presentation of drinkers

into the hospital setting. No SROI studies were identified in community or non-healthcare settings, which, arguably, may be better suited to evaluate interventions attempting early prevention of alcohol misuse.

The identified literature also conducted minimal exploration of drinking in underage groups. Only Ingels *et al.* (232) evaluated an intervention targeting adolescents and Sassi *et al. (211)* included a school-based programme within their multiple interventions. Whilst not targeted at young people, three of the SAPM reports included an underage population of either 11-17 years (208, 214) or 16-17 years (213) for sub-group analysis of the alcohol policies evaluated. If early prevention of alcohol misuse is a key priority, more research focussed in this area would be recommended.

4.6 Summary

This review has identified a stronger prevalence of alternative methods of economic evaluations being used in recent years with CUA remaining the modal choice. Some improvement has been observed with reference to the methodological challenges introduced by Weatherly *et al.* (48), such as the use of modelling alongside RCTs, incorporation of broader outcomes, the use of sector-specific QALYs, and the consideration of equity implications in a small number of studies. However, gaps remain with reference to addressing methodological challenges, such as the non-existent use of WTP values for outcomes, and also with reference to the evidence base for preventive alcohol interventions, for instance, interventions targeting underage drinkers.

Finally, the review findings implied four areas for further empirical consideration in this thesis: PHDM's interest in CUA and alternative techniques, the use of priority-setting techniques within local public health contexts, the importance to PHDMs of equity considerations in economic evaluations, and the importance of incorporating non-health outcomes.

Chapter 5. The use of health economic tools by public health decision-makers: a qualitative study

This chapter reports on a qualitative investigation of the understanding and use of health economic tools to aid decision-making by PHDMs in North-East England. The study guided the further empirical work presented later in this thesis in the following ways: (i) further informing the selection of economic evaluation methods to present at the workshop (described in Chapter Ten), and (ii) affirming the decision to focus the remainder of this thesis on economic evaluation methods to the exclusion of other priority-setting techniques. This chapter also addresses the second overarching research question of the thesis outlined in Chapter One. This chapter, therefore, explores the extent of use and understanding of health economic tools, and investigates barriers to the use of health economic evidence by PHDMs.

The first section of this chapter provides the rationale for conducting the qualitative study, followed by an outline of the aims and objectives of the interviews in section 5.2. Section 5.3 describes the methods followed in order to recruit participants and collect data. It also describes and justifies the analytical approach used to synthesise findings. This is followed by a presentation of the study findings in section 5.4 and the final section discusses these findings in relation to current literature and explores the implications for further research.

5.1 Introduction

Public health in England has experienced radical change within the past decade. As described in Chapter One, English public health departments transferred from the NHS back into local government in April 2013. Consequently, public health decisions are being made by agents operating in a different culture compared to that of the previous 40 years, who have priorities that extend beyond population health to the broader well-being of the populace. The relocation of the function brought questions over how resource allocation decisions are made for public health interventions, and by whom, into sharp focus.

The move of public health back to LAs has triggered qualitative research into the process and context of decision-making in public health. Willmott *et al.* (246) explored the experiences of DsPH who have been advocating for public health investment in LAs amidst cuts to local government funding. The study raised the importance of having reliable evidence of the impact of investment, particularly that which demonstrates savings to the LA. The authors

additionally identified a need for research on the impact of public health initiatives on other sectors within the LA remit. Oliver & de Vocht (247) surveyed policy-makers in Greater Manchester in order to understand how evidence is defined and used to influence policy. Their findings revealed extensive use, and attributed value, of local data. Such data includes epidemiological, historical and interpersonal information. Academic research evidence, such as meta-analyses, systematic reviews and clinical trial reports, was reportedly less influential in the sample of 82 surveyed. However, in this study no mention was made of health economic evidence.

Furthermore, Wye *et al.* (248) also identified a preference for local data over national data or research evidence. Their study highlighted the highly pragmatic nature of local commissioners, with a preference for using evidence able to create compelling cases for action. Wye *et al.* (248) concluded that *"we may need to adapt our role as researchers"* (p.10) in order to produce evidence of use to commissioners. Marks *et al.* (53) similarly reported the importance of local knowledge in public health priority-setting. The authors also highlighted the impact of the LA context on priority-setting decisions in public health from needing decisions to be accountable to the local electorate to prioritising public health initiatives within the LA's broader responsibility for health and well-being.

Several other studies examined the use of evidence in local public health decision-making (249-251), reporting similar findings to those noted earlier (53, 246-248). A common finding from all of this research is that researchers need to take greater responsibility for producing relevant evidence by gaining a deeper understanding of decision-makers' requirements, rather than assuming policy-makers will dedicate time and effort to develop the skills necessary to comprehend the evidence provided. Due to this distinction between the presentation and grasp of evidence, Denford *et al.* (250) reported it being underused by public health practitioners.

The existing literature proposes the importance of understanding the evidence requirements of those with decision-making roles and observes a divide between the information provided and what is perceived by decision-makers to be relevant. Whilst the existing qualitative evidence base in this area does not consider health economic tools specifically, one could expect a similar case of asymmetry amongst health economic researchers and PHDMs.

In order to investigate the use of health economic tools by PHDMs and their needs with respect to the information provided by health economic evidence, a qualitative exploration

was undertaken using semi-structured interviews with local PHDMs in North-East England. Furthermore, which, if any, of the tools used to generate such evidence may prove most beneficial to PHDMs work was also investigated.

5.2 Aims and objectives

The previous chapter (Chapter Four) identified four points for further consideration based on the current evidence base on the use of economic evaluations of public health interventions: (i) exploring PHDMs' interest in CUA and alternative techniques, (ii) exploring the use of priority-setting techniques within local public health contexts, (iii) exploring the importance to PHDMs of equity considerations in economic evaluations, and (iv) understanding the relevance to public health decision-making of incorporating non-health outcomes. These four points contributed to shaping the discussions with PHDMs during the semi-structured interviews. Points (i) and (ii) were addressed by objective two, below, and points (iii) and (iv) were incorporated into objective four.

The aim of the qualitative research reported in this chapter was to address the second overarching research question of the thesis. In so doing, the study intended to explore the extent of PHDMs' use, knowledge, and perceived barriers to use of health economic evidence for decision-making. Furthermore, the study aimed to identify whether one or more health economic evaluation tools could be identified as most beneficial to current local PHDMs. Five objectives were addressed to meet these aims:

Objective 1: To understand the local public health decision-making context and how that impacts on the use of economic evaluation evidence

Objective 2: To explore current understanding and knowledge of health economic evaluation tools amongst PHDMs

Objective 3: To explore how extensively PHDMs use economic evaluation evidence to aid decision-making

Objective 4: To identify information requirements for public health decision-making

Objective 5: To explore barriers to the use of health economic evidence to aid decision-making as perceived by PHDMs.

5.3 Methods

This qualitative study used a thematic framework analysis of semi-structured one-to-one in-depth interviews with PHDMs in the North-East region of England. This chapter reports the

study according to the best-practice Standards for Reporting Qualitative Research (SRQR) checklist (252); the completed SRQR checklist for this study is reported in Appendix D.

5.3.1 *Study population and sample selection*

The study population for the semi-structured interviews comprised individuals with the capacity to influence public health decisions in North-East England. A purposive sample of individuals was recruited to cover a diverse range of roles within public health decision-making. A purposive sample is a non-random selection of information-rich cases for in-depth study (253). Purposive sampling can provide a resource-effective means of accessing information relevant to exploring the question of study. Given the wide range of individuals in roles which influence local public health decision-making (54), purposive sampling provided an efficient means of gathering pertinent data and minimised the risk of bias from the exclusion of relevant participants.

A sampling frame determines individuals who are eligible for inclusion in a qualitative study, taken from the study population (254). The sampling frame for this study was divided into three branches of individuals from North-East England: public health specialists, public health practitioners, and LA councillors with a health brief. Specialists held roles such as DsPH and public health consultants who had undertaken specialist public health training. Practitioner roles included commissioning and leading speciality area portfolios (e.g. drugs and alcohol). Individuals in these roles do not typically receive the equivalent specialist training as DsPH or consultants, although, they remain involved in public health decisions via the commissioning and oversight of public health services. Councillors with a health brief were the elected leads for a public health-related portfolio and were actively engaged in public health decision-making in their LAs.

Reaching an appropriate sample size in qualitative research is typically decided retrospectively on the basis of achieving “redundancy” or “saturation coverage” of data, a point at which no new themes emerge during data collection and concepts are repeated multiple times (255). Consequently, an approximate sample size of between 15 and 20 individuals was estimated to be sufficient with the final sample anticipated to be driven by the data and individuals’ willingness and availability to participate (256).

A purposive sample of individuals from the sample frame described earlier was identified from local council web pages, contacts known to my supervisory team, and individuals

suggested by the interviewees themselves. Individuals selected for recruitment were sent a copy of the study information sheet (available in Appendix E) along with an invitation to participate via e-mail. Individuals who did not respond to the e-mail after two weeks were sent a follow-up email. Non-response to the follow-up email was taken to mean that the study was not of interest to the individual concerned and no further contact was made. Recruitment ran from October 2016 to August 2017.

5.3.2 *Pilot testing*

Prior to the commencement of the interviews, four pre-study pilot interviews were conducted with two public health registrars, one member of Newcastle City Council's Health and Wellbeing Board (Wellbeing for life), and a member of a local alcohol campaign group which collaborates with LAs on relevant campaigns. The aim of the pilot interviews was to develop the topic guide, test the interview schedule and interview techniques, and to aid in the identification of potential participants for the main study. The data from the pilot interviews were not included in the final analysis, although, they provided useful contextual information which helped guide the main study interviews.

5.3.3 *Data collection*

In-depth semi-structured interviews were conducted using open-ended questions, enabling full accounts of the views and experiences of participants to be expressed (257). Semi-structured interviews follow a predetermined set of questions which can be asked in any order to allow participants flexibility in their responses and interviews are designed to *"unfold in a conversational manner"* (258, p.143). An interview schedule was developed to guide the interviews, which encapsulated a topic guide for questions covering the areas outlined in the study objectives (section 5.2). Development of the topic guide drew on data from pre-study pilot interviews. The topic guide comprised of six topic areas which covered each of the study objectives; these are outlined in Table 5.1. The interview schedule containing the final version of the topic guide is reproduced in Appendix F.

The flexibility of the semi-structured interview method allowed additional questions to be posed in order to explore emergent issues and experiences that may not have been anticipated prior to conducting the interviews. Consequently, whilst each interview followed the topic guide, the precise questions and ordering differed between interviews according to the interviewee's own responses to the predetermined questions.

Table 5.1 The five topics addressed by the topic guide

Study objective	Topic from topic guide
1	Role in decision-making
1	Decision-making process
2	Understanding of health economic evidence
3	Use and opinions of health economic tools
4	Information required to inform decisions
5	Barriers to use of health economic tools

All the interviews were conducted either in person or over the telephone according to the interviewees' preferences. Each interviewee was sent an electronic "glossary of definitions" document (Appendix G), which covered simple definitions of economic evaluation and priority-setting tools which would be discussed during the interview. Interview participants were invited to read through the short document in advance of the interview; reassurance was given that prior knowledge of the tools listed in the glossary was not a necessary requirement for taking part in the interview.

The interviews were audio recorded and each recording was transcribed verbatim by SH. Following transcription, each recording was replayed to ensure congruence between the written transcripts and audio. Transcripts were anonymised replacing interview participant names with a study identification number; geographical locations and LA names were also removed from transcripts. Audio recordings were subsequently deleted following transcription to maintain participant anonymity.

5.3.4 *Data analysis*

Analysis of the interview transcripts followed a thematic framework approach (259, 260). The framework method was developed for social policy research but has seen recent popularity in health research and is a common approach to thematic analysis of semi-structured interview data (260).

The framework method provides a systematic approach to analysis and thematic analysis is a method to identify, analyse and report patterns within qualitative data (261). Whilst the current study adopted a thematic approach, comparisons between individuals in the three branches of the sampling frame (described in section 5.3.1) are also discussed. The analysis

entailed six stages (familiarisation, coding, developing a thematic framework, indexing, charting, and mapping and interpretation). They are each described below.

Step 1: *Familiarisation*

The initial stage of analysis is familiarisation with the material gathered to place instincts about the data that emerged during data collection in the context of the material as a whole (259). Immersion in the data was achieved through transcription of interviews, re-reading transcripts, and making notes on emerging patterns in interviewees' responses.

Step 2: *Coding*

A code in qualitative thematic analysis refers to a concise label which displays the essence of a meaningful segment of raw data (262). Codes convey the analyst's interpretation of importance within data, regarding interviewees' actions, behaviours, incidents, beliefs, or emotions, or can refer to impressionistic elements observed by the analyst (e.g. a participant misusing terminology) (260). Transcripts were carefully examined and codes applied to sections of text until all the data were coded. Supplementary coding of two (10%) randomly selected transcripts was conducted by a qualitative expert within my supervisory team, Emily Henderson (EH), early in the coding process. A discussion of the codes developed by EH and I directed the coding process for the remainder of the transcripts.

Step 3: *Developing the framework*

A thematic framework was developed using the codes generated in Step 2. Codes were examined both within and between transcripts and were refined based on their similarities. Refined codes reflecting views or experiences on a similar issue or phenomena were categorised in groups to form themes (260). The framework then comprised a set of distinct themes, each containing several common codes. Specialist qualitative software (Nvivo11 QSR International) was used during the coding process. Use of computer software is not essential but can be beneficial for editing and refining existing codes.

Step 4: *Indexing*

The framework was developed using codes from a small sample of interview transcripts. The collated codes were then applied to the subsequent transcripts (this process is formally referred to as indexing (259)). During the indexing process, additional codes emerged as new perspectives and experiences were revealed by interview participants. The framework was,

thus, a fluid entity that was refined during the indexing process until all transcripts had been analysed (259). All transcripts were revisited once using the final, refined framework to ensure the indexing was consistent across all transcripts. To validate the indexing process, a copy of the framework and an indexed transcript were distributed amongst my supervisory team for examination.

Step 5: *Charting*

A framework matrix was generated to contain the coded data from each transcript. A separate matrix was created for each theme in the thematic framework. This process is formally referred to as charting (259, 260). Each row in a matrix represented an interview participant and each column designated a code within the theme. An example of the matrix layout is presented in Table 5.2. Transcript data belonging to each code was summarised from the transcripts and recorded in the appropriate cell of the matrix (represented by the shaded area in Table 5.2). The use of Nvivo simplified this task as the exact transcript text belonging to each code could be easily identified. Charting the data enabled it to be viewed as a whole and allowed cases to be compared and contrasted by reading down each column (259).

Step 6: *Interpretation*

The conclusive step in any qualitative analysis is interpreting the findings. The data matrix and additional notes made throughout the analysis process were examined to synthesise the data in order to produce a narrative which addressed the study objectives (section 5.2). Key objectives and features of qualitative research, such as mapping the range of phenomena, defining concepts, finding associations, and providing explanations (259) were attempted to synthesise the data. Discussions with my supervisory team throughout the synthesis process were beneficial to encourage critical reflection on the interpretation of the data, as recommended for qualitative research (260).

Table 5.2 An example of a framework matrix

	Theme 1 Context		
Interviewee	**Code 1** Budget cuts	**Code 2** Political influence	**Code 3** Cost savings
Interviewee 1 (Councillor)			
Interviewee 2 (Practitioner)			
Interviewee 3 (Specialist)			
Interviewee 4 (Practitioner)			

5.3.5 *Characteristics of study sample*

A total of 20 individuals were invited to take part in the study; 15 agreed to be interviewed, three declined, and two did not respond after a follow-up invitation. Between October 2016 and August 2017, 18 interviews were conducted, which included three follow-up interviews with three respondents. The characteristics of the study sample (n=15) are reported in Table 5.3. Interview length was on average 48 minutes (range: 28 – 66 minutes).

Table 5.3 Table of study sample characteristics

Characteristic	N (%)
Gender	
Male	8 (53%)
Female	7 (47%)
Role	
Specialist	6 (40%)
Practitioner	6 (40%)
Councillor	3 (20%)
Location	
Newcastle	3 (20%)
Gateshead	3 (20%)
Durham	1 (7%)
Sunderland	2 (13%)
South Tyneside	1 (7%)
Teesside	2 (13%)
Northumberland	2 (13%)
*Hackney, London**	1 (7%)
Mode of interview	
In person	11 (73%)
Over the phone	4 (27%)

*This participant was included purposefully, despite not being located in North-East England, after meeting at a workshop organised by Fuse (The Centre for Translational Research in Public Health) owing to his revealed experience with priority-setting methods at the workshop. The participant met all other inclusion criteria.

5.3.6 *Ethical approval and consent*

Ethical approval was granted for this study by Newcastle University Research Office (Reference: 1640/2015). Written consent to participate in the study and to be audio recorded was confirmed at the start of each interview. A copy of the consent form can be viewed in Appendix H.

5.4 Results

The results are presented in the following sub-sections according to each of the study objectives: the local public health decision-making context (section 5.4.1), current understanding and knowledge of health economic evaluation tools amongst PHDMs (section 5.4.2), PHDMs' use of economic evaluation evidence to aid decision-making (section 5.4.3),

informational requirements for public health decision-making (section 5.4.4), and barriers to the use of health economic evidence to aid decision-making (section 5.4.5).

5.4.1 *Objective 1: The local public health decision-making context*

Two of the domains from the thematic framework were relevant to exploring the context of local public health decision-making *(Context* and *Integration with wider systems)*.

Whilst politics has always been a factor affecting decisions in healthcare and public health, interview participants reported that post-transfer of public health to local government *"the politics is much more direct"* [Interviewee 7, Specialist].

The directness of the politics emanates largely from the influence of elected members. They were described as the key decision-makers in a LA, responsible for all major decisions regarding public health expenditure, including the public health budget. However, a discord between the NHS and LAs with regards to addressing issues of public health was expressed by one of the specialists.

> *"I still worry that local government at senior levels doesn't quite understand the business of health. And yet they have responsibilities for commissioning health services. And in the same way…equally, the NHS clearly doesn't understand local government."* [Interviewee 10, Specialist]

The opinion of interviewee 10 was, in part, shared by one of the elected members.

> *"…one thing that a lot of people in the NHS don't get about local government is local and government. You know, we're not a national organisation, therefore, it's not a one-size fits all approach. The other bit is about it's a form of government. And government means that we actually, we're a political organisation."* [Interviewee 11, Councillor]

Given the role of elected members in the decision-making process, public health officers (including both specialists and practitioners) must consider, and take account of, members' priorities in order to gain support for public health initiatives. Public health officers' recommendations must, therefore, be considered *"palatable to elected members"* [Interviewee 6, Practitioner], an issue which became apparent during discussions on certain alcohol policies for which evidence exhibited positive health impacts, yet the political support was not forthcoming due to other concerns.

> *"…there's been I think quite a lot of…a lot of negativity from politicians about it. I think they [elected members] are kind of worried about impact in particular on tourism industry so they are quite keen to understand what is the nature of the benefits in terms of impact on the system?"* [Interviewee 9, Specialist]

Financial pressure on LAs has also been a relevant contextual factor evident in public health decision-making. Significant reductions in the financial settlement with local government since 2010 have instilled a cost-saving mentality in elected members and officers alike. Despite public health teams bringing a ring-fenced public health grant with them following the function's transfer, austerity remains a major concern throughout LAs, as one councillor explained:

> *"Saving money [is] at the forefront of council decisions currently"* [Interviewee 12, Councillor]

Consequently, public health officers described feeling under pressure to demonstrate cost-savings, or specifically *"illustrate actual cashable return"* [Interviewee 7, Specialist] from their actions.

As a result of financial constraints, the trade-off between implementing policies focussed on prevention on the one hand and programmes for treatment of current ill-health on the other, has heightened.

> *"Everybody knows that prevention's the right thing to do but how do you do prevention when you're managing the fires that are happening now?"* [Interviewee 1, Specialist]

The difficulty in diverting funding towards preventive activities was additionally described from a political viewpoint, where it was considered *"brave to put money in prevention because you don't see the results straight away"* [Interviewee 13, Councillor].

Austerity measures were not, however, always viewed in a distinctly negative light. For some respondents they were seen as a facilitator for change.

> *"But it [austerity] has also, it does also create some opportunities for thinking differently about the way things are done."* [Interviewee 5, Specialist]

Other consequences of austerity measures were described, such as the role of the public health department as a rescue service for projects the council can no longer afford. Several interviewees offered examples of services, not traditionally belonging to the remit of the

public health department, obtaining funding via the public health grant in order to prevent their abandonment.

> *"I think people are finding creative ways of spending the public health grant on all things we wouldn't have traditionally spent it on and wouldn't necessarily want to spend it on either"* [Interviewee 8, Practitioner]

Whilst not approved of by some officers, this was not always met with pessimism. The opportunity to broaden the remit of public health departments towards a greater focus on health determinants that was more closely aligned with the origins and purpose of public health was welcomed. One interviewee described the importance of the integration of the public health department with the wider council remit as follows:

> *"…the LA is the public health body rather than the public health department"* [Interviewee 5, Specialist]

5.4.2 *Objective 2: Current understanding and knowledge of health economic evaluation tools amongst PHDMs*

The second objective examines PHDMs' comprehension of health economic concepts and familiarity with the various health economic tools (i.e. economic evaluation methods and priority-setting techniques) available to aid decision-making. Data from the *understanding and knowledge* theme from the framework are drawn on to illustrate this discussion.

Interviewees of all roles displayed a general appreciation of principles of efficiency and health economic concepts, including councillors who expressed a desire for reassurance that action is cost-effective.

> *"I would ask for when things are presented to me I would sort of say, "is there evidence that it's cost-effective?"* [Interviewee 12, Councillor]

However, while the concepts underpinning health economic tools, described in the glossary of terms presented to each interviewee, were often professed to be familiar, the health economic vocabulary was not necessarily known.

> *"You know some of the bits in there [in the glossary of definitions], I won't say all of them, that I thought "yeah that goes through your thinking" but I didn't know you'd call it this particular thing."* [Interviewee 11, Councillor]

In some cases, the disparity between an interviewee's use of a concept and its use by an expert was apparent to the individual.

> *"I'm not a public health professional so when I say "cost-effective" I might mean something slightly different to what an economist or erm a health economist might mean."* [Interviewee 12, Councillor]

On the other hand, it was expressed that all public health team members should have a basic understanding of health economics.

> *"…you know everyone's [colleagues in public health] got a level of knowledge around health economics and broader evaluation"* [Interviewee 9, Specialist]

There did, however, appear to be a division between interviewees in different roles regarding their understanding of health economics. Specialists were better able to demonstrate comprehension of the tools and how they might be used than practitioners, who claimed to be familiar with the terms but did not display a true understanding of concepts when discussed further. The quotation below from a practitioner exemplifies the limited ability to differentiate between the nuances of the different tools available.

> *"Everybody has their own definitions of them [health economic tools], they all mean the same thing really, don't they? Pretty much."* [Interviewee 2, Practitioner]

Additionally, specialists displayed familiarity with certain tools, such as CUA, and accurate knowledge of other tools, such as CCA.

> *"I suppose I've leant on other bits of work I've done in other areas…so what we know maybe about cost per QALY, so the cost-utility analysis, but then also I suppose …otherwise think of it as cost-consequence analysis, you know here are the costs and here is a range of benefits and then we'll make some subjective judgements between the services about, to agree if that delivers value for money."* [Interviewee 9, Specialist]

Specialists' knowledge of other methods, such as CBA and SROI, was admittedly scarce.

> *"Erm, I'm less familiar with the broader categories you've got on your cheat sheet here because we don't do those at the moment."* [Interviewee 5, Specialist]

However, non-specialists' use of health economic terms suggested minimal command of the tools' methodological nuances and terminology was used colloquially to describe any benefit resulting from an action.

"So it's that thing of that cost-benefit analysis of like, you do work upstream that will create benefit for you further down the line, maybe in terms of people don't fall into a ditch" [Interviewee 6, Practitioner]

5.4.3 *Objective 3: PHDMs' use of economic evaluation evidence to aid decision-making*

Following closely on from PHDMs' understanding of health economic tools is their use in decision-making. This third objective draws on data from the *Economic evaluation use* theme of the framework to consider how economic evaluation evidence is used to inform current public health decisions.

Health economic data was regarded as important. In response to a question on how much of a place economic evaluation has in public health, one interview participant replied:

"Oh, critical place. We need to do more. More and more and more" [Interviewee 10, Specialist]

LAs' financial situation was cited as a reason to focus attention on determining the value of public health programmes.

"And actually, cost-effectiveness is an important one given the budget challenges that we've got at the moment" [Interviewee 1, Specialist]

However, despite an aspiration to use economic evaluation evidence, in practice it was not always deemed appropriate to aid decisions on the ground, therefore, limiting its use in practical matters around commissioning services.

"So that's sort of an esoteric, academic way of thinking about things then actually… Sometimes there's a gap between marrying up you know, what the evidence says and how you actually make something work in a practical way." [Interviewee 6, Practitioner]

Where health economic evidence was used, the tools used appeared to differ between roles in the public health department. The previous section (5.4.2) touched on specialists' use of CUA, which was often via consultation of reports produced by NICE. However, it was acknowledged that QALYs, as a measurement of value, have become less beneficial since public health departments transferred to local government.

> *"…we tend to use NICE quite a lot to make a case and they tend to use the QALYs, you know and that sort of thing. But again, I think NICE is more understood in NHS terms than necessarily in local government."* [Interviewee 1, Specialist]

> *"Well at the moment we use cost-utility analysis because those data tend to be available and they are helpful, but they are less helpful in a broader context than the NHS"* [Interviewee 5, Specialist]

Despite claims that CUA is less helpful in LA settings, the preceding quotation describes using it over other tools typically on the basis of availability; a situation in which CUA was held applicable was when conversing with stakeholders with a clinical background, such as Clinical Commissioning Groups (CCGs). However, one Specialist suggested a desire by NICE to incorporate alternative methodologies despite a current paucity of relevant evidence.

> *"…they're [NICE] wanting for public health to move towards broader, much less focussed, more cost-benefit or cost-consequence analysis. I haven't really, I haven't particularly seen that there's been a great shift in that way, yet… I mean yeah, there aren't many cost-benefit analyses that I have to say I've found I've been able to use and I think mainly that's...the lack of them."* [Interviewee 9, Specialist]

On the other hand, practitioners most commonly referred to using ROI/SROI and CBA. It transpired that these interviewees were referring to evaluations integrated into a commissioning tool for drug and alcohol treatment services as part of the National Drug Treatment Monitoring System (NDTMS).

> *"So we'd definitely be using the cost-benefit analysis, social return on investment, definitely. We'd be using cost-effectiveness analysis… and those are all tools that have been developed specifically to use alongside the datasets that we use."* [Interviewee 14, Practitioner]

Part of the NDTMS toolkit embedded a CEA and CBA, and recently a separate SROI tool. Access to the toolkit is strictly monitored, however, a discussion with a PHE employee who helped develop the toolkit indicated that the CBA uses a monetised value of the QALY in its calculations.

The glossary of definitions was also used to gain insight into interviewees' opinions of the potential for adopting health economic methods that may not be currently in use. Specialists discussed interest in CBA and SROI due to their ability to cover a wider scope of benefits.

"Well, I think cost-effectiveness analysis is very important. I also think cost-benefit analysis the more of those the better. I think social return on investment is very important and I would like to see more emphasis given on that. …because I think it [CBA], for me, my understanding of it is it's more, looks at more parameters than simply cost-effective analysis." [Interviewee 10, Specialist]

Some interviewees, however, had some concerns about CBA:

"Erm, but it obviously doesn't account for quality-adjusted life within that so er, I think the QALY is the better tool to use." [Interviewee 7, Specialist]

The preceding quotation indicates a limit to the understanding of CBA, since a CBA has the potential to capture all relevant benefit arising from some action. Providing the benefit valuation is designed appropriately, a CBA has the ability to account for quality-adjusted life. For instance, stated preferences could be elicited for the outcome of a programme in which changes to quality of life are explicitly defined in the scenario being valued (see Chapter Three for detail of stated preference elicitation).

An alternative method for incorporating a wide range of outcomes is MCDA, which was reportedly used in some LAs, and was typically appreciated by those interviewees experienced with the priority-setting technique. The following quotation reflects both the positive and negative experiences of the process required to conduct an MCDA.

"…it increased everyone's knowledge of everyone else's work areas, it provided actually some really good summaries…of what value are we getting from our interventions, it's informed our budget and recommendations to the, to cabinet and to the portfolio holder. It's made us much clearer about the evidence base, and I think it's given a kind of confidence within the team that everyone has been involved in assessing the value of different services and programmes…I guess the difficulties were the kind of time and resource it takes" [Interviewee 9, Specialist]

5.4.4 *Objective 4: Informational requirements for public health decision-making*

The fourth objective of this study was to elicit information considered necessary, or highly beneficial, to aiding public health decisions. The intention behind this objective was to ascertain whether any of the existing economic evaluation tools are best-suited to fulfil the

requirements expressed by PHDMs or whether current tools could be adapted to provide relevant information.

Some elements of beneficial evidence have already been addressed in the previous sections, such as returns from investment to other LA departments (section 5.4.1) and demonstrable cost savings to the council compared to savings to other health-care sectors like the NHS.

> *"…all of these savings on health, well that's brilliant, but they're all savings which will be achieved by the NHS and the CCG. Where are the savings to the LA?"* [Interviewee 8, Practitioner]

Addressing health inequalities was frequently declared as a top priority for the council. The ability to differentiate the impact of an intervention on various subgroups within society was, thus, reported as important.

> *"The resources are scarcer and scarcer and scarcer, we've got to be targeting where they're going to have the most impact in terms of reducing inequalities."* [Interviewee 11, Councillor]

Considering long-term impact was also described as imperative to PHDMs. The difficulty promoting prevention was highlighted earlier, in section 5.4.1, and is reiterated again here.

> *"…but we're better off putting our focus there even though we know that's really long-term stuff. You know, it doesn't have that immediate impact but actually if we're thinking about population health that's really what we need to be thinking of."* [Interviewee 1, Specialist]

The quote below from a councillor exemplifies the need for reliable evidence on the long-term impact of interventions to be available, particularly in the current financial climate, to make a case for preventive interventions, which may not realise benefits until some future point in time.

> *"…the challenge there is for public health teams to explain why erm, investing in prevention is a sensible thing to do when money is tight"* [Interviewee 12, Councillor]

5.4.5 *Objective 5: Barriers to the use of health economic evidence to aid decision-making*

The final objective of this study was to examine barriers reported, or implied, to the use of health economic tools to aid public health decision-making. The findings reported in this section were identified in the final domain of the thematic framework, *barriers to economic*

evaluation. Several of the barriers presented here have been identified in earlier sections of this chapter so will only be covered briefly.

Objective 3 (section 5.4.3) discussed the availability of evidence precluding the use of certain health economic tools. For example, specialists reported limited published evidence of CBAs or SROIs, therefore, resulting in sub-optimal use of those tools, as interpreted by interviewees' proposed interest in alternative evaluative methods. The evidence available, additionally, was reported as not always relevant to LA decisions, thus, reducing its benefit to local decisions.

> *"…it's about taking that evidence from academics and NICE and Cochrane review and all that and using all that and localising it because when we have commissioned some services that are based on massive literature reviews and evidence, I think they don't then necessarily work at a local level"* [Interviewee 8, Specialist]

Section 5.4.2 examined the limited understanding of health economic tools amongst some PHDMs. One interviewee even expressed concern that important skills have been lost amongst some public health officers he had encountered.

> *"…in terms of literature searching, reviewing evidence, critically appraising evidence, understanding of the hierarchy of evidence. Some of the things that I sort of took for granted a few times…I've possibly just been a bit surprised that they haven't necessarily had the knowledge and understanding…we get so bogged down with politics, in not having enough time to spend looking at published evidence that I think, you know some of these skills do get lost."* [Interviewee 9, Specialist]

The above quotation implies that not only is the narrow appreciation of health economic tools amongst many public health officers a barrier to their use, but a further obstacle may also lie with more basic evidence interpretation skills.

If the barriers of evidence availability and capacity to effectively use the evidence to inform decisions are overcome, that still does not mean that economic evaluation evidence will ultimately direct action. There remain further factors which determine public health decisions, notably, politics and priorities.

The political context discussed earlier in section 5.4.1 described a decision-making hierarchy in LAs, with elected members being the ultimate decision-makers. The approach to evidence-informed decision-making by elected members, however, is not necessarily the

same as the one understood by public health officers. As one specialist described, elected members respond to different types of evidence, potentially leading to disconnect if economic evaluation evidence conflicts with other evidence privileged by elected members.

> *"I mean elected members are very much guided by the information they get from their officers as well as their own opinions and what they get from the public...I think it's probably different evidence as well. So, it's not so much kind of, clinical evidence or solid public health evidence, it also incorporates a kind of, what does it mean to citizens? What does it mean to members of the public?"* [Interviewee 7, Specialist]

Thus, in addition to ensuring political acceptability of proposals, as discussed in section 5.4.1, officers must also present evidence provided using health economic tools in a way which is acceptable to elected members and appeals to their preferences. Examples were provided during the interviews of instances where elected members' personal priorities have influenced the funding of interventions which public health officers did not recommend or opposed based on effectiveness grounds. The following quote from a specialist described one experience of recommending the withdrawal of a popular service for which evidence on its effectiveness and cost-effectiveness did not support its continuation.

> *"…if a service is very popular with members you either can't touch it because you're never going to get the decision through or you're going to have to do a lot of work with members to educate them as to why you're doing that, get them on board"* [Interviewee 15, Specialist]

Furthermore, even if health economic evidence is available and has been identified by public health officers, there was a perception that the evidence was superfluous to the outcome of the decision. In the case reported by Interviewee 9, below, the economic evidence didn't necessarily support implementation of the service compared to alternative options. However, a case had already been made to direct funding towards the programme, therefore, it went ahead regardless of the economic evaluation evidence.

> *"...but broadly it [health economic evidence] was mainly from the NICE guideline that I used, mainly, cost-utility analysis erm, information but I suppose the interesting thing is I'm not quite sure what kind of impact it had. Essentially the agenda...there was to get a specialised service up and running because he'd put in a business case for it and he*

wanted to make sure this helped. Well, it came up somewhere in the middle...Well, there was new money for it, the new money was earmarked for this.

Yeah [the outcome had already been decided], you know...the degree to which it made an impact I don't know." [Interviewee 9, Specialist]

5.5 Discussion

This chapter reports a qualitative investigation with PHDMs. The findings reported in this chapter address the second overarching research question of the thesis via a discussion of the extent of use and understanding of health economic tools, and the identification of barriers to the use of health economic evidence, by PHDMs. The study additionally explored the needs of PHDMs with regards to health economic evidence and investigated which economic evaluation methods may prove most useful in assisting decision-making. The study investigated these issues using semi-structured interviews with 15 individuals from a range of roles relevant to public health decision-making in LAs in the North-East of England.

Variation in evidence use and knowledge of health economic tools was identified between individuals of different roles in the LAs. Additionally, throughout the interviews, the political nature of LAs was referenced as heavily impacting the evidence base used to inform decisions. Several informational requirements, addressing the four points for consideration identified by the systematic review in Chapter Four (see section 5.2), were identified alongside a range of barriers to the use of economic evaluation hailing from the political context and the supply of relevant evidence. These results are discussed in detail below, drawing on comparisons with existing literature. Finally, implications for further research are provided.

5.5.1 *Summary of findings and their relation to existing literature*

An overarching finding from this qualitative study is the complexity surrounding the decision-making processes in LAs. Public health decisions draw on a spectrum of evidence from academic research, public opinion, and political preferences. Additionally, the multi-sectoral nature of LAs, which are responsible for a myriad of functions besides public health such as education, transport, and planning, necessitates public health decisions being made in a context which considers, and takes into account, their impact on broader council objectives. This stems from an understanding that funding to one area of the LA forgoes funding in another; the cuts to local government budgets since 2010, commented on

extensively by interviewees, stresses the opportunity cost of funding initiatives for any LA department to an even greater extent.

Economic evaluation has emerged and developed to provide information on the relative value of different courses of action, and has been used to inform healthcare decisions made in a budget-constrained environment by organisations such as NICE (11). NICE have also drawn on economic evaluation evidence to inform their public health guidance (31). On paper, economic evaluation should also be a valuable aid to local public health decisions. However, the findings reported in this chapter show that whilst health economic evidence is viewed as important, and public health officers aspire to use it more, several barriers prevent its optimal use.

The complex process of decision-making in public health creates obstacles to the use of economic evaluation evidence. To begin with, the direct political influence impacting on public health decisions since the transfer of responsibility back to LAs in England has been referenced extensively by interviewees throughout this study. Factors such as council priorities, elected members' preferences, and distinct features of the local area are all reported to have a bearing on public health decisions. These additional criteria were reported as at times coming into conflict with each other if the political will to enact a policy did not align with the evidence supplied by public health officers, or if the health economic evidence suggested disinvestment in a programme which had significant backing from the public and other stakeholders, whose preferences elected members valued greatly.

Existing literature reports similar findings with respect to the discovery that economic evaluation evidence in its current format insufficiently provides information required to guide local public health decisions. Several commentators (56, 247, 248, 263) have reported discrepancies between the evidence available, often reporting outcomes on a national scale, and the local evidence valued by local decision-makers. Views were expressed in the current study, and in wider literature, that translating existing evidence into the practical needs of local authorities is difficult and can act as a barrier to the use of academic evidence, including economic evaluation evidence, in decision-making. Furthermore, a review examining evidence use in English LAs arrived at the conclusion that available evidence is insufficiently sensitive to the heterogeneity of local decision-making cultures (249), echoing the findings from this study that greater emphasis is needed to provide locally relevant information as opposed to national level data.

However, previous literature examining the application of economic evaluation evidence in local healthcare decision-making when public health remained the responsibility of the NHS (264) also demonstrated limited use of economic evaluation approaches. Therefore, whilst the direct political context of current public health decision-making has undoubtedly affected the use of health economics in LAs, the onus cannot be placed solely on political influence. The unique complexities of decision-making in a public health context also play a large role in health economic evidence use.

Reflections from the interview participants revealed a narrow use of health economic evidence and provide insight into how tools may be improved to add value to public health decisions. Unique aspects of public health interventions relevant to decision-making were revealed during the interviews, several of which concern the four areas outlined for further consideration in Chapter Four. Two of the points were closely linked: PHDMs' use and opinion of CUA and whether incorporating broader outcomes is important. The findings from this study suggest that PHDMs, most often specialists, do use evidence from CUA, yet express views that this tool is perhaps insufficient for public health decisions due to the multi-sectoral nature of the interventions they are scrutinising. Health economic methods that are better suited to incorporating a wider breadth of outcomes were viewed as providing value in the LA context, yet the absence of evidence utilising those methods was cited as the main barrier. The lack of exposure to published evidence such as CBA and SROI certainly had an impact on interviewees' familiarity with the tools; however, interest in using the tools if available was expressed in several instances.

Another consideration from the systematic review in the previous chapter was the use of priority-setting techniques, which have featured minimally in the published literature. Two priority-setting techniques were discussed during the interviews, PBMA and MCDA, with only MCDA reported to have been used. Several of the interview participants from different LAs (Interviewees 7, 9, 12, and 15) described undertaking an MCDA in their public health department and all remarked favourably on the tool, principally because it provided a transparent and systematic tool to deal with difficult decisions, such as areas where disinvestment is needed to adjust to budget cuts.

Despite the time and resource commitment reported by interviewees to complete an MCDA, an advantage of priority-setting techniques is the breadth of criteria that can be considered, thus enabling a broad spectrum of outcomes to be compared. Several comments were

raised during the interviews regarding the value gained from the process of undertaking an MCDA as much as the outcomes from the exercise, such as engaging the wider public health team and providing transparency to decision-making practices. Much of the priority-setting action observed locally was part of the "*Shifting the Gravity of Spending*" research project, which explored priority-setting techniques in LA public health teams (54, 58). Furthermore, a prioritisation framework, which draws primarily on MCDA techniques, has recently been developed by PHE (265) and evaluated as an extension of the previously mentioned research project (266). The framework was developed for in-house use within teams based in LAs, therefore, suggesting that the hypothesis stated in Chapter Four, that the published literature base underestimates the amount of priority-setting work being conducted, is likely correct.

The importance of incorporating impacts on equity in economic evaluations was the final point for consideration from Chapter Four. Reducing inequity was confirmed in this study as a top priority for LAs and the work around public health particularly in ensuring access to services and targeting scarce funding towards members of the population most in need. Economic evaluation evidence that is able to account for heterogeneity in the population relevant to health inequalities, perhaps by socio-economic status or other identifiable criteria, could be beneficial for targeting scarce resources if value differentials are identified between the different sub-populations. Incorporation of impacts on equity in economic evaluation evidence would require subjective judgements on the part of decision-makers where trade-offs are identified between maximising overall population benefit and addressing equity issues. Tools such as MCDA can assist in making these judgements via the act of consensual weighting of decision criteria, therefore, reducing some of the subjective nature of difficult decisions.

A complementary qualitative study, published since the completion of the study reported in this chapter, examined the use of health economics in public health in an English LA (56). The study, conducted in the West Midlands between May 2016 and June 2017, identified similar findings to those reported in this chapter with regards to the politicised decision-making context, the requirement for a broader scope in outcomes evaluated, and limited knowledge of health economics amongst those working in LAs. Frew & Breheny (56) focused on the context of decision-making and the barriers to use of economic evidence, with less regard paid to the appropriateness of specific economic evaluation or priority-setting

methods, which this chapter has sought to explore. The complementary findings from Frew & Breheny (56), however, usefully serve to validate those reported in this chapter.

Given the interest expressed by PHDMs to incorporate economic evidence to a greater extent in their decision-making and the inadequacies of the current evidence base reported in this study and others (56, 249), it is imperative that research progresses in this area to establish how to make health economic evidence more accessible to PHDMs. The findings discussed here justify the continued exploration of economic evaluation methods and ways in which that evidence is disseminated that are better able to address the complexities of local public health decision-making processes.

5.5.2 *Reflexive statement*

The nature of qualitative interviews involves an *"exchange between researcher and the researched"* (267, p.111). As such, the impact of the researcher on the ensuing conversations cannot be overlooked. Each interview participant was aware that the study was interested in examining the use of economic evaluation evidence, and may, therefore, have felt pressure, however unintentional, to over-emphasise their interest in health economics or over-state familiarity with tools in order to appear better-informed (268). By way of minimising any imposed bias, interviewees were reassured that no prior knowledge of the health economic concepts discussed was a pre-requisite to being interviewed.

The knowledge that the interview findings would contribute to the writing of this thesis, and that the interviews were being audio-recorded, may have led respondents to be less candid in their responses, particularly councillors who are spokespeople for their local council. However, the assurance that all data would be used anonymously with any identifiable features removed should have reduced this effect.

Additionally, the themes drawn out from the transcript data may have been influenced by my background in health economics. Data discussions with the non-health economist members of my supervisory team were intended to mitigate against this risk, and to dilute any unintended focus on select issues at the expense of alternatively valuable insights revealed by the data.

5.5.3 *Strengths and limitations*

In exploring the demand for economic evidence in public health decision-making, Kneale *et al.* (249) note that relevant studies lacked exploration of the awareness and use of available decision-support tools. This empirical study has addressed an identified gap in the literature and has provided a unique investigation of PHDMs' knowledge and awareness of a range of health economic decision tools.

This study included interviewees from a range of LA roles, from specialist public health officers to locally elected members of the council, which ensured the viewpoints of individuals from all perspectives of the local decision-making process could be explored. Purposive sampling, however, ensured that breadth was not sought at the expense of sufficient depth.

However, this study is also subject to possible limitations. Firstly, the findings are a manifestation of the perspectives of decision-makers from LAs in one region in England (with the exception of one councillor). LAs across the country are subject to unique contextual factors based on their specific political leadership and decision-making processes have been demonstrated to be particular or specific to each organisation (53). As a result, this study cannot generalise results to other areas in England. However, as noted above, the findings from Frew & Breheny (56) suggest that LAs across England may share similarities in their decision-making cultures and evidence use, therefore, the findings reported in the chapter should have a wider relevance.

Secondly, the relatively small sample size of 15 interview participants may be considered a potential limitation. Recruitment for the qualitative research was, however, determined by reaching "redundancy" (255) regarding the study objectives. With regards to the sample of specialists and practitioners, redundancy was considered to have been achieved when no new codes emerged in the transcripts. Data from the sub-group of councillors were less convergent, therefore, saturation of the perspectives of elected members cannot be claimed. Nevertheless, this component of the doctoral research was intended as a modest study to surface issues for subsequent inquiry. To that extent, the 15 interview participants were sufficient to generate information to meet the purposes of the study and provided direction for the doctoral project.

5.5.4 *Implications for further research*

The findings from the study reported in this chapter have implications for the direction of the remainder of this doctoral research.

Given the recent and substantial research into priority-setting tools through the *Shifting the Gravity of Spending* project (58, 266), the remainder of the research in this doctoral project concentrates exclusively on methods of economic evaluation. This decision was based, additionally, on the findings from this empirical study, which suggest the use of priority-setting tools is an interactive process conducted within public health teams. The use of existing evidence, including economic evaluations, is part of the MCDA process. Therefore, improving the evidence base of economic evaluations, which should, in turn, enable effective prioritisation of resources, was deemed the most beneficial course of action to aid public health decision-making.

No single method of economic evaluation prevailed with regards to preferences expressed by the interview participants. However, interest was expressed in exploring alternative methodologies, which are better suited to examine the multi-sectoral context of local public health decisions. Therefore, CBA and SROI were identified as suitable candidates for further exploration, neither of which were identified as prominent in the economic evaluation of public health intervention literature featured in the systematic review reported in Chapter Four. CCA additionally holds promise for examining multi-sectoral impacts of public health programmes, due to its ability to present a broad range of costs and effects, although its lack of an aggregated outcome may cause concern for PHDMs. Consequently, these three economic evaluation frameworks would benefit from further exploration, especially given PHDMs' reported limited exposure to them.

Interviewees also expressed a desire for locally relevant information, therefore, evaluations conducted in Part Two of this thesis should consider local analyses where possible. Furthermore, the limited appreciation of the nuances of health economic tools imply communication of economic evaluation evidence should be presented in a suitable non-technical fashion to enable engagement from PHDMs.

5.5.5 *Summary*

This study provided a qualitative exploration of the use of various economic evaluation and priority-setting methods in multiple LAs in North-East England. This research was timely given the transfer of public health responsibilities back to local government in 2013 and PHE's current interest in advocating prioritisation frameworks.

The interview study reported varying degrees of understanding of health economic tools amongst the interview participants and narrow use of published economic evaluations, typically due to issues regarding literature availability. Toolkits designed specifically for the commissioning of drug and alcohol treatment services were claimed to report CEA and CBA (using monetised QALYs) outcomes. However, the quality of these tools could not be verified, and their coverage is severely limited compared to the wider public health remit. The political culture of LAs was noted to have affected decision-making in public health from both an evidence standpoint, i.e. necessitating novel sources of evidence, and a procedural standpoint, i.e. ensuring recommendations are acceptable, and in keeping with the priorities, of elected council members. Furthermore, interviewees reported it necessary to incorporate broader outcomes into economic evaluations and emphasise demonstrating cost-savings to the council due to unprecedented fiscal constraints.

This study has provided relevant evidence for subsequent empirical work reported in Part Two. The remainder of this thesis focuses exclusively on economic evaluation methods, in particular exploring the relevance to PHDMs of CBA, SROI, and CCA compared to CUA, which was identified as the most commonly used source of economic evidence currently.

PART II. Economic evaluation empirical work

Chapter 6. Development of the contingent valuation survey

The qualitative study reported in the previous chapter did not identify a preferred health economic tool for evaluating public health interventions. However, interest was expressed by the PHDM interviewees in exploring tools that can incorporate a more holistic range of outcomes than solely health benefits, as these would better appeal to the multi-sectoral context of LAs. Frameworks such as SROI (84), CBA (81), and CCA (113) were identified as potentially relevant economic evaluation methods to meet these needs.

The PHDMs interviewed were, however, largely unfamiliar with the nuances of these tools. To establish whether these largely unknown techniques would be valuable to public health decision-making, an initial empirical study was conducted that involved introducing examples of the methods to PHDMs. The aim of the remaining empirical work presented in this thesis is, therefore, to produce exemplars of each of the economic evaluation methods shortlisted as potentially beneficial and to present PHDMs with the resulting evidence. Feedback on the appropriateness and usability of each tool for decision-making can be assessed after exposure to each method.

The systematic review reported in Chapter Four identified a paucity of studies utilising SROI, CBA, or CCA methods. Although a small number of CBA studies were identified, none valued benefits according to a method that is consistent with economic welfare theory, as outlined in Chapter Two, in which CBA is grounded. To measure benefit in a way that captures the holistic nature of many public health interventions, valuing benefit using stated preference techniques may be preferable to monetising QALYs, which only capture health-related outcomes (269).

Consequently, it was necessary to conduct new economic evaluations using the methods mentioned earlier. To conduct a CBA capturing a holistic measure of benefit, a study to first elicit stated preferences was required. Chapter Three detailed methods able to elicit stated preferences: CV surveys and DCEs. A CV survey was chosen for the empirical work conducted for this thesis since an overall measure of benefit was sufficient for conducting a CBA to present to PHDMs and a CV survey can be less cognitively burdensome for respondents than a DCE (160).

The qualitative study also suggested that PHDMs were more familiar with CUAs than alternative methods of economic evaluation and, additionally, CUA was the most prominent

type of economic evaluation identified in the systematic review reported in Chapter Four. Therefore, presenting a CUA alongside the SROI, CBA and CCA could provide a reference point to compare the alternative methods against. For a fair comparison of the four techniques, basing each evaluation on the same case study is preferable. Doing so ensures that PHDMs' assessment of the methods is influenced only by factors related to the techniques and any preference based on the intervention evaluated is avoided.

The first section of this chapter, therefore, outlines the case study used for all four evaluations. Section 6.2 introduces the survey setting and sample. Section 6.3 outlines the development of the CV survey and the pre-testing and pilot stages. Section 6.4 illustrates the structure of the overall survey. Section 6.5 outlines the ethical approval for the survey and section 6.6 summarises the chapter. The results of the CV study are reported in Chapter Seven.

6.1 The case study: SIPS Jr HIGH

The Screening and Intervention Programme for Sensible Drinking (SIPS) Jr HIGH study (59) is an RCT evaluating an ASBI programme in a school setting for young people aged 14-15 (Year 10 in English school years). The trial compares ASBI for risky drinkers with standard practice for alcohol issues in schools in four locations in England (North-East, North-West, South-East and London).

Secondary prevention interventions are adopted during the early stages of a disease to treat and minimise the damage incurred. ASBI is considered secondary prevention because it specifically targets those who already consume alcohol, potentially increasing the salience of intervention compared with prevention that is aimed at all individuals regardless of alcohol consumption status (270). The effectiveness of ASBI at reducing alcohol consumption across a broad definition of young people (10-21 years) has been demonstrated previously (62). However, no ASBI studies had been conducted previously in a UK school setting; therefore, the SIPS Jr HIGH trial aimed to evaluate the effectiveness and cost-effectiveness of this form of secondary prevention in a select group of young people in a school setting.

6.1.1 *The intervention*

Students were eligible to take part in the trial if they were aged 14-15 and scored positively on the Adolescent Single Alcohol Question (A-SAQ)[5], demonstrating risky drinking behaviour. The A-SAQ was developed by the SIPS project team as a modified version of the Single Alcohol Screening Question (271) that is suitable for adolescents. Eligible and consenting students were randomised into either the control group or intervention group. The control group received a healthy lifestyles leaflet, which provided information on healthy eating and physical activity. The lifestyles leaflet contained no information on alcohol and students were not provided with any feedback from their alcohol screening. The intervention group received a brief intervention with a trained learning mentor, which comprised of a 30-minute interactive session. During the session, students were provided with feedback on their alcohol screen results and engaged in an interactive discussion intended to raise awareness of the risks of alcohol consumption and encourage students to consider motivations for altering their behaviour around alcohol.

6.1.2 *Economic analysis*

Students were followed-up after 12-months to examine whether the intervention had affected the alcohol consumption of the students in the trial. Additionally, a range of secondary outcomes such as alcohol-related problems, smoking status, health-related quality of life (measured using the EQ-5D-3L[6]) and psychological wellbeing were also assessed (for the full list of outcome measures see Giles *et al.,* 2019 (59)).

A health economic evaluation was undertaken as part of the SIPS Jr HIGH project to examine the cost-effectiveness of the ASBI intervention. The economic evaluation consisted of both a CCA and CUA using QALYs (derived from responses by the students to the EQ-5D-3L) as the outcome measure. The CUA was conducted from the perspective of the UK public sector, therefore, costs (and savings) associated with service use were estimated based on self-reported use of health, social, and other public sector services (as is common practice in economic evaluation (272)). The collection of data related to public sector service use for the

[5] A positive score was a response to the A-SAQ of any of the following regarding frequency of alcohol consumption over the past six-months: '4 or more times but not every month', 'at least once a month but not every week', 'every week but not every day', or 'every day'

[6] The EQ-5D-3L questionnaire was chosen over the version specifically designed for young people (EQ-5D-Y) because the EQ-5D-Y currently has no value sets which are necessary for the calculation of QALYs and EQ-5D-3L state it is possible to use the EQ-5L-3L for young people over the age of 12 years (94).

calculation of costs made this project an attractive study for a CBA and SROI because this information could be utilised to inform evaluations considering a holistic measure of outcomes.

The pre-existing CUA and CCA evaluations (59) were adapted in order to present to the PHDMs alongside the novel CBA and SROI evaluations. The CV study that was conducted to provide a measure of benefit for the CBA used data collected from the trial to create the hypothetical scenarios from which WTP can be elicited. The remainder of this chapter outlines the development process of the CV study.

6.2 CV survey setting and sample

Hanemann and Arrow (273, 274) recommend that CV surveys are conducted face-to-face for optimum reliability, however, the practicalities of conducting interviews in person often limit sample size. A relatively large sample (between several hundred and several thousand) is generally necessary to elicit a mean WTP value that is generalisable to the population of interest, particularly for public goods (275). A compromise was sought for this study by using an online survey. The online format enabled data collection from a large and varied sample and allowed sufficient information to be provided to participants in a cost-effective manner.

An external market research company specialising in hosting online surveys (ResearchNow[7]) was used to host the CV survey as they were able to access a broad sample of the UK population. Furthermore, they have been used to host CV surveys by other researchers, e.g. Somers *et al.,* 2019 (276). The accessibility of the survey was tested on a variety of mediums (mobile phones, tablets, laptops, and desktop computers) using a range of operating systems (Microsoft, Macintosh, Apple OS and Android) to ensure that participants would be able to complete the survey on whichever device they own in order to reduce non-response and sample selection biases (124).

6.2.1 *Study sample*

A UK general adult (over 18 years old) population sample was chosen for the CV study, although this sample would not be direct recipients of the ASBI. Obtaining WTP valuations from Year 10 students would be have been complex for several reasons. Firstly, 14 and 15-year-old Year 10 students in the UK are unaccustomed to paying for services due to a greater

[7] ResearchNow is now known as Dynata (https://www.dynata.com)

proportion of goods offered free to young people under the age of 18 and in full-time education (e.g. medication prescriptions, eye-sight tests etc.). Additionally, this population is unlikely to have a significant financial budget which is crucial to provide a concept of opportunity cost when considering WTP. Finally, accessing survey participants under the age of 18 using an online survey may not have been appropriate from either the viewpoint of obtaining appropriate consent or ensuring understanding of the stated-preference task. Therefore, alternative means such as face-to-face interviews would have been required, which are resource-intensive and, as stated in section 6.2, would have restricted the sample size.

The choice of a general population sample has implications for the perspective of the WTP study because the CV survey respondents are not the population who would receive the intervention (Year 10 school students). WTP valuations can reflect different values depending on the relation of the respondent to the good being valued; they can either reflect use or non-use values (1, 124). Since the intervention recipients are not included in the CV study sample the majority of WTP values elicited are expected to reflect non-use values, although recipients may also perceive the intervention to generate personal value indirectly through positive externalities arising from the intervention outcomes. Nevertheless, a large component of WTP responses are anticipated to be driven by altruistic, non-use value (1, 277), either by recipients who have adolescent or young children and perceive future value for their offspring, or non-parents who wish for the intervention to be available for young people in general. Fortunately, stated preference techniques are suitable for eliciting both use and non-use values (1) and the ability of CV to obtain non-use value is, in fact, an asset quoted by supporters of the method (124).

6.2.2 *Sample size calculation*

A Bayesian approach to calculating sample sizes for CV studies has been suggested as superior to an ad hoc decision based on available resources (278), however, prior knowledge of WTP values is required for this approach. No published studies could be found eliciting a general population WTP for an ASBI, therefore, the literature was examined to find examples of CV studies from the broader area of public health, including alcohol treatment. Since it was explained in section 6.2 that the mode of survey conduction may have an impact on the sample size collected, literature specifically recruiting online samples was examined to

ensure relevance to this study. Sample sizes ranged from approximately n=312 (279) to n=2146 (280).

Mitchell & Carson (124) set out an alternative approach to determine sample size which can be applied when priors are unknown. Their approach is based on three factors: deviation from true WTP (Δ), relative error (V) and confidence levels (1-α). Equation 6.1 outlines the sample size calculation where Z represents the Z-score from a standard normal distribution Z∼N (0,1) for a given confidence level (1- α).

$$\left[\frac{Z\widehat{V}}{\Delta}\right]^2 \qquad \textit{(6.1)}$$

If no prior evidence is available, the authors (124) recommend assuming a value of 2 for relative error (V). Therefore, taking a significance level of α=0.05, equivalent to a two-sided 95% confidence interval, and deviation from true WTP Δ=0.15, the mid-point of reasonable values (0.05-0.3) suggested by Mitchell & Carson (124), a sample size can be estimated. Substituting the values stated above into Equation 6.1 results in a sample size of 683 (Equation 6.2), which is located within the range of sample sizes identified in the similar literature reported earlier.

$$\left[\frac{1.96*2}{0.15}\right]^2 = 683 \qquad \textit{(6.2)}$$

Mitchell & Carson (124) recommend inflating the sample size to allow for non-responses and protest responses. Since non-responses do not contribute to the sample quota collected by the online survey, considering these was not necessary to estimate the sample size. Protest responses, however, can be difficult to anticipate in advance. Of the similar CV literature reported earlier, only one study (279) reported evidence of protest responses, relaying that 3.2% of their sample gave protest zero responses. Alternatively, other health-related CV studies have reported proportions of protest zeros of between 6%-10% (281) and approximately 24% (282). Given this range in protest zeros, an average of the identified proportions of protesting (~12%) was used to estimate the inflated sample size. A sample size of approximately 765 complete responses was, therefore, considered appropriate for this study.

6.3 Developing the survey, pre-testing and piloting

A series of hypothetical scenarios were developed to describe the intervention and its outcomes. The CV survey aimed to establish the general public's WTP for the ASBI programme contingent on specified outcomes. The outcomes presented in the scenarios were based on those obtained from the SIPS Jr HIGH trial (section 6.1).

The trial primary outcome was alcohol intake over the previous month, measured as units of alcohol. Data were also collected on secondary outcomes, which included both clinical and non-clinical effects. The purpose of conducting a CBA was to include a holistic composition of outcomes, including those broader than health; therefore, the trial primary outcome and two of the secondary outcomes were incorporated in the scenario. The secondary outcomes included were police arrests and school absenteeism. These outcomes were considered relevant in order to incorporate examples of multi-sectoral impacts (e.g. the criminal justice sector), represent the broader impact on young people's general wellbeing, and may also be perceived by survey respondents to produce positive externalities for society if reducing arrests and absenteeism are considered proxy measures for lessening anti-social behaviour.

School absenteeism was classified as five or more days missed from school over six months. Therefore, a student was only considered to have been absent from school if they reported missing a minimum of five days from school (the equivalent of one-week of missed school). The minimum criterion of five days was a distinction made by the SIPS Jr HIGH researchers based on literature which suggests that recurrent absenteeism produces the most damaging effects for student performance compared to infrequent absenteeism (283, 284). Sälzer and Heine (285) additionally demonstrate that young people with regular truancy achieve lower scores on numeracy and literacy tests compared to their peers.

The SIPS Jr HIGH trial had not been powered to detect significant differences in the secondary outcomes, therefore, the arrest and absenteeism outcomes presented in the CV scenario should be considered illustrative rather than definitive. However, the outcomes used were considered appropriate for generating the hypothetical CV scenarios to ultimately conduct a CBA that is comparable to the CUA and CCA evaluations conducted as part of the trial.

Three different scenarios for the intervention were developed. The first scenario depicted outcomes directly obtained from the trial results; the second scenario depicted the same

outcomes as Scenario 1 but also included potential long-run outcomes; the third scenario presented the same information as Scenario 2 but with a greater improvement in alcohol reduction. The purpose of including three scenarios was to examine whether the explicit recognition of long-run outcomes increases the value of the ASBI to respondents (i.e. Scenario 2 compared to Scenario 1) and whether an improvement in a health-related outcome (reduction in alcohol consumption) increases the value of the ASBI compared to only improvement in non-health outcomes (arrests and school absenteeism, i.e. Scenario 3 compared to Scenario 2).

Longer term outcomes were estimated using available literature on long-run effects of school absenteeism and arrests in young people. Findings from several studies (286-291) were adapted to present potential long-run outcomes from the intervention with respect to the likelihood of being arrested as an adult, the likelihood of completing further education and future earnings potential. Due to the uncertainty of long-run outcomes and heterogeneity identified in the literature in terms of long-run impact, the outcomes included in the CV scenarios were presented as increased or decreased chances rather than reporting absolute figures so as not to misrepresent the potential long-run effects of the intervention.

6.3.1 *Pre-testing*

Several rounds of testing the CV survey were conducted to determine that the payment vehicle and elicitation method were appropriate, to finalise the design of the hypothetical scenarios, and to ensure clarity of the survey instructions. Since the CV survey was to be administered online, respondents would have no opportunity to ask for points of clarification, therefore, it was imperative that the instructions were unambiguous to ensure the validity of the survey outcomes.

The first round of pre-testing was conducted with two colleagues who were familiar with the CV approach, in order to obtain expert opinions on the methodological aspects of the survey, such as the elicitation method and payment vehicle. The survey was then tested on five non-economist colleagues and eight members of the general public to ensure the scenarios were relatable and that the instructions could be clearly understood. Following completion of the survey design, incorporating feedback from the pre-testing stage, a pilot test was conducted via a soft-launch of the online survey. This final pilot was critical to

assessing how the general public engaged with the online survey since the pre-testing had been completed face-to-face using printed materials rather than the online version.

6.3.2 *Testing scenario presentation*

Initially, the intervention outcomes were presented textually, either as a percentage change, a proportion of students achieving an outcome, or descriptively (e.g. “significantly fewer arrests”). Feedback from the pre-testing indicated that none of the presentation variants successfully provided information that was easily understood. One of the volunteers testing the material suggested representing the information graphically may be more intuitive to respondents.

The scenarios were adapted to present the outcomes in the form of simple bar charts, which illustrated the change from baseline to 12-month follow-up for both the intervention and control groups. The graphic version was tested alongside the original text versions for comparison. Feedback indicated a clear preference for the graphic version; however, lay public volunteers, who were unaccustomed with trials, were confused about why changes in outcomes were seen in both the control and outcome groups. As a result, some lay volunteers reported mistrusting the results. Consultation with the supervisory team led to the decision to include only the difference between control and intervention groups at 12-month follow-up as this is the most poignant outcome and should minimise confusion for respondents, especially during the online survey when explanation and reassurance cannot be provided. A small reduction in accuracy of the outcomes reported was considered a suitable trade-off to create a scenario which was understandable to respondents and which may elicit greater engagement with the survey.

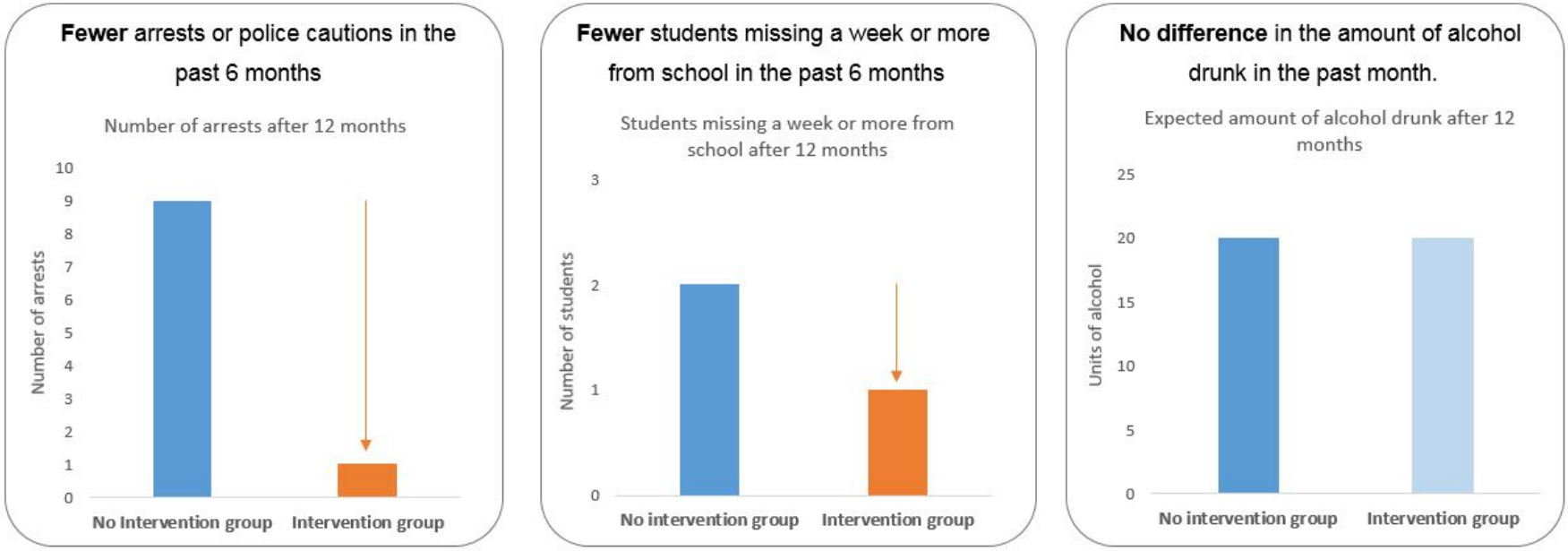
Throughout the pre-tests, particularly those with members of the public, the wording of the scenarios, instructions and background information was assessed for readability. Any text that was unfamiliar or unclear was altered and advice was sought from the pre-test volunteers as to appropriate alternative wording. One issue raised during the pre-testing stage was that the similarity of Scenarios 2 and 3 led to several pre-testers failing to notice that Scenario 3 depicted a change in alcohol consumption compared to Scenarios 1 and 2. To make this distinction clear, the colours used in the bar charts, which displayed each outcome, were varied. Where a difference in outcome is observed between the control and intervention groups (i.e. for arrests and school absenteeism in all scenarios and alcohol

consumption in Scenario 3 only), the control group was displayed in blue and the intervention group in orange. Whereas, in Scenarios 1 and 2, where no difference is observed in the consumption of alcohol between the intervention and control groups, both bars were coloured in shades of blue. This made it easier for respondents to visually discern a difference in the alcohol consumption outcome in Scenario 3 compared to Scenario 1 and Scenario 2. The final versions of the scenarios are displayed Figures 6.1 to 6.3.

Figure 6.1 Scenario one as shown to survey respondents

Twelve months after the brief alcohol intervention (screening plus a 30-minute interactive session), the following outcomes could be expected for risky drinking young people who received the intervention *compared with those who do not receive the intervention*.

Outcomes are based on high quality research of how well the intervention works.

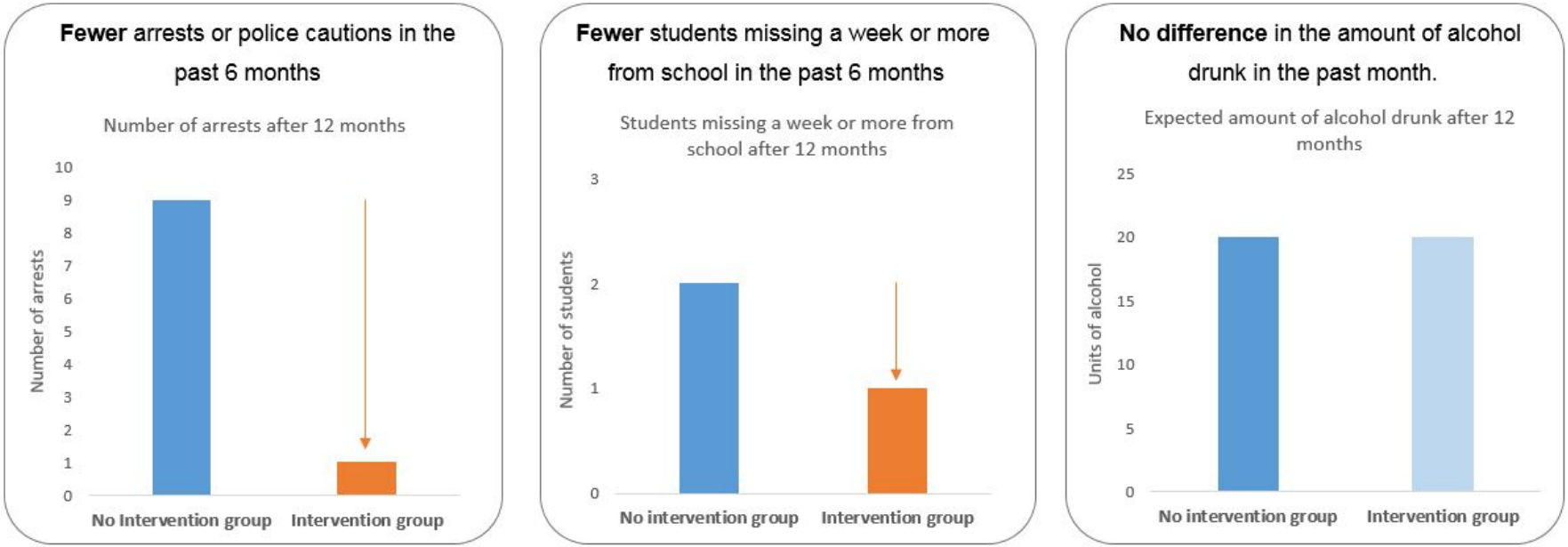


Long run outcomes (more than 12 months after the intervention) are unknown.

Figure 6.2 Scenario two as shown to survey respondents

Twelve months after the brief alcohol intervention (screening plus a 30-minute interactive session), the following outcomes could be expected for risky drinking young people who received the intervention *compared with those who do not receive the intervention*.

Outcomes are based on high quality research of how well the intervention works.



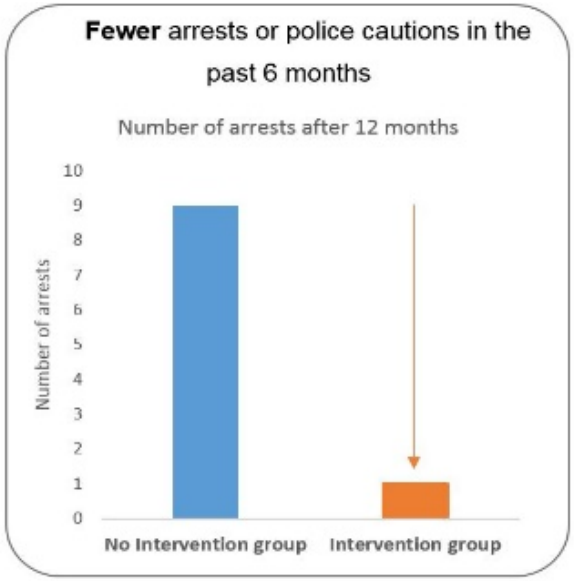
In the longer term (up to twenty years after the intervention), research has suggested that the following outcomes could be expected for those who received the intervention compared to those who did not:

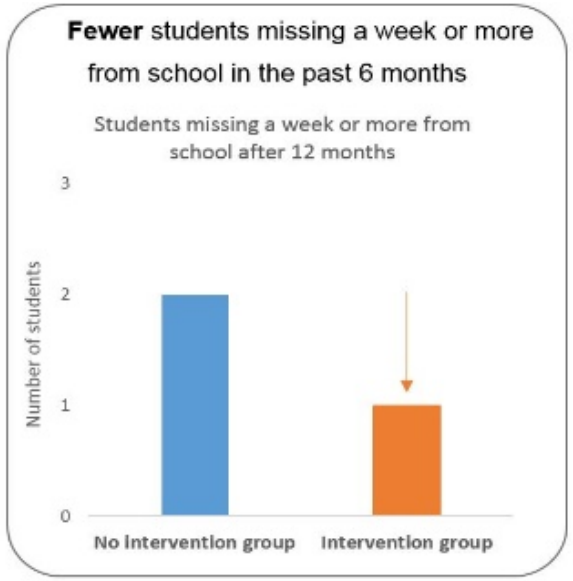
- **Less likely** to be arrested as an adult
- **More likely** to complete high school and start a course at University
- **Receive higher** average wages from a combined effect of fewer arrests as a young person and less time missed from school which is linked to better educational achievement and higher employment rates

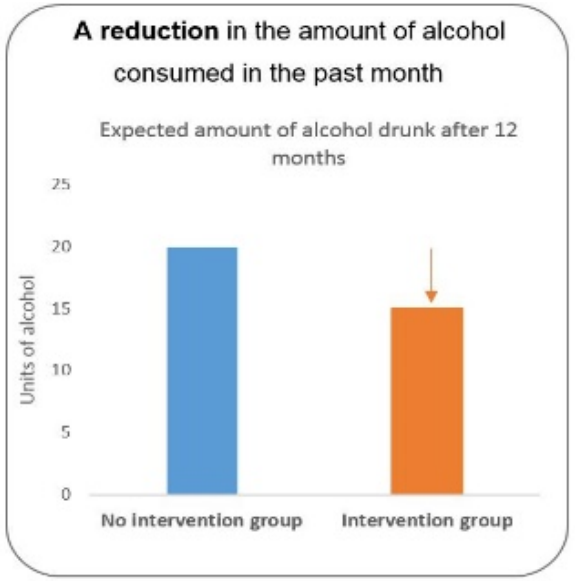
Figure 6.3 Scenario three as shown to survey respondent

Twelve months after the brief alcohol intervention (screening plus a 30-minute interactive session), the following outcomes could be expected for risky drinking young people who received the intervention *compared with those who do not receive the intervention*.

Outcomes are based on high quality research of how well the intervention works.







In the longer term (up to twenty years after the intervention), research has suggested that the following outcomes could be expected for those who received the intervention compared to those who did not:

- **Less likely** to be arrested as an adult
- **More likely** to complete high school and start a course at University
- **Receive higher** average wages from a combined effect of fewer arrests as a young person and less time missed from school which is linked to better educational achievement and higher employment rates

6.3.3 *Testing the payment vehicle*

The choice of additional monthly taxation as the payment vehicle was affirmed during the pre-test with expert colleagues. This mode of payment was chosen on the basis that the intervention being valued is a public health good, which would typically be funded via general taxation in the UK. Individuals in the UK do not routinely pay at the point of access for public health services and the majority of the general adult population are familiar with paying taxes in some form. In keeping with this cultural context, it was, therefore, considered most appropriate to frame payments as an additional monthly tax contribution which would be used directly to fund the ASBI.

The outcomes of the ASBI in the hypothetical scenario reflected the intervention impact after 12 months of implementation. Thus, to remain consistent with the scenario outcomes, a one-year time horizon was also used for the payment vehicle. The one-year duration was compared with a four-year duration[8] during the pre-test stage; a greater likelihood of WTP was observed with a one-year payment period, thus supporting the monthly tax for one year.

The payment question posed to respondents following exposure to each hypothetical scenario was as follows:

> *"Thinking about the intervention and outcomes that have just been shown to you, would you be willing to pay anything for the brief alcohol intervention to be provided to 'Year 10' students in schools in the UK? A payment would be made in the form of extra monthly taxation for one year which would be used to directly fund the intervention."*

A "yes" response would take respondents to the WTP task, which is described in section 6.3.4 below, and a "no" response would end the WTP task (see section 6.3.5 for detail on responses indicating unwillingness to pay).

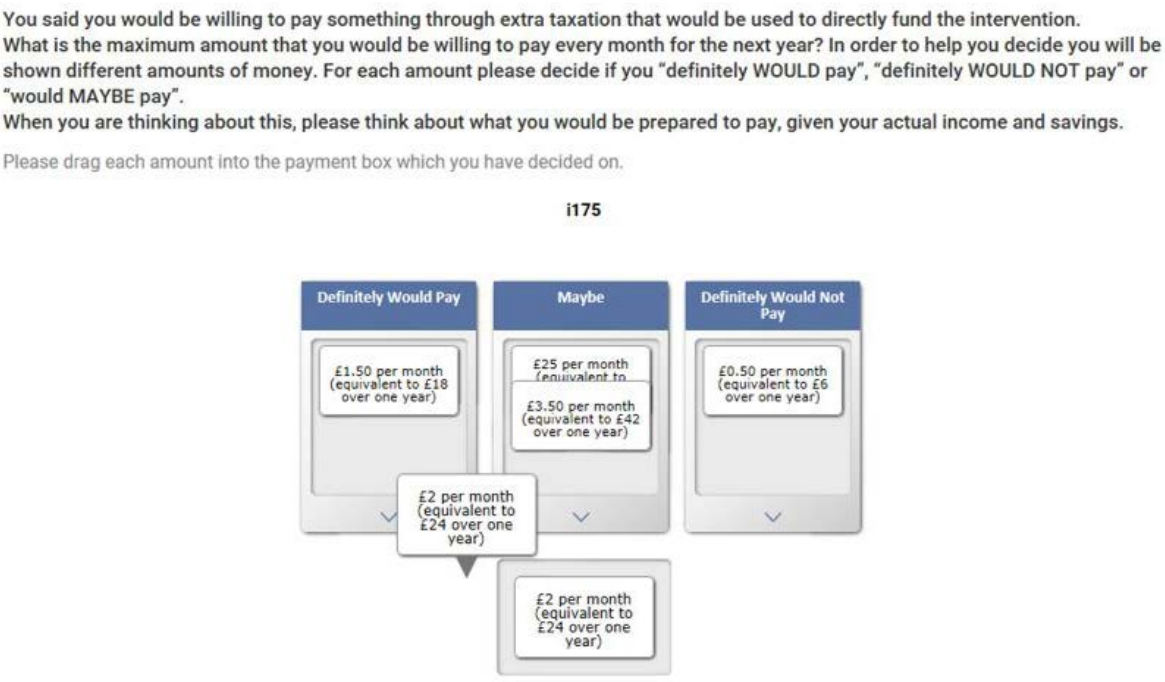
6.3.4 *Testing the elicitation method*

The pre-testing phase with expert colleagues affirmed the choice of elicitation method as the random card sort followed by an open-ended question to obtain a maximum WTP value

[8] Four years is the duration of a term for a locally elected member of council. This period would, therefore, reflect the time-frame that individual elected members may consider for an investment, given the uncertainty of re-election at the end of each term.

(see Chapter Three for detail on possible alternative elicitation methods). The initial range of values presented during the random card sort (per month: £0.50, £1, £1.50, £2, £5, £7.50, £10, £15, £20, £30, £50 and £100) was examined in all pre-tests and, consequently, additional values were added to the lower end of the random card sort scale and the highest value was removed. Doing so aimed to improve the sensitivity of WTP responses since the majority of the pre-tests provided WTP values in the lower region of the random card sort scale. The final range of random card sort values used in the survey was: (per month) £0.50, £1, £1.50, £2, £3.50, £5, £7.50, £12.50, £15, £25 and £50. It was clearly stated in the survey that the duration of tax payments would be one year; to ensure this was not overlooked by respondents, the annual equivalent value of each monthly payment was also indicated on each random card sort payment card (see Figure 6.4).

Figure 6.4 A screenshot of the random card sort task showing a £2 payment card appearing in a random order and the participant moving it to their preferred box: Definitely would pay



*This example demonstrates an illogical placement of cards (£0.50 in the "Definitely would not pay" box where higher values have been placed in the "Definitely would pay" box). This would be identified once all cards had been sorted and the participant would be asked to check their card placements before moving on.

The random card sort was designed to display each of the payment card values in a random order. Respondents who had previously stated that they would be willing to pay for the intervention were instructed to place each payment card in a box relative to the

respondent's certainty of willingness or unwillingness to pay. A reminder was also given to consider what would be feasible to pay given the respondent's budget constraint, to ensure the opportunity cost of their valuation was considered.

> *"You said you would be willing to pay something through extra taxation that would be used to directly fund the intervention.*
>
> *What is the maximum amount that you would be willing to pay every month for the next year? In order to help you decide you will be shown different amounts of money. For each amount please decide if you "definitely WOULD pay", "definitely WOULD NOT pay" and "would MAYBE pay".*
>
> *When you are thinking about this, please think about what you would be prepared to pay, given your actual income and savings."*

The random card sort task is illustrated in Figure 6.4. The online survey was programmed to identify illogical responses during the random card sort task, such as placing a higher value in the "definitely would pay" box than in the "definitely would not pay" box. Respondents would be unable to complete the random card sort task until changes were made to ensure their card sort was logical. Figure 6.4 provides an example of an illogical placement of cards.

Following a logical placement of cards, respondents would be asked an open-ended WTP question, guided by their responses to the random card sort. The open-ended question would allow respondents to state their maximum WTP, bounded by the values they placed in the "definitely would pay" and "definitely would not pay" boxes.

> *"From the list of amounts, you said the highest amount you "definitely WOULD pay" per month is [£X] and the lowest amount you "definitely WOULD NOT pay" per month is [£X].*
>
> *What is the MAXIMUM value you would be willing to pay as extra monthly taxation for the intervention? It could be one of these amounts or something in between.*

The open-ended question displayed above allowed respondents to state any value, within the given bounds, to generate a continuous value of WTP. Alternative instructions were given for the open-ended payment if no cards had been placed in either the "definitely would pay" or "definitely would not pay" boxes during the random card sort. The alternative versions can be viewed in Appendix I, section I.1.

6.3.5 *Distinguishing unwillingness to pay from protesting*

Participants may overstate or understate their WTP for a good for strategic reasons. Strategic action is referred to as *protesting* when respondents (i) state unwillingness to pay despite valuing the good positively, or (ii) offer excessively high or low values which do not represent the respondent's true WTP (130, 292). Individuals may provide protest responses of Type 1 if, for example, they believe that someone else should pay (e.g. the Government), and Type 2 if, for example, they believe that providing a large WTP value will result in a good being provided (130).

In order to distinguish protest responses of Type 1 from zero valuations reflecting a genuine perception that the ASBI lacks value, it is recommended that respondents be asked to justify their unwillingness to pay (124, 275). The motivation for a value of zero can indicate whether the response is a true zero valuation or an act of protest. During the payment question described in section 6.3.3, if a respondent stated that they were not willing to pay anything for the intervention, the WTP task would end and the respondent would be asked to select the reason for their unwillingness to pay from a list of options. Box 6.1 lists the justification options that were developed to distinguish protest and true zero WTP valuations. The options were tested and amended using responses provided during the survey pre-test stages.

Box 6.1 Unwillingness to pay justification options

1. Other interventions are more valuable
2. I am not concerned about the issue of risky drinking in young people
3. I think the intervention is valuable, but I cannot afford it
4. I think the intervention is valuable, but I do not think it should be funded from taxes
5. Other tax payers, who are better off, should pay for it
6. Parents/guardians of 'Year 10' students should pay for it.
7. Other (please specify)

Responses 1 and 2 in Box 6.1 indicate that a respondent does not value the ASBI positively and response 3 indicates inability (as opposed to unwillingness) to pay, therefore, these responses are considered to represent true zero valuations. Identifying protest responses is a challenging task, due to the lack of a standard definition for what constitutes a protest response (281, 292-294). Frey *et al.* state that "*A protest bid is defined as not stating the true WTP value for the good in question for whatever reason"* (293, p.2).

Responses 4-6 were considered to reflect protest action since they indicate that a respondent values the ASBI positively but has other motivations behind their unwillingness to pay, therefore, respondents selecting these responses are not stating their true WTP value. This categorisation is, however, subjective to the researcher, a reported issue in the literature regarding the treatment of protest responses (293, p.2). Response 4 indicates that respondents value the intervention positively but object to the vehicle of tax payments. Response 5 also indicates that respondents value the intervention (since they believe it should be funded by someone, namely wealthier taxpayers) but object to being responsible for this funding. Although this response could be seen as similar to being unable to afford to pay anything, response 5 is subtly different because the vehicle of tax payment is introduced. It was expected that respondents whose motivation for unwillingness to pay is purely inability to pay would select response 3, thus, response 5 was considered to represent protest action. Response 6 was also interpreted as respondents indicating that the ASBI has value (since they believe someone should fund it, namely parents or guardians of Year 10 students), however, they object to being asked to contribute towards implementation of the ASBI. As with response 5, genuine lack of interest in the issue of risky drinking in young people was expected to be identified from response 2. Thus, response 6 was considered a protest response.

Examination of the free-text specifications to response 7 was required to determine whether these responses reflect a true or protest zero valuation.

Protest responses of Type 2 can be more difficult to identify, however offering respondents the opportunity to elaborate qualitatively on the amount they are willing to pay is recommended (130) to attempt to distinguish unreasonable WTP values that may indicate strategic responding. Optional comment boxes were, therefore, included in the CV survey to elicit this information.

6.3.6 *The final pilot test*

The final piloting stage was a soft launch of the online survey. The online version was completed by approximately 5% (n=32) of the desired survey sample size (see section 6.2.2). The results from the soft launch indicated public engagement with the survey and whether the payment vehicle was acceptable to most respondents.

Two-thirds (n=19) of the respondents left comments in an optional comment box, which asked for elaboration on respondents' willingness or unwillingness to pay for the intervention. This was perceived as a good level of engagement with the survey since this section was not mandatory. One comment indicated mistrust in the motivation behind the survey. Consequently, an extra sentence was added to the information page to explain that the survey was being used exclusively for research purposes by researchers at Newcastle University and would not be used to increase taxes.

There did not appear to be any significant objection to the tax payment vehicle from those who were unwilling to pay. However, information was not available regarding non-responses because data from only completed surveys were collected.

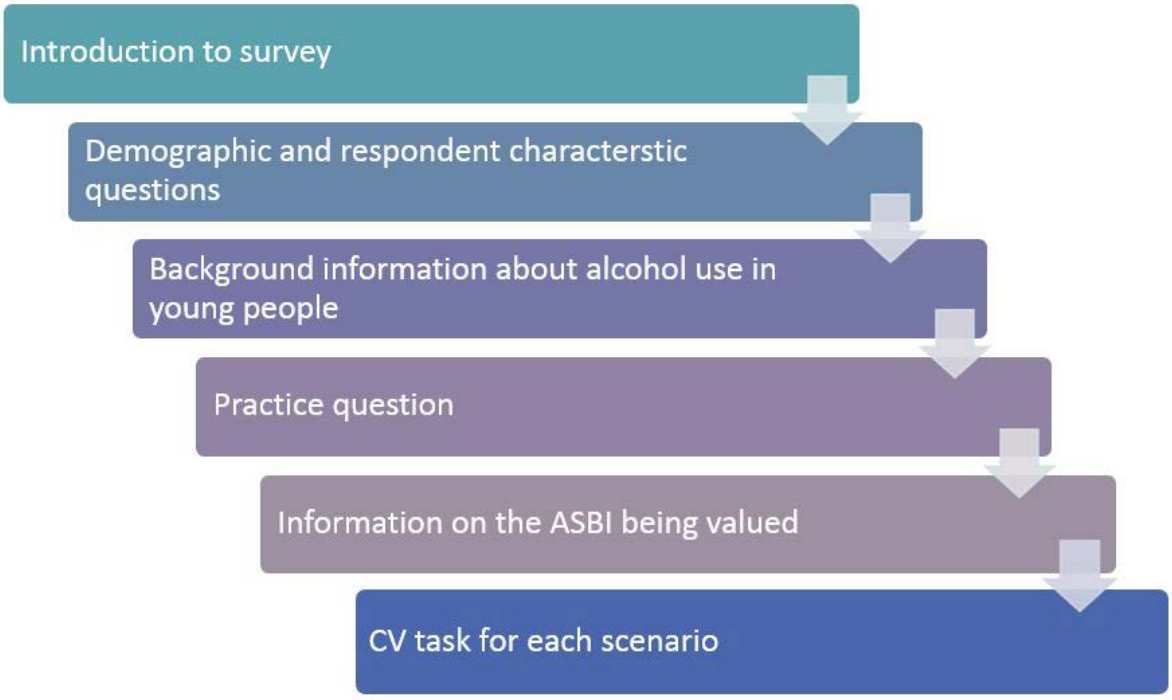
Following the soft launch, no changes were made to the scenarios or payment vehicle.

6.4 Survey structure

In addition to the WTP questions, a short questionnaire was presented to respondents to elicit demographic information and information that may be salient to WTP responses, such as whether an individual is a parent or guardian of a child under the age of 16 and the respondent's alcohol consumption. Standard demographic details were requested such as age, location (by UK region), gender, household income, employment status, marital status, and education level. Respondent characteristics would be used as covariates during the WTP data analysis to test for predictors of WTP and to test the theoretical validity of the CV survey (see Chapter Seven for details on the WTP analysis).

The structure of the online survey is outlined in Figure 6.5. The practice question used a more compact version of the random card sort and a hypothetical scenario unrelated to ASBIs. The purpose of the practice question was to introduce the elicitation method to participants in order to reduce errors in the main CV survey that may be caused due to unfamiliarity with the WTP task. The full survey, including the practice question, can be viewed in Appendix I.

Figure 6.5 The structure of the online survey



6.5 Ethics

Ethical approval for the CV study was granted by the Newcastle University ethics committee (Reference number: 4106/2018) on 26/02/18. Respondents were provided with information at the start of the survey which explained the purpose of the survey, who was conducting the research, that participation was voluntary, and that responses would be anonymous. Continuation past the information page was considered informed consent to take part in the study.

6.6 Summary

This chapter has outlined the case study (the SIPS Jr HIGH trial) that provided the context for the CV task. The development process required to generate a reliable CV survey has also been detailed, including justification of the online setting and the required sample size. The description of each stage of the online survey has been outlined including designing the hypothetical scenarios, determining the payment vehicle and elicitation method, and putting measures in place to distinguish protest responses. Each stage of the survey has been pre-tested and piloted, and this process has also been recorded in the chapter. The following chapter (Chapter Seven) provides details on the analysis of the WTP data collected and discusses the survey results.

Chapter 7. Contingent valuation study

Chapter Six outlined the development of the CV survey and the approach taken to collect data. This chapter details the results of the online data collection and the analysis that was conducted on the WTP data.

The first section of this chapter outlines the objectives of the data analysis of the CV survey. Section 7.2 outlines the methods of analysis employed to explore the data. Section 7.3 presents the results obtained from each analytical objective. Section 7.4 discusses the results in relation to the existing literature and outlines the limitations of the study and areas for further research. The final section provides a concluding summary of the study.

7.1 Analytical objectives

The overarching aim of this CV study was to obtain a WTP benefit value for the ASBI programme for a CBA. In addition to this aim, the analysis of the WTP data had four objectives:

1. Examine descriptive statistics of WTP for use in the cost-benefit analysis (CBA)
2. Examine the difference between WTP for the three scenarios
3. Examine predictors of WTP to test theoretical validity of the CV survey
4. Identify protest responses and examine the characteristics of respondents who protest

7.1.1 *Objective 1*

The first objective was relatively simple since maximum WTP was elicited via an open-ended question following the random card sort task. The WTP values could, therefore, be treated as continuous variables necessitating no adjustment to ascertain mean and median values to be used in the CBA (295).

7.1.2 *Objective 2*

The second objective aimed to examine the relative value that the study sample, representative of the UK general public, placed on non-health and health impacts of the ASBI, and long-run outcomes compared with short-run outcomes. Both the consideration of effects broader than health and long-run outcomes are elements of public health interventions that have been documented as methodologically challenging to address in economic evaluations, as discussed in previous chapters (Chapters One and Four). Whether

these factors place any weight on public valuations of interventions was, therefore, of interest to explore. Supposing that these factors do affect valuations, it was hypothesised that:

i. WTP values would be larger for Scenario 3 than Scenarios 1 and 2 because there was a reduction in the consumption of alcohol
ii. WTP values would be larger for Scenario 2 than Scenario 1 because long-run outcomes were included which may present the outcomes with greater certainty and impact

7.1.3 *Objective 3*

The third objective was to examine the characteristics of respondents that might predict WTP. Certain theoretical expectations can be tested during this examination, for example, characteristics such as income and education level are generally theorised to impact WTP whilst factors specific to this study, such as being a parent, guardian or grandparent of a child under 16 years of age, may be associated with higher WTP. The aim of the third objective was to examine whether the WTP results reflect the a priori theoretical expectations.

The hypothesised direction of impact on WTP of each characteristic is outlined below. Details on the mechanism for impact and theoretical evidence are provided in Table 7.1.

Higher income was theorised to positively impact WTP via ability to pay; that is to say respondents with higher household income would be expected to offer higher WTP values. Higher educated respondents were similarly theorised to offer greater WTP for health-related interventions due to either greater health literacy or, in the context of the current study, greater risk aversion. The association between education and risk may be more relevant given the breadth of outcomes in the hypothetical scenarios.

Other demographic characteristics may affect respondents' WTP, however, the direction of the effect is more ambiguous. The CV literature valuing other public or healthcare goods has demonstrated contrasting results regarding the effect of gender, age, and marital status on WTP, indicating that these factors are study specific (276, 279, 296, 297). Married or cohabiting individuals, however, may indicate greater likelihood of WTP due to greater financial security compared to single respondents.

Table 7.1 Mechanism and theoretical evidence for impact of predictors on WTP

	Predictor	Hypothesised direction of impact on WTP	Mechanism for impact
Demographic characteristics	Income	Positive	Ability to pay. It has been previously documented that individuals on higher incomes would be expected to pay more than those on low incomes (298-300).
	Education	Positive	Greater health literacy and risk awareness. Higher educated individuals may have a greater awareness of the importance of good health. Additionally, alcohol consumption can lead to risky behaviour. Individuals who are more risk averse may be expected to pay more for an intervention that reduces risky behaviour, such as engaging in activities which results in being arrested. Jung *et al.* (301) show a positive relationship between education level and risk aversion
	Married/ cohabiting	Positive for likelihood of WTP	Married or cohabiting individuals may experience greater financial security if two incomes are contributing to one household.
	Gender	Ambiguous	Insufficient evidence of direction
	Age	Ambiguous	Insufficient evidence of direction
Study specific characteristics	Parent or guardian of a child <16 years	Positive	Stronger altruistic motivation from parents compared with non-parents. Additionally, the potential for indirect benefit to parents from a possible positive change in their child's behaviour as a result of the ASBI.
	Drinking frequency	Ambiguous	Frequent drinkers may regret the choices they made regarding alcohol consumption and, therefore, support early intervention; alternatively, they may not view their drinking behaviour as problematic and thus, not value the ASBI at all. Non-drinkers, who perhaps choose not to drink due to the effects they perceive from alcohol consumption, may support the ASBI and express greater WTP than drinkers of other frequencies.

Two additional factors specifically related to this study were hypothesised to have an effect on WTP. Firstly, whether a respondent is a parent, guardian or grandparent of a child under the age of 16 was expected to be associated positively with WTP since parents of children who could benefit from the ASBI may value it more than either parents of older children who could not benefit, or individuals without children[9].

Secondly, drinking frequency was hypothesised to impact WTP, however, the direction is ambiguous. Non-drinkers may be expected to offer higher WTP in support of the ASBI due to their own drinking preferences, while frequent drinkers' preferences may depend on their attitude towards their current level of alcohol consumption.

7.1.4 *Objective 4*

The final objective was to identify protest responses and examine the characteristics of protest responders. Examining the characteristics of respondents who protest is important to (i) identify characteristics that may predict these responses in future in order to minimise protesting, and (ii) consider whether the sample of protesters is different to the sample of non-protesters. The latter point is particularly important if protest responses are removed from the sample prior to analysis since sample bias may be introduced if there is a significant divergence in characteristics between protesters and non-protesters (302).

Common practice in the literature has been to remove protest responses from study samples (124, 294, 303, 304) since it is argued that protest responses are not true reflections of the value a protest responder places on the good in question (294, 302). It is valid to consider that protesters who offer a zero value of WTP may value the ASBI positively, yet state unwillingness to pay due to a rejection of the method of payment or objection to the survey. Therefore, removing protest zeros from the sample can reduce downwards bias introduced to estimates of WTP by the "false" zero valuations (302). Conversely, removing protesters offering extremely high WTP values, due to the perception that stating value far above their true value may secure implementation of the intervention, can reduce upwards bias on WTP.

[9] For simplicity and brevity, henceforth, parents, guardians, or grandparents of children **over** 16 years of age and individuals without children will be referred to as "non-parents". Furthermore, parents, guardians, or grandparents of children **under** the age of 16 will be referred to simply as "parents".

7.2 Methods

Each of the objectives outlined in section 7.1 require a particular analytical approach to examine the data. Each of these approaches will be outlined in turn below. All analysis was conducted using Stata 14 (305).

7.2.1 *Objective 1: Estimate measures of central tendency*

Mean and median WTP was estimated to obtain a measure of benefit that can be used to populate the CBA (reported in Chapter Eight). Prior to estimating WTP, responses that were identified as protests, using the approaches discussed in Chapter Six, were removed from the sample, as discussed in section 7.1.4. To examine the uncertainty of the estimate of mean WTP, confidence intervals at the 95% level (95% CI) and standard errors[10] were estimated by bootstrapping the sample values using 10,000 repetitions[11].

Sensitivity analysis on the mean and median estimates of WTP was also conducted to examine the effect of trimming the data to exclude very large values which may distort the mean WTP, and to include protest responses in the sample since an alternative method of addressing protest responses is to include them in the sample as zero values (302).

7.2.2 *Objective 2: Examining differences in WTP between each scenario*

In order to examine whether WTP values differed significantly between each of the three scenarios, the non-parametric Skillings-Mack test (307) was used. Distributions of WTP data are often skewed (308), therefore, a non-parametric test, which does not require normality assumptions, was chosen. The Skillings-Mack test is able to account for many ties or equal ranks, which would be the case if there are a substantial number of true zero valuations or equivalent WTP values.

[10] The standard error of the mean measures the distance of the sample mean from the likely true population mean. This measure can be used to validate the accuracy of the mean estimate calculated from the sample data. The 95% confidence interval represents a range of values that have a 95% chance of containing the true population mean.

[11] The appropriate number of bootstrap repetitions was determined following the methods reported by Statacorp (306). The use of 10,000 repetitions produced a minimised difference in the standard errors and confidence intervals of two different bootstrap random-number seeds when compared to 12,500 repetitions and 7,500 repetitions.

To further investigate the two hypotheses stated in section 7.1.2, pairwise comparisons of WTP were conducted between the scenarios using the non-parametric Wilcoxon matched-pairs signed-rank test.

7.2.3 *Objective 3: Examining predictors of WTP*

Regression analysis was used to examine predictors of WTP for each hypothetical scenario. Manipulation of the raw data was necessary to facilitate the regression analysis. The majority of the demographic data collected was ordinal or categorical, with a relatively large number of categories to increase response precision. For example, household income had been collected using eight income range categories plus an additional two for "unknown" and "prefer not to say". The "tick-box" method of data collection is particularly useful for data such as income because individuals may not know, or wish to reveal, their precise income; therefore, providing income ranges can provide some ambiguity and potentially encourage fewer non-responses (see Appendix I for the full survey text). In order to make the data more manageable and meaningful during the regression analysis, the ordinal categories of household income, age, educational achievement, and drinking frequency were compressed into three categories per variable. Dummy variables were created for the remaining characteristics: gender, marital status, living in the UK outside of England, and parental status. Box 7.1 lists the final variables used in the regression analysis[12].

Several econometric models can be used to examine WTP data and choosing the most appropriate model depends on the elicitation method used and the distribution of the data. Limited dependent models, such as Tobit, selectivity, or two-part models, are recommended by Donaldson *et al.* 1998 (309) for WTP data elicited via open-ended elicitation methods which contain a significant proportion of zero valuations. As discussed in Chapter Six, although the random card sort technique was employed initially to assist respondents' WTP decisions, maximum WTP was ultimately elicited via an open-ended question. Therefore, analytical approaches considered appropriate for data elicited via open-ended methods were used. A two-part model was employed to examine determinants of WTP for the ASBI due to its ability to account for values of zero WTP.

[12] Data were collected on employment status, however this was excluded as an independent variable in the final analysis due to strong collinearity with income (χ^2 test, $p < 0.000$).

To examine whether the results were sensitive to model specification sensitivity analysis was conducted using two regression approaches commonly used to analyse WTP data, Tobit and log-transformed ordinary least squares (OLS). Tobit models account for censored data and are therefore able to include values of zero WTP. Log OLS transforms the dependent variable (i.e. WTP) onto a log scale to reduce non-normality in the distribution of residuals, which is a necessary condition for unbiased OLS models[13]. However, zero values cannot be included in the analytic sample because "0" cannot be log transformed.

Box 7.1 Variable specification for regression analysis

Dummy variables	
PARENT_GUARDIAN	Parent, guardian or grand-parent of a child under 16 years of age; 1 for parent, 0 for non-parents
MALE	Gender; 1 for males, 0 for females
MARRIED	Marital status; 1 for married/cohabiting, 0 for single/divorced/widowed
NONDRINK	Frequency of alcohol consumption; 1 for drinker, 0 for non-drinker (categorised as consuming alcohol less than 1-2 times per month)
NON_ENGLISH	UK location; 1 for Scotland, Wales or Northern Ireland, 0 for England
Ordinal variables	
AGE	Age; 0 for under 35, 1 for 35-65, 2 for over 65
EDUCATION	Highest level of education; 0 for no formal qualifications, 1 for school level qualifications, 2 for higher education qualification
INCOME	Household income; 0 for less than £20,000, 1 for £20,000-£40,000, 2 for over £40,000

Two-part models are particularly adept at analysing data with a cluster of zero values by dividing the analysis into two parts, as their name suggests. The first part considers the dependent variable as binary to analyse the probability of observing either zero or a positive value. The second part of the model analyses the effect of covariates on the dependent variable, conditional on the probability that the outcome is positive from the first part of the model. The second part of a two-part model, thus, treats the dependent variable as positive

[13] Log transformation was applied to WTP because tests for heteroscedasticity (Breusch-Pagan (310) and Cook-Weisberg (311)) on the raw WTP data suggested errors were heteroskedastic ($p<0.05$). Following log transformation, heteroscedasticity was no longer indicated using the same tests ($p>0.05$).

and continuous. A two-part model is suited to examining the determinants of both the likelihood of WTP and the value of WTP, given that the value is positive[14].

The first part of the two-part model employed a logit regression[15] and the second part employed a Generalised Linear Model (GLM). A GLM was chosen for the second part in order to account for skewness in the data since this approach has been recommended as more robust than alternatives, such as log-transformed OLS, for skewed data distributions[16] (314).

A limited set of model specification tests are available for the two-part model (312), however the Pearson χ^2 goodness-of-fit test was employed to test the specification of the logistic regression.

7.2.4 *Objective 4: Identifying and examining characteristics of protesters*

Protest responses were identified using the two approaches outlined in Chapter Six. Firstly, the justifications for unwillingness to pay were examined to identify Type 1 protest responses (respondents reporting unwillingness to pay despite valuing the good positively). Secondly, qualitative responses from respondents justifying their WTP decision were examined to identify Type 2 protest responses (respondents offering excessively high or low values which do not represent true WTP). Particular attention was paid to respondents offering high WTP values (considered over £20 per month, equivalent to over £240 per annum). Respondents identified as protesters were compared with non-protesting respondents for each scenario separately using logistic regression. A dummy variable was created to distinguish protest respondents and non-protest respondents for each hypothetical scenario. The protest dummy variable was held as the dependent variable during the regression analysis in order to estimate the odds of protesting dependent on the respondent's characteristics, collected from the demographic survey questions. Statistical significance on any of the resultant odds ratios was examined to test whether the sample of protesters differed from the sample of non-protesters in a way that is unlikely to be caused by chance.

[14] For further detail on the two-part model see Belotti *et al.,* 2015 (312).

[15] A probit model would have been equally acceptable, either specification is recommended for use in a two-part model (312). A logit model was chosen in order to present the outcome of the regression in odds-ratios, via the 'or' option in Stata.

[16] The GLM used a log link and a gamma distribution. The most appropriate link function and distribution family for the WTP data were identified using a Box cox approach to test the link function and the Modified Park test to determine the distribution family (313)

7.3 Results

The CV survey was conducted in June 2018. Responses were obtained from 766 people from the UK general public, which met the intended sample size of 765 (see Chapter Six). Table 7.2 reports the characteristics of the study sample in addition to the characteristics of three sub-samples: (i) respondents who provided a positive WTP value for all three scenarios, (ii) respondents who were consistently unwilling to pay for the ASBI in all scenarios and offered true zero valuations and (iii) respondents consistently protesting in all scenarios. The sample was selected to be representative of the UK adult population in 2018 according to location, gender, and age. A little over one-quarter (27%) of the sample were parents and a small majority (57%) reported consuming alcohol on either a moderate or frequent basis (see Box 7.1 for categorisation criteria).

Table 7.2 Demographic characteristics of survey sample and sub-samples of respondents

Characteristic	All respondents (n=766)		Positive respondents* (n=200)		True zero respondents* (n=105)		Protest respondents* (n=221)	
	N	% of sample	N	% of sample	N	% of sample	N	% of sample
Gender								
Male	377	49%	97	48.5%	45	43%	114	52%
Female	389	51%	103	51.5%	60	57%	107	48%
Age								
Under 35	204	27%	50	25%	26	25%	40	18%
35-65	409	53%	114	57%	66	63%	115	52%
Over 65	153	20%	36	18%	13	12%	66	30%
Marital status								
Single/divorced/ widowed	324	42%	66	33%	56	53%	99	45%
Married/cohabiting	438	57%	134	67%	49	47%	121	55%
Employment status								
Employed	416	40%	121	61%	52	50%	117	53%
Unemployed	305	54%	69	35%	46	44%	91	41%
Educational level								
No formal qualifications	35	5%	8	4%	6	6%	9	4%
School level qualifications	375	49%	89	45%	55	52%	116	57%
Higher education qualifications	353	46%	101	51%	44	42%	96	43%
Annual household income								
Less than £20,000	229	30%	46	23%	44	42%	69	31%
£20,000 - £40,000	276	36%	71	36%	32	30%	74	33%
Greater than £40,000	198	26%	72	36%	17	16%	56	26%
UK Location								
England	644	84%	164	82%	88	84%	195	88%
Scotland, Wales or Northern Ireland	122	16%	36	18%	17	16%	26	12%
Parent/guardian of a child under 16								
Parent	204	27%	62	31%	33	31%	42	19%
Non-parent	562	73%	138	69%	72	69%	179	81%
Alcohol consumption								
Drinker	436	57%	119	60%	61	58%	126	57%
Non-drinker	328	43%	81	41%	44	42%	95	43%

*Sub-samples of respondents consistently offering a positive WTP value, true zero, or protest zero for all three scenarios

7.3.1 *Descriptive statistics of WTP*

The first analytical objective was to examine the survey results descriptively. As commonly found in WTP data (130), the distribution of WTP was highly right-skewed and remained so after removing protest responses (skewness = 3.34 - 4.25 across the scenarios, excluding protest zeros) (see Appendix J for histograms of the data demonstrating skew). Additionally, a large proportion of the sample (n=476 - 514, 62% - 67% across the scenarios) was unwilling to pay for the intervention. Once protest zero responses were identified, the proportion of the entire sample (n=766) that were true zero responses was still substantial for all scenarios (n=178 - 202, 23% - 26%). Since protest zeros were removed from the samples prior to analysis, the proportion of zero WTP values in the analysed samples (n=454 - 466) remained substantial (39% - 45%) producing a large spike in the WTP distribution at zero (protest responses are discussed in detail in section 7.3.4).

The mean and median values of WTP for each scenario are reported in Table 7.3. Mean annual WTP for the ASBI is, therefore, estimated at between £65 and £68, scenario dependent. The mean values are substantially larger than the median (£24), due to the right skew of the data.

The results of the sensitivity analyses conducted to explore the impact of excluding large values and including protest responses are also reported in Table 7.3. Trimming the highest 1% of responses reduced mean WTP slightly to between £59 and £62 and had no effect on the median values. The trimmed mean WTP for Scenario 2 (£59), however, was less than Scenario 1 (£61), contrary to the direction of untrimmed means. Including responses coded as protest responses in the base case analysis had the most substantial impact on WTP results, reducing the mean to between £41 and £43, scenario dependent. The much greater proportion of zero values in the sensitivity analysis including protest responses also has a substantial effect on median WTP reducing it to £0 for all scenarios.

Table 7.3 Mean and median WTP for base case and sensitivity analyses

		Scenario 1	Scenario 2	Scenario 3
WTP, £	N	454	462	466
	Mean (SD)	65 (126)	66 (119)	68 (112)
	Median (IQR)	24 (0-87)	24 (0-90)	24 (0-90)
	95% CI	54 - 77	55 - 77	58 - 78
Sensitivity analysis				
WTP, £	N	450	458	462
Largest 1% values trimmed	Mean (SD)	61 (108)	59 (91)	62 (93)
	Median (IQR)	24 (0-84)	24 (0-90)	24 (0-90)
	95% CI	51 - 71	51 - 68	54 - 71
WTP, £	N	766	766	766
Including protest responses	Mean (SD)	41 (111)	42 (107)	43 (103)
	Median (IQR)	0 (0-36)	0 (0-48)	0 (0-60)
	95% CI	33 - 48	35 - 50	36 - 51

SD=standard deviation, IQR=interquartile range, SE=standard error of the mean, 95% CI=95% confidence interval. SE and 95% CI estimated using 10,000 bootstrap repetitions

7.3.2 *Comparison of WTP between scenarios*

The second analytical objective was to examine the difference between mean WTP for each scenario, in order to consider the effect that long-run outcomes and a reduction of alcohol consumption had on respondents' valuation of the ASBI. The Skillings-Mack test rejected the null hypothesis that the values of WTP for each scenario were equal (p=0.039), indicating that the WTP values from the three scenarios were different from each other.

These results were further explored to consider the two hypotheses outlined in section 7.1.2. Pairwise comparisons provided no evidence to support the first hypothesis that Scenario 3 would elicit larger WTP values than Scenarios 1 and 2. Neither test could reject the null hypothesis that distributions of WTP were equal between pairs of scenarios (p=0.648 and p=0.241 for Scenario 3 compared with 2 and 1, respectively). The Wilcoxon signed rank test did, however, support the second hypothesis that Scenario 2 would elicit larger WTP values than Scenario 1. The null hypothesis of the test, that the distribution of WTP values was equal between Scenarios 1 and 2, was rejected (p=0.021) and the sign of

ranked pair comparisons indicated that Scenario 2 elicited larger WTP values more frequently than Scenario 1.

7.3.3 *Examine predictors of WTP*

The third objective of this study was to examine predictors of WTP and test the theoretical validity of the CV study. Section 7.1.3 outlined several expectations with regards to how respondents might be expected to value the ASBI based on their characteristics.

The results from the regression analysis of WTP for each scenario are reported in Table 7.4; the first part of the model reported outcomes as odds ratios whilst the second part reported coefficients representing percentage changes[17]. Statistically significant (at the 1%, 5% and 10% levels) results are displayed in bold. The odds ratios from the logit model report the odds that a positive WTP value is observed (compared with a zero value) given exposure to each covariate compared to its base factor. The coefficients reported from the conditional GLM model estimate the effect of each covariate on the value of WTP offered, conditional that the value is positive.

The results of the Pearson χ^2 goodness-of-fit test are all greater than 0.05, therefore, the logit model cannot be rejected. This indicates the logit regression model is an appropriate fit for the data.

The first notable observation about the predictors is the difference between which covariates we have evidence as being associated with WTP for the first and second parts of the two-part model for each scenario. The results from the regression analyses suggest that the decision to pay a positive amount (compared with zero) and the absolute value to place on the ASBI are influenced by different factors.

[17] Due to the log link, the GLM produces an exponential mean which allows the coefficients to be interpreted as percentage changes. To identify the percentage change of a factor covariate the following formula must be applied: $\exp(\beta) - 1$

Table 7.4 Two-part model regression results for each scenario – dependent variable WTP

Variable[A]	Scenario 1		Scenario 2		Scenario 3	
	Logit	Conditional GLM[B]	Logit	Conditional GLM[B]	Logit	Conditional GLM[B]
	Odds ratio *(S.E.)*	Coefficient *(S.E.)*	Odds ratio *(S.E.)*	Coefficient *(S.E.)*	Odds ratio *(S.E.)*	Coefficient *(S.E.)*
EDUCATION						
No formal qualifications	1.236 *(0.690)*	0.329 *(0.416)*	1.062 *(0.636)*	-0.060 *(0.401)*	1.190 *(0.678)*	-0.097 *(0.379)*
Higher education	1.015 *(0.218)*	-0.016 *(0.158)*	1.062 *(0.235)*	0.017 *(0.137)*	1.432 *(0.316)*	-0.098 *(0.132)*
NON_ENGLISH						
Non-English UK nations	1.156 *(0.323)*	**-0.359*** *(0.190)*	0.817 *(0.231)*	-0.264 *(0.177)*	1.037 *(0.288)*	**-0.311*** *(0.162)*
AGE						
Under 35	0.983 *(0.241)*	**0.465**** *(0.187)*	1.107 *(0.273)*	0.177 *(0.164)*	1.181 *(0.299)*	**0.293*** *(0.160)*
Over 65	0.946 *(0.283)*	**-0.339*** *(0.206)*	1.434 *(0.464)*	**-0.402**** *(0.180)*	1.061 *(0.331)*	**-0.326*** *(0.178)*
MALE						
Male	1.349 *(0.281)*	**0.273*** *(0.144)*	0.914 *(0.195)*	0.013 *(0.134)*	0.882 *(0.190)*	0.207 *(0.128)*
INCOME						
Less than £20,000	0.776 *(0.201)*	-0.148 *(0.194)*	**0.601*** *(0.156)*	0.029 *(0.175)*	0.785 *(0.203)*	0.159 *(0.168)*
Over £40,000	**1.604*** *(0.418)*	0.054 *(0.185)*	**2.126***** *(0.592)*	0.216 *(0.162)*	**2.348***** *(0.658)*	0.219 *(0.152)*
PARENT_GUARDIAN						
Parent	1.022 *(0.245)*	0.101 *(0.171)*	**0.626*** *(0.159)*	0.188 *(0.156)*	**0.575**** *(0.144)*	0.146 *(0.148)*
MARRIED						
Married / Cohabiting	**1.523*** *(0.384)*	0.153 *(0.188)*	**2.389***** *(0.606)*	-0.066 *(0.170)*	**2.318***** *(0.603)*	0.087 *(0.160)*
NONDRINK						
No alcohol	1.022 *(0.218)*	0.046 *(0.162)*	1.117 *(0.245)*	-0.197 *(0.146)*	1.439 *(0.322)*	-0.059 *(0.141)*
Observations	414	231	422	248	426	262
Pr>chi2 goodness of fit test	0.428	-	0.301	-	0.533	-

[A] Base factors excluded from regression: (education) school level qualifications, (location) England, (age) 35-65, (gender) female, (income) £20,000-£40,000, (parent/guardian) non-parent, (marriage status) single/divorced/widowed, (drinking frequency) alcohol consumer

[B] Coefficients from the second (conditional) part of the 2PM

*** p<0.01, ** p<0.05, * p<0.1

The odds of offering a positive value of WTP for the ASBI increase for respondents with a household income over £40,000 compared to respondents with a household income of £20,000-£40,000 for all three scenarios. The predictive association is slight ($p<0.1$) in Scenario 1 but strong ($p<0.01$) in Scenarios 2 and 3. In Scenario 2 there is also slight evidence of a ($p<0.1$) reduction in odds of households with an annual income of less than £20,000 offering to pay a positive value of WTP compared with respondents with a household income of £20,000-£40,000. There is no evidence that the value of WTP offered is associated with income in any of the scenarios.

There is no evidence that parental status is a predictor for WTP in Scenario 1 ($p>0.1$). However, in Scenarios 2 and 3, parental status is weakly and moderately ($p<0.1$ and $p<0.05$, respectively) associated with reduced odds of WTP for the ASBI. There is no evidence that parental status is a predictor of the value offered, given positive WTP, for any of the scenarios.

The final predictor of the probability of offering a positive value of WTP is being married or cohabiting compared to being single, divorced or widowed. There is strong evidence ($p<0.01$) that being married or cohabiting has an odds ratio that is different to 1 in Scenarios 2 and 3. Whilst there is weak evidence ($p<0.1$) that the odds ratio for WTP is greater for married or cohabiting respondents compared with single, divorced or widowed respondents for Scenario 1.

With respect to the value of WTP offered, conditional that it is positive, there is moderate to weak evidence that respondents of retirement age or older (over 65) have a WTP that is less in all three scenarios compared with respondents aged 35-65 ($p<0.1$ for Scenarios 1 and 3, $p<0.05$ for Scenario 2). Additionally, males and younger respondents (under 35 years) are willing to pay more ($p<0.1$ and $p<0.05$, respectively) than females and respondents in the 35-65 age range for Scenario 1. The same association is observed for Scenario 3 for younger respondents, however, the evidence is weak ($p<0.1$). There is no evidence that gender and younger age are predictors of WTP value for Scenario 2.

There is weak evidence for Scenarios 1 and 3 that the respondent's location is a predictor of WTP. Living in either Scotland, Wales or Northern Ireland is weakly associated with a lower value of WTP compared with respondents living in England ($p<0.1$).

Sensitivity analysis was run using Tobit and log OLS regressions to test for sensitivity to model specification. Both models used in the sensitivity analysis report only the effect of covariates on the value of WTP offered for the entire sample. This differs from the two-part model which reports the effect of covariates on WTP values given that the observations are positive. Additionally, whilst the Tobit model is able to account for the existence of zero values, log OLS cannot. Thus substantially reducing the sample size analysed in the log OLS model compared to either the Tobit or two-part models. Therefore, the outputs of the two-part model are not directly comparable with either the Tobit or log OLS. However, no evidence was found that additional covariates in either of the regression analyses were predictors of WTP.

The Tobit model results reported moderate evidence that married respondents, compared to single, divorced or widowed respondents, are predictors for higher WTP in Scenarios 2 and 3 (p=0.022 and p=0.011, respectively). Furthermore, there is strong evidence that having an annual household income greater than £40,000, compared to an income of £20,000-£40,000, is associated with higher WTP in Scenarios 2 and 3 (p=0.003 and p=0.001, respectively). These findings are similar to those of the two-part model reported above (which identified both of these covariates as predictors for WTP).

There is weak evidence from Scenario 3 only that younger respondents (under 35) also paid more than respondents in the 35-65 age range (p=0.089). The Tobit regression analysis of Scenario 1 identified only being male as being a predictor of WTP (p=0.034), this is similar to the two-part model which provided weak evidence that male gender was a predictor or WTP for Scenario 1. The Tobit regression provided evidence that fewer covariates were predictors of WTP than the two-part model, however the pattern of predictors identified in the Tobit model is similar to that of the two-part model.

The log OLS model did not provide any evidence that there were any predictors of WTP for Scenario 3. However, for Scenario 1 there was moderate to weak evidence that WTP increased for younger respondents (under 35) (p=0.019) and for males (p=0.083) respectively. Older respondents (over 65) were associated with paying comparatively less than 35-65-year-old respondents in Scenario 2 (p=0.053). The log OLS model provided evidence that far fewer factors were predictors of WTP than either the Tobit or two-part models. However, the factors that appeared to be important in both the OLS and Tobit models were also predictors in the two-part model.

There is no evidence that the WTP results are sensitive to model specification. No findings from the sensitivity analysis contradict the base-case findings using the two-part model. The full regression outputs for the sensitivity analyses are reported in Appendix K.

7.3.4 *Examining protest responses*

The fourth objective of this chapter was to examine the reasons given for protest responses and identify whether the sample of protesters differed from the sample of non-protesters. Firstly, Table 7.5 reports the reasons reported for unwillingness to pay for the ASBI (see Chapter Six for justification behind selection of protest and true-zero responses). Responses considered to represent a protest of Type 1 (respondents reporting unwillingness to pay despite valuing the good positively) are displayed in bold in Table 7.5. The free-text responses to "Other" were examined and only those indicating a protest to the survey or method of payment were considered protest responses. Examples of "Other" responses categorised as protests are:

> *"state should pay for health through general taxation."*
>
> *"The money to pay for this should come from the tax that is already levied on alcohol (sic)"*

Examples of "Other" responses considered true zeros were:

> *"there are no definite long-term benefits, so I don't think it's worth it"*
>
> *"Cause (sic) I only get money for a week"*

In total, 54% - 65% of "Other" responses were categorised as protests across the three scenarios.

Table 7.5 Reasons reported by survey respondents for unwillingness to pay for the ASBI

Reason for unwillingness to pay	Scenario 1	Scenario 2	Scenario 3
	N (%)	N %	N %
Other interventions are more valuable	57 (11)	48 (10)	42 (9)
I am not concerned about the issue of risky drinking in young people	40 (8)	48 (10)	42 (9)
I think the intervention is valuable, but I cannot afford it	91 (18)	81 (16)	83 (17)
I think the intervention is valuable, but I do not think it should be funded from taxes	96 (19)	86 (17)	93 (19)
Other taxpayers, who are better off, should pay for it	21 (4)	23 (5)	27 (6)
Parents/guardians of 'Year 10' students should pay for it	177 (34)	171 (35)	165 (35)
Other	32 (6)	34 (7)	24 (5)
Total true zeros	202 (39)	189 (38)	178 (37)
Total protest zeros	312 (61)	302 (62)	298 (63)
Total protest and true zeros	514 (100)	491 (100)	476 (100)

Regardless of the scenario, the most commonly cited reason for unwillingness to pay was "parents and guardians of 'Year 10' students should pay for it".

The proportion of all zero responses (n=476 – 514) that were categorised as protests was fairly constant across the three scenarios at approximately 60% (Table 7.5). The proportion of the whole sample (n=766) categorised as offering protest zero responses was approximately 40% for each scenario.

Protest responses of Type 2 (respondents offering excessively high or low values which do not represent true WTP) were identified via examination of the qualitative comments provided by respondents following the payment task for each scenario. Whilst it is difficult to detect protest responses of Type 2, attempts were made to identify whether extremely high WTP could be considered a protest response by considering the justification given by the respondent. Many of the respondents offering the highest WTP values (over £60 per month) did not provide any comments, therefore, precluding a decision on whether these values represent protest or true responses. Two respondents, however, could be identified as

protesting based on their responses. Both respondents indicated that their valuation of the ASBI was nominally high and perhaps did not reflect a true valuation of the ASBI but rather strategic action due to their strong support for intervention to help young people. Two example responses from the identified protesters are:

"you can't put a price on a students (sic) life"

"YOU CANNOT PUT APRICE (sic) ON HEALTH" [capitalisation by respondent]

The results of the logistic regression that was run to examine differences between the samples of protesters and non-protesters identified evidence of differences on three variables: parental status, age and location (for Scenario 3 only). In all three scenarios, there was evidence of reduced odds of parents protesting compared with non-parents ($p<0.01$ for Scenario 1 and $p<0.05$ for Scenarios 2 and 3). For all three scenarios there was evidence that respondents over 65 years old were more likely to offer protest responses ($p<0.1$, $p<0.05$ and $p<0.1$, respectively), whilst in Scenario 2 only, there was evidence that respondents under 35 were also less likely to protest ($p<0.05$).In Scenario 3 only there was weak evidence that non-English respondents were less likely to protest ($p<0.1$).

No other evidence of differences between the protest and non-protest respondents was identified from the logistic regression; therefore, for the most part, the samples are very similar. However, it could not be stated that the samples of protesters and non-protesters are homogenous and as such the WTP value estimates excluding protest respondents may be biased towards the views of parents and respondents below retirement age.

7.4 Discussion

The aim of this CV study was predominantly to obtain a WTP value for the ASBI which could be used to populate a CBA. The inclusion of three separate scenarios also allowed inferences to be made about the impact of certain elements of the intervention, such as the impact of health-related outcomes compared to non-health related outcomes and the inclusion of long-run outcomes compared to short-run outcomes. The theoretical validity of the survey has also been examined alongside the potential for bias of the WTP introduced by the study sample. Each of these points is discussed in greater detail below.

7.4.1 *Estimates of WTP and the presence of zero valuations*

The survey elicited a positive mean value of WTP for the ASBI. The standard errors and 95% CIs calculated for the mean values of WTP suggest a high likelihood that the population mean annual WTP is positive and in the region of £50-£80. The sensitivity analyses conducted indicate that mean and median WTP are not extremely sensitive to the removal of the largest 1% of values. Furthermore, the uncertainty around the mean reported by the 95% CIs is similar to the base analysis for each scenario.

Including protest responses substantially reduced both the mean and median values of WTP, although, the lower bound of the 95% CIs remained greater than zero across all scenarios. Thus, accounting for uncertainty around the mean WTP estimate, the evidence suggests that mean WTP remains positive when protest responses are included as true WTP responses. This observation is due to the high proportion of the sample who were unwilling to pay anything for the ASBI. Including protest zeros, approximately two-thirds of the sample were unwilling to pay for the intervention, substantially reducing the mean WTP.

The proportion of all zero valuations identified in this study, including both protest zeros and true zeros (62%-67%), is larger than has been reported previously in healthcare related CV surveys. Smith *et al.* (315) conducted a review of WTP studies and, where data were available, reported the proportion of zero responses in their identified studies. Two of the 18 studies reporting relevant data had proportions of zero responses greater than 60%, whilst the majority reported 0%-5% (315). The authors (315) suggest that large numbers of zero-responses may be due to self-completion of surveys from respondents since surveys conducted face-to-face elicited lower proportions of zero valuations. This could provide an explanation for the current study since the survey was completed online, however, given the reasons reported for unwillingness to pay it is likely that other factors were responsible.

The survey sample did not constitute potential recipients of the ASBI, therefore, as discussed in Chapter Six, the values elicited would be non-use values and, thus, largely rely on altruistic valuations. Olsen and Donaldson (282), however, found that in a CV survey of healthcare programmes, altruistic motivations were greater predictors of WTP than “selfish” motivations (p.6), therefore, relying on altruism may not necessarily be a problem. Hence, the zero responses may be attributed to factors specific to the ASBI, such as its focus on

adolescents, rather than necessarily methodological issues with the survey design. Protest zero responses are discussed later in section 7.4.4.

7.4.2 *Inferences from the WTP differentials between scenarios*

The WTP values elicited for the three scenarios were identified as differing from each other by the Skillings-Mack test. The pair-wise comparisons show that this is largely driven by the comparison between the WTP responses to Scenarios 1 and 2. This result suggests that the inclusion of long-run outcomes in Scenario 2 had an impact on respondents' WTP for the ASBI, as initially hypothesised (section 7.1.2). Whether this is due to the perception of greater certainty in the outcomes when long-run effects are presented, or whether respondents have a genuine preference for outcomes in the longer term would require further investigation to determine.

A similar finding was not, however, identified for the comparison between Scenario 3 and Scenarios 1 and 2, suggesting that the reduction in alcohol consumption (included only in Scenario 3) did not increase WTP. The results do not provide evidence to suggest that health outcomes (i.e. a reduction in alcohol consumption) are valued by the UK public greater than non-health outcomes (i.e. reductions in arrests and missed school). This finding is contrary to the initial hypothesis. However, since the CV scenarios included three outcomes from the ASBI programme, this finding may be due to an embedding phenomenon. An argument follows that the sum of parts (i.e. each of the outcomes) valued individually may be greater than the sum of the whole (i.e. the scenario presenting all three outcomes simultaneously), and if this is the case in a CV study presenting several attributes it can be considered an embedding phenomenon (1). This effect can also be referred to as *sub-additivity bias* (316) or *part-whole bias* (317).

Alternatively, the lack of significant differences between Scenario 3 and the other scenarios may represent a lack of scope on the alcohol-consumption outcome (1). Scope effects refer to increases in WTP for a good when the size of benefit increases. Scope insensitivity in WTP studies has been an area of contention in the CV literature (318) and to some opponents of CV methodology, lack of scope effects in WTP cast doubt on the suitability of the method (319). However, with reference to the reduction in alcohol consumption in this study, it may simply be the case that the modest change in the alcohol-consumption outcome was insufficient to test scope sensitivity.

7.4.3 *Theoretical validity and predictors of WTP*

Section 7.2.3 outlined theoretical expectations regarding the effects of income, education, marital status, parental status and frequency of alcohol consumption. The two-part model regression analysis partly confirmed the a priori expectation regarding income such that respondents from higher-income households were more likely to offer to pay for the ASBI than moderate-income households and the reverse was observed for respondents from the lowest income households. However, with respect to the value of WTP offered by respondents who were willing-to-pay for the ASBI, there was no evidence that values offered varied as household income varied. This indicates that once a decision has been made to pay a positive amount, income does not greatly influence how much a respondent values the intervention.

Education did not appear to be a predictor of WTP, suggesting that neither risk aversion nor higher health-literacy via the proxy of education level had an impact on WTP. This may be a result of the sample population since the study sample constituted non-users of the intervention; the theoretical expectations for the impact of education may only be appropriate to predict influence on use value.

Marital status was hypothesised to influence the likelihood of WTP positively due to the potential for greater financial security in a two-person household, although, no expectation was posited for the impact on the value of WTP. This theoretical expectation appears to be validated by the results from the regression analysis since there was strong evidence that the odds of paying something for the ASBI were increased ($p<0.01$) if a respondent was married or cohabiting, yet there was no evidence of an influence on the positive value offered.

The predictive ability of the parent covariate on WTP is also observed in the regression analysis, however, in the opposite direction to that hypothesised. It was expected a priori that parents would be WTP more than non-parents due to potentially stronger altruistic motivations; however, where there was evidence of a difference (i.e. in Scenarios 2 and 3) the odds of WTP for the intervention were lower for parents than non-parents.

There are several possible explanations for this finding. Firstly, parents may be more sceptical of the intervention given their knowledge of their own children's behaviour resulting in greater unwillingness to pay for the ASBI. Secondly, if parents in the sample do

not believe their children exhibit risky drinking behaviour, they may not view the intervention as beneficial. Thirdly, the finding may be the result of sample selection bias introduced by the removal of protest responders who were more likely to be non-parents, thus skewing the analytic sample population. However, the resultant proportion of parents in the analytic samples (30-31%) is not substantially different to the proportion of parents in the whole sample from which data were collected (27%). Therefore, potential sample bias is unlikely to be uniquely responsible for the association between parental status and unwillingness to pay for the ASBI. Fourthly, parents may have less disposable income compared to non-parents once spending on children is accounted for. Therefore, an income effect that is separate to the overall household income effect examined in this study may also contribute to this finding. Data were not collected on spending patterns; therefore, this hypothesis could not be examined.

Finally, the expected direction of influence on WTP of alcohol consumption frequency was ambiguous prior to conducting the analysis. However, the results provide no evidence that this factor was a predictor of either WTP, or the value offered.

With the exception of the parental status covariate, the CV survey findings generally appear theoretically valid in relation to the expected influence of factors on WTP. Other demographic factors (age, gender and location) were also significantly associated with WTP for which there were no theoretical expectations due to contrasting effects in the related literature.

The ASBI also appears to be more valuable to respondents under the age of 35 and to males, although evidence for the latter is weak. Arrest rates in the UK are higher for males than females (320), therefore, the reduction in arrests outcome of the ASBI may be more salient to male than female respondents, leading to slightly increased WTP. Additionally, given the ASBI is targeted at adolescents aged 14-15, younger respondents may recall their own experiences, or those of others, with regards to drinking at that age to a greater extent than respondents over the age of 35. Consequently, the intervention and the reported outcomes may resonate more strongly with respondents closer in age to the potential ASBI recipients and, therefore, elicit higher WTP. The converse may also explain why respondents of retirement age (over 65) had a lower WTP compared with the 35-65 group.

The sensitivity analysis conducted using the Tobit and Log OLS indicates that the data were sensitive to model choice. However, due to the distribution of the data, the two-part model

appeared better able to detect predictive factors of WTP by considering and reporting the two related decisions of WTP and the value offered separately. Whilst the Tobit model also considers WTP conditional on the probability of the outcome being positive (321), the regression output does not provide detailed feedback on the factors contributing to the two choices involved in WTP decisions. The log OLS regression was unable to account for zero values at all and so was considered less appropriate for the data. This was reflected in the far fewer covariates having evidence that they were predicting WTP in the log OLS model. The two-part model could, therefore, be argued to be the most appropriate model choice for the data obtained. However, a more conservative evaluation of predictors of WTP would consider only those identified by both the Tobit and the two-part models. Considering such a perspective indicates that being married and having an annual household income of over £40,000 were the most consistently influential factors of WTP for the ASBI.

7.4.4 *Protest responses*

The proportion of the whole sample (n=766) who offered protests responses was approximately 40%. A meta-analysis of protest responses in environmental CV studies (322) estimated a mean proportion of samples offering protest responses of approximately 18%, although, the maximum identified proportion was 59%. Meyerhoff *et al.* (322) also examined protest responses according to factors such as the payment vehicle, thus, allowing a comparison to be drawn between the CV study reported in this chapter and a large body of literature. For studies using a tax payment vehicle, the range of protest proportions identified by Meyerhoff *et al.* was 1.5% - 47.7%, therefore, whilst at the high end of this range, the current study is still within the bounds of other similar literature. Nevertheless, compared with the proportion of protest zeros reported in similar health-related CV literature (between 3% and 24% of the total sample) (279, 281, 282), substantially more protest responses were identified in this study. It is important, therefore, to examine whether the high rate of protesting is a result of inappropriate survey design, or whether factors specific to the intervention can explain these findings.

The use of tax-based payment vehicles is often considered a contributing factor for eliciting protest responses (293, 315, 323), however, objection to paying additional tax made up only 19% of the reasons given for unwillingness to pay for the ASBI (Table 7.5). The most commonly cited reason for a zero valuation was “Parents/guardians of ‘Year 10’ students should pay for it”, making up approximately 35% of justifications for unwillingness to pay.

This justification is a greater reflection on the intervention being valued than, necessarily, issues with elements of the survey design. The concept that it is the responsibility of parents rather than the public to fund the ASBI is specific to this study where the recipients of the intervention are adolescents rather than the general public. Additionally, the intervention is one aimed at reducing alcohol consumption, a public health issue which has been demonstrated to elicit low WTP due to connotations of voluntary risk behaviour, e.g. Pellegrini & Jeanrenaud., 2001, cited in Jeanrenaud & Pellegrini., 2007 (324). Therefore, a combination of negative perceptions of individuals engaging in harmful alcohol consumption and the intervention recipients being adolescents, who are presumed to be under the guardianship of parents, are probable explanatory factors for the majority of protest responses. Therefore, design issues with the CV survey can be expected to contribute only partly to the high proportion of protest responses.

The choice of eliciting WTP from a general public sample likely contributed to the number of protesters, given that the most common reason for unwillingness to pay was that parents should fund the ASBI. Additionally, approximately three-quarters of the sample (73%) were non-parents. Balancing the proportion of parents in the sample may have reduced the number of protest votes, however, the study sample would no longer be representative of the UK population.

The logistic regression analysis comparing protestors to non-protestors indicated that the sample of protest respondents differed from non-protest respondents. It is not unusual for heterogeneity to be identified between samples of protest and non-protest respondents (293) and, unsurprisingly, the strongest indicator of a protester was being a non-parent. Heterogeneity between the two samples can be a cause of concern where the treatment of protest responses is to remove them from the sample, due to the potential for bias to be introduced affecting mean WTP (302). Examination of the proportion of parents in both the whole sample and analytic sample (excluding protesters), however, revealed that whilst there was evidence of a difference between both samples, the importance of that difference is debatable as the difference was not large (approximately 4 percentage points). Therefore, bias from removing protestors from the analytic sample may not be as great a cause of concern as the literature indicates.

The direction of impact on mean WTP of excluding protest responses from the analytic sample is ambiguous without further research to elicit protesters' true WTP for the ASBI,

therefore, whether the current estimates of mean WTP illustrate a conservative or overestimate of the mean cannot be established with any certainty.

A final point for consideration is whether the categorisation of protest responses was appropriate. As discussed in Chapter Six, a standard definition for what constitutes a protest response is lacking amongst CV researchers (281, 292-294). Therefore, the judgement regarding which justifications for unwillingness to pay should be identified as protests (see section 7.3.4) may not be universally agreed upon. Responses 5 and 6, "Other tax payers, who are better off, should pay for it" and "Parents/guardians of 'Year 10' students should pay for it", respectively, may be queried in particular.

Response 6 could be questioned on the basis that non-parents who state that parents should pay for the ASBI may be reflecting that the programme offers no altruistic or option value (1, 277) to them; therefore, unwillingness to pay is a reflection of their values, rather than a protest. However, this does not account for existence value, in which a good represents value to an individual merely due to its existence regardless of any intention or ability to use it (151, 152). Without further investigation, distinguishing whether a response that "Parents/guardians of 'Year 10' students should pay for it" constitutes a true zero, reflecting a lack of altruistic or option value, or a protest response, on the basis that the respondent actually holds some existence value for the programme, cannot be determined. However, it is possible that by assuming all unwillingness to pay justification responses of "Parents/guardians of 'Year 10' students should pay for it" represent protest zeros, the proportion of protest responses has been overestimated.

Making an alternative assumption to that made for the base-case results of this study, that the two justifications for unwillingness to pay discussed here in fact reflect true zero valuations, the consequence for mean WTP can be examined. The number of observations included in the analytic sample increase from between n=454 – 466 to a range of n=650 – 658 across all three scenarios. Mean WTP reduces, as expected, due to the larger proportion of zero valuations in the analytic sample. Mean WTP under the new assumption ranges from approximately £45 (SD: 126; 95% CI: 37 to 54) to £48 (SD: 112; 95% CI: 41 to 56). Therefore, whilst the new assumption would have an impact on WTP, the conclusion that WTP is on average positive would not change.

7.4.5 *Study limitations*

The CV study presented in this chapter represents a novel valuation of a school-based ASBI. Nevertheless, this study has limitations.

The CV survey elicited a large proportion of zero valuations for the intervention, which has two significant implications for this study. Firstly, the distribution of the resulting WTP data was extremely skewed and precluded the use of standard regression analysis methods, such as OLS, to examine the predictors of WTP. A two-part model was used to address the unique distribution of data, however, the sample sizes for the second, conditional part of the model for all scenarios were relatively small (n=233 – 264) due to the prominence of zero valuations. Therefore, the predictive ability of the model was significantly reduced compared to the ideal sample size calculated a priori (n=683, see Chapter Six). The much larger proportion of protest responses identified (approximately 40%) than anticipated a priori (approximately 12%, see Chapter Six) is partly responsible for this. The sample size was inflated based on the expectation of a much lower proportion of protest responses. Given the proportion observed, a more appropriate sample size would have been approximately 950[18].

Additionally, the relatively large proportion of respondents who were unwilling to pay for the ASBI could have implications for policy-makers; if the intervention were to be implemented based on the results of this CV study it would seem that a relatively large proportion of the UK general public may not be in support of the intervention, assuming generalisability of the study to the UK population. However, mean and median WTP, accounting for true zero valuations, was positive, therefore, an argument could be made that the intervention is still valued positively even accounting for those who do not value it at all.

The decision to use an additional tax contribution as the payment vehicle may be considered a limitation since, although not the primary reason for protests responses, it was still a motivator for almost 20% of zero valuations (section 7.3.4). An alternative payment vehicle, such as voluntary payments may be preferable to responders and elicit fewer protest responses.

[18] This is calculated by inflating 683 by 40%

Similarly, the use of a non-user sample population to value the ASBI was partly identified as a key motivator for unwillingness to pay for the intervention. The valuation of the ASBI relied predominantly on altruistic behaviour, which, whilst has been shown to have a positive impact on WTP in previous studies (282), did not appear to resonate well with the survey sample in this study. Altruistic behaviour per se may not be the root of the problem in this case, but rather the intervention case study. As discussed earlier in section 7.4.4, respondents may have taken issue with the intervention being centred on alcohol misuse or the fact that the recipients would be a small population in society. Therefore, the choice of case study to use for the study may be considered a limitation. Had a different case study been used which had a broader impact on the UK population, a larger sample of positive WTP responses may have been elicited. Alternatively, as suggested in section 7.4.4, a different study sample could have been selected, for instance, parents of children under 16 or adolescents themselves. The latter of these suggestions, however, would bring additional complexities to the CV survey, as discussed in Chapter Six, therefore, may not be a viable or reliable alternative.

Closely related to the previous two points, the method of addressing protest zeros could be viewed as a limiting factor of this study. There remains a lack of consensus about the appropriate treatment of protest responses (293) since the commonly used method of exclusion from the analytic sample can lead to sample selection problems (325). Sample-selection models (323, 325) and multiple imputation methods (281) have been suggested as alternative approaches to treating protest zeros. Sample selection models can suffer from convergence problems (325), a problem confirmed during an attempt to employ such a model using the data from the current study, therefore, could not be utilised. The multiple imputation method described by Pennington and colleagues (281) requires that data are missing at random. As reported in section 7.3.4, this was not the case with in the protest sample, therefore, this method would unlikely have been viable.

The two-part model used to examine predictors of WTP is also subject to limitations. Limited tests for model specification or heteroscedasticity can be conducted using the two-part model; therefore, it is difficult to establish model appropriateness using usual statistical tests. The first part of the model was tested for specification by running a separate logit regression, however, the second part of the model is conditional on the first part, and

therefore, testing a separate GLM of the data would not have been equitable to conducting statistical tests on the two-part model.

Additionally, the data in the current study were modelled as continuous data, since maximum WTP was ultimately asked as an open-ended question. However, whilst the random card sort employed prior to the open-ended question was intended to assist in respondents' open-ended valuation by providing some context for their valuations, the random card sort may have had a greater influence on respondents' maximum WTP. If respondents consistently reported their maximum WTP as the highest payment card value they stated they "Definitely WOULD pay" during the random card sort, rather than stating a value between that and the lowest value stated that they "Definitely WOULD NOT pay", the resulting WTP data could be considered interval rather than continuous (1). Consequently, analysing the data using interval regression could be an appropriate alternative. Standard interval regression, however, may be unable to account for the substantial proportion of zero WTP values present in the current dataset. Alternatively, Donaldson and colleagues (321) recommend a grouped data regression model for WTP data with a limited dependent variable elicited via the payment scale method.

Finally, limitations may be observed with the descriptions of the hypothetical scenarios used in the CV survey. The outcomes reported in scenarios 1 and 2 (reduced arrests and school absenteeism rates in the intervention group and no change in alcohol consumption) may appear counter-intuitive and are likely a result of limited power in the SIPS Jr HIGH trial to detect a significant change in secondary outcomes, as discussed in Chapter Six. This point was reflected in a small number of free-text responses from respondents unwilling to pay for the ASBI who reported a lack of trust in the outcomes presented. However, Scenario 3 was presented as a potentially more intuitive scenario where an observable reduction in alcohol consumption was evident in addition to the other outcomes. If the reported intervention effects in Scenarios 1 and 2 were perceived as counter-intuitive and responsible for reduced WTP, a significant increase in WTP would be expected for Scenario 3. This was not the case; therefore, it does not appear that the scenario descriptions caused any observable problems.

The lack of difference in WTP observed between Scenario 3 (incorporating a reduction in alcohol consumption) and Scenarios 1 and 2 (incorporating no change in alcohol consumption) may be due to scope effects or embedding phenomena, as discussed in

section 7.4.2. Although, this result may also be related to the presentation of the scenarios. Whilst efforts were made during the pre-test and piloting phases of survey development to ensure Scenarios 2 and 3 were visibly and clearly distinguishable (see Chapter Six) it is possible that some respondents did not observe the important difference in the alcohol consumption outcome. This could, therefore, explain the lack of evidence for a difference between the values of WTP offered for Scenarios 2 and 3.

7.4.6 *Further research*

Four areas for further research can be identified from the results of this survey. Each of these has been identified in the previous section and they are summarised below:

1. Explore the impact of alternative payment vehicles on WTP and proportions of respondents unwilling to pay for the intervention
2. Explore whether an alternative sample with a greater connection to the ASBI, for example, parents of adolescents, would offer fewer zero valuations and elicit significantly different values of WTP
3. Address protest responses using alternative methods, such as multiple imputation or sample selection models
4. Analyse WTP assuming data is interval rather than continuous using either interval regression of grouped data regression models

7.5 Summary

This study has used the CV method to obtain a positive mean value of WTP for a school-based ASBI for Year 10 students. Responses were obtained from a study sample representative of the UK general population stratified by age, gender, and location and the sample size satisfied the estimated sample size calculated in Chapter Six. However, a large proportion of the study sample was unwilling to pay for the ASBI, which has potential implications for policy regarding the implementation of the intervention. Nevertheless, accounting for non-protest zero responses, a positive mean value of WTP of between £65 and £68 was estimated, which can be used to populate a CBA evaluation of the ASBI programme.

Protest responses were identified in the data. The majority of protest responses were driven by the perception that parents of Year 10 students should fund the intervention. The motivation for protesting was, thus, largely attributed to factors specific to the case study,

rather than elements of survey design, although, use of an additional tax contribution as the payment vehicle also contributed to a proportion of the protest responses.

Key predictors of WTP were identified as income, marital status, and age. The direction of the effect of income and marital status validates theoretical expectations. The effect of age on WTP was ambiguous a priori. The theoretical expectation for the impact of parental status, was, however, contradictory to that identified in the data.

The mean value of WTP for Scenario 1 (£65) is taken forward for use in a CBA, which is reported in the following chapter (Chapter Eight).

Chapter 8. Economic evaluations of the alcohol screening and brief intervention from the SIPS Jr HIGH trial

8.1 Introduction

This chapter outlines the conduct of two novel economic analyses of the ASBI programme for Year 10 students that was examined within the SIPS Jr HIGH trial (59). Whilst a CUA and a CCA were conducted alongside the trial (59), the qualitative interviews with PHDMs in Chapter Five indicated that more holistic economic evaluation methods may be of relevance and interest to PHDMs. Therefore, in response to these conjectures from PHDMs reported in Chapter Five, additional analyses using alternative economic evaluation methods were undertaken. This chapter reports two economic evaluation studies in turn; firstly, a CBA, and secondly, an SROI analysis. The results of these studies, alongside the existing CUA and CCA conducted within the SIPS Jr HIGH trial, could then be presented to PHDMs in a workshop for an informed comparison of methods.

Details of the workshop are discussed later in Chapter Nine. The current chapter reports the two novel economic evaluations. The CBA is presented first, followed by the SROI analysis. The results and limitations of each evaluation are discussed separately. The final section of this chapter compares the two evaluations and draws out implications of their findings for the evidence to be presented at the workshop.

8.2 Cost-benefit analysis

As described in Chapter Three, stated preference techniques can be used to value benefits in monetary terms. One of the most widely known stated preference techniques is CV (326) and it is this method that was used to obtain a monetary value of benefit for the school-based ASBI programme from the SIPS Jr HIGH trial case study. The previous chapter outlines the results of the WTP study for three scenarios depicting slight variations in outcomes of the ASBI. Scenario 1 was designed to reflect the outcomes observed from the SIPS Jr HIGH trial. In order to ensure comparability with the economic evaluations conducted as part of the trial (CUA and CCA) (59), which were also presented to PHDMs, the WTP results for Scenario 1 were used to populate a CBA of the ASBI programme.

The net benefit outcome from a CBA can be calculated using two approaches. The first approach is to examine net benefit on an individual level. Examples of this approach are most commonly found when conducting evaluations of healthcare, e.g. medical procedures

such as spinal surgery (327) or programmes of care (328). In the two aforementioned examples, both WTP and cost are measured at the patient level, therefore, the net benefit is calculated as mean WTP per patient minus mean cost per patient. The second approach is to aggregate WTP and cost at a higher level, for example, to a population level. This approach is commonly used in CBAs in the environmental field where action is valued for a public good or landmark to which action may impact society broadly, for example, water quality protection (329). This approach has also been used for the evaluation of public health initiatives, for example, strategies to prevent the West Nile Virus (330).

The CBA reported in this chapter uses the latter approach and aggregates benefit to a broad societal scale. The reasons for this are twofold. Firstly, if implemented as a public health intervention the ASBI would likely be funded by taxes paid for by the general public, therefore, the WTP survey elicited values in the form of an additional tax contribution. Aggregating WTP to a societal level is, therefore, consistent with the approach of eliciting monetary value adopted in the CV study. A similar approach was taken by Eisen-Hecht *et al.* (329) in their CBA of water quality protection. Secondly, WTP was elicited for the ASBI programme based on overall trial outcomes, whilst the costs from the trial are measured at the level of an individual risky-drinking Year 10 student. Therefore, in order to ensure that the costs and benefits in the CBA are comparable, it was necessary to aggregate both the costs and the WTP value to an equivalent level.

8.2.1 *Methods of analysis*

Two analyses were conducted for the CBA to assess methodological uncertainty (331). A simple analysis was conducted initially, followed by a more complex analysis to compare whether complexity of analytic methods affected the results or conclusions drawn from the CBA. The simple analysis formed the example shown to PHDMs at the workshop (see Chapter Nine for details of the workshop). The PHDMs were non-economists who, based on the findings from the qualitative study reported in Chapter Five, were expected to be unfamiliar with CBA. Therefore, it was considered prudent to minimise the technicality of the reporting of the analysis in order to enable an accurate description of the methods used to obtain the results of the CBA presented at the workshop. Henceforth, the simple analysis will be referred to as the "primary analysis" and the more complex analysis will be referred to as the "validation analysis".

The main difference between the primary and validation analyses was the treatment of the trial cost data. In the primary analysis, the difference in costs between the intervention and control groups was calculated using the raw costs collected over the 12-month follow-up period. This approach was commonly used in the 1990s and early 2000s (332, 333). The act of randomisation in RCTs is intended to distribute factors that may influence the outcomes of a trial evenly amongst control and intervention groups (334). Thus, theoretically, mean differences in costs and effects at follow-up should reflect differences incurred by the intervention. Even with randomisation, however, multivariable analyses can be employed in order to improve the power of between-group differences in costs by explaining cost variation due to factors other than the intervention (335). Therefore, econometric techniques can be used to adjust trial costs for baseline differences in a range of participant and contextual factors (332). This approach was followed in the validation analysis. The validation analysis is perhaps more consistent with contemporary economic evaluations conducted alongside trials (335).

A UK public sector perspective was taken for both methods of analysis and a time horizon of 12-months was considered in keeping with the time-horizon reported in the economic evaluation conducted within the SIPS Jr HIGH trial (59). The methods of analysis for each of the primary and validation analyses are described below.

8.2.2 *Primary analysis*

Benefits were estimated as annual mean WTP for the intervention to be provided in schools in the UK. The counterfactual was that the intervention was not to be provided in UK schools and the WTP for this was assumed to be zero. The mean WTP was used in preference to the median in accordance with best practice for CBA (336). Mean WTP at the individual level for Scenario 1 of the CV study was £65.

As stated in section 8.2, it was necessary to aggregate WTP to a societal level. Since WTP has been elicited using additional tax contributions as the payment vehicle, the number of UK taxpayers was used as the scale factor for aggregating WTP to a societal level[19]. Government statistics estimate the number of UK individual taxpayers in 2017/18 to be approximately 30.8 million (337).

[19] The use of the tax-paying population as a scale factor for WTP has been reported in previous CBA studies, e.g. Eisen-Hecht *et al.,* 2002 (329)

Costs for the CBA were taken from the economic evaluation conducted for the SIPS Jr HIGH trial. The costs calculated for the within-trial CUA consisted of: delivery costs related to intervention delivery, and resource costs related to subsequent use of health, social care and public services. Intervention delivery costs included: costs of materials for both screening and the interactive sessions, costs of training learning mentors to deliver the interactive sessions, and costs of learning mentor time to deliver the interactive sessions. Subsequent service use costs included those attributed to: GP visits, A&E visits, non-A&E hospital visits, social worker visits, school nurse visits, arrests and missed school days. The trial health economist identified unit costs for each service from trial records and online published sources (59).

The costs associated with school absenteeism were calculated based on reduced future earnings as a result of time missed from school (59); these costs are, therefore, indirect costs considering longer-term outcomes from the intervention. No short-run costs were attributed to school absence. Scenario 1 of the CV survey explicitly stipulated only short-run outcomes of the ASBI identified during the trial, consequently, only short-run costs associated with the time-horizon of the trial were included in the primary analysis. Furthermore, absenteeism in older age groups is potentially less important that absenteeism in younger children (below the age considered in SIPS Jr HIGH trial) (283). Thus, the costs attributed to school absenteeism in the Year 10 age group may overestimate the value of missed school for the trial participants. The costs to an individual of reduced future earnings associated with school absenteeism were, therefore, excluded from the CBA base case analysis.

In order to estimate costs on a level that would be comparable with the WTP value, as explained previously in section 8.2, costs were scaled by the number of risky drinking Year 10 students in the UK. The trial found that, on average, approximately 23.5% of the Year 10 school children screened positive for risky drinking behaviour (59); for the purposes of both the CV survey and the CBA it was assumed that this figure was representative of the proportion across the UK. The total number of Year 10 school children enrolled across the UK (or equivalent school group for children aged 14-15 in Scotland, Northern Ireland and Wales) was estimated using published data on school pupil numbers by school year for each nation in the UK in January 2018 (338-341). Approximately 702,740 pupils were identified in the published data leading to approximately 165,144 pupils estimated to exhibit risky

drinking behaviour in the UK, according to the definition of risky drinking adopted by the SIPS Jr HIGH trial (59).

The difference in mean total cost per risky-drinking child at 12-months follow-up between the intervention and control groups was estimated using a two-sided, unequal t-test, which is the most common parametric test of raw costs (335)[20]. Bootstrapping using 2500 repetitions was employed to estimate 95% CIs around the difference in costs estimated in the t-test. The use of the non-parametric bootstrap technique to estimate uncertainty around parameters from trial data is a well-established method (335, 342). A minimum of 1000 bootstrap repetitions is commonly recommended to estimate confidence intervals (343), however, the precise number required to produce reliable estimates is dependent on the complexity of the estimator used (344); more complicated estimators require more replications[21]. The t-test is a relatively simple estimator; therefore, a moderate number of bootstrap repetitions was sufficient. On the other hand, regression analysis is considerably more complex, therefore, bootstrapping estimates in the validation analysis required a larger number of bootstrap repetitions (see section 8.2.3).

Mean WTP value was also bootstrapped using 2500 repetitions. Thus, datasets were generated via bootstrapping for each of the benefit and cost parameters that had an equal number of observations (n=2500). This was required for the calculation of net societal benefit (NSB) [22] described below. All analyses were conducted using Stata 14 (305).

A second perspective for the aggregation of costs and benefits was also considered for the CBA to be presented alongside the national level aggregation as a scenario analysis (331). An LA perspective was chosen for the scenario analysis to simulate the level at which local

[20] In moderately large samples where the comparator samples (i.e. control and intervention groups) are of similar size and similar skewness, use of a t-test produces robust estimates even when violating the assumption of normality (335, p.93). The size of control and intervention groups in this analysis were similar (n=185 and n=161) as was skewness (3.03 and 2.44), therefore, the t-test was considered a suitable analytical tool for examining the raw cost difference between the intervention and control groups.

[21] The appropriate number of bootstrap repetitions was determined following the methods reported by Statacorp (306), as discussed in footnote 11. The use of normal or bias-corrected outcomes was determined by examining histograms and standardised normal probability plots of the bootstrap estimates. If the distribution of bootstrap repetitions was non-normal then bias-corrected values were used.

[22] The term “net societal benefit” is used over the standard terms “net benefit” or “net present value” to indicate (i) the societal level of the analysis as opposed to the individual level analysis conducted in many healthcare CBAs (see section 8.2), and (ii) reflect the fact that no discounting was required due to the 12-month time horizon.

public health decisions are made in England. As the CBA was to be presented at a workshop in North-East England, Newcastle City Council was chosen for the alternative analytic perspective. LA funding accrues mainly from council taxes and business rates (8). In order to improve relatability of the simulation to public health officers at the workshop, and to present a "worst-case scenario" in which the WTP value is scaled by a smaller factor relative to costs, it was assumed that any tax increases on a local level would be made progressively. Therefore, it was assumed that only households in the three highest council tax bands (F-H, inclusive) would be eligible to pay the additional tax estimated in the WTP study. The most recently published figures at the time of analysis (2017)[23] on UK households contributing to council tax show that there are 4020 households in Newcastle in the three highest bands (345).

One of the national locations included in the SIPS Jr HIGH trial was North-East England and the trial estimated a higher than average proportion of Year 10 students screened positive for risky drinking behaviour (27.4%) in this area. The number of Year 10 pupils in schools in Newcastle was estimated based on the average size of Year 10 groups in UK state schools (n=158 pupils) and the approximate size of independent school year groups in Newcastle (n=125 pupils) (346) multiplied by the number of secondary/senior state and independent schools in Newcastle (n=12 and n=6, respectively) (347)[24]. This provided an estimate of 2646 Year 10 pupils in the Newcastle LA area. Taking the local proportion of risky drinkers established in the trial (27.4%), approximately 725 students in Newcastle were estimated to exhibit the SIPS Jr HIGH definition of risky drinking behaviour.

Consequently, the values of mean WTP and mean difference in costs between the intervention and control groups were scaled by the number of households in higher council tax bands (n=4020) and the number of risky drinking Year 10 students in Newcastle (n=725) for the scenario analysis using the LA perspective.

Costs and benefits were combined to calculate the NSB. The bootstrap estimates of the difference in mean costs between the control and intervention arms and the bootstrap estimates of the mean WTP value were combined into one dataset of 2500 observations for each parameter. The NSB was estimated by subtracting each estimate of mean cost

[23] An updated version for 2018 has since been released, however the 2017 values were the most recent during the time of analysis.

[24] Special education schools excluded. Only schools taking pupils of age 14-15 included.

difference from each estimate of mean WTP. The mean of the 2500 NSB calculations was then calculated and 95% CIs around this value were estimated with further bootstrapping.

The approach taken to calculate NSB replicates the stochastic process commonly applied to costs and outcomes in cost-effectiveness analyses of clinical trial data (348). Whilst point estimates of mean WTP and mean cost difference could have been used to calculate the NSB, this would not have allowed for characterisation of the uncertainty due to sampling variation around each of the parameters and the final outcome of the CBA, the NSB. On the other hand, employing a stochastic analysis using the bootstrap technique enabled uncertainty around the NSB value to be examined using the non-parametric distributions of each of the cost and outcome parameters. Consequently, a cost-benefit plane could be produced to graphically display the uncertainty around NSB. The same process was used for both the national and LA level analyses; the cost-benefit planes associated with each analysis are displayed in Figure 8.1 and Figure 8.2.

Finally, further sensitivity analyses were conducted in order to estimate parameter uncertainty. Assumptions around key parameters (e.g. costs and outcomes) are often made due to limitations in the available data or ambiguity over the most appropriate value to use (349). Therefore, in order to estimate how significantly the assumptions made have affected the economic evaluation results, sensitivity analysis reproduces the analysis with different, possible values for uncertain parameters or using different assumptions (342).

Simple one-way analysis was applied to both the national and LA level analyses. This is the most common form of sensitivity analysis (342) in which values of one parameter are changed whilst other parameters remain at their baseline values (331). The sensitivity analyses assessed truncated values of WTP and costs at the 95th percentile to reduce the influence of very large values in both parameters. An extreme scenario analysis (331, 342) was conducted to examine the impact of the assumption that the appropriate scale factor for WTP was the UK tax base. An alternative scenario used parents of Year 10 students as the scale factor for WTP in order to demonstrate a scenario where the intervention is funded from parental contributions rather than general, or local, taxation. This scenario was deemed appropriate given that the intervention's positive externalities may fall largely on individuals closest to the students, such as their parents, and reflects a common view expressed by the CV survey respondents who were asked to value the ASBI (see Chapter Seven).

Additionally, two additional sensitivity analyses were conducted which were not presented to PHDMs. The additional analyses were (i) adjusting WTP to 2016 currency values and (ii) including indirect school absenteeism costs. The costs estimated for the CUA and CCA within trial economic evaluations were reported in 2016 prices and, in order to maintain consistency with the CUA and CCA evaluations, the price year of the costs was not altered. The WTP survey was conducted in 2018, therefore, introducing a small difference between the currency years of the costs and outcomes. It is not uncommon for WTP to be elicited at a later time than the price year reported for costs when a CBA is conducted alongside a clinical trial, e.g. Sanghera *et al.*, 2015 (350), Haefeli *et al.*, 2008 (327), and Ramsay *et al.,* 2018 (351). There is no precedent for adjusting WTP values in line with costs, therefore, the base-case analysis did not adjust WTP. Although, arguably this approach is inconsistent with standard practice for an economic evaluation in which the reporting of all costs in the same price year is recommended (182).

Although it has not been done in previous evaluations this could be thought to be a methodological weakness of such studies. Therefore, a sensitivity analysis was conducted adjusting the WTP value to 2016 currency values using the Consumer Price Index (CPI) (352). Costs in economic evaluations are commonly adjusted using the Hospital and Community Health Services pay and prices index which combines two indices, the pay cost index and the health service cost index, in order to capture inflation on both staff costs and medical goods and services (353, p.216), whereas the overall CPI was used for WTP to reflect the fact that the CV survey respondents providing WTP values would have been considering their spending on all goods and services when valuing the ASBI.

The second additional sensitivity analysis included costs calculated during the SIPS Jr HIGH trial associated with school absenteeism. For the sensitivity analysis, costs indirectly attributed to school absenteeism that were calculated for the CUA that was part of the SIPS Jr HIGH trial were added to the total costs used in the base-case analysis. This analysis could be thought of as exploratory as the data on costs indirectly attributed to school absenteeism are uncertain.

Other than the scenario analysis presenting the LA perspective in which alternative values for both costs and outcomes were analysed, no further multiway sensitivity analyses were conducted.

8.2.3 *Methodological validation analysis*

The validation analysis altered only the estimation of costs, therefore, the methods applied to estimate benefit remain as described in the previous section.

As described in section 8.2.1, econometric techniques can be applied to cost data to adjust for between-group baseline differences in the distribution of costs that may be present despite randomisation. Other participant characteristics and contextual factors related to the intervention setting may also influence costs, for example, males are typically arrested more frequently than females (354), thus, participant gender may impact on the costs associated with arrests independently of intervention effects. Therefore, in order to account for cost variation between the intervention and control groups that could be explained by alternative factors, a multivariable analysis was conducted (332).

The cost estimator employed was a generalised linear model[25] (GLM). A GLM was employed over alternative models, such as ordinary least squares (OLS) regression, due to the skewed distribution of the cost data. GLMs have been demonstrated to behave well when estimating predicted means of healthcare cost data (313, 356, 357) and it has been reported that they produce more robust estimates for skewed data than applying OLS to log-transformed costs (313), which is an alternative method to deal with skewed data. For these reasons, GLMs are well-accepted estimators of costs when analysing clinical trial data (332).

Predicted total costs at follow-up were estimated adjusting for: trial arm, gender, location, baseline AUDIT (Alcohol Use Disorders Identification Test) score (358), race, smoking status at baseline, baseline EQ-5D-3L score (359), a dummy variable indicating that 12-month follow-up was either 30 days earlier or later than expected, and baseline total resource costs. The trial arm covariate was included in order to estimate the predicted difference in cost between control and intervention groups. Baseline costs were included to adjust for any differences in costs at the start of the trial between groups despite randomisation, as explained in the previous section. The dummy variable for late or early follow-up was included to adjust for the fact that some students were followed up after only 11 months or after 13 months rather than the intended 12 months in order to limit the impact of improper

[25] The GLM used a log link and a gamma distribution. The most appropriate link function and distribution family for the cost data were identified using a Boxcox approach to test the link function and the Modified Park test to determine the distribution family (313). This combination of link function and family is recommended as a suitable approach by the ISPOR Good Research Practices Task Force (355).

adherence to trial protocol on costs. The remaining covariates were selected as potential predictors for differences in costs independent of the intervention. For example, participants with higher AUDIT scores at baseline may be more likely to be arrested or use healthcare services more frequently than participants with lower AUDIT scores.

The regression-adjusted mean difference in total cost per risky-drinking student at follow-up between the intervention and control groups was estimated using Stata's "margins" command. When applied to a binary covariate (i.e. the variable denoting intervention or control), the margins command displays the mean incremental difference between the values of the binary covariate (i.e. the difference between intervention and control groups' costs)[26]. The value of the coefficient on the study arm covariate cannot be used directly to determine the incremental costs between the two groups since the use of a log link in the GLM estimator presents regression coefficients as percentage changes rather than absolute values. Obtaining estimations using the margins command is recommended to identify the outcome of incremental changes in binary covariates when using a GLM estimator (313). Bootstrapping using 20,000 repetitions[27] was employed to estimate 95% CIs around the difference in costs estimated by the GLM regression analysis.

The same procedures were followed to calculate the NSB and conduct sensitivity analyses as described in section 8.2.2. Cost-benefit planes were also generated using the stochastic analysis outcomes (See Figures 8.1 and 8.2).

8.2.4 *Primary analysis results*

The resource costs related to subsequent use of health, social care, and public services over the 12-month follow-up period of the SIPS Jr HIGH trial are displayed in Table 8.1. The mean delivery cost of the intervention across all students in the intervention arm was £22.20 (95% CI £21.80 to £22.60).

[26] Specifically, the margins command estimates the difference in expected values of the dependent variable at each of the binary covariate values (i.e. for each of intervention and control) keeping all other covariates fixed (313)

[27] As described in section 8.2.2, the appropriate number of bootstrap repetitions is determined by the size and variation of the sample. The sample size included in the regression analysis is smaller than the raw data (due to the exclusion from the analysis of observations with missing data on independent variables), therefore, a greater number of bootstrap repetitions were required to estimate robust standard errors and CIs. The procedure as described in footnote 21 was employed to determine the appropriate number of bootstrap repetitions.

Table 8.1 Mean costs and cost differences between intervention and control groups for each area of resource use over the 12-month follow-up period

Resource	Cost (£), mean (n)		Cost difference (£), mean (95% CI)
	Intervention	Control	Intervention – Control
GP visits	98 (173)	125 (190)	-26 (-63 to 9)
Social worker visits	27 (179)	9 (192)	18 (-12 to 49)
School nurse visits	83 (175)	54 (191)	28 (-20 to 78)
Hospital admissions	200 (178)	161 (193)	40 (-68 to 147)
A&E attendance	91 (178)	76 (193)	16 (-23 to 54)
Arrests	0.1 (179)	1 (194)	-0.9 (-1.6 to -0.1)
School absenteeism	1134 (181)	2083 (197)	-950 (-4600 to 2700)
Total resource cost A	501 (169)	403 (184)	98 (-66 to 262)
Total resource cost B	1715 (169)	2634 (184)	-919 (-4848 to 3010)

95% CI = confidence interval at 5% level
Costs over 12 months, not adjusted for participant characteristics
Total resource cost A = excluding school absenteeism costs for base-case analysis, Total resource cost B = including school absenteeism costs for sensitivity analysis
Mean costs, cost differences and CIs of the difference between the intervention and control groups are based on a comparison of means using an independent samples t-test with unequal variances

The results of the base case analysis are presented below in Tables 8.2 to 8.4. The aggregation of mean WTP values used for both the national and LA level analyses is displayed in Table 8.2. The mean WTP value for the ASBI intervention prior to aggregation was £65. The pre-aggregated mean difference in costs per risky drinking student between the intervention and control groups determined by the two-sided, unequal t-test was £117.85. The 95% CI around the WTP value in both the national and LA cases is positive, indicating strong likelihood that the benefit value is greater than zero. On the other hand, Table 8.3 outlines the aggregation of costs for each level of analysis and the 95% CIs around the mean cost difference cross zero. Whilst the point estimates of the aggregated cost differences are positive (£19.5 million and £85,442), indicating that the intervention arm is, on average, costlier than the control arm, the CIs indicate that there is a substantial possibility that the intervention is, on average, less costly than control.

The CBA results in the form of NSB for both the national level analysis and LA scenario analysis are displayed in Table 8.4. Both analyses result in a positive NSB with confidence intervals which do not cross zero. The NSB from the societal perspective is approximately £2

billion, whilst the NSB from the LA perspective is substantially smaller as would be expected given the difference in magnitude of the scale factors used for both costs and WTP. However, with an NSB of approximately £180,000, the LA scenario still indicates worth in the implementation of the ASBI. The outcome of the CBA could alternatively be reported as a BCR, in which the value of benefit is divided by the cost (see Chapter Three for further detail). Using this alternative metric, the outcome for the national level analysis would be 103 (Equation 8.1) and the LA scenario analysis would be 3.07 (Equation 8.2).

$$BCR = \frac{2.007\ billion}{19.5\ million} = 103 \qquad \textbf{(8.1)}$$

$$BCR = \frac{261{,}930}{85{,}442} = 3.07 \qquad \textbf{(8.2)}$$

The outcomes of the one-way sensitivity analyses adjusting costs and outcomes independently are reported in Table 8.5. There is no change in the direction of either costs or outcomes in any of the three sensitivity analyses shown to PHDMs (Year 10 parents providing the WTP scale factor and truncating WTP and costs at the 95% percentile). The only scenario which changes the direction of outcome is including school absenteeism costs. In this case, the mean cost difference is negative, indicating that the intervention arm is less costly on average than the control arm. However, reflecting the base-case analysis, the 95% CIs around the cost parameter in all five sensitivity analyses cross zero; this indicates substantial uncertainty around the cost estimate with regards to which trial arm is costlier. Whereas, the lower bound of all confidence intervals around the NSB outcome in each sensitivity analysis remain positive, suggesting with confidence that implementing the ASBI is worthwhile considering from either a national or local societal perspective.

Table 8.2 Willingness to pay values scaled for both national and LA level base analyses

	Individual level from CV study		National level			Local authority level		
	Mean (SE)	[95% CI]	Scale factor, n	Aggregated mean (SE)	[95% CI]	Scale factor, n	Aggregated mean (SE)	[95% CI]
WTP, £	65 (6)	[54 to 77]	30,800,000	2.007 billion (182 million)	[1.649 billion to 2.364 billion]	4,020	261,930 (24,022)	[214,846 to 309,015]

SE = Standard error; 95% CI = confidence interval at the 5% significance level; n = number of individuals
CIs and SEs calculated using bootstrapping

Table 8.3 Difference in cost between intervention and control arms scaled for both national and local level analyses

	Individual level from trial		National level			Local authority level		
	Mean (SE)	[95% CI]	Scale factor, n	Aggregated mean (SE)	[95% CI]	Scale factor, n	Aggregated mean (SE)	[95% CI]
Difference in cost intervention - control, £	117.85 (85.64)	[-286.39 to 50.69]	165,144	19.5 million (14 million)	[-8.05 million to 46.98 million]	725	85,442 (61,635)	[-35,362 to 206,245]

SE = Standard error; 95% CI = confidence interval at the 5% significance level; n = number of individuals
CIs and SEs calculated using bootstrapping

Table 8.4 Net societal benefit calculation for both national and local level analyses

National level				Local authority level			
Benefit, £	Cost, £	NSB (SE), £	[95% CI]	Benefit, £	Cost, £	NSB (SE), £	[95% CI]
2.007 billion	19.50 million	1.982 billion (3.68 million)	[1.975 billion to 1.989 billion]	264,629	85,442	177,000 (1,259)	[174,532 to 179,468]

NSB = Net Societal Benefit; SE = Standard error; 95% CI = confidence interval at the 5% significance level

Table 8.5 Sensitivity analyses of NSB for both national and local level analyses

Sensitivity analyses	National level			Local authority level		
	Benefit (WTP), £ (95% CI)	Incremental cost, £ (95% CI)	NSB, £ (95% CI)	Benefit (WTP), £ (95% CI[a])	Incremental cost, £ (95% CI[a])	NSB, £ (95% CI[a])
Year 10 parents WTP scale factor	45.8 million (37.6 m to 53.9 m)	19.5 million (-8.05 m to 46.98 m)	26.3 million (25.7 m to 26.8 m)	172,404 (142,000 to 203,000)	85,442 (-35,000 to 206,000)	86,774 (84,000 to 89,000)
WTP truncated	1.319 billion (1.164 b to 1.494 b)	19.5 million (-8.05 m to 46.98 m)	1.298 billion (1.295 b to 1.301 b)	172,169 (150,000 to 194,000)	85,442 (-35,000 to 206,000)	86,773 (84,000 to 89,000)
Cost truncated	2.007 billion (1.649 b to 2.364 b)	1.18 million (-14.8 m to 17.2 m)	2.000 billion (1.993 b to 2.007 b)	261,930 (215,000 to 309,000)	5,216 (-65,000 to 76,000)	256,023 (254,000 to 258,000)
2016 WTP values	1.914 billion (1.597 b to 2.282 b)	19.5 million (-8.05 m to 46.98 m)	1.889 billion (1.882 b to 1.896 b)	249,810 (208,000 to 298,000)	85,442 (-35,000 to 206,000)	163,968 (161,000 to 166,000)
Including school absenteeism costs	2.007 billion (1.649 b to 2.364 b)	-138.4 m (-740 m to 501 m)	2.148 billion (2.133 b to 2.164 b)	261,930 (215,000 to 309,000)	-607,612 (-3.25 m to 2.20 m)	907,846 (849,000 to 969,000)

95% CI = confidence interval at the 5% significance level; m = million; b = billion

[a] to nearest £1000

The cost-benefit planes for both the national and LA analyses are displayed in Figure 8.1 and Figure 8.2, respectively. The crossed lines in the plane illustrate the 95% CIs around the cost difference and WTP parameters, as calculated by the bootstrap repetitions. All points are in the north-east and south-east quadrants of the plane demonstrating that benefit was positive in all repetitions. The majority of the repetitions are in the north-east quadrant (92%), indicating a greater likelihood of the intervention being costlier, however, a small cluster of bootstrap observations can be identified in the south-east quadrant (8%), representing instances where the intervention arm is less costly than control. Neither of the cost-benefit planes provide evidence of any extreme outliers; the majority of repetitions are clustered closely together around mean NSB.

Figure 8.1 Cost-benefit plane for national level primary analysis with 95% CI illustrated

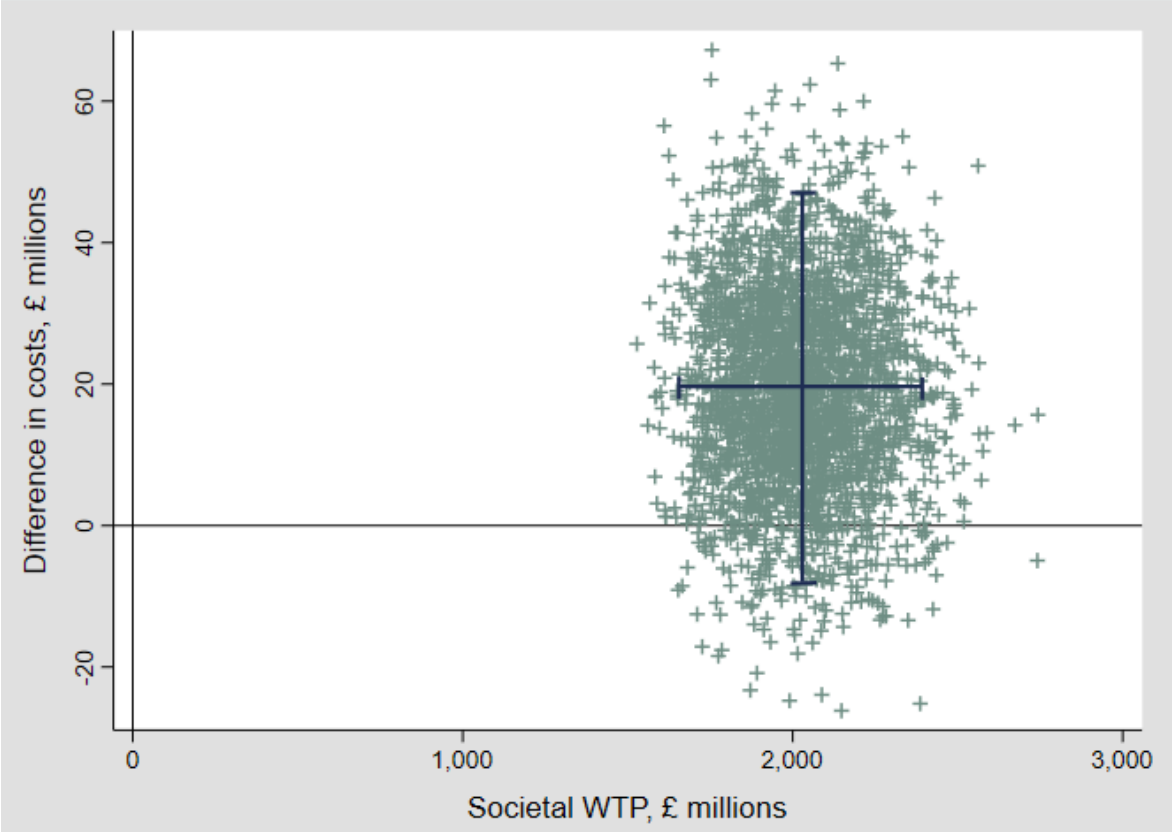
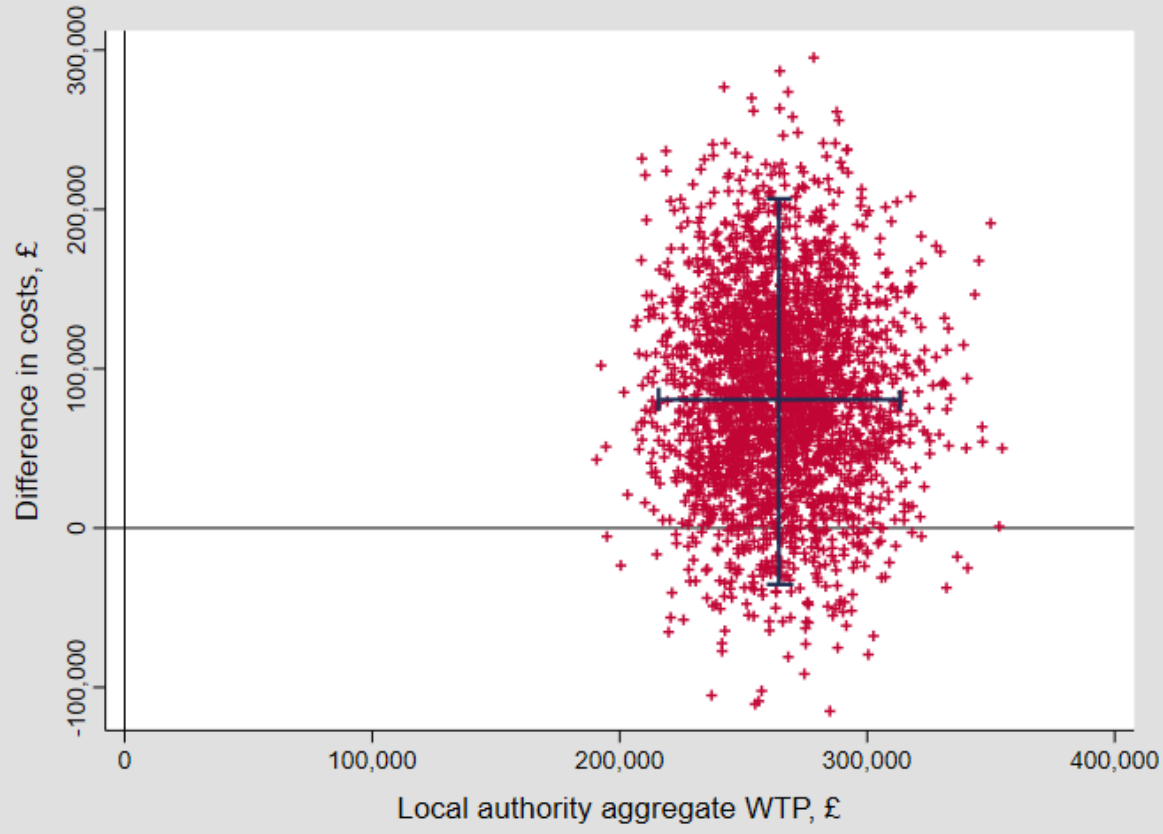


Figure 8.2 Cost-benefit plane for LA scenario analysis with 95% CIs illustrated



8.2.5 *Validation analysis results*

The results of the validation analyses using regression-adjusted costs are displayed in Table 8.6 and Table 8.7. No change was made to the WTP values, therefore, the data displayed in Table 8.2 represents the benefit values also used for the validation analyses. The mean difference in costs between the intervention and control groups predicted by the GLM regression model is £86.28 (Table 8.6), which is smaller than the difference estimated using the unadjusted follow-up values. This is expected given that the regression analysis accounts for some of the variance in costs influenced by non-intervention factors. Therefore, the predicted cost difference can be attributed more closely to direct impacts of the intervention than the difference estimated using unadjusted costs. Similarly to the primary analysis, the 95% CI around costs crosses zero, suggesting that although the mean point estimate of the cost difference suggests that the intervention group costs are greater than the control group costs, it is possible that the control arm is costlier than the intervention arm.

The values of NSB estimated in the validated analyses are fairly similar to those estimated in the base analyses. Both the national level and LA scenario analyses have larger NSB values compared to the primary analysis (£1.989 billion compared to £1.982 billion and £192,398 compared to £177,000, respectively) due to the smaller difference in costs estimated by the GLM regression. The 95% CIs around the NSB estimates are of a similar range to those in the primary analysis and none cross zero. The BCRs of the national and LA level analyses are 141 and 4.19, respectively. As with the NSB, the ratios are slightly larger than the primary analysis due to the smaller cost difference relative to the WTP benefit value.

Table 8.6 GLM regression-adjusted difference in cost between intervention and control arms scaled for both national and local level analyses

	Individual level from trial		National level			Local authority level		
	Mean (SE)	[95% CI]	Scale factor, n	Aggregated mean (SE)	[95% CI]	Scale factor, n	Aggregated mean (SE)	[95% CI]
Difference in cost intervention - control, £	86.28 (115.08)	[-139.26 to 311.83]	165,144	14.25 million (49.7 million)	[-23.97 million to 56.15 million]	725	62,556 (223,078)	[-102,163 to 249,795]

SE = Standard error; 95% CI = confidence interval at the 5% significance level; n = number of individuals
CIs and SEs calculated using bootstrapping

Table 8.7 Net societal benefit calculation for both national and local level analyses using GLM regression-adjusted costs

National level				Local authority level			
Benefit, £	Cost, £	NSB, £ (SE)	[95% CI]	Benefit, £	Cost, £	NSB, £ (SE)	[95% CI]
2.007 billion	14.25 million	1.989 billion (1.75 million)	[1.985 billion to 1.992 billion]	261,930	62,556	192,398 (1,722)	[189,027 to 198,570]

NSB = Net Societal Benefit; SE = Standard error; 95% CI = confidence interval at the 5% significance level

The one-way sensitivity analyses of the validated analysis is displayed in Table 8.8. The sensitivity analyses results are largely similar to those reported using the primary method of analysis but with slightly larger NSB values as a result of the smaller cost difference estimated by the GLM regression analysis (£86.28 compared to £117.85). There are, however, two notable differences between the sensitivity analyses using the primary and validation methods of analysis. Firstly, the mean estimate of the difference in regression-adjusted costs truncated at the 95% percentile is negative, suggesting that the control arm is costlier than the intervention arm after the largest 5% of values are removed from the analysis. However, the 95% CIs indicate a lot of uncertainty around this point estimate, therefore, there is a substantial possibility that the intervention arm is in fact costlier on average. Secondly, whilst the direction of input parameters and NSB of the analysis including school absenteeism costs is the same using both the primary and validation methods of analyses, the NSB at both the national and LA levels is considerably larger in the validation analysis and there is considerably more uncertainty around the NSB based on the CIs. The 95% CIs around NSB in the validation analysis do not cross zero, however the intervals are much larger compared to those estimated using the less complex estimator in the primary analysis. Explanation for these findings is proposed in section 8.2.6.

Table 8.8 Sensitivity analyses of net societal benefit for both national and local level analyses using GLM regression-adjusted costs

Sensitivity analyses	National level			Local authority level		
	Benefit (WTP), £ (95% CI)	Incremental cost, £ (95% CI)	NSB, £ (95% CI)	Benefit (WTP), £ (95% CI[a])	Incremental cost, £ (95% CI[a])	NSB, £ (95% CI[a])
Year 10 parents WTP scale factor	45.79 million (37.71 m to 53.86 m)	14.25 million (-23.97 m to 56.15 m)	29.76 million (28.90 m to 30.62 m)	172,404 (144,000 to 205,000)	62,556 (-102,000 to 250,000)	102,931 (99,000 to 106,000)
WTP truncated	1.319 billion (1.152 b to 1.486 b)	14.25 million (-23.97 m to 56.15 m)	1.303 billion (1.301 b to 1.305 b)	172,169 (150,000 to 194,000)	62,556 (-102,000 to 250,000)	102,779 (99,000 to 106,000)
Cost truncated	2.007 billion (1.649 b to 2.364 b)	-3.63 million (-28.5 m to 21.2 m)	2.009 billion (2.006 b to 2.012 b)	261,930 (215,000 to 309,000)	-15,930 (-125,000 to 94,000)	279,914 (279,000 to 281,000)
2016 WTP values	1.914 billion (1.576 b to 2.252 b)	14.25 million (-23.97 m to 56.15 m)	1.896 billion (1.893 b to 1.899 b)	249,811 (206,000 to 294,000)	62,556 (-102,000 to 250,000)	180,286 (177,000 to 184,000)
Including school absenteeism costs	2.007 billion (1.649 b to 2.364 b)	-28.35 million (-445 m to 291 m)	3.426 billion (1.963 b to 7.551 b)	261,930 (215,000 to 309,000)	-124,463 (-1.97 m to 1.28 m)	58.40 million (305,000 to 219 m)

95% CI = confidence interval at the 5% significance level; m = million; b = billion

[a] to nearest £1000

The cost-benefit planes using the bootstrap repetitions of the validation analysis are displayed in Figures 8.3 and 8.4 below. Unlike the primary analysis, there are a number of visible outliers on the cost parameter estimated by the GLM regression. However, as with the primary analysis, all points lie in either the north-east or south-east quadrants. A greater proportion reside in the north-east quadrant (81%), in which the intervention arm is costlier than the control arm, than the south-east quadrant (19%). Creating a cost-benefit plane of the sensitivity analyses in which costs are truncated at the 95th percentile removes the outlier observations visible in Figure 8.3 and Figure 8.4. The distribution of bootstrapped observations in the truncated cost-benefit planes (see Figure L.1 and Figure L.2 in Appendix L) appear much more similar to those in Figures 8.1 and 8.2, suggesting that the majority of the bootstrapped estimates are similar in both the validation and base analyses.

Figure 8.3 Cost-benefit plane for national level validation analysis with 95% CI illustrated

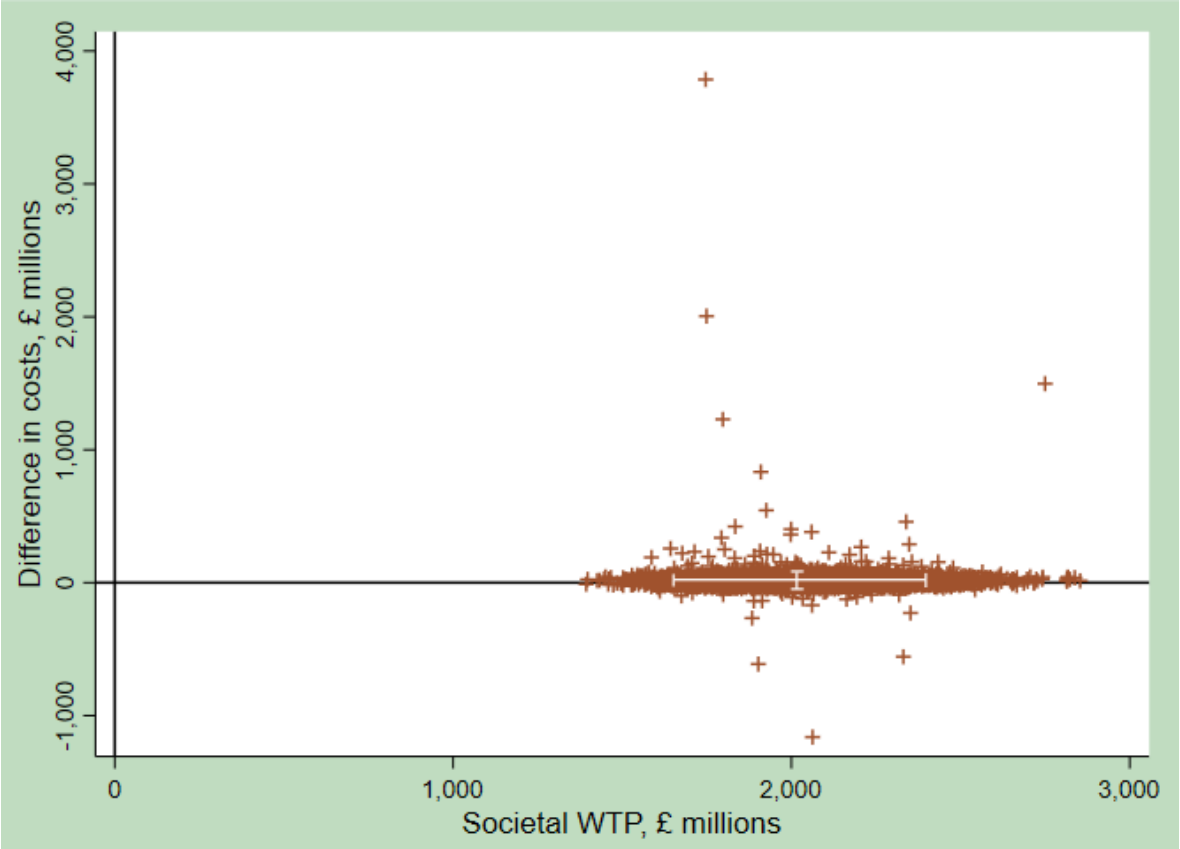
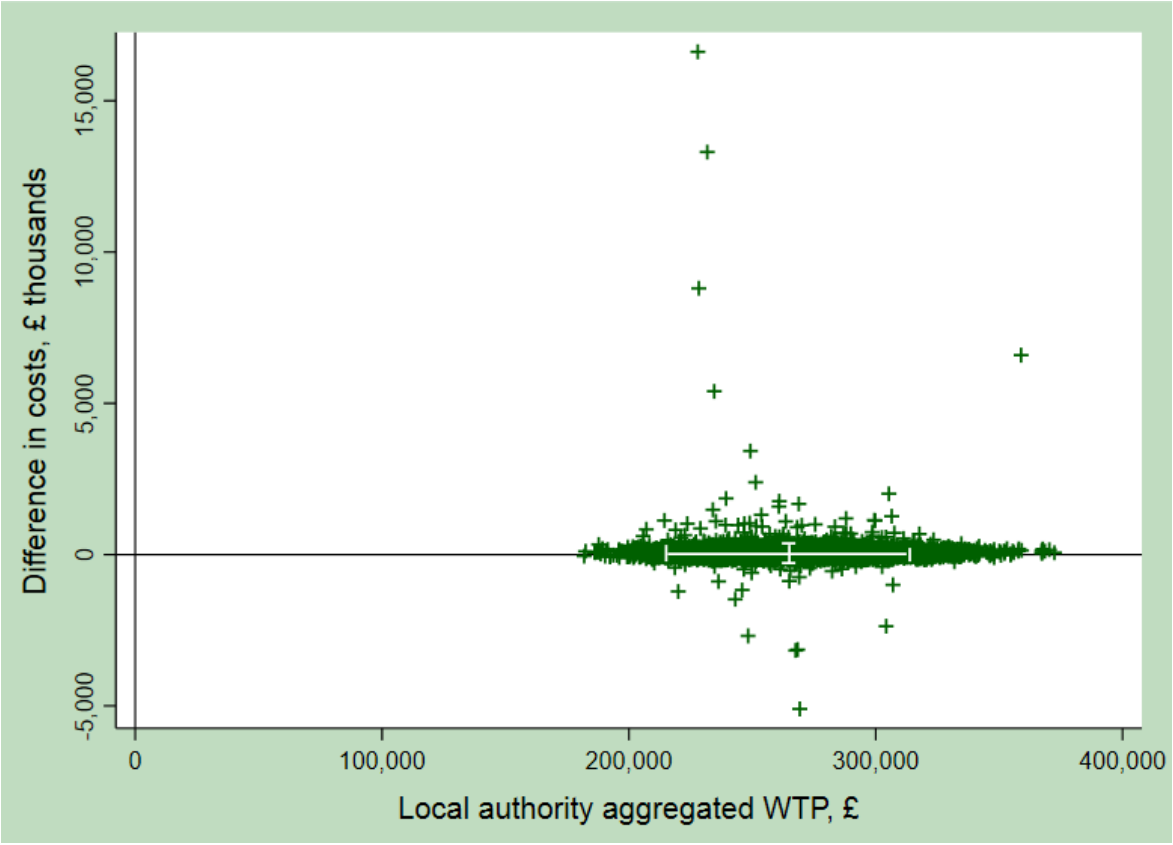


Figure 8.4 Cost-benefit plane for LA validation analysis with 95% CI illustrated



8.2.6 *Discussion*

The results of the CBA indicate that the ASBI has a positive net societal benefit at both a national and LA level. The 95% CIs around the NSB outcomes do not cross zero in any of the base case or sensitivity analyses. This indicates that, given the assumptions made in the analysis, the benefits, as valued by the UK general public, outweigh the incremental costs of providing the ASBI compared with standard practice.

The magnitude of the NSB estimated in this study is worthy of discussion here. The NSB at the national level in the primary analysis was estimated at approximately £2 billion, which is equivalent to approximately 0.1% of the UK Gross Domestic Product in 2018[28]. The LA scenario NSB was estimated at approximately £180,000 which is approximately 0.8% of the public health ring-fenced grant awarded to Newcastle City Council in 2018[29]. These figures are not inconsequential; therefore, their feasibility must be considered. The main cause of these large values is the disparity in scale factors used for the cost and benefit parameters.

[28] According to the International Monetary Fund, UK Gross Domestic Product for 2018 was approximately £2.033 trillion (360)

[29] The public health ring-fenced grant in 2018 was approximately £23.5 million (361)

The use of the entire tax base as a scale factor for the benefit side of the equation at the national level could be argued against; however, the LA scenario utilised a more conservative scale factor of only the top three highest council banded households. Additionally, sensitivity analysis was conducted in which a plausible minimum scale factor of only parents of risky-drinking young children is used. In each of these scenarios the NSB is still positive, suggesting that regardless of which scale factor is used, the CBA results are robust to the conclusion that the ASBI has positive benefit for society compared to standard care.

The other four sensitivity analyses also conclude that the NSB is positive and that this outcome is robust regardless of various assumptions. Nevertheless, the two observations regarding the sensitivity analyses conducted in the validation analysis raised in section 8.2.5 require further discussion. Firstly, the change in direction of incremental costs following truncation at the 95th percentile in the validation analysis can be explained by a combination of two factors. Firstly, the smaller difference between estimated mean costs for the two trial arms after accounting for alternative sources of cost variance in the regression analysis, and secondly, almost twice as many observations in the top 5th percentile of costs were from the intervention arm compared to the control arm (n=11 and n=6, respectively). Thus, once the top 5th percentile of costs were removed, the impact on mean costs was greater in the intervention arm compared with the control arm. This had a substantial impact on the resulting cost difference between the two arms regardless of which analytic method was used. In the primary analysis the cost difference reduced to approximately £1 million compared to £19 million in the base case at the national level, and to approximately £5000 from over £85,000 in the base-case at the LA level. Due to the smaller initial difference in mean costs in the validation analysis compared to the primary analysis (£86.28 compared to £117.85), it is likely that the truncation of costs resulted in the mean costs for the intervention arm reducing to below those of the control arm in the validation analysis.

Secondly, the inclusion of missed school costs introduced a lot of variance to the total cost variable which had a profound impact on the regression analysis of costs in the validation analysis. The measures of skewness and kurtosis of the distribution of total costs prior to the inclusion of school absenteeism were 2.75 and 10.70, respectively. For comparison, a Gaussian normal distribution would have a value of 0 for skew and 3 for kurtosis. However, once school absenteeism costs are included, the total cost distribution is skewed much

further to the right resulting in a skewness measure of 10.56 and kurtosis of 112.66. Therefore, despite having chosen the GLM estimator for its ability to deal with skewed distributions, it is possible that the resulting distribution was too greatly skewed even for the GLM to provide efficient estimates. This can be observed from the results of the sensitivity analysis, in which the presence of a very small proportion (<1%) of extreme outliers in the bootstrap repetitions of cost difference have drastically affected the mean value of NSB.

In both the national and local level analyses, the values of NSB for the sensitivity analysis including school absenteeism costs calculated using each pair of bootstrap repetitions for benefit and cost is far greater than the outcome of the point estimate of benefit minus the point estimate of cost difference as a result of the extremely large outliers that lie outside of the range of the 95% CIs. This explanation was verified by observing the difference in the NSB after removing outlier values[30]. This resulted in point estimates of mean NSB of £2.008 billion and £344,103 for the national and local level scenarios, respectively. These values align much more closely to the values that would be observed by calculating NSB from point estimates of mean benefit and cost (£2.035 billion and £386,393, respectively).

The inclusion of costs associated with school absenteeism resulted in some extreme outcomes due to the skewed distribution, particularly in the validation analysis. Whilst less extreme in the primary analysis, the NSB outcomes were still considerably larger than the base-case analysis, specifically for the LA scenario where NSB was five times larger than the base-case value. Whilst the choice to exclude school absenteeism costs from the base-case analysis were made on theoretical grounds, the extreme skew their addition introduced to the distribution of costs justifies this decision on statistical grounds. The sample size of the unadjusted costs sample (n=345) is unlikely to be sufficiently large to overcome the extreme skew introduced disproportionately between trial arms, therefore, t-tests might not be trusted to provide robust estimations of difference in costs. Additionally, as was demonstrated above, the regression analysis was also unable to provide reliable estimates.

Nevertheless, the overall conclusion regarding the ASBI does not change depending on whether school absenteeism costs are included or excluded from the base-case analysis. The base-case analysis presented to the PHDMs could, therefore, be considered a conservative analysis. Typically, demonstrating a favourable outcome from a conservative analysis is

[30] Outliers were defined as greater than two standard deviations from the mean. Two out of 17,031 observations removed in total.

preferred to that of an optimistic analysis since decision-makers can have greater confidence that the outcome is truly efficient if the results show the intervention to be favourable compared with the comparator in the worst-case scenario.

Overall, the validation analysis demonstrated that the outcome of the base-case CBA does not change dramatically depending on whether a simple or complex method of analysis was employed and the conclusion that the ASBI provides positive net benefit to society is stable. Additionally, the conclusions drawn by three sensitivity analyses presented to the PHDMs do not change. Therefore, from the point of view of presenting the simple model of the economic evaluation at the workshop, the results presented in this chapter confirm that this was a suitable decision.

A comparison of the results identified in the current study against existing literature is difficult given the paucity of CBAs that have been published, even within the broader area of alcohol treatment or general prevention of alcohol misuse. The pool of literature is further restricted when comparing with evaluations that have used stated preference methods to value benefits (see Chapter Four, in which no studies were identified using CV methods). Comparing the current study results with one of the few CBA studies conducted in a similar area, which also aggregated costs and benefits on a national level, De Wit *et al.* (237) conducted CBAs of three alcohol policies (tax increases, alcohol outlet density reductions, and an advertising ban). Their analysis found an aggregated social benefit across the Dutch population of between €4 billion - €12 billion in 2013 euros, policy dependent (equivalent to approximately £3.83 billion - £10.4 billion in 2018 pounds sterling).

It is difficult to directly compare De Wit *et al.'s* (237) results with those found in the current study, since the interventions and methods of calculating benefits differ; however, the social benefit value of the least effective intervention from the De Wit *et al.* (237) study is almost twice as large as the national-level NSB identified for the ASBI. Yet, given that De Wit *et al.* (231) report the net present benefit over 50 years and the current study considers the net benefit of the ASBI over only one year, it would be expected that the societal benefit of the Dutch alcohol policy would be larger than the societal benefit of the ASBI by more than a factor of two, even accounting for a four-fold difference in population size between the Netherlands and the UK[31]. This suggests that the estimate of the ASBI using all UK tax payers

[31] De Wit *et al.* (237) report the size of the population in the Netherlands to be approximately 14.5 million, compared to the size of the UK population considered in the current study (approximately 66 million).

as a scale factor may be overestimating the benefit to society of the ASBI, as discussed previously in this section.

8.2.7 *Strengths and limitations*

A strength of the current study is its novelty as a CBA of an alcohol prevention intervention, particularly one in which stated preference measures have been used to value the intervention. However, there are some limitations with the study. Firstly, no alternative estimators were used to attempt to model costs when those associated with school absenteeism were included. Alternative econometric techniques may be better suited to extremely skewed distributions. However, the underlying problem is associated with the fact that clinical trials are not powered to detect significant differences in economic outcomes, therefore, econometrics may not be able to truly address the issue of insufficient data in certain outcomes (335).

Additionally, excluding the sensitivity analysis demonstrating the impact of school absenteeism costs from the report shown to PHDMs could be considered a limitation. However, as explained in section 8.2.6, the analysis excluding school absenteeism costs is conservative, therefore, their inclusion would have strengthened the conclusions from the CBA; as such, there is no reason to suspect that the exclusion of this sensitivity analysis would have altered PHDMs' impression of the CBA.

Thirdly, the choice of the UK tax base as a scale factor could be considered a limitation, given that this method produced extremely large NSB. However, the sensitivity analyses presented alongside the base-case analysis provide sufficient indication that the conclusion remains the same even with a reduced scale factor. Similarly, for the LA analysis the use of the total households in the top three council-tax bands (n=4020) may be considered a small number (approximately 3% of all households in Newcastle). However, this offers a conservative analysis, since the benefit value is directly proportional to the size of the scale factor. Consequently, using the number of households in the top four or five council-tax bands would only strengthen the results reported here.

Finally, costs for the national-level analysis were calculated based on the assumption that the average proportion of risky drinking Year 10 students across the UK is consistent with the average proportion identified in the SIPS Jr HIGH trial (23.5%). This average value was presented as fixed in the CV survey and elicited WTP values on such an assumption, whereas

in reality the trial identified a distribution of proportions amongst the trial site locations (59). This distribution in proportions was exploited for the LA level analysis in which the higher proportion of Year 10 risky drinkers for the North-East region (27.4%) was used to calculate the costs in a Newcastle LA scenario. A caveat of the LA analysis is, therefore, that mean WTP remains consistent when a slightly larger proportion of Year 10 students are categorised as risky drinkers.

8.3 Social return on investment

SROI has become increasingly promoted in UK local government (33), which possibly explains the interest in the method by PHDMs interviewed during the qualitative study in Chapter Five. SROI has been promoted in policy-making arenas as a more pragmatic version of CBA from the perspective of an investor in health services (362), however, since it does not have the same theoretical grounding as CBA, health economists may still prefer CBA over SROI. Nevertheless, SROI has been proposed as potentially more relatable for policy makers at local or national levels because of the language familiarity of a pound-for-pound return on investment (33, p.296).

At the time of conducting the SROI of the ASBI, no SROI literature in alcohol prevention had been identified, since then only one study on that topic has been identified, which was based in Thailand (233) (see Chapter Four). Minimal studies were, therefore, available to provide relevant guidance for methods. Additionally, where SROI studies in other areas could be identified (e.g. housing (165)) the methods could not be guaranteed to be consistent with other SROI studies. Methods for SROI have not yet been standardised (33) and whilst a select few guidance documents are available (84, 164) there exists no guidance or examples of an SROI conducted using clinical trial data. Therefore, this study represents the best possible interpretation of SROI methodology applied to the available data from the SIPS Jr HIGH trial.

A cornerstone of SROI is broad stakeholder engagement throughout the process of evaluation from identifying outcomes, verifying appropriate indicators, and, where needed, extracting proxy financial values. Due to resource availability, stakeholder engagement was limited in the current study to that conducted for SIPS Jr HIGH economic evaluation and examination of the findings of the qualitative interview component of the project (59). Therefore, the evaluation presented here may be considered a limited SROI with respect to broad stakeholder engagement.

8.3.1 *Stages of SROI analysis*

Chapter Three reports the six stages of conducting an SROI. The methods used for the current study are outlined below for each stage of analysis (stages 1-5).

Stage 1: Establishing scope and identifying key stakeholders

An evaluative SROI was conducted using outcomes from the SIPS Jr HIGH trial. A broad range of stakeholders was considered: The National Health Service & Personal Social Services (NHS & PSS), the LA, the students exhibiting risky drinking behaviour, and their parents. Qualitative interviews were conducted with parents of students in the trial as part of the SIPS Jr HIGH study (59). The qualitative findings were reviewed to identify any relevant outcomes for parent stakeholders, however, none were identified. Consequently, parents were removed for consideration as stakeholders in the SROI.

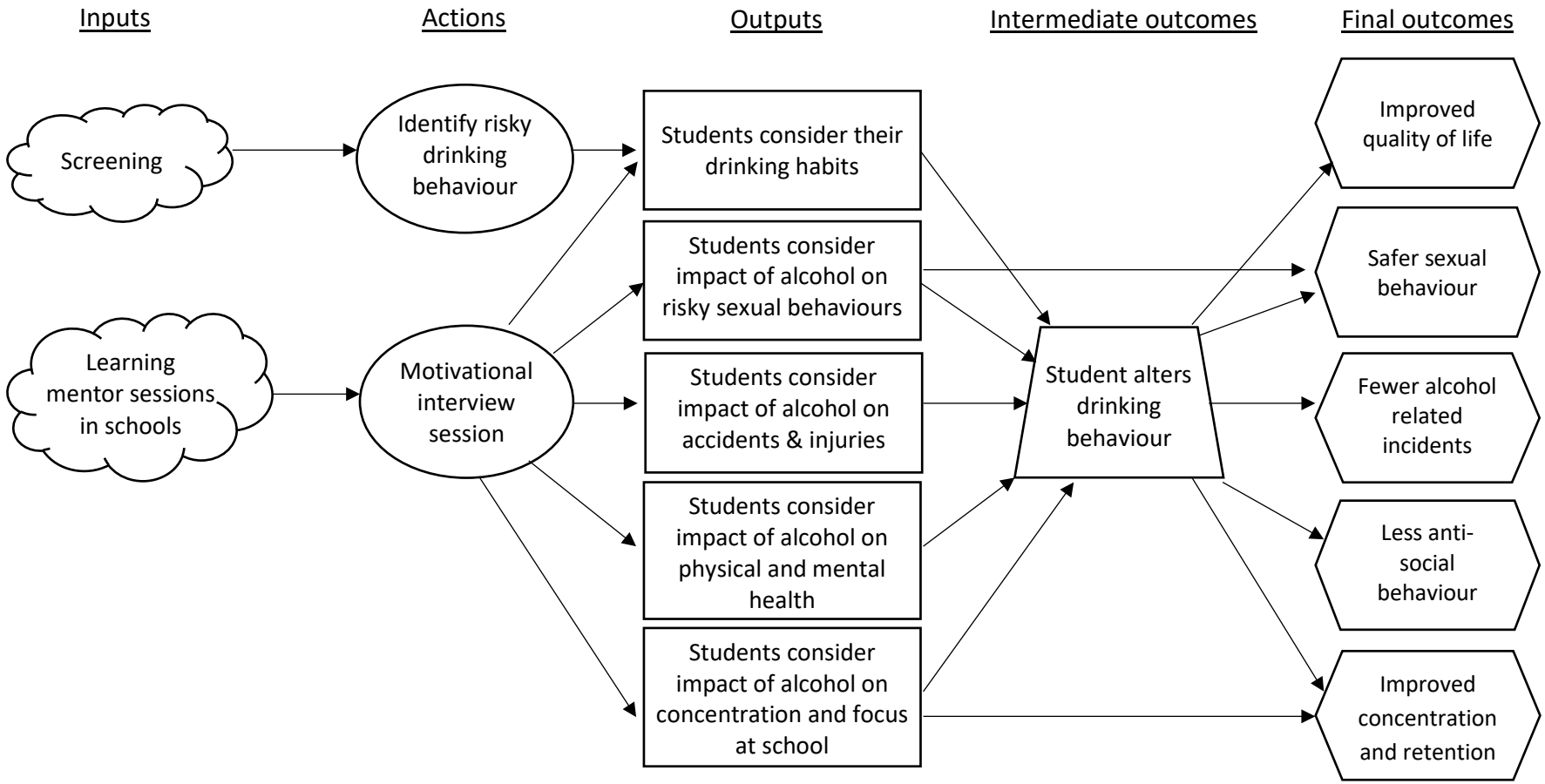
Stage 2: Mapping outcomes

Examination of the ASBI intervention as reported by the SIPS Jr HIGH trial and the outcomes of the trial assisted the outcomes mapping stage of the SROI. An impact map was developed to outline the linkages between the inputs and actions of the trial and final outcomes. In many cases, the final outcomes arrive via the intermediate outcome of reduced alcohol consumption, however, as the motivational interview component of the ASBI discussed a range of behaviours linked to alcohol consumption, it was proposed that several final outcomes could have also been affected independently of the intermediate outcome. For example, discussing the impact of alcohol consumption on concentration at school and the outcomes of this in relation to their educational attainment may have encouraged students to improve their engagement in class regardless of whether their alcohol consumption changed. The impact map is illustrated in Figure 8.5. The final outcomes were:

1. Improvement in health-related quality of life
2. Safer sexual behaviour
3. Fewer alcohol related accidents and injuries
4. A reduction in anti-social behaviour
5. Improved focus and concentration at school

The final outcomes were anticipated to have an effect on several stakeholders; the different effects were captured and accounted for separately when values were placed on outcomes in the following stage.

Figure 8.5 Impact map of the ASBI on all stakeholders



Stage 3: Establish indicators and applying values

In order to place a monetary value on each outcome, relevant indicators have to be identified in order to account for the magnitude of each area of impact. The indicators identified for each outcome are displayed in Table 8.9. For the majority of the indicators chosen, data were available from the trial. However, data for the indicators for a reduction in risky sexual behaviour were not collected by the trial. The only related data collected were "engaging in regretted sexual intercourse" or "engaging in sexual intercourse without a condom" (59). Neither of these data alone could indicate whether negative repercussions occurred either with unwanted pregnancies or sexually transmitted infections. Additionally, the data collected were reported to be inconsistent, for instance students stating at follow-up that they have "Never engaged in regretted sexual intercourse" yet reporting to have done so at baseline.

Table 8.9 Indicators identified for each outcome of the SROI, by stakeholder

Stakeholder	Outcome	Indicator
Student	Improved quality of life (via both physical and mental health improvements)	Improvement in EQ-5D-3L score
	Improved focus and concentration at school	Less school absenteeism [A]
NHS & PSS	Fewer accidents and injuries	Fewer secondary care visits (including both A&E and in-patient hospital care)
	Improvement in student's quality of life	Fewer visits to GP
		Fewer visits to school nurse [B]
	Less anti-social behaviour	Fewer social worker visits
Local Authority	Less anti-social behaviour	Fewer arrests and police cautions

[A] Ideally, improvements in educational attainment would have been used as an indicator for improved concentration and focus in school, however, these data were not available, therefore, a less direct indicator of school absenteeism was considered.

[B] Students may visit a school nurse to discuss physical health/mental health issues or to seek advice regarding substance misuse rather than visit a GP. School nurses may refer students to a GP if necessary, therefore, including both school nurse and GP visits is necessary to account for the full impact of the intervention on student's HRQoL.

The ideal indicator to capture the impact of the intervention on improved focus and concentration at school would be educational outcomes (e.g. exam results). However, data

were not available on exam results or other measures of educational achievement within the 12-month time frame of the trial follow-up. An indirect measure of improved focus and concentration in school might be school absenteeism, since it could be expected that students would attend school more frequently if they were better engaged in their classes; boredom and disengagement with teachers or subjects has been recognised as one of the key reasons for absenteeism amongst secondary school students (363). Since data were collected on days missed from school, this indicator was chosen.

In order to value the impact that the ASBI intervention had on each of the identified outcomes, it was necessary to identify appropriate proxy financial values for each indicator. For each of the health and social service indicators, the cost of an attendance was used as a proxy for the financial value of the saving to the NHS and PSS resulting from a reduced use of services; this approach has been recommended for valuing outcomes in an SROI (33). As part of the SIPS Jr HIGH trial health economics component, unit costs had been identified from various sources for the health and social care, LA and student-related indicators listed above (see Table 20 in Giles *et al.,* 2019 (59) for details on the sources used for each element of impact). The costs identified for each of those elements as part of the trial were used to populate the SROI where relevant. The savings in costs associated with arrests were taken from expert opinion as appropriate cost data were not available (59).

A cost was attributed to substantial school absenteeism (i.e. five or more days missed from school in six-months) as part of the SIPS Jr HIGH trial. As discussed in section 8.2.2*,* the cost attributed to school absenteeism represented the present value of reduced future earnings as a result of lower educational attainment, therefore, capturing the impact of the intervention on educational attainment but using an alternative indicator. As an indirect cost, it was excluded from the CBA base-case analysis, however, the process used to place a value on school absenteeism is similar to the human capital approach of valuing health outcomes in which future productivity gains, based on future earnings, are considered an appropriate proxy value (alongside reductions in healthcare costs) for the accrued benefit (118). Although this approach to valuing health outcomes has largely been replaced by alternative methods that are less biased towards high-earners (81), the consideration of future productivity gains could be argued an appropriate proxy value for the impact of the ASBI on educational attainment specifically. Consequently, the cost estimated by the trial

health economist for substantial school absenteeism was used as the proxy financial value of improved educational attainment.

No relevant valuations could be identified for the indicators available from the trial data relating to risky sexual behaviour. In order to place a value on "regretted sexual intercourse" eliciting a value via stated preferences could have enabled a value to be placed on the extent to which individuals would value avoiding regretted sexual intercourse. However, the resources were not available to undertake the necessary study and no relevant literature could be identified. Alternatively, costs related to teenage pregnancy or the contraction of sexually transmitted diseases could be roughly estimated from the trial data on intercourse without a condom. However, the proxy values attributed to these indicators would largely consist of additional health service use, which would result in double counting given that data collected on health service use during the trial was not restricted to attendances directly related to alcohol use. Therefore, any savings in costs attributed to reduced risky sexual behaviour were considered to be captured by health service use and were not estimated separately.

The indicator assigned to the outcome of improvement in quality of life was the EQ-5D-3L. Participants' EQ-5D-3L scores were combined with a time component representing the duration over which any improvement (or deterioration) in health-related quality of life was observed in order to estimate QALY gain over the 12-month period as a result of the ASBI. Monetary values can be applied to QALYs in order to estimate a proxy financial value for health improvement (164) and the value recommended for use in SROI by PHE is £60,000 per QALY[32] (164), a value originally recommended by the Department of Health (242).

Step 4: Establishing impact

The penultimate stage of the SROI analysis is to apply each of the financial proxy values to the indicators and account for deadweight, displacement, attribution, and drop-off to avoid over-estimation of impact (see Chapter Three for details on each of these elements). Deadweight refers to the counterfactual of any impact as a result of the intervention. The recommended approach to capture deadweight is via comparison with a control group (33).

[32] This value is three-times the value of a QALY for which incremental cost-effectiveness ratios of interventions are considered cost-effective (£20,000 per QALY) by the National Institute for Health and Care Excellence for evaluations using cost-utility analysis (10).

The difference in impact between the intervention and control groups at 12-month follow-up was, therefore, estimated in order to account for deadweight.

Displacement does not always occur, it is mostly considered relevant to interventions aimed at reducing crime where crime may simply move from the area receiving the intervention to another (33). Whilst impacts related to crime (i.e. number of arrests) were included in the SROI presented in this study, the impact on crime was measured at the individual participant level rather than, for example, levels of anti-social behaviour in a particular location. Consequently, it was not considered relevant to account for displacement for any of the outcomes measured for this SROI.

The attribution of the intervention to the observed outcomes was considered to be accounted for by virtue of the RCT setting (see section 8.2.1). Therefore, the attribution of the intervention to outcomes was considered to be 100%. Following the protocol described in section 8.2.1, this was examined by also conducting a methodological validation analysis using proxy values that were adjusted for baseline characteristics (see Appendix M).

Finally, no drop-off was accounted for since the time-horizon of the SROI study was 12-months, thus, it was unnecessary to account for future depreciation of outcomes resulting from diminishing returns of the intervention.

The calculation of the total impact value follows Equation 3.5 in Chapter Three. For the impact on health or social services resources, arrests, and school absenteeism the financial proxy value was represented by the mean cost of each resource for the group receiving the ASBI in the trial[33]. The value of deadweight for each resource was the mean cost for the control group of the trial. In order to calculate a total monetary value of the impact for each resource, the mean cost was then multiplied by the quantity of impact, which was represented by the number of students in the group receiving the intervention (n=210). This reflects an intention-to-treat design, which is recommended for pragmatic clinical trial data (335). As the intervention and control groups in the trial differed in size (n=210 and n=233, respectively), the total value of deadweight was estimated using the same number of students as the group receiving the intervention (n=210) to ensure the value of deadweight truly reflected the counterfactual of the intervention's impact.

[33] Mean cost was calculated as the sum of the unit cost of the resource multiplied by quantity of resource use for each participant, divided by the number of participants with non-missing data at follow-up.

The total monetary value of impact for each resource was then estimated by subtracting the value calculated for the group that received the intervention from value calculated for the control group to obtain the incremental impact of the intervention over what would have occurred in the absence of any intervention. The resulting value represented the financial saving from a reduction in resource use as a result of implementing the ASBI. A reduction in resource costs would be considered a positive outcome, therefore, total impact was calculated in this manner so that the estimated value of impact would be a positive integer if the intervention had a favourable effect on outcomes.

The approach taken to calculate total impact was considered more appropriate than the standard SROI approach of estimating impact as the unit cost multiplied by the quantity of impact adjusting for deadweight as a percentage (33) since deadweight was not estimated as a percentage but via the observation of outcomes from a designated control group within the SIPS Jr HIGH trial.

The impact of the ASBI on health was calculated by first multiplying mean QALY gain over the follow-up period for both intervention and control groups by £60,000 to obtain a monetary value of QALYs for each group. The corresponding values were then multiplied by the number of students in the intervention arm (n=210), as explained previously, in order to estimate a total financial value of QALY gain at follow-up for each group. Contrary to the impact on resources, a positive outcome for QALY gain is represented by a greater value for the intervention group compared to the control group. Thus, in contrast to the calculation for total resource impact, the overall health impact was calculated by subtracting the financial value of QALY gain of the control group from the intervention group. Thus, if the intervention generated greater QALY gain than would be generated in the absence of intervention, the total financial value of impact would be a positive integer.

Step 5: Calculating the SROI

The SROI calculation is outlined in Equation 3.5 in Chapter Three and requires the present value of total impact to be divided by the total investment value. As explained previously, no discounting was necessary to obtain the present value of impacts as a 12-month time horizon was used. The total present value of impact was estimated by summing the total values of impact on resources and impact on health.

Intervention delivery cost, the components of which are outlined in section 8.2.2, constituted the value of investment. The total investment value was calculated in a similar manner to impact described earlier; the mean cost for each element of intervention delivery was multiplied by the number of students in the intervention arm (n=210).

8.3.2 *Sensitivity analysis*

As with other forms of economic evaluation, sensitivity analysis is recommended to examine whether SROI results are robust to alternative assumptions regarding input parameters (33). Two sensitivity analyses were conducted to examine (i) the impact of one of the financial proxies (the monetary value of a QALY) and (ii) a component of resource impact (school absenteeism).

The monetary value used to calculate the financial value of health gain was £60,000 per QALY. However, there are several alternative values of a QALY used in economic evaluation which could be used to value the health gains associated with the intervention. NICE consider a value of £20,000 per QALY or less to be considered cost-effective and a recent estimation of the value of a QALY by Claxton *et al.,* 2015 (364) estimated a value of £12,936 per QALY. Both values were used in sensitivity analyses of the financial proxy of health gain.

The costs estimated to be associated with school absenteeism were based on an assumption of reduced literacy and numeracy skills as a result of missed school. However, returns to education have been demonstrated much to be much greater in primary school than secondary school (283). Therefore, the value attributed to school absence in secondary school may be overestimating the impact of that absence. Consequently, sensitivity analysis explored a scenario where school absenteeism in the Year 10 students did not compromise their education outcomes and future earnings. For this analysis, resource use associated with absenteeism was removed from the impact value.

8.3.3 *Results*

The intervention delivery costs outlined are in Table 8.10. This table shows that the value of investment was £4,666.

Table 8.10 Calculation of investment value

Delivery cost elements	Intervention group			Control group	Delivery cost for intervention group*
	Mean cost per student	Students	Total*	Mean cost per student	
Interview materials	£1.58	210	£332	£ 0	£332
Training costs for learning mentors and screening	£14.20	210	£2,982	£ 0	£2,982
Learning mentor time	£6.44	210	£1,352	£ 0	£1,352
Total					**£4,666**

*Values rounded to nearest £

The calculation of the financial impact of the ASBI on resource use is outlined in Table 8.11. The net financial impact of a reduction in service use is positive, indicating that the total financial value of resource use was lower for the intervention group at 12-month follow-up than what would have occurred without intervention (i.e. deadweight, demonstrated by the control group). This outcome is comparable with the cost difference illustrated in the CBA sensitivity analysis in which school absenteeism costs were included (see Table 8.5), since resources related to time missed from school were considered appropriate to include in the base analysis of the SROI as proxy values for improved concentration in school.

Table 8.11 Calculation of net financial value of reduced resource use outcomes

Outcomes		Impact			Deadweight			Net financial value of reduction in resource use
		Mean proxy value per student	Students	Total	Mean proxy value per student	Students	Total	
Health services	GP visits	£98	210	£20,616	£125	210	£26,261	£5,645
	Secondary care	£293	210	£61,510	£227	210	£47,565	-£13,945
	School nurse visits	£83	210	£17,366	£54	210	£11,281	-£6,086
Social services	Social worker visits	£27	210	£5,678	£9	210	£1,805	-£3,874
Students	>5 missed school days	£1,134	210	£238,105	£2,083	210	£437,533	£199,428
Local authority	Arrests	£0.12	210	£26	£1.02	210	£214	£189
Total impact on reduction in services								**£181,357**

*Values rounded to nearest £, except for arrests where values were extremely small.

The second part of the impact value is determined by the financial value of health gained as a result of the intervention, using monetary values of QALYs gained. Both the impact and deadweight values are positive, indicating that both the intervention and control groups realised an improvement in health at follow-up compared to baseline; however, the net financial value of health improvement is negative (Table 8.12), indicating that the financial value of health gain was lower for the intervention group compared to the control group. Although a slight difference in mean QALY gain in favour of the control group was identified during the SIPS Jr HIGH trial, the 95% CI crossed zero (95% CI around QALY difference: -0.019 to 0.011) (59) indicating an absence of evidence for a difference.

Table 8.12 Calculation of financial value of health improvements

Health improvement	Impact			Deadweight			Net financial value of health improvement
	Mean financial value of QALY gain per student	Students	Total	Mean financial value of QALY gain per student	Students	Total	
QALYs	£21,780	210	£4,573,800	£22,020	210	£4,624,200	**-£50,400**

The total impact constitutes the combined net financial values of resource impact and health impact. The net present value of impact is £130,957 (i.e. £181,357 + -£50,400) and the value of investment is £4,666. Equation 8.3 outlines the SROI calculation. The results indicate that every £1 invested in the ASBI generates approximately £28 in social value. As the return is greater than one, the SROI favours implementing the ASBI from the viewpoint of students, the NHS & PSS, and the LA.

$$\text{SROI} = \frac{\text{Net present impact value}}{\text{Net present investment value}} = \frac{£130,957}{£4,666} = £28.07 \qquad \textit{(8.3)}$$

The results of the two sets of sensitivity analyses are displayed in Tables 8.13 and 8.14. The first sensitivity analysis investigated the impact of changing the financial value of a QALY on the overall SROI. Using either the value of a QALY considered cost-effective by NICE (£20,000) or the value of a QALY estimated by Claxton *et al.,* 2015 (364) to represent the opportunity cost of spending in the healthcare sector (£12,936) results in a larger SROI compared to baseline (Table 8.13). However, the SROI remains positive regardless of the value used.

Table 8.13 Sensitivity analysis for alternative financial values of a QALY

	Impact		Deadweight				
	Mean financial value of QALY gain per student	**Total for 210 students**	**Mean financial value of QALY gain per student**	**Total for 210 students**	**Net financial value of health im-provement**	**Net present impact value**	**SROI ratio, £**
£20,000 (NICE)	£7,240	£1,520,400	£7,320	£1,537,200	-£16,800	£164,557	**35:1**
£12,936 (Claxton *et al.*)	£4,683	£983,395	£4,735	£994,261	-£10,866	£170,491	**37:1**

The second sensitivity analysis, which examines the impact of excluding the value associated with school absenteeism, however, results in a change in the direction of the SROI (Table 8.14). Removing the financial value of missed school from the impact calculation results in the intervention group exhibiting a larger financial value associated with resource use compared to the control group. Therefore, once deadweight is accounted for, the net financial value of a reduction in resource use is negative; in other words, there is not a net reduction in resource use as a result of the ASBI once school absenteeism is removed from consideration. The SROI ratio becomes approximately -£15:1; i.e. a loss of £15 per £1 invested. This result reflects the base-case analysis of the CBA in which the value of resource use associated with school absenteeism is excluded (Table 8.3). In the base-case CBA, the value of resources used in the intervention group are higher on average compared to the control group.

Table 8.14 Sensitivity analysis excluding missed school days from impact

Impact	**Deadweight**	**Net financial value of reduction in resource use**	**Net financial value of health improvement**	**Total impact**	**SROI ratio, £**
£105,197	£87,126	-£18,071	-£50,400	-£68,471	**-15:1**

8.3.4 *Discussion*

The SROI results indicate a positive return from implementing the ASBI, which is driven by savings from resource use rather than improvement in health outcomes. However, the sensitivity analysis in which the outcome of reduced school absenteeism is excluded from the impact value alters the SROI outcome substantially. This sensitivity analysis was conducted due to concerns that the value attributed to time missed from school in the base-case may over-estimate the true financial impact of school absenteeism. Under the assumption that there is no impact on the educational attainment and future earnings of a Year 10 student who misses more than 5 days of school in 6 months, the return from the ASBI is negative, which demonstrates that the ASBI does not provide social value.

The sensitivity analysis on school absenteeism explored a "worst-case" scenario of no impact. Whilst it may be argued that the value attributed to school absenteeism in the base-analysis is an over-estimate (£102,612 per student missing more than 5 days of school in 6-months (59)), assuming no impact on educational attainment of school absenteeism is likely an under-estimate of the outcome, based on the findings of studies of truancy and missed school on educational attainment (285, 365). Therefore, the true value of school absenteeism may lie somewhere between zero and £102,612. Probabilistic analysis of the value attributed to school absenteeism could provide an indication of the minimum value of school absenteeism at which the SROI outcome would remain positive. Decision-makers could then use their own judgement on whether they believe that value is an appropriate financial proxy.

Nevertheless, the findings of the current study suggest that the SROI outcome is very sensitive to changes in the value associated with school absenteeism. Sensitivity of outcomes to assumptions is not uncommon in SROI given the extent to which assumptions are necessarily made during an SROI, from attributing appropriate financial proxy values to defining the extent of benefit. Hex and Tatlock (366) noted that over a range of 15 SROI evaluations that they had conducted, each was highly sensitive to assumptions. Thus, due to this uncertainty their results should be viewed as illustrative of the sort of returns that are possible rather than definitive. A similar approach could be recommended for the results of the current analysis, although, the use of data collected from a clinical trial may have resulted in more robust data on the impact of the ASBI. The data used to populate the SROI evaluation in this study was directly observed, rather than estimated from stakeholder

opinion or via estimates of impact, particularly with reference to the percentage of the deadweight for each outcome.

The sensitivity analysis of the financial valuation of a QALY improved the SROI value. Usually, the opposite effect would be expected from attributing a lower financial value to health improvement. However, in the current study, the health impact of the deadweight is larger than the impact from the intervention because the control group had on average marginally higher QALYs at 12-month follow-up compared with the intervention arm (59). Therefore, using a lower financial value of the QALY reduces the negative net financial impact on health, which increases the numerator of the SROI equation (see Equation 3.5) resulting in a larger SROI ratio.

As with the CBA, it is difficult to provide a comparison of the current SROI study in relation to similar studies since no published analyses have been identified. Only one SROI in the broader alcohol prevention field was identified in Chapter Four (233) which calculated an SROI ratio of 2:1 Thai baht. Whilst positive, this is a smaller return than that identified in the current study; however, a criticism of SROI as an evaluative approach is that the extensive use of stakeholder involvement makes comparison of SROI ratios across different interventions and studies inappropriate due to the variation in indicators and outcomes used across studies (33). Therefore, comparing the SROI ratio outcomes of the current study and that by Tanaree *et al.,* 2019 (233) is of limited benefit, however, both studies show positive returns to alcohol prevention.

8.3.5 *Strengths and limitations*

This study represents a novel approach to evaluating an ASBI programme as no prior SROIs of this intervention have been identified in the literature. Nevertheless, the study has some limitations.

Firstly, as discussed in section 8.3, limited stakeholder involvement was undertaken for the SROI. Indirect stakeholder involvement was considered via the examination of the results of a qualitative study with students and parents of students in the SIPS Jr HIGH trial (59). However, the qualitative study was not conducted with the aim of extracting relevant outcomes for an SROI, therefore, the qualitative findings were of limited use and did not highlight any additional outcomes from parent stakeholders. Consequently, the identification of outcomes to include in the SROI was driven by examination of the

description of the motivational interview component of the intervention and the outcomes that would potentially arise based on the areas of discussion included in the motivational interview. This was combined with an examination of the actual outcomes collected during the trial.

As with the qualitative interviews conducted within the trial, the outcomes for which data were collected during the trial were not chosen with an SROI in mind, therefore, the final outcomes included in the SROI were limited and perhaps not as broad as would be ideal for a SROI. The use of clinical trial data as a basis for an SROI is, therefore, also a limitation. It is quite unorthodox to conduct a trial-based SROI and there is no guidance for how this should be approached. As such, best-practice guidelines for conducting an SROI (84, 164) were followed as closely as practicable within this setting; although, it should be noted that methods for conducting SROI evaluations are yet to become standardised (33). SROI is still a relatively new mode of evaluation, particularly within public health, therefore, there is no explicit "best-practice" reference case to follow. Nevertheless, it is possible that additional outcomes may have been identified with broader stakeholder involvement, for example, outcomes for parents, teachers, or schools. Consequently, the SROI conducted in this study is likely to be conservative with regards to the societal impact of the ASBI. However, the objective guiding the conduction of the SROI was to present an example of the method to PHDMs and indicate examples of outcomes that could be included to demonstrate a holistic evaluation. The SROI conducted in the current study was able to achieve this objective, albeit with a perhaps narrower collection of outcomes than may be available.

The use of the raw 12-month follow-up costs as proxy financial values for resource use could also be considered a limitation, for the same reasons that were discussed for providing a methodological validation analysis of the CBA earlier in section 8.2.1. However, conducting a similar validation analysis was not appropriate for the SROI. Due to an extremely large proportion of zero observations for some of the outcomes included in the evaluation (e.g. school absenteeism, arrests and social worker visits) it was not possible to obtain robust, adjusted estimates of these resources using regression analyses for each resource separately. The only possible approach would have been to either combine all resource outcomes and adjust the total value of outcomes (using similar methods to the adjustment of costs for the CBA, reported in section 8.2.3) or, alternatively, exclude the outcomes for which robust estimates could not be obtained.

The latter option would result in an arbitrary exclusion of relevant outcomes, therefore, was not considered. The former was possible in this particular case study since the same proportions of attribution and drop-off (100% and 0%, respectively) were applied to all outcomes in this study, however, had this not been the case it would have been inappropriate to combine all outcomes. As a validation of the primary method used for the SROI, an analysis was conducted using adjusted values. However, this has not been reported in this chapter since it is not in keeping with the approach to analysis that underpins SROI, in which each outcome is evaluated separately. Part of the benefit of SROI is the ability to view the disaggregated impact of various outcomes of an intervention and combine these to provide a measure of return per pound invested. Using a combined resource outcome does not provide this same level of granularity, therefore, would not have suitably showcased the advantages of the SROI methodology. For completeness, therefore, see Appendix M for the SROI using combined, adjusted values.

Finally, the sensitivity analysis conducted in the current study explored the impacts of a relatively modest number of assumptions on the SROI ratio. Alternative analyses could have been done, such as examining the assumption of 100% attribution and exploring alternative values of financial proxies. Best and worst-case scenarios could have been explored using lower and upper limits of unit costs where these were identified (see Table 20 in Giles *et al.,* 2019 (59)), however for brevity of the evaluation during the workshop, the number of sensitivity analyses were minimised. Furthermore, neither probabilistic nor stochastic sensitivity analyses were conducted. Whilst sensitivity analysis is recommended in the limited SROI guidance literature (84), only deterministic analysis related to examining the impact of assumptions, such as the rate of attribution, or financial proxy values. Complex sensitivity analyses are not discussed in the guidance, which reflects the lower expectation for rigorous analysis compared to other methods of economic evaluation that was highlighted in Chapter Three (83).

Therefore, the whole SROI analysis was based upon point estimates and deterministic analysis. This means that the joint uncertainty caused by imprecision in estimates has not been considered. For a contemporary economic evaluation this would be considered unacceptable. Methods guides and decision-making organisations making use of economic evaluations throughout the world demand that probabilistic sensitivity analysis is performed, e.g. NICE (10), CADTH (235), Washington Panel on Cost-Effectiveness in Health and Medicine

(86) and the ISPOR Society for Medical Decision Making task force (367). It might be considered unfair to assign methods guidance from another set of techniques to SROI but the underlying rationale for adopting these techniques was the potential for making biased conclusions in their absence. Reporting SROI analyses probabilistically should be explored and could have been done here had this been deemed appropriate. However, in keeping with current guidelines, and to maintain relative simplicity for the workshop, only deterministic analysis was conducted and reported in this study.

8.4 Economic evaluations discussion

Two economic evaluations are reported in this chapter, both evaluating the ASBI using predominantly data collected as part of the SIPS Jr HIGH trial. However, the ways in which each evaluation used the SIPS Jr HIGH data differed. This point is important to raise given the similarity in reporting of the results of both analyses as a BCR and an SROI ratio. As discussed in Chapter Three, the contents of the numerator and denominator of each ratio differ, which makes direct comparisons between the results derived from the CBA and SROI challenging, although one would expect the results to lead to generally similar conclusions. For example, the CBA considered the value of resource use, including intervention delivery costs for the intervention group, as a cost and measured intervention benefit using stated preference techniques. Whereas, the SROI considered the value of resource use as an impact of the intervention alongside a monetised value of QALY gains and considered the intervention delivery costs as the intervention investment.

Additionally, the base-case of the CBA excluded costs associated with school absenteeism due to concerns that the values were inappropriate for the evaluation (see section 8.2.2 for details), whilst the value of school absenteeism was considered a relevant financial proxy for the outcome of increased focus and concentration in school, therefore, was included in the base-case of the SROI evaluation (see section 8.3.1).

The base-case outcomes for both the CBA and the SROI evaluations favour implementation of the ASBI, reporting both a positive NSB and positive SROI ratio. However, the CBA results are much more robust to assumptions made during the evaluation than the SROI since excluding the value of reduced school absenteeism from the SROI results in a negative ratio, whereas the NSB remains positive in all variants of sensitivity analyses conducted (inclusive of whether school absenteeism costs are included or not). This difference in the evaluations

can be explained by (i) the benefit measure used for the CBA and (ii) the narrow range of outcomes considered for the SROI.

The WTP benefit measure used in the CBA measured the broader value to the general public of reduced arrests and school absenteeism from the intervention in addition to capturing any financial savings that may have occurred from reduced service use in the cost side of the evaluation. On the other hand, the SROI purely considered the financial savings of these resources and did not include any broader, non-financial benefits. Therefore, removing the financial savings associated with reduced school absenteeism had a greater impact on the SROI than the CBA since the CBA also accounted for the more holistic value of this outcome. As noted in section 8.3.5, due to the limited stakeholder involvement, the current study reports a conservative SROI of the ASBI, therefore, a more comprehensive evaluation would include additional outcomes, which could include the non-financial value to the various stakeholders (e.g. students and parents) of these outcomes. It could be expected that if stated preference measures had been used to additionally capture the value of the ASBI impact from these stakeholders, the SROI ratio would likely be greater, and perhaps less sensitive, to the assumption of the financial return from school absenteeism.

With respect to the use of the evaluations reported above for the workshop with PHDMs, the primary analysis of the CBA, in which the unadjusted 12-month follow-up costs were used, was the example of the CBA used for the workshop. This is consistent with the SROI, which used the same raw values. Additionally, unadjusted follow-up outcomes with respect to arrests and school absenteeism were reflected in the hypothetical scenarios used in the CV survey (see Chapter Six), the result of which was used to generate the benefit measure for the CBA. Therefore, the primary analysis is consistent with the WTP value via its use of unadjusted follow-up costs. Regardless, the outcomes of the CBA using adjusted cost values would not have changed the conclusions drawn about the ASBI, therefore, the choice of which is the better approach is academic. As such, the simpler analysis could be argued to be the best analysis to present to a non-economist audience, given the relatively short time available for PHDMs to digest the information provided to them in the workshop setting.

8.5 Summary

This chapter reports the conduction of two economic evaluations: a CBA and an SROI, using a combination of data collected from the SIPS Jr HIGH trial and independently generated data from the CV survey. The base-case results of these evaluations demonstrate positive

benefits to society from the ASBI, on both a national and a local level. Sensitivity analyses indicate that the CBA base-case results are robust; however, the SROI base-case is highly sensitive to the inclusion of the financial value of school absenteeism.

Both evaluations are considered conservative analyses of the ASBI; however, both evaluations fulfil the objective of providing examples of alternative methods of economic evaluation that are available, yet underused, to evaluate public health interventions. Consequently, both the CBA and SROI evaluations were presented to PHDMs at the workshop, which is detailed in Chapter Nine.

Chapter 9. Preparing economic evaluation materials for the workshop with PHDMs

This chapter outlines the preparation of the economic evaluation evidence that was presented to PHDMs at the workshop, the details and findings of which are reported in Chapter Ten. Section 9.1 outlines the choice of economic evaluation techniques that were presented at the workshop. Section 9.2 provides the rationale for adapting each of the economic evaluation methods for the purposes of the workshop, followed by details of the adaptations made to each evaluation method in turn. The final section summarises the content of the chapter.

9.1 Introduction

The previous chapter outlined the two novel economic evaluations: the CBA and the SROI analysis. The evaluations were conducted in order to present evidence of these methods of economic evaluation to PHDMs at a workshop. In addition to evidence of both a CBA and an SROI analysis, examples of two other economic evaluation techniques were also presented at the workshop: a CUA and a CCA.

These four methods of economic evaluation were chosen for inclusion at the workshop following the findings of the qualitative interview study reported in Chapter Five. The interview findings identified no clear preference for one health economic evaluation technique; however, interest was expressed in methods that are better suited to addressing the multi-sectoral nature of public health. In addition, the qualitative study revealed that specialist PHDMs were familiar with CUA and the systematic review reported in Chapter Four identified CUA as the most commonly utilised method for the evaluation of alcohol prevention interventions. Therefore, CBA, SROI, and CCA were chosen to provide examples of alternative methods able to incorporate outcomes broader than health. CUA was chosen to present alongside the other methods to represent the status-quo of economic evaluation in this area and to facilitate a comparison with these alternative methods.

The economic evaluations were conducted using the same case study, discussed in Chapter Six, as the basis of the evaluation. The intervention of interest is the school-based ASBI evaluated in the SIPS Jr HIGH trial (59). Using the same case study for each evaluation allowed PHDMs to compare each evaluation on an even footing.

9.2 Developing workshop materials

In order to make the evidence appropriate for the workshop audience, who were non-health economists, each of the four economic evaluations of the ASBI were summarised and simplified into reports. The reports were less technical than would be found in a typical economic evaluation report or journal article and contained a short introduction, an explanation of the methods used to conduct the evaluation, base-case results, and results from sensitivity analyses where these were conducted. In contrast to published reports of economic evaluations, a discussion section was not included in the reports prepared for the workshop in order to minimise the information burden on PHDMs at the workshop and to reduce the time required for PHDMs to familiarise themselves with the information provided. Additionally, since the aim of the workshop was to elicit feedback on the methods of evaluation and observe PHDMs' interpretation of the data presented it was not deemed appropriate to include a discussion of the results in the reports as this may have biased the feedback from the PHDMs by providing interpretations that may not have aligned with the workshop attendees' own.

The evaluations from which the workshop evidence was based on for both the CBA and SROI are reported in the previous chapter. Both a CUA and CCA were conducted alongside the SIPS Jr HIGH trial and have been reported in the trial publication (59). Permission was obtained from the trial management team to use the data collected in the trial for this study. The within trial CUA (59) was conducted to a high quality, using standard methods for the economic analysis of clinical trial data (9, 335); therefore, it was not considered necessary to repeat the CUA. The adaptations of the trial based CUA for the workshop evidence were, thus, purely presentational (see section 9.2.1 for details).

The CCA did not conduct any analysis to aggregate outcomes since CCA by definition reports costs and outcomes in a disaggregated format (113, 368) to allow decision-makers to form their own conclusions about an intervention based on the outcomes of most relevance to them. Therefore, as with the CUA, no further analysis was conducted in order to present evidence of a CCA at the workshop, any changes from the CCA reported for the SIPS Jr HIGH trial were merely presentational (see section 9.2.2 for details).

In addition to the four economic evaluation reports, a brief outline of the SIPS Jr HIGH case study was presented to the PHDMs to provide context for the economic evaluations. Since the same case study was used for each evaluation, the information on the SIPS Jr HIGH trial

was provided separately in order to exclude an introduction to the trial in each report. This prevented unnecessary repetition in each of the pieces of economic evaluation evidence.

Each report was examined for technical accuracy by one of my health economist supervisors, Luke Vale (LV). Additionally, two non-health economist individuals examined the text to assess the level of technicality of the text. One member had a basic knowledge of statistics and one was a retired GP. They had a basic understanding of the concepts of statistical analysis and of cost-efficiency that was expected of the workshop attendees. Minor changes to the wording of the text were made following the consultations.

9.2.1 *Adapting the cost-utility analysis*

The main adaptation of the original results was the description of uncertainty around the ICER results. In the original CUA conducted by the SIPS Jr HIGH health economist (59), a stochastic analysis of the estimates of the difference in costs and QALY gains between the control and intervention groups was conducted using bootstrapping to examine the joint distribution of cost and effectiveness of the intervention versus control. Standard practice for presenting the results of stochastic analysis of incremental costs and outcomes was followed (9, 335) by the trial health economist. The results of the bootstrapped estimates were reported using a cost-effectiveness plane in which each bootstrap iteration was represented as a dot in the relevant quadrant of the plane (i.e. costlier and more effective, costlier and less effective, less costly and less effective, and less costly and more effective). Additionally, a cost-effectiveness acceptability curve (CEAC) was generated to estimate the probability that the ASBI is cost-effective at a range of possible cost-per-QALY thresholds (59).

The qualitative interview study (Chapter Five) found that only specialist PHDMs were familiar with CUA. Therefore, it was presumed that the majority of workshop attendees would be unfamiliar with either cost-effectiveness planes or CEACs. Presenting the stochastic analysis results using the standard approaches was, therefore, considered to be potentially overly complex for the workshop. As an alternative, a pie chart was used to represent the proportion of the bootstrap repetitions of the ICER that fell into each quadrant of the cost-effectiveness plane. Each section of the pie chart was clearly labelled to show what it represented (e.g. ICER outcomes which are less costly and less effective).

The trial CEAC representing the probability of the intervention being cost-effective at different cost-per-QALY thresholds was also omitted and replaced with a textual description of key outcomes.

9.2.2 *Adapting the cost-consequence analysis*

The CCA reported for the SIPS Jr HIGH trial was presented as a balance sheet in which costs and outcomes were presented according to whether they favoured current practice, favoured implementing the ASBI, or did not favour either option (59). The outcomes presented in the within-trial CCA included outcomes from the CUA (i.e. QALY gains) in addition to secondary outcomes collected for the trial and analyses by the trial statisticians (59). An alternative way of presenting a CCA balance sheet is to list all costs and outcomes for each alternative being evaluated (i.e. the control group compared to the intervention group) so that all consequences can be compared to all costs without explicitly specifying whether the outcome favours a particular action (113, 368). The latter format was chosen for the presentation of the CCA evidence for the workshop, which enabled more information to be presented for each outcome compared to the format used for the within-trial CCA (59).

The CCA presented a range of outcomes collected during the SIPS Jr HIGH trial (QALY gains, alcohol intake in units, number of drinking days, mental wellbeing, Adolescent Single Alcohol Question (A-SAQ) score, smoking behaviour, and energy drinks consumption) and costs related to intervention delivery and resource use (including healthcare, social care, local authority and student resources). The data was presented as either mean or median values for the intervention group at follow-up, the mean or median values for control group at follow-up, and the difference between the two groups adjusted for baseline values and sample characteristics. The data used to populate the CCA was taken from results of the CUA and statistical analyses that were completed by the trial health economist and trial statisticians (59).

No deterministic sensitivity analysis was included within the CCA, however uncertainty around the point estimates reported was represented using 95% CIs and interquartile ranges where appropriate to provide an idea of the spread of the data.

9.2.3 *Adapting the CBA*

The previous chapter reports the CBA of the ASBI. Stochastic analysis was conducted on the cost difference between control and intervention groups and the mean WTP value in order to explore the distribution of the NSB. Following guidance for the presentation of stochastic analyses in CBA (130), cost-benefit planes were produced for both the national and local authority level analyses (see Chapter Eight). For the same reasons as explained in section 9.2.1, the cost-benefit planes were not included in the report of the CBA generated for the workshop. Textual descriptions of the outcomes of the stochastic analyses were used to represent the proportion of the bootstrap repetitions that were present in the relevant quadrants of the cost-benefit plane. In order to examine whether the pie-chart representation of the stochastic analysis used for the CUA report was preferred to a simple textual explanation, a pie-chart was not included in the CBA report. Additionally, since the bootstrap repetitions only populated two of the four available quadrants of the cost-benefit plane, a pie-chart was not considered to provide any additional information to the text.

As explained in Chapter Eight (see section 8.2), the more simplistic version of analysis was used to form the CBA report for the workshop in order to reduce the technicality of the explanation of the analysis undertaken whilst remaining accurate. Additionally, following the justification provided in Chapter Eight (see section 8.2.2), only three of the five deterministic sensitivity analyses reported in the analysis in Chapter Eight were included in the final report. The deterministic sensitivity analyses included in the final report were scaling WTP to parents of Year 10 students, trimmed WTP, and trimmed costs. As demonstrated in Chapter Eight, neither the two additional sensitivity analyses reported in the previous chapter (including school absenteeism costs and using 2016 WTP values) nor the more complex methodological analysis resulted in NSB outcomes that would change the overall conclusion around the efficiency of the ASBI.

9.2.4 *Adapting the social return on investment*

The SROI analysis report compiled for the workshop represented a simplified version compared to the reporting of the analysis in Chapter Eight. The six stages of a SROI (see Chapter Three and Chapter Eight) were not described in the workshop report and development of the outcomes and indicators was also excluded. This decision was made to minimise the information burden on workshop attendees and minimise the time required to

digest the information presented. Methods pertaining to the calculation of the impact and investment values, were however, included. The two deterministic sensitivity analyses (altering the monetary value of QALY gains and excluding school absenteeism costs) were both included in the workshop report as described in Chapter Eight (section 8.3).

9.3 Summary

This chapter has outlined the adaptations made to each of the four original economic evaluations conducted either previously in this thesis (CBA and SROI, in Chapter Eight) or as part of the within-trial economic evaluation for the SIPS Jr HIGH trial (CUA and CCA, see the final report by Giles *et al.,* 2019 (59)). Each of the original analyses were summarised into condensed reports of less than 2000 words. The reports were necessarily reduced compared to the original reports in order to ensure the feasibility of presenting, and allowing discussions on, four different economic evaluation methods to non-health economists in a session lasting under two hours.

The simplification of the analyses in terms of both the language used and the explanation of statistical methods could be considered a limitation of the reports since the process of simplification necessarily resulted in a loss of analytical detail. However, given the anticipated level of understanding of the economic evaluation methods informed by the qualitative findings in Chapter Five, it was considered a worthwhile trade-off to sacrifice some analytical detail to improve the likelihood of engagement from the workshop attendees.

The findings from the workshop, including a post hoc discussion of the limitations of the simplification of the reports following the workshop, are presented in Chapter Ten. Copies of the final reports and the brief introduction to the SIPS Jr HIGH case study that were presented at the workshop can be found in Appendix N - Appendix R.

PART III. Workshop findings and final discussion

Chapter 10. Workshop with public health decision-makers

The previous chapter (Chapter Nine) described the preparation of each of the economic evaluation reports that were shown to the PHDMs at the workshop. This chapter builds on the previous chapter and discusses the format of the workshop and the feedback elicited from the PHDMs about the economic evaluations that were presented to them. The findings from the workshop reported in this chapter address the third overarching research question of the thesis, outlined in Chapter One. This chapter, thus, explores whether a particular method of economic evaluation, or combination of methods, can be identified as most beneficial to PHDMs for their decision-making needs.

The first section of this chapter outlines the aims and objectives of the workshop. Section 10.2 provides details of the workshop itself, followed by an outline of the feedback from the PHDMs on each method of economic evaluation in section 10.3. Section 10.4 discusses the findings and outlines the implications for methods of economic appraisal of public health interventions and the presentation of such evidence, implications for practice, and further research. The final section summarises the chapter.

10.1 Introduction

The aim of the workshop was to explore PHDMs' opinions about each of the economic evaluation reports in order to examine whether a particular method is most appropriate to aid public health decision-making. The qualitative study in Chapter Five explored the views of a sample of PHDMs about economic evaluation methods. However, any preferences displayed by the interviewees were based largely on descriptions of each method rather than experience of using the available economic evaluation techniques. The interviews revealed that the knowledge and experience of economic evaluation methods was quite limited amongst the PHDMs. The workshop, therefore, served to introduce PHDMs to examples of each method in order to facilitate a more informed discussion of the merits of each method from the point of view of public health decision-making in practice. This enabled the generation of feedback on the relative relevance and usefulness of each economic evaluation for informing public health decisions.

The objectives of the workshop study were as follows:

1. Elicit the views of PHDMs on how useful the information provided by each method is for decision-making in practice
2. Identify whether there is a consensus amongst PHDMs on a preferred economic evaluation method
3. Explore the presentation of evidence to non-health economist decision-makers and identify areas in which this can be improved to engage PHDMs with health economic evidence

10.2 Details of the workshop

A workshop was hosted for PHDMs in North-East England in September 2018. The session consisted of an initial, hour-long training session by SH which introduced workshop attendees to economic evaluation methods. The training was followed by an interactive group session which was facilitated by three health economists (LV, Jo Gray (JG) and SH) and lasted approximately two hours. The group work was intended to stimulate discussion about the use, and potential for use, of economic evaluations in public health decision-making.

PHDMs were defined as anyone working in the field of public health who has influence on public health policy, commissioning or funding decisions and included commissioners, public health consultants, data analysts, public health specialists (e.g. drugs and alcohol specialty leads), and elected members of local councils. Attendees self-selected themselves as PHDMs and no exclusion criteria were set for attending the session.

No prior knowledge of health economics was expected or necessary to attend the workshop. The initial training session was, therefore, important to provide attendees with some basic information on economic evaluation to prepare them for the group work. The training session provided information on the different methods of economic evaluation, how each may be used, and some brief instruction on critically appraising economic evaluation evidence.

During the group work phase, workshop attendees selected themselves into one of three groups. Each group was facilitated by one of the health economist facilitators who were able to answer any questions and provide explanations if anything was unclear to participants. The short reports of each of the four economic evaluations discussed in Chapter Nine (CUA, CCA, CBA, and SROI, see Appendix O - Appendix R for copies of the reports) were presented

in turn to the attendees alongside information about the case study. Reports were presented in the same order to each of the groups. The CUA was presented first as this was considered to be the most familiar report for the workshop attendees and would allow for a discussion of the "status quo" of economic evaluations in public health before moving on to alternative, and potentially less familiar, methods. The CCA was presented second, followed by the CBA and finally the SROI. This order was chosen to keep the evaluations which use monetary outcomes (CBA and SROI) together.

Each evaluation was discussed separately by the groups; the discussions were assisted via the use of crib sheets which posed four questions to consider for each of the four methods presented:

- How easy is the information to understand?
- How appropriate is the information for your decision-making needs?
- Is the information useful to you? If not, what additional information would you like to see?
- Imagine you have been given £20,000 to allocate to public health services. If you were undertaking a prioritisation exercise to decide how to allocate this additional funding, would the information presented be appropriate for use as part of that?

The final question on the crib sheet was designed to place the discussion in a familiar and credible setting for the PHDMs. Setting the discussion within the context of a prioritisation exercise had been suggested by a public health specialist at PHE who has experience with priority setting and the use of health economic evidence in public health settings. A meeting had been arranged with the specialist at PHE to discuss the workshop utilising his expertise and knowledge of PHDMs in order to frame the event in a manner that would be engaging to attendees.

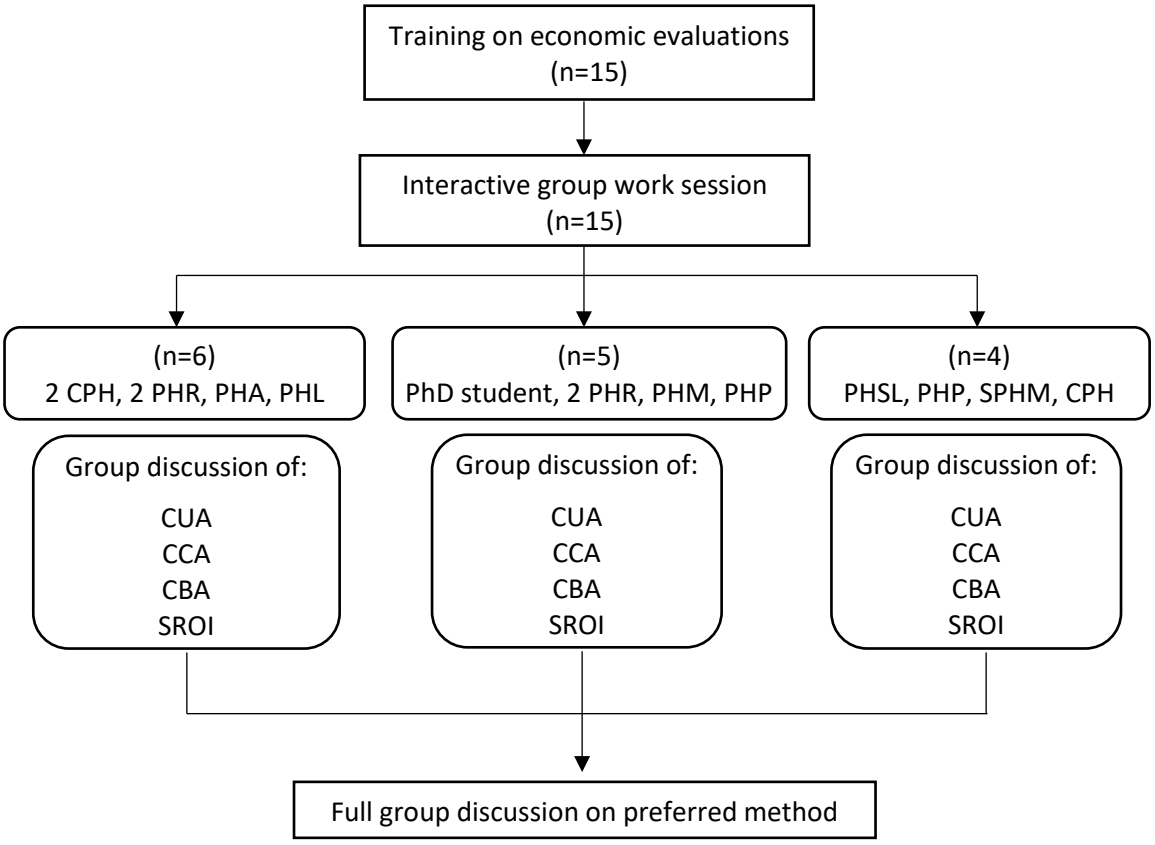
The £20,000 figure used in the final question on the crib sheet was chosen to represent an amount which was sufficiently large to require careful consideration of its allocation, yet sufficiently small to ensure that opportunity costs would be incurred in regard to other public health programmes if it was allocated to implement a new programme. Expert opinion from health economists with specialist knowledge in public health prioritisation (LV and JG) was elicited to finalise the figure presented on the crib sheet (£20,000).

At the end of the interactive group work session the three groups, each comprising between four and six attendees, compared the four evaluations to consider which of the methods

they perceived would be most beneficial to assist decision-making in their line of work. Each group ranked the methods in order of preference and relayed their ranking to the rest of the workshop attendees, which allowed for further discussion regarding the justification for each group's most and least preferred economic evaluation tool.

Feedback regarding the evaluations was captured by the facilitators who took notes of group discussions, comments, and answers to the questions on each crib sheet. Notes were also taken from the overall discussion with all three groups at the end of the group work phase. The notes from each group were collated following the workshop. A summary of the feedback is presented in section 10.3 and the structure of the workshop is outlined below in Figure 10.1.

Figure 10.1 Flow chart of the structure of the workshop with PHDMs



Legend:

CPH = consultant in public health
PHA = public health analyst
PHM = public health manager
PHSL = public health speciality lead
PHR = public health registrar
PHL = public health lecturer
PHP = public health practitioner
SPHM = senior public health manager

10.2.1 *Workshop recruitment*

The workshop was advertised as a Fuse[34] workshop on "Critical appraisal of health economic decision tools for public health decision-makers" and was promoted via the Fuse network, which includes academics and practitioners who work in, or have an interest in, public health (see Appendix S for the flyer used to promote the workshop). The workshop was also publicised through the PHE continuing professional development network and directly advertised to those interviewed in the qualitative interview study (Chapter Five). The workshop was marketed to a wide range of individuals although the flier specifically targeted LA officers and anyone involved in public health decision-making. Both members of public health departments and members of other LA departments were invited, for example those in transport or planning departments, since officers in these areas may also be involved in decisions regarding the implementation of interventions which have an impact on public health (e.g. planning cycle routes or modernisation of council housing).

Fifteen PHDMs attended the workshop. The attendees had varied backgrounds. There were three consultants in public health, a public health specialist/lecturer in public health, four public health specialty registrars, two public health practitioners, a public health analyst, a public health manager, a public health senior manager, a public health programme lead, and a PhD student in health economics who had worked previously as a public health practitioner outside of the UK. Nine different LAs from the North-East of England were represented in addition to an officer from PHE.

10.2.2 *Ethical approval and consent*

Ethical approval for the study was granted by Newcastle University Ethics Committee (REF: 6738/2018) on 16/07/2018. Written consent was obtained on the day from the workshop attendees to use their feedback for this study (see Appendix T for the consent form).

10.3 Workshop feedback from PHDMs

During the group work, each economic evaluation was discussed in turn for approximately 20 minutes. Feedback was invited from the attendees on both the content of each piece of evidence and the manner in which the information was presented. The discussions explored what, if any, further desirable information would be useful that was not necessarily related

[34] The Centre for Translational Research in Public Health - http://www.fuse.ac.uk/

to the choice of methodology. Such feedback is particularly illuminating for health economists with an interest in presenting their findings to decision-makers since the guidance for the presentation of economic evaluation evidence (e.g. the CHEERS checklist (182), the Drummond checklist (9), and the Washington Panel (86)) may not represent the optimal way in which to translate information to PHDMs. The feedback in relation to each of the workshop objectives (section 10.1) is presented below with further discussion of it, and its implications, presented in section 10.4.

10.3.1 *Objective 1: The usefulness of the information provided by each method for decision-making in practice*

Cost-utility analysis

General feedback about the CUA was neither strongly for nor against the method. However, the programme lead and senior manager both felt it was overly complex and lacking in information relevant to them. For example, they were unsure how to interpret the ICER and questioned whether QALYs were the most appropriate outcome measure for the intervention considering the age group of the population involved. Additionally, both PHDMs reflected on their experience that QALYs were not used in LA decision-making, therefore, it would be difficult to make an implementation decision on the basis of the results from a CUA.

The remainder of the workshop attendees expressed the view that CUA could be used for decommissioning services but would only be of use if similar analyses were available for alternative interventions to facilitate a comparison. Concern was also expressed about the appropriateness of the QALY for the case study. Several attendees agreed with the programme lead and a senior manager that the CUA evidence was not intuitive, particularly the reporting of the certainty of the ICER outcome. It was agreed that the CUA evidence overall would only be appropriate in a LA setting if translated into a narrative that would engage other LA stakeholders, suggesting the need for some form of knowledge brokerage between health economists and PHDMs.

Notwithstanding some of the difficulty experienced by PHDMs in grasping the CUA, the concluding remarks of the report, in which the ICER was compared to the NICE threshold for cost-effectiveness, were greatly appreciated. Attendees expressed that more emphasis could have been placed on this over some of the other information that was provided in the

report. The PHDMs liked the idea that the ICER could be used to compare cost-effectiveness directly with that of other interventions for which a CUA has been conducted through the provision of a reference point, i.e. the NICE cost-effectiveness threshold.

Cost-consequence analysis

Two elements of the CCA appeared to make it attractive to PHDMs: firstly, the clear and simple presentation of the CCA in one table; and secondly, the inclusion of a broad array of outcomes and costs would allow decision-makers from public health departments to engage with others within the LA who may have an interest in alternative outcomes. It was noted that it could be a particularly useful tool to take to Health and Wellbeing Board meetings; non-LA partners could engage with the information since at least one of the outcomes included should resonate with them.

Whilst the flexibility of CCA was admired, some workshop attendees felt the lack of aggregation of results had to be treated with caution as it left the information open to interpretation.

Cost-benefit analysis

Attendees were confused over what the WTP outcome represented and questioned why this was used as an outcome measure. Some expressed the view that benefits represented in this manner drew attention away from effectiveness and focused it more on how the intervention would be funded. This view shows a misunderstanding of how effectiveness data was used to derive the WTP values. This could have been a fault of the presentation of the outcomes or insufficient explanation.

In addition, mistrust in the validity of the WTP outcome was also a concern. Further explanation was desired on how the WTP value was derived in order to trust the measure. Since WTP was an unfamiliar outcome measure to the PHDMs, “fear of the unknown” potentially played a role in the mistrust conveyed during the workshop. Concern was expressed over the public’s ability to value interventions and the attendees were not convinced of the robustness of the technique given the apparent subjectivity of the outcome measure. Some thought that other decision-makers within the wider LA would challenge the methodology.

One group did, however, see a redeeming feature in the publicly elicited value of WTP. The penchant of elected members for involving the public in decisions, given their role to serve their electorate, aligns with the concept of valuing an intervention using public views. Some of the workshop attendees, therefore, stated that CBA may be welcomed by their elected members on the basis of the evaluation reflecting public opinion. Unfortunately, no elected council members were in attendance at the workshop to confirm this statement. Further research would be necessary in order to examine whether WTP measures would truly be welcomed by council members.

Social return on investment

The PHDMs described the SROI results as generally familiar and user-friendly, particularly compared to the other methods presented. Nevertheless, areas for improvement were also reported. The evaluation was presented in more detail than it transpired that the PHDMs were accustomed to. For example, PHE produced an ROI report covering many preventive interventions which included limited detail on the analysis conducted to generate the final ROI outcome displayed[35] (35). Some of the terminology used as part of an SROI evaluation, such as "deadweight", was unfamiliar to PHDMs, who did not appear familiar with the methodology underpinning SROI. The concept that "deadweight" (the financial value of the intervention counter-factual) was represented by the SIPS Jr HIGH trial control group was not well understood by attendees from one group, who expressed interest in comparing the SROI of the intervention to that of the control group, despite having it explained that the SROI represented the incremental impacts of the intervention compared to control. This may have been partly due to the unconventionality of conducting an SROI analysis of a clinical trial. However, the concept of deadweight reflecting the counter-factual and its use in SROI remains consistent regardless of the nature of the intervention under evaluation. Additionally, some attendees were confused over how QALY gains were presented in the SROI.

[35]New ROI reports have been produced by PHE in which the methods of analysis are discussed in detail (e.g. Jayatunga., 2018 (369) and Optimity Advisors Ltd., 2018 (370)), however, these reports were published either after the workshop or very soon prior to the workshop, therefore, PHDMs in attendance may not have been aware of these pieces of evidence. Additionally, the analyses were ROI analyses rather than SROI, thus, the analytic methods differ slightly. Terms such as "deadweight" and "attribution", which are associated specifically with SROI methodology, are not present in the reports.

In terms of practical use, one group attested to the SROI being most appropriate at a higher policy level, rather than at an individual LA level, and that some indication of how soon after implementation returns would be realised could be beneficial. Again, this suggests some misunderstanding of the methodology since returns represent those expected within the time horizon of the evaluation, which was stated as 12-months. However, it was opined by members of another group that if SROI was to become a key form of evidence for decision-making, appropriate training on critical appraisal of the method would be necessary.

Additional information requirements

Part of the group work was intended to identify any additional information that would be required to improve the quality of evidence for decision-making purposes. There were three main aspects on which the attendees required greater information: impact on inequalities, longer-term benefits, and sectoral analysis of savings.

PHDMs requested information on the ASBI's impact on different populations, perhaps as additional sensitivity analysis. This was not featured in the reports due to the constraints of the original data collection. However, given the interest in this information it should be considered by those designing economic evaluations and projects examining public health interventions in future.

The ASBI case study was conducted with young people, and many of the benefits associated with these types of behaviour change interventions are unlikely to be realised for several years. The workshop attendees were aware of this fact and wanted evidence of the effectiveness of the intervention on longer-term outcomes than the 12-month outcomes provided by the trial. This is a common theme for many public health interventions which are termed "preventative", where beneficial outcomes are largely intended to occur at some future time-point. Thus, being able to demonstrate long-term outcomes, either via longer trials (which is unlikely due to practicalities and cost) or modelling using data from longitudinal evidence (more likely yet potentially not as robust as a long-term trial) would be greatly beneficial to public health decision-making.

Finally, there are often spill-over effects from public health outcomes into other sectors (both health, e.g. the NHS, and non-health) and vice versa. Given that LAs are responsible for multiple sectors, one of the workshop groups expressed that it would be pertinent for economic evaluations to be able to demonstrate where savings fall in other sectors and be

able to provide this analysis within health economic evaluations. This could be extended to include costs (where savings do not occur) and non-financial benefits (where these are measurable).

10.3.2 *Objective 2: Identifying consensus amongst PHDMs around preferences for a specific economic evaluation method*

Following a review of each of the four economic evaluation methods (CUA, CCA, CBA, and SROI), the overall consensus from all three groups was that SROI was the most preferred method. Each of the groups liked the presentation of the information in which each of the investment and impact elements were clearly broken down to demonstrate the effect each component had on the final SROI value. Feedback was particularly positive regarding the disaggregated presentation of health sector outcomes and outcomes for other sectors in monetary terms. Attendees noted that this could be beneficial in discussions with officers from other LA sectors beyond the public health department or with elected members of council. Including the broader spectrum of outcomes was welcomed, especially when compared to CUA which focused solely on health consequences.

As complementary evidence, CCA was also ranked highly by all groups. There was consensus that it was a valuable tool for providing a broad range of information but, at the same time, a recognition that on its own it would be insufficient for decision-making. As a supplement to other forms of evidence, however, CCA was highly favoured as it was able to "fill in the gaps" of information not provided by the alternative evaluations. Two of the groups envisaged scope to combine an SROI and a CCA by presenting the SROI results as one outcome within the CCA. Such a venture would require endeavours to ensure that no double counting occurred between the impacts and investments included in the SROI and the disaggregated results reported in the remainder of the CCA. Nevertheless, from the practical perspective of providing desirable information to decision-makers it would be possible to present evidence of a SROI alongside additional outcomes which could not be included within the SROI analysis.

A minority of workshop attendees (a programme lead and a senior manager in public health) preferred CBA due to the fact they considered it to report a "bottom line" via the overall net societal benefit (NSB). This description of the CBA outcome could have two interpretations. According to the Oxford English dictionary, a "bottom line" can be defined as *"the last line of*

a profit and loss account, showing the final profit (or loss); (also more generally) a net profit." (371). Whether the NSB truly refers to a bottom line in this sense is debateable. The NSB resulting from the CBA did not reflect direct financial benefit, since the monetary benefit derived from the CV survey is an indirect monetary outcome reflecting value to society. Alternatively, the term "bottom line" may have been used in a more colloquial sense, referring to a single, final outcome. In the latter interpretation, the attendees would merely be expressing a preference for having a single figure outcome compared to the format of the outcome from either a CUA or SROI, which both report ratios.

The same attendees also found the outcomes of the CBA easy to understand and in a relatable format (money). Two-thirds of the PHDMs, however, had an opposing view and ranked CBA as their least preferred evaluation method. The driving force behind this preference was a lack of familiarity with the terminology used in the reporting of the evidence (e.g. "willingness to pay"), which they felt made it difficult to engage with.

Overall, one of the groups fed back that a "one-tool-fits-all" approach was unlikely to be the best approach, as they considered different methods to be relevant to different decision-makers. The reality of public health decision-making was that decisions are made collaboratively with decision-makers of varying roles and expertise. Therefore, it was felt that different methods would resonate better than others with different audiences. For example, a director of public health may prefer CCA, whilst other LA officers may prefer SROI and NHS stakeholders would likely prefer CEA or CUA.

10.3.3 *Objective 3: Explore the presentation of evidence and identify areas in which this can be improved*

Feedback from the group work regarding the presentation of the evidence produced some common themes across all, or most of, the methods of economic evaluation. The findings from this aspect of the evidence are pertinent to health economists conducting economic evaluations of public health interventions who intend their work to reach, and be utilised by, decision-makers in practice.

Firstly, the PHDMs desired to see a full break-down of costs included in the evaluations. Knowing the total cost was only beneficial to their decision-making to a limited extent, and they wanted greater information on where the costs were being incurred and whether certain costs were substantial compared to others. This detail would typically be presented

in an academic journal article or health technology assessment report. It had, however, been excluded from the workshop reports for reasons of brevity.

Secondly, there was a general view that the CUA and CBA reports were “too wordy”. This both increased the time the attendees took to read through the evidence, compared to the SROI and CCA, but also added to the cognitive burden of deciphering the results of the evaluations. Results tables were welcomed as these clearly demonstrated the key outcomes.

Additionally, some of the attempts to present information graphically were unsuccessful. As explained in Chapter Nine, the results from bootstrapping the incremental costs and QALY gains were presented via a pie chart in the CUA report for the workshop, rather than on a cost-effectiveness plane or using the data to create a CEAC. However, this adaptation did not appear to improve comprehension of the concept of bootstrapping the cost and QALY outcomes and caused confusion.

Thirdly, sensitivity analysis was greatly appreciated. Feedback from both the CBA and CUA was that the sensitivity analysis was extremely useful in indicating certainty of the evaluations’ findings which would help PHDMs decide whether the evidence is sufficient to base decisions on. In the CBA in particular, the presentation of a local level analysis was well received as it enabled the PHDMs to anticipate the impact of the intervention in a context relevant to them.

Finally, recognising the limitations of PHDMs to be able to interpret health economic evidence, feedback was expressed for plain English explanations of the evidence. Attempts had been made to minimise the use of jargon in the reports but according to the feedback from the groups this could still be improved upon to further engage PHDMs with economic evaluation evidence and ensure that analyses are appropriately interpreted. Issues with comprehension were identified particularly with the CUA and for certain elements of the SROI. The terms “deadweight” and “attribution” in the SROI analysis were not well understood and would require either greater explanation or alternative language to convey their meaning. Although in general the presentation of the SROI was preferred by most of the PHDMs, one of the groups was overwhelmed by the information and stated their preference for more concise reporting of the results.

10.4 Discussion

This chapter reports the findings from a workshop with a range of PHDMs in which four pieces of economic evaluation evidence (CUA, CCA, CBA, and SROI) were appraised for their potential merits in aiding public health decisions. The feedback obtained during the workshop assisted in addressing the third overarching research question of the thesis, which aimed to explore whether a particular method of economic evaluation, or combination of methods, is most beneficial to PHDMs. Overall, the findings from the workshop could not identify a single approach to economic evaluation which would be most beneficial to PHDMs. SROI was preferred by a majority of the attendees, nevertheless, there were several limitations identified regarding the practical use of the method which would require more research to address prior to recommending it as principally beneficial for public health decision-making. However, a favoured approach of including a CCA in combination with another evaluation was reported. The remainder of this section will discuss the implications of the workshop findings for the conduct and presentation of economic evaluations in more detail.

10.4.1 *Implications for methods of economic evaluation*

A finding which arose most notably during group discussions around the SROI and CCA evaluations was that the PHDMs appreciated the inclusion of a broad range of benefits in evaluations. Attendees felt that this provided more useful and compelling evidence of effectiveness compared with focusing solely on health benefits. Although this feedback was primarily given for the SROI analysis and the CCA, it is one of the key academic arguments for including CBA in the arsenal of economic evaluations of public health interventions (47). Consequently, this provides some evidence of convergence between theoretical arguments for CBA and requirements in practice.

Despite this, however, preferences for using CBA in practice were weak. Many of the PHDMs questioned the validity of the benefit measure, expressing mistrust with the WTP survey technique. Concern was voiced over the public's ability to value interventions and the robustness of the technique given the apparent subjectivity of the outcome measure. Concern over whether stated preference techniques are appropriate methods for evaluating goods using individual preferences is not uncommon; debate is ongoing between proponents and opponents of CV in the academic literature on this issue (see Borzykowski *et*

al., 2018 (318), Mitchell & Carson., 2013 (124), Diamond & Hausman., 1994 (303) or Hausman., 2012 (319) for some examples and discussions of criticisms against CV).

However, a further potential explanation for the PHDMs' responses to WTP might in part be due to "fear of the unknown", as suggested earlier in this chapter (section 10.3.1). A lack of trust in the public valuation of the ASBI may be attributed to a lack of understanding of the CV process and apprehension towards an unfamiliar tool which bears little resemblance, on the face of it, to outcome measures which PHDMs may be accustomed to and which may be perceived as more objective measures, such as alcohol consumption or years of life gained. Despite covering CBA in the short training session at the beginning of the workshop, stated preference measures were covered only briefly. Additionally, limited explanation of the CV study was provided in the CBA report on the basis of brevity for the workshop. Whether a "fear of the unknown" contributes to some of the attendees' concern regarding public valuations of intervention outcomes is unclear currently and further research could examine these issues to shed greater light on whether information from CBAs offers value to PHDMs.

On the other hand, outcome measures such as QALYs, which have been more commonly used in evaluations of public health interventions (372), were more familiar to the attendees. QALYs as a general outcome measure were not queried by the majority of the PHDMs, although several suggested that they may not be the most appropriate measure for the school based ASBI. This view could be unique to the choice of case study for the workshop, although, as remarked by members of one of the workshop groups, QALYs are not considered in LA decisions. Therefore, QALYs seem less relevant to current public health decision-making at a local level.

This disparity in trust between WTP measures and QALY measures, on the basis of "subjectivity" is illuminating and reveals a limited knowledge of the QALY measure. For instance, estimating QALY gains relies on some level of subjectivity, via public valuation of health states extracted from either the EQ-5D or another generic, or condition-specific, quality of life survey. The difference between QALYs and WTP values is that in a CBA the subjective valuation is brought to the forefront of the evaluation rather than masquerading behind the veil of a QALY, which end users may feel they understand.

A minority of the PHDMs observed positive potential for the WTP measure via the hypothesis that elected members may appreciate its elicitation of public viewpoints.

However, in order for WTP to be relevant in a specific LA context, evaluations would ideally use WTP data captured from the local population, which would have consequences for the generalisability and transferability of data to other localities. An alternative approach could be to conduct a survey that focuses on obtaining a sample of the general public's WTP from the area of interest or at the very least ensuring that there is a sub-sample from that area that is sufficient to provide robust results. If this is not possible then methods of weighting the general population WTP measure more heavily towards the values stated by the local population could also be explored.

Both the CBA and CUA did, however, receive positive feedback for their aggregated results. A minority of the workshop attendees resonated with the NSB outcome of the CBA and reflected that the use of monetary units was relevant and familiar to LA officers. The monetary representation of the SROI outcome ratio was also viewed positively for the same reason. On the other hand, the comparative ability of the ICER reported in the CUA, particularly to a clearly stated reference point for cost-effectiveness such as the NICE cost-effectiveness threshold, was compelling to many attendees.

A similar finding was reported by Phillips *et al*., 2011 (19), nevertheless, it warrants further consideration since Owen *et al*. (372, 373) demonstrated that the vast majority of public health interventions fall well below the current cost-effectiveness threshold posited by NICE for healthcare technologies (£20,000-£30,000 per QALY) or even the lower threshold recently revealed for use by the Department of Health for their impact assessments (£15,000 per QALY) (374). In such a case, comparing ICERs of alternative interventions to a threshold value for cost-effectiveness of healthcare technologies may be moot since the majority of interventions would all be considered cost-effective when compared to the current threshold.

Additionally, in England the threshold value is supposedly representative of the opportunity cost to the healthcare budget of funding a specific course of action (although the current threshold is based on no empirical evidence of this opportunity cost and recent endeavours to empirically calculate a threshold suggest a much lower value (364, 375)). Therefore, unless there can be a flow of funds from the healthcare budget into public health, comparisons to the threshold posited by NICE is perhaps irrelevant for LAs. Recent research has been conducted to examine the marginal productivity of public health expenditure in English LAs and has calculated a cost per QALY estimate for the public health grant of around

£3,800 per QALY (376). If acknowledged as a relevant threshold, there may be a stronger argument for the use of CUA in public health to aid efficient resource allocation within the constraint of the public health grant. Although, these findings are only preliminary and the issue with reliance on health maximisation as a decision criterion in the current CUA framework remains. Furthermore, following a removal of the public health grant, public health expenditure would be explicitly competing with other LA responsibilities. Thus, further work would be required to explore an appropriate cost per QALY threshold in a setting where the opportunity cost of public health expenditure explicitly falls on non-public health activities.

The relative comparability between ICERs of alternative interventions may be of greater relevance to PHDMs. Following this argument, CUA is not the only method which is able to provide this level of comparability; the NSB of substitute interventions may also be a relevant yardstick for examining relative efficiency (33, 80). It could be argued that CBA goes one step further than ICERs by allowing comparison of public health programmes in other sectors of the LA. Since the use of monetised outcomes does not rely on the necessity of benefit to be driven by health outcomes, which would be necessary to estimate cost-effectiveness using current QALY measures, CBA may offer greater applicability within the broader field of public health than CUA.

The pertinence of this was demonstrated in feedback requesting sector specific reporting of costs and outcomes. QALYs currently are not able to cross the border from healthcare to other sectors, although some recent explorations have been made to examine the possibility of cross-sector comparisons using minor adjustments in QALY measurement (24). The suggested adjustments, however, limit measurement of outcomes to the space of “quality of life” which may not encapsulate all benefits related to other sectors e.g. environmental improvements from physical activity interventions which reduce car use.

On the other hand, the same feature of comparability does not hold true for SROI analysis. Due to the extensive use of stakeholder engagement to assign both relevant outcomes for analysis and the financial proxy values, the subjectivity of each SROI analysis is considered too extensive to allow for robust comparisons of SROI ratios (33). SROI can be used generally to inform whether an investment may be worthwhile, however, its use to assist allocating resources may be limited.

Finally, SROI has only been introduced to the field of public health in recent years (33) with efforts from PHE to guide the conduct and collection of data for SROI analyses only coming to the fore since 2015 (164, 377). The method's novelty places it at risk of misinterpretation if PHDMs do not receive appropriate training to understand the method and its limitations as a decision-aid. Following the preference for the method during the workshop, it is ever more important that anyone with a desire to use SROI analysis as a decision-aid understands the limits of the tool and is able to accurately interpret its outcomes. This point was raised by one group in the workshop but has also been addressed elsewhere (378). The main area of concern involves the interpretation of what constitutes the "return" in a SROI and how that relates to cost-savings. SROI extends the benefit space beyond pure financial returns, yet this is often not well comprehended, and misinterpretation occurs when the outcome of an SROI is believed to represent a cash return.

10.4.2 *Implications for the presentation of economic evaluations*

Emphasis was placed on the presentation of information during the group discussions. This was evident strongly in the positive feedback for the CCA and SROI evaluations. Consequently, it may have been favourable presentation of the evidence, rather than the actual methodology, that drove the attendees' preferences for SROI and CCA. However, it was recognised that CCA would only be valuable alongside other tools as supplementary evidence. It should be feasible to include CCAs alongside other evaluations in future analysis if this is viewed as beneficial to decision-making by PHDMs. Weatherly *et al*. (48) in fact recommend a CCA be conducted prior to other valuation methods when evaluating public health programmes due to the broad associated costs and consequences.

One of the foremost findings from this study was the extent of trade-off between presenting sufficient information to be useful and convincingly reliable for decision-making on the one hand and providing a concise and navigable report to a non-expert audience on the other. Although considerable effort was made to balance these competing elements in the production of the material for the workshop, the feedback illustrates that there is still room for improvement.

The CUA prompted the most emphatic comments on the difficulty of following parts of the report, particularly the alternative presentation of the bootstrapped incremental costs and QALYs, as discussed earlier (section 10.3.3). The PHDMs did not report any difficulty in

understanding similar information presented purely as text in the CBA. It remains unclear whether presenting the cost-effectiveness plane would have been understood by the PHDMs. The feedback suggests that replacing it with an alternative figure did not add value to the report and reporting outcomes textually was sufficient to translate the data from the stochastic analysis.

Furthermore, heterogeneity in the level of detail requested for inclusion in the reports was identified. As reported in section 10.3.3, plain English and concise summaries of key findings were requested. On the other hand, public health consultants and registrars in particular sought further information on the statistical methods that had been used. This demonstrates the difficulty of providing evidence that meets the informational requirements of a range of PHDMs simultaneously and implies that the reporting of results should be able to account for this heterogeneity.

A suggestion may be to produce reports in a format that allows those readers who desire limited detail to easily access the key findings whilst also providing access to sufficient analytic detail for the more technically minded decision-makers. This could potentially be operationalised via the use of plain English executive summaries which include key results in a tabular form and discuss key implications of the outcomes. The detail of the analysis could, thus, be incorporated into the main body of the report so that it is available for those who desire to scrutinise the methods more carefully. In reports for PHDMs, however, it would be prudent to make few assumptions regarding end users' knowledge of health economic specific terminology. Consequently, lay terminology could be used where possible and clear definitions for all technical terms should be included in the report.

The feedback from the workshop attendees suggests a need for improved knowledge transfer from academia and other research institutions to public health policy-makers. Hunter *et al*. (266) identified issues of a similar nature when evaluating the prioritisation framework recently developed by PHE (265). Many of the public policy makers testing out the prioritisation framework tool reported difficulty in interpreting the technical terminology and commented on a lack of supporting documentation or clear definition of terminology. Knowledge transfer between academic and policy-makers is a currently discussed issue (e.g. Hunter., 2019 (379), Wilson & Sheldon*.,* 2019 (380), and McAteer *et al*., 2018 (381)) and is a foremost focus of research institutions such as Fuse (382). Including a focus on economic

evaluation evidence specifically in these knowledge transfer discussions should be encouraged.

An element of the reports which attendees found beneficial was the presentation of the CBA results as a local scenario. An important consideration for public health economic evaluations, therefore, is whether more locally relevant analyses can be conducted. It would be extremely resource-intensive to conduct separate evaluations for all LAs, however, authors could ensure sufficient detail is given about the characteristics of the population and context of the evaluation for decision-makers to assess whether it may be relevant to their local demographic. Additionally, more commonplace sensitivity analysis to model potential outcomes for changes in population characteristics (e.g. the proportion of risky drinkers in the population) could be encouraged to promote greater resonance with individual LA decisions.

Sensitivity analysis could also be extended to different population groups to assess the potential impact on inequalities, such as: health, socio-economic, or any other basis particularly relevant to the intervention. Chapter Four identified that analyses of this kind are rare in the evaluation of alcohol prevention interventions and may be often overlooked due to lack of available data; it would be worth exploring whether modelling can be used when direct empirical data are unavailable. Whilst this would add complexity and require greater resources on the part of the economists conducting the evaluations, findings from this study indicate that such information would be valued by PHDMs.

10.4.3 *Implications for practice*

The findings reported in this chapter have several implications for health economists performing evaluations of public health interventions. Firstly, with reference to the presentation of evaluations, the findings indicate the importance of improving knowledge transfer between academic research and PHDMs. Regardless of the method of economic evaluation used, the current format in which evidence is presented is not optimal for end-users of the information.

Secondly, health economists may consider conducting SROI analyses where appropriate, for instance where programmes are expected to have considerable social impacts beyond health. A caveat of this implication, however, is that appropriate training on interpreting

SROI analyses and understanding their limitations should be provided to PHDMs to facilitate the appropriate use of SROI findings.

Thirdly, CCA should be presented alongside other modes of evaluation in order to provide information on outcomes that it may not be possible to include in the main economic evaluation. This can allow multiple stakeholders to obtain relevant information from the evidence.

Finally, when possible, scenario analysis using local data or assumptions that reflect local demographics should be conducted if end-users are expected to be LA officers. If data unavailability renders local scenario analysis not possible, detailed reporting of demographic data of the study sample should be commonplace in order to assist PHDMs assess the relevance of the evidence to their locality.

10.4.4 *Implications for further research*

The discussion of the implications of the workshop findings in sections 10.4.1 and 10.4.2 identified several areas for further research. Firstly, further examination is required with respect to whether CBA would be more acceptable to PHDMs if they received relevant training and if reports included more extensive explanation of the WTP measure and how WTP was elicited for the evaluation. Since this was the major focus of the attendees' concern, yet they expressed interest in other elements of the evaluation such as the incorporation of broader benefits, there may remain scope for CBA if trust is gained in the measure of benefit.

Secondly, since no elected members attended the workshop, an exploration of their views of WTP as a measure of benefit could be undertaken to examine whether the feedback regarding elected members potentially preferring publicly valued outcomes holds merit. If the hypothesis holds true, this would favour the argument for the use of CBA for local public health purposes.

Thirdly, further research should be conducted with PHDMs following adaptations to the presentation of the economic evaluations as specified in section 10.4.2, particularly around SROI. Reassessment of PHDMs' ranking of methods should then be considered to examine whether the preferences reported in this chapter remain consistent.

Finally, the impact of the case study on preferences for methods should also be examined since it is not possible to evaluate many public health interventions using RCTs. However, the SROI analysis was the most strongly impacted of all the techniques by the use of trial data from the perspective of the PHDMs, who misinterpreted the comparison of the control group and intervention group. Therefore, the use of a non-trial-based case study may strengthen the case for SROI if interpretation of the outcomes can be improved.

10.4.5 *Strengths and limitations*

The greatest strength of this study is the active engagement with PHDMs in order to examine economic evaluation tools from their perspective as end-users. This approach moves the discussion on the future of health economic tools for public health decision-making away from a purely academic conversation towards one embedded in practice. The engagement of public health decision-makers was beneficial to both the PHDMs themselves, via an introduction to health economic decision-aids, and to health economists who ultimately want to produce evidence that has practical value in addition to academic credibility.

The workshop itself was successful in bringing together individuals from a range of public health decision-making roles who were representative of the majority of LAs in North-East England. The variety of responsibilities and localities strengthened the study as it allowed for rich and diverse discussions representing the needs of different decision-making levels and local priorities. However, despite the diverse range of attendees, the results are limited to a relatively small sample of decision-makers from one region in England. The attendees also self-selected to attend the workshop so may not fully represent all PHDMs with the results potentially biased towards individuals who have a predisposed interest in using economic evaluation. However, research examining similar questions has been recently published (57, 58), which identify similar concerns with regards to the unsuitability of current economic evaluation evidence, and the way it is presented, for public health decision-making. Therefore, the findings from this empirical study appear to be suitably generalisable to other LAs in England.

Each workshop group was facilitated by a health economist who was likely to have had some influence on the feedback given by the attendees. Facilitators were instructed to minimise their impact on group opinions and remove themselves from the discussions other than to

guide attendees towards the discussion points on the crib sheet. Nevertheless, their presence and explanation of concepts may have had unintended consequences in guiding the groups' opinions. Additionally, the workshop groups were mixed in terms of seniority of individuals and more senior decision-makers may have "crowded out" the views of more junior members to some extent.

The case study chosen for the evaluations could also be considered a limitation as it reported the outcomes of an RCT studying an intervention on a specific population (Year 10 school students). As discussed in section 10.4.4, further research should be conducted using alternative case studies in order to examine whether the results from this workshop are transferable to other cases.

The potential for bias from the order in which the four reports were presented may also pose a limitation on the results of the workshop discussions. In the context of surveys, order bias has been demonstrated to influence responses on occasion (383). This is typically a potential concern when researchers wish for each option to be evaluated independently (124). In the context of the current study, when discussing the merits of each evaluation, there was potential for workshop attendees' views of the economic evaluation under review to be influenced by the evidence which had been discussed prior. Since the SROI report was presented last to each of the groups, the PHDMs had the opportunity to compare it against each of the other methods. Presenting the economic evaluations separately was deemed essential to minimise cognitive overload and to ensure equal discussion time was allocated to each method, however, this may have introduced unintended bias to the discussion. Prior to the full group discussion (see Figure 10.1), however, each group was instructed to reconsider each method before revealing their preferences. This was intended to minimise any bias from the order of reports by allowing each method to be considered relative to all others.

Finally, the presentation of the reports produced for the workshop may have influenced the feedback for the methods from the attendees. The simplification of data for the non-economist audience resulted in the portrayal of limited information, which could have contributed to confusion when interpreting the evaluations. Therefore, the reports themselves could be considered a limitation in this study. However, this study was a first attempt at understanding what, and how, to present economic evaluation information to

PHDMs. Therefore, identifying and capturing aspects of the evaluations which were not well comprehended is positive in terms of identifying areas for future improvement.

10.5 Summary

The aim of this study was to explore PHDMs' opinions about each of the economic evaluation reports in order to address the third overarching research question of the thesis which considers whether a particular method is most appropriate to aid public health decision-making. A general preferential consensus for SROI was identified; CCA was additionally reported as a beneficial supplement to any other evaluation. CBA and CCA were less favoured, largely due to their measures of outcome. WTP was not viewed as a reliable outcome measure, and the validity of the QALY was also questioned. Despite the preference for SROI, some concern still remained over its use and interpretation by PHDMs and certain attendees voiced the opinion that a "one-tool-fits-all" approach to economic evaluation is unlikely to be appropriate, given the range of expertise and interests of PHDMs. Therefore, a particularly beneficial method of economic evaluation could not be identified, however, CCA in combination with another evaluation did appear to be a favoured and potentially beneficial approach which would aid PHDMs' decision-making needs.

Regardless of method, however, the upshot of the feedback is that much greater effort is required to present economic evaluation data suitably for PHDMs as non-health economists. Further research is required in order to establish how best to translate economic evaluation evidence to its end users, the PHDMs. In order to provide appropriate information for PHDMs of differing roles and knowledge requirements, health economists may need to emphasise producing simplified, plain English reports of their findings alongside detailed reports which can provide transparency of methodological rigour. The inclusion of CCA more regularly alongside other economic evaluations may also contribute towards providing information with which varied stakeholders can engage.

The workshop reported in this chapter concludes the empirical research for this doctoral thesis. The next and final chapter discusses the overall findings of this body of work.

Chapter 11. Discussion and conclusions

This chapter discusses the findings of the thesis. The first section outlines the structure of the thesis and research questions addressed. Section 11.2 outlines the contributions made to the existing literature from this research study; each research question is addressed individually. Section 11.3 presents implications for practice while section 11.4 outlines the strengths and limitations of this body of work. Areas for further research are proposed in section 11.5 with the final section concluding the thesis.

11.1 Thesis aims and outline

This thesis set out to identify the most beneficial health economic evaluation tool(s) to meet the decision-making needs of PHDMs. In order to achieve this aim, three key research questions were posed:

1. With respect to current economic evaluation and priority-setting tools:
    a. What evidence is currently available and which methods are used by the health economic research community to evaluate public health interventions?
    b. Does the quality of evidence produced meet recommendations for health economic evaluations of public health interventions from the available guidance?
2. With respect to the use of health economic evidence by PHDMs:
    a. To what extent is health economic evidence used by PHDMs to aid decision-making?
    b. To what extent do PHDMs have sufficient knowledge of health economic tools to appropriately use the available evidence?
    c. What barriers do PHDMs perceive exist to the use of health economic evidence as it is currently produced?
3. Is a particular method of economic evaluations, or combination of methods, most beneficial to PHDMs for their decision-making needs?

The empirical studies reported in this thesis were designed to explore each of the above research questions. Prior to reporting the findings of the empirical studies, the theoretical underpinnings of the health economic techniques examined in the thesis were outlined (Chapter Two). The methodology relevant to each of the economic evaluation and priority-setting methods initially considered for exploration in this doctoral study was then

presented (Chapter Three). A systematic review was conducted (Chapter Four) in order to address the first research question. The review provided baseline knowledge of the methods of economic evaluation that have been used. A qualitative exploration of PHDMs' use and understanding of health economic tools was carried out (Chapter Five) in order to address research question two. Based on the findings of the qualitative exploration, and other research which was being undertaken around the same time focusing on priority-setting tools in public health settings (58), the remainder of this doctoral research was dedicated solely to economic evaluation methods.

Four economic evaluation methods were chosen for further exploration (CUA, CCA, CBA, and SROI) based on the findings from the interviews with PHDMs. Due to a lack of published CBA and SROI studies, it was necessary to conduct two novel evaluations in order to provide the necessary exposure for an informed discussion on the merits of alternative methodologies with PHDMs. Chapters Six and Seven report the preparation, conduct and analysis of the CV survey employed to obtain an estimate of WTP for use in the CBA. The conduct of the novel CBA and SROI analyses was reported in Chapter Eight. Feedback on the alternative methods of economic evaluation was sought during a workshop with PHDMs. Chapter Nine reports on the preparation of evidence reports for the final empirical study of this thesis, the workshop with the PHDMs. The workshop and its findings, which addressed research question three, are discussed in Chapter Ten.

Each of the economic evaluations presented at the workshop were based on the same public health case study to enable an unbiased comparison of the methods, with respect to the intervention under evaluation. The case study was a school-based ASBI programme for students in Year 10 (aged 14-15 years) (see Giles *et al.,* 2019 (59) and Chapter Six for further details on the case study). Preventing misuse of alcohol is an important public health issue in England, particularly for young people (59). Alcohol misuse in young people is associated with health problems (60) and also broader social issues such as anti-social behaviour (61), reduced educational attainment (62), and risky sexual behaviour (62) to name but a few. A case study examining the impact of an intervention to reduce alcohol consumption in young people, therefore, provided an example of a complex public health issue. Consequently, this case study was relevant to exploring which method of economic evaluation is most appropriate for evaluating complex public health interventions.

11.2 Contribution of the thesis to existing literature

This thesis contributes to the existing literature on economic evaluation methods for evaluating public health interventions. The precise contribution of the research conducted to the knowledge base is explained below. Each research question, listed above, is discussed in turn.

11.2.1 *Research question one – What evidence of economic evaluation and priority-setting tools is currently available for public health interventions and does it meet methodological recommendations?*

Over the past 10 to 15 years, calls have been made for the expansion of the economic evaluation framework for public health appraisal. Commentators such as Kelly *et al.,* 2005 (47), Edwards *et al.,* 2013 (17) and Weatherly *et al.,* 2009 (48) have suggested that greater use of methods such as CBA and CCA could account for the broader impacts of public health programmes. NICE also updated their guidance for methods of public health appraisal to consider method such as CBA and CCA alongside their preferred method, CUA, in addition to recommending public health and societal perspectives for analysis (10).

Additionally, economic evaluation methods outside of the standard health economist's toolbox such as ROI and SROI, alongside prioritisation tools such as PBMA and MCDA, have become a focus of attention in the public health field, e.g. Banke-Thomas *et al.,* 2015 (32), Edwards *et al.,* 2014 (39) and Hunter *et al.,* 2016 (54). Notably, in light of this focus, PHE recently developed a prioritisation framework based on MCDA techniques (265).

However, the literature examining whether these advances in recommendations for a broader framework of evaluation and prioritisation had translated to updated research agendas remains sparse. Previous literature reviews that had examined economic evaluations of public health interventions either focussed on narrow evidence sources (e.g. Owen *et al.,* 2012 (372) who looked only at NICE guidance) or were outdated (e.g. Weatherly *et al.,* 2009 (48)).

The systematic review reported in Chapter Four aimed to fill this gap in the knowledge base. The review findings identified that CUA remained the most prominent method of evaluating alcohol prevention interventions until March 2019. However, since 2014 a change was observed in the prominence of CUAs. The proportion of CUA studies declined and CBAs (although, the limitations of these are discussed in the next paragraph) were produced with

slightly greater regularity. The first MCDA and SROI studies in the alcohol prevention field were identified in 2018 and 2019, respectively. Nevertheless, it is too early to conclude whether these findings mark the beginning of a trend for greater heterogeneity in methods used to evaluate and prioritise public health interventions. A review update in the near future would be beneficial in order to draw conclusions on the development of economic evaluation methods for public health interventions.

In answer to the second part of research question one, examining the methodological quality of economic evaluations, improvement was observed in some methodological areas as recommended by previous scholars (e.g. Weatherly *et al.,* 2009 (48) and Edwards *et al.,* 2013 (17)). For example: the use of modelling alongside RCTs, incorporation of broader outcomes, the use of sector specific QALYs, and the consideration of equity implications in a small number of studies. However, these improvements were not consistently observed across all studies and areas for improvement remain, for example the non-existent use of WTP values in CBAs and lack of incorporation of intersectoral costs and consequences. None of the CBAs identified in the review used methods of monetising benefit that fit with the welfarist theoretical grounding of CBA discussed in Chapter Two (i.e. measuring benefit as WTP derived from either stated or revealed preferences). Rather, all CBAs identified monetised QALYs to obtain a monetary health benefit measure. Additionally, whilst CCAs offer potential to report costs and outcomes from different sectors without the challenge of appropriately aggregating them (114), only one study using such methodology in the examination of alcohol prevention interventions was identified between 2006 and 2019.

The review reported in Chapter Four provides good evidence in answer to research question one, however, it is possible that some grey literature remained unidentified. Additionally, potentially relevant published literature in other areas that were not investigated in this review, such as transport economics or sexual health, may exist. Furthermore, alcohol prevention interventions are only one element of public health, therefore, it would be unwise to conclude the findings here definitively reflect the status of economic evaluations of public health in general. However, recent reviews of economic evaluations of physical activity interventions (384) and childhood and adolescent obesity interventions (385) reported similar findings to those reported in Chapter Four. Furthermore, Owen *et al.* recently updated their 2012 review (372) of NICE public health guidelines for the years 2011-2016 (373) and again in 2019 where they combined updated information with their findings

from earlier reviews (21). In the updated reviews, Owen *et al.* explicitly examined methods of economic evaluation used in NICE guidelines in the review update. Whilst their review is limited only to guidelines produced by NICE, it covers a broad range of public health interventions and their results similarly identify a large majority of CUA and only a handful of either CBA, CCA, or CEA evaluations (21, 373). This is not altogether unsurprising given the historical emphasis on CUA in NICE public health appraisal methods guides (31, 386); however, the inclusion of alternative economic evaluation methods in the guidance demonstrates that the new NICE guidelines for public health appraisal, with respect to accepted evaluations, are being heeded (10) if only to a small extent currently. Consequently, there is reason to believe that the findings regarding alcohol prevention interventions are widespread amongst evaluations of all public health interventions.

The findings from the systematic review reported in this thesis were published in 2017 (114). The impact of this research on the wider literature can be observed in the aforementioned review by Cochrane *et al*., 2019 (384), which cited the published findings. Cochrane *et al*. (384) report recommendations for future economic evaluations of physical activity interventions and cite recommendations reported in the published version of the review conducted for this doctoral research, such as greater use of CCA to report multi-sectoral outcomes and the reporting of information to allow for equity impact to be considered.

11.2.2 *Research question two – Examining use, knowledge, and barriers/enablers to use of health economic evidence by PHDMs*

The existing literature has identified a divide between the evidence produced by researchers and the information required for decision-making by PHDMs (246-250, 387). At the onset of the qualitative study reported in Chapter Five (October 2016), the existing literature focused on evidence in general and touched only minimally on health economic evidence, with the exception of research conducted in LA (or Welsh local health board) settings on the use of priority-setting tools (54, 388). There was a paucity, however, of literature examining the use of economic evaluation evidence in public health decision-making. The qualitative interview study thus aimed to address this gap in knowledge by exploring the use of health economics to aid public health decision-making and to consider the barriers and enablers to its use.

This doctoral research has identified that health economic tools are not extensively used by PHDMs in LAs. Several reasons for this were inferred from the qualitative study's findings

(Chapter Five). Firstly, the setting in which public health decisions are made within LAs is complex and decisions are based on a multitude of factors, with health economic evidence playing only a small part of the decision-making framework. Other scholars examining public health decision-making practices have recently reported similar findings (57, 379, 380). Secondly, limited availability and relevance of economic evaluation evidence was stated as a barrier to its use. Mirroring the findings of the systematic review (Chapter Four) a minority of interviewees reported that where evidence was available it typically consisted of CUAs looking at outcomes on a national scale. The available evidence was reported as generally insufficient for PHDMs' purposes as neither local contexts were accounted for, nor were broader social impacts of public health programmes, which were noted to provide value within the context of a LA.

The interview findings also identified varying degrees of knowledge of health economic tools amongst PHDMs of different roles. Overall, however, understanding of the majority of economic evaluation tools was limited. These findings were further validated during the workshop with PHDMs (Chapter Ten) which also identified gaps in PHDMs' knowledge of economic evaluation. This resulted in issues when interpreting the evidence and also in trusting the methods (e.g. the WTP measure). Recently published research by Frew & Breheny (56) further serves to validate this point as their study similarly identified knowledge limitations around health economic terminology. Frew & Breheny's (56) findings lend support for the argument that more extensive training in health economics for PHDMs may be necessitated to ensure economic evaluation evidence can be used effectively for decision-making in LAs.

Nevertheless, the onus should not only be on PHDMs to become better acquainted with health economic tools. Health economists should also assume responsibility for ensuring that the presentation of information is appropriate for non-health economists. As discussed in Chapter Ten, the presentation of economic evaluation information in non-technical formats is essential to facilitate greater engagement with health economic evidence. The need to improve knowledge transfer from academia and other research institutions to public health policy-makers is a currently discussed issue (e.g. Hunter., 2019 (379), Wilson & Sheldon*.,* 2019 (380), and McAteer *et al*., 2018 (381)) and is a foremost focus of research institutions such as Fuse (382). However, the findings from this thesis serve to demonstrate that health economics is not immune to issues with knowledge transfer and further research

should be conducted in order to bridge the gap between the production of evidence and its use in public health decision-making.

In addition to the barriers to the use of economic evaluation already discussed, the qualitative study also identified areas in which health economic evidence was perceived by PHDMs to be lacking in relevant information. Areas such as the impact of interventions on inequalities and long-term outcomes were reported as important points for consideration. Considering the equity implications of public health programmes within economic evaluations has been an area of discussion in the existing literature (17, 48) and was also addressed in Chapter Four. Therefore, the qualitative study findings serve as further endorsement for consideration in this area. Whether it is appropriate and feasible to include such considerations within an economic evaluation framework (see Asaria *et al.,* 2016 (27) and Cookson *et al.,* 2017 (26) for recent research into how this might be possible), or whether prioritisation frameworks such as MCDA might better address these considerations via the inclusion of equity implications as a decision criterion, remains an area for further deliberation.

Inroads to research examining the use of health economic tools in LAs have been made recently which complement the findings of this thesis with respect to the question of how health economic evidence is used by PHDMs. Firstly, a substantive research project examining the use of priority-setting tools was concluded (58) and a follow-on programme of research assessing the use and impact of the prioritisation framework developed by PHE was undertaken (266). The political environment in which public health decisions are currently made was reported in both studies (58, 266) as a barrier to the adoption of prioritisation tools; the same issue was reported by the interviewees in Chapter Five with regards to economic evaluation evidence. Additionally, the earlier cited study by Frew & Breheny (56) published findings which complement those reported in this thesis with respect to the context of current public health decision-making in England and Wales and barriers to the use of economic evaluation in such settings. Given the emerging research in this subject area, the findings reported in this thesis in answer to my second research question (section 11.1) are demonstrably relevant and topical.

11.2.3 *Research question three – Is a particular method most beneficial to PHDMs?*

Following the conduction of a CBA and SROI analysis (Chapter Eight) of the ASBI programme evaluated in the SIPS Jr HIGH trial (59) evidence reports for four methods of economic evaluation (CUA, CCA, CBA, and SROI) were presented to PHDMs in a workshop and the PHDMs' views of the merits and demerits of each method with respect to their day-to-day decision-making were obtained. The workshop findings (Chapter Ten) were considered in order to answer the final thesis research question of whether a method of economic evaluation can be identified as most beneficial to meet PHDMs' needs.

The workshop findings were not sufficient to conclude that a single technique should be chosen as a primary method of evaluation for public health interventions. PHDMs in attendance at the workshop cautioned that a "one-tool-fits-all" approach was unlikely to be the best approach, as they considered the information from different methods to be relevant to different decision-makers. However, the suggestions from the existing guidance for public health appraisal (10, 17, 47) and the findings from systematic reviews exploring the same issue (48, 114, 384) were confirmed to some extent during the workshop. For example, PHDMs favoured CCA and perceived it to be welcomed in public health decision-making in order to facilitate the consideration of outcomes relevant to different stakeholders. A caveat was placed on this stating that CCA would be beneficial but only as an adjunct to another method of evaluation. This reflects the disadvantage of CCA that it does not produce a final aggregated outcome (see Chapter Three).

The workshop discussions (Chapter Ten) revealed that neither CBA nor CUA were generally favoured, although elements from each did provide value. The lack of favour towards CBA was based largely on mistrust of the WTP benefit measure, however, the use of monetised outcomes was viewed positively as being relatable to LA members. QALYs, on the other hand, were reported as less useful with attendees commenting that the measure is not used in LA decision-making. Notwithstanding this, the comparative ability of the ICER outcome was considered beneficial. An overall preference for SROI was revealed amongst the majority of the PHDMs. The fondness for SROI was attributed to both methodological and presentational features. Methodologically, workshop attendees expressed the merits of the monetary units of outcome and the inclusion of impacts relevant to various LA sectors in addition to health outcomes via monetised QALYs. From the perspective of the presentation,

the PHDMs liked the disaggregated approach to presenting each element of impact which allowed for a relative comparison of how each element contributed to the final SROI ratio.

However, recommending SROI as the most beneficial method of economic evaluation based on the preference displayed at the workshop should come with a caution attached. First, a minority of the PHDMs expressed views that SROI would be more useful at a national level than at the local level. If this is a commonly shared viewpoint amongst PHDMs, recommending SROI as a primary method of economic evaluation would negate the objective of improving local public health decision-making. However, this consideration was not raised by the majority of the workshop attendees, hence further research would need to be conducted specifically examining SROI in LA decision-making contexts in order to determine its relevance to local decisions.

Furthermore, some of the workshop attendees failed to fully understand parts of the SROI analysis report, for instance misunderstanding the role of "deadweight" (represented via the trial control group) and failing to interpret the 12-month time horizon of the study. Whilst some of these issues could be attributed to the approach taken to present the report at the workshop, it should not be dismissed that this is a relatively new method in the public health field to with which PHDMs are largely unfamiliar (a point reported by the workshop attendees). Other common misinterpretations of SROI evidence have been discussed, such as misinterpreting the SROI ratio as a direct cashable return rather than a combined return which includes both financial savings from reduced resource use and a monetary measure of valued impact (378). Therefore, if PHDMs observe value in the use of SROI analysis, formalised training on the method, its limitations, and how to critically appraise SROI analyses should to be provided to PHDMs.

A very recently published Delphi study was conducted with local PHDMs to measure agreement on a number of economic evaluation methodological elements (57). This study was also unable to recommend a particular economic evaluation methodology that would be most appropriate to aid local public health decisions, due to large heterogeneity in responses to preferences for evaluation outcomes. However, several of the findings from the Delphi study validate the preferences expressed within the workshop conducted for this research. For example, Frew & Breheny (57) identified high levels of agreement on aspects such as ensuring evaluations are relevant to a local context, ensuring that costs and effects are transparently reported for different sectors and population subgroups, and preference

for a broad evaluative framework that captures both health and non-health costs and outcomes.

In order to definitively answer research question three, further research needs to be conducted. As stated in Chapter Ten, the preference for SROI may be due to specific presentational choices made when preparing the evidence for the workshop (Chapter Nine). Therefore, using the feedback from the workshop attendees regarding their preferred mode of presentation, information should be reformatted, and feedback obtained to identify whether the mode of presentation or the information reported most influences preferences for specific methodologies.

11.3 Implications of thesis findings

The research findings reported in this thesis have a number of implications for both health economists conducting evaluations of public health interventions and decision-makers as end users of the information.

Whilst no distinct method of economic evaluation can be robustly recommended for the appraisal of public health interventions, this research has strengthened the argument for broader analyses than are currently used for the majority of evaluations (i.e. either CEA or CUA, see Hill *et al.,* 2017 (114) and Cochrane *et al.,* 2019 (384)). This validates recommendations that have been made in academic circles for over a decade. This thesis suggests that health economists should re-evaluate the status quo of conducting CUA when evaluating public health programmes and consider whether an alternative approach, such as CBA or SROI, may be relevant in that case. At a minimum, researchers should consider using broader analytic perspectives, such as a public health or societal perspective, and consider the scope of sensitivity analyses conducted to provide information relevant to different population subgroups and locally relevant data where possible. In doing so, health economic evidence can provide relevant information for the consideration of intervention impact on inequalities and whether a programme may be suited to the locality in which decision-makers reside.

Furthermore, health economists should strive to report CCAs as a secondary analysis in all evaluations in order to clearly provide relevant information for a range of stakeholders.

An additional finding from the workshop study (Chapter Ten) was that improvement is required with respect to the presentation of health economic information. Whilst this issue

has been considered previously (389), limited research has been conducted in conjunction with decision-makers, particularly those in the public health field; the guidance identified by Sullivan *et al.,* 2015 (389) was typically guiding the reporting of evaluations for HTA or pharmaceutical reimbursement.

Heterogeneity in information requirements was observed during the workshop; some PHDMs desired greater detail whilst others sought greater simplification. Common amongst all attendees, however, was desire for transparency with regards to the data included in the analysis and reporting in plain English. Hence, health economists should consider how best to deliver the findings from their evaluations. Whilst publishing in academic journals may remain necessary for dissemination to peers, this is unlikely to be an appropriate form of dissemination for PHDMs. Frew & Breheny (57) similarly identified high levels of agreement from PHDMs regarding dissemination in the form of short briefings and local reports. Therefore, in addition to journal publication, the findings from this thesis recommend that health economists prepare additional reports for use in decision-making practice. Furthermore, greater attention should be paid to presentational styles tailored to different stakeholders.

This thesis also provided insight into the level of knowledge that PHDMs hold around health economic concepts. In order for health economic evidence to be most beneficial to public health decision-making, PHDMs of all levels would benefit from more training opportunities in health economics alongside the efforts described earlier from health economists to improve the translation of information to end-users. Whilst some coverage of health economics is included in public health specialty registrar training, it may be insufficient for the day-to-day needs of decision-makers. Additionally, decision-makers in other roles, such as commissioners and practitioners, may not have had similar training opportunities. Therefore, the findings from this thesis suggest that education on how to interpret economic evaluation evidence and to understand the limitations of different approaches should be more widely provided, and taken up, by those in a public health decision-making capacity.

11.4 Strengths and limitations

Directly involving PHDMs was a key strength in this research. This project was intended to have a positive impact on the practice of local public health decision-making; therefore, it was essential that the views of PHDMs were sought and included throughout. Although the concept of economic evaluation methodology is largely academic, the findings from this

research have enabled the concept to be placed in the context to which end-users would use the information. In doing so, the question of what the most beneficial economic evaluation method is to aid public health decision-making could be considered more appropriately, acknowledging that health economics is only one factor in public health decisions.

Two unique economic evaluations were also conducted within this doctoral project, a CBA and an SROI. As was identified in the systematic review (Chapter Four), CBAs using benefits elicited via stated preference techniques are rare in the public health field. This thesis, thus, provides a novel evaluation in the domain of public health. In conducting the CBA evaluation, limited guidance was identified within the health and public health field on how to aggregate costs and outcomes on a national scale, therefore, direction was taken from the environmental health literature. The potential limitations of the approach taken were discussed in Chapter Eight. The lessons learned from conducting the CBA are important; firstly, to provide examples to future researchers, and secondly, to acknowledge that standardised guidance on conducting CBAs of public health interventions is required.

SROI analyses of public health programmes are also infrequently published, therefore, the evaluation reported in this thesis offers a novel evaluation of a school based ASBI. Similar to the CBA, important lessons were learned during the course of the evaluation which may be beneficial for future researchers. Firstly, conducting an SROI using trial-based data is slightly unconventional; if future trial-based SROIs are to be conducted, appropriate stakeholder engagement should be planned from the beginning of the trial to ensure appropriate outcomes can be identified and relevant data collected. Secondly, standardised guidance for SROI is yet to be developed (33), therefore, this should be prioritised and should include guidance on conducting SROI in various contexts, including alongside trials.

Finally, presenting evidence of economic evaluations, including those unfamiliar to PHDMs was novel. To my knowledge, no other studies have explored PHDMs' views of different economic evaluation methods using this approach. Exposing PHDMs to evidence they may encounter if the information was made available to them, as opposed to discussing the methods in theory, adds strength to the findings of this research.

There are also limitations to the research conducted in this thesis. Some of these related to each empirical study have been stated earlier in this chapter and within the empirical chapters individually (see Chapters Four, Five, Seven, Eight and Ten), nevertheless, there are also broader limitations to the overall doctoral research. Firstly, the case study chosen (i.e.

the SIPS Jr HIGH trial) could be viewed a limitation as a targeted prevention intervention evaluated within a trial. However, one of the complexities of public health programmes is their diversity, therefore, no single case study could cover all potential formats in which public health programmes may be realised. Additionally, the case study enabled lessons to be learned with regards to conducting SROI analyses and CBAs using trial data.

Secondly, it was not possible to address all of the methodological challenges identified in Chapter Four using the data available, for example there was insufficient data on socioeconomic status of trial participants to provide subgroup analysis relevant to health inequalities. When planning future evaluations, the collection of relevant socioeconomic data should be proposed at the beginning of the study to ensure relevant subgroup analysis can be conducted.

Finally, the choice of the stated preference technique used to obtain a measure of WTP could also be a limiting factor, particularly with hindsight from the reactions of the workshop attendees to the WTP measure. Whether an alternative method such as a DCE would have been perceived less controversially is unsure, however, the indirect mode of eliciting a WTP value may seem less "subjective" than the direct elicitation from a CV survey. Nevertheless, prior to the workshop, a CV study was chosen as it was considered less cognitively burdensome for respondents.

11.5 Future research

Some areas for future research have been identified earlier in this chapter such as improving knowledge transfer between health economists and PHDMs (section 11.2.2) and further exploration of the benefit to local decision-making of SROI (section 11.2.3). Both of these aspects of future research call for PHDM stakeholder involvement, in which academics work with the end-users of economic evaluation evidence to further explore how health economic information can be best produced and used in public health decision-making practice.

Additionally, in response to PHDMs initial reactions to WTP measures, research to examine whether CBA would be acceptable to PHDMs using any stated preference measure of benefit (e.g. elicited via either CV or DCE) would be beneficial. Furthermore, alternative methods of measuring benefit such as using monetised QALYs for different sectors (e.g. health and crime, as was conducted in some reports identified in Chapter Four (207, 208, 212-215)) could be explored. Although this would result in CBAs that do not follow the welfarist

grounding in which CBA emerged (see Chapter Two), this approach was identified in the systematic review (Chapter Four) as most commonly used for CBAs of public health interventions. Therefore, this could be considered if PHDMs perceive merit in the approach. An alternative approach for consideration may also be social CBA, which uses a non-preference based approach to eliciting benefit measures by attaching a monetary value to individuals' reported "life satisfaction" (390). Caution has been advised with this version of CBA, however, due to implausibly high valuations arising from the life-satisfaction approach (391). In order to consider social CBA an appropriate alternative to CBA as presented in this thesis, further work on the validity of life-satisfaction valuations should be conducted.

Another area in which further methodological exploration would be beneficial lies with SROI and whether it could be, to any extent, standardised in order to improve comparability of the method. Some form of a "reference case" for outcomes recommended to be explored in SROI analyses of public health programmes could perhaps be introduced to make evaluations more comparable, however, this may contradict the underlying methodology of SROI.

Furthermore, Chapter Seven outlined four areas of further research specific to the empirical work conducted for examining the WTP outcomes (section 7.4.6). Points for further exploration included: investigating the impact of alternative payment vehicles on WTP outcomes, examining the impact on the proportion of zero valuations of a different survey sample, using alternative methods to address protest responses, and re-analysing the WTP taking the assumption of interval, rather than continuous, data.

Finally, a suggestion was made in section 11.3 that health economists should strive to produce CCAs as secondary analyses in order to provide relevant information to a range of stakeholders. Since this is not currently the norm (judging by the paucity of CCAs identified in the systematic review in Chapter Four), research into the feasibility and acceptability of this may be beneficial. It may be prudent to explore the attitudes of health economists who are accustomed to conducting CEAs or CUAs towards such a request. Barriers may exist for health economists habituated to producing CEAs and CUAs, who may be less familiar with capturing a broad range of costs and outcomes, which would be necessary for the production of secondary CCAs.

11.6 Concluding remarks

The findings reported in this thesis offer insight into areas worth pursuing further, such as the transfer of research from academia to end-users and addressing methodological aspects of evaluation that are challenging due to the complexities of public health programmes compared to other health technologies. Yet, more research is needed. Fortuitously, recent endeavours by PHE (265), academics (e.g. Brown *et al*., 2017 (58), Frew & Breheny., 2019 (57), and Edwards & McIntosh., 2019 (55)), and those bridging the gap between academia and policy (e.g. Hunter., 2019 (379), Wilson & Sheldon., 2019 (380), Cheetham *et al.* (382) and McAteer *et al.,* 2018 (381)) are bringing to the fore exploration into areas for further research identified in this thesis. This demonstrates research momentum in these fields and suggests that further investigation may be welcomed and encouraged.

The complexity of decision-making in public health in England currently has been demonstrated in this thesis and addressing this will require more than a single doctoral research investigation. However, the findings reported in this dissertation can be used as stepping stones to reach the goal of improving the landscape of evidence-informed public health decision-making.

The findings in this thesis indicate that there may not be a single method of economic evaluation that should be recommended for all public health programmes; however, a number of recommendations for the economic evaluations of public health interventions can be made based on this research. Firstly, evaluations should routinely incorporate a CCA as a secondary analysis, particularly in cases where the primary analysis is a CUA, which would enable the provision of relevant information to a variety of stakeholders involved in public health decision-making. Such a recommendation goes beyond that set forth by NICE in their recent guidance on public health appraisal (10) which states that secondary analyses, such as CCA, are accepted rather than necessarily expected. Secondly, health economists should strive to expand sensitivity analyses where practicable to provide information of an intervention's impact on different population subgroups. Such an endeavour could improve the relevance of economic evaluation evidence to the current local contexts of public health decision-making.

As has been highlighted during this research project, when addressing the complex system that is public health, a "one-tool-fits-all" approach is unlikely to be appropriate. However, it has also been identified that the status-quo of relying predominantly on CUA to evaluate

public health interventions may not be appropriate in many instances. PHDMs require evidence that resonates with, and appeals to, the multi-sectoral nature of LAs. However, the precise methodology to achieve this remains to be determined. The natural choice from a theoretical economic perspective would be CBA due to the method's ability to incorporate all costs and outcomes on whomsoever they fall, however, in practice it was not necessarily favoured.

Areas for further research that have been identified during the course of this doctoral project have been outlined. These include efforts to explore how best to present health economic evidence to be engaging to the variety of stakeholders involved in public health decision-making and to ensure guidance for methods such as SROI and CBA, which are infrequently used to evaluate public health programmes by the academic community currently, are standardised to improve the quality of evaluations. Misinterpreted or poor-quality evidence may be worse than no evidence at all. Therefore, efforts to improve the interpretation of evidence by end-users (from both supply and demand side initiatives) and to ensure the production of high-quality, transparent, and trustworthy research is a key imperative moving forward.

Appendices

Appendix A. Full search strategies for systematic review

The search strategies used in each database search for the original systematic review (January 2006- May 2016) reported in Chapter Four are listed in Table A.1 - Table A.6 below. The databases searched were NHS EED, Medline, Embase, PsychINFO, CINAHL, and Scopus.

Table A.1. Search terms for NHS EED database

((Alcohol* or drink* or intoxica* or beer or wine)) IN NHSEED
MeSH DESCRIPTOR Drinking Behavior EXPLODE ALL TREES
MeSH DESCRIPTOR Alcohol Drinking EXPLODE ALL TREES WITH QUALIFIER PC IN NHSEED
MeSH DESCRIPTOR Alcoholic Beverages EXPLODE ALL TREES
(Drink* behavio*) IN NHSEED
("Alcohol* use disorder*") OR ("alcohol* abuse") OR ("alcohol* beverage*") IN NHSEED
#1 OR #2 OR #3 OR #4 OR #5 OR #6
(PBMA or "option appraisal" or "priority setting" or "return on investment" or ROI) IN NHSEED
#7 AND #8
(#7) IN NHSEED FROM 2006 TO 2015

Table A.2. Search strategy for Medline database

Economics/
exp "costs and cost analysis"/
Economics, Dental/
exp economics, hospital/
Economics, Medical/
Economics, Nursing/
Economics, Pharmaceutical/
(economic$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab.
(expenditure$ not energy).ti,ab.
value for money.ti,ab.
budget$.ti,ab.
or/1-11
((energy or oxygen) adj cost).ti,ab.
(metabolic adj cost).ti,ab.
((energy or oxygen) adj expenditure).ti,ab.
or/13-15
12 not 16
letter.pt.
editorial.pt.
historical article.pt.
or/18-20
17 not 21
exp animals/ not humans/
22 not 23
bmj.jn.
"cochrane database of systematic reviews".jn.
health technology assessment winchester england.jn.
or/25-27
24 not 28
Decision Making, Organizational/mt [Methods]
exp Resource Allocation/mt [Methods]
(MCDA or PBMA).ti,ab.
"option appraisal".ti,ab.
"multi$ criteria decision analys$".ti,ab.
"program$ budget$ marginal analys$".ti,ab.
(Priority?setting adj2 method$).ti,ab.
"social return on investment".ti,ab.
(SROI or ROI).ti,ab.
"return on investment".ti,ab.

or/30-39
29 or 40
(intoxica$ or beer or wine).ti,ab.
*drinking behavior/
Alcoholic Beverages/
*Binge Drinking/pc [Prevention & Control]
Alcohol Drinking/pc [Prevention & Control]
*Alcoholism/pc [Prevention & Control]
("Drink$ behavio$" or "binge drink$").ti,ab.
(Alcohol$ adj2 ("use disorder$" or abuse or beverage$ or addiction$ or consumption or drink$)).ti,ab.
or/42-49
41 and 50
limit 51 to (english language and humans and yr="2015 -Current")
remove duplicates from 52

Table A.3. Search strategy for Embase database

Health Economics/
exp Economic Evaluation/
exp Healthcare Cost/
pharmacoeconomics/
1 or 2 or 3 or 4
(econom$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$).ti,ab.
(expenditure$ not energy).ti,ab.
(value adj2 money).ti,ab.
budget$.ti,ab.
6 or 7 or 8 or 9
5 or 10
letter.pt.
editorial.pt.
note.pt.
12 or 13 or 14
11 not 15
(metabolic adj cost).ti,ab.
((energy or oxygen) adj cost).ti,ab.
((energy or oxygen) adj expenditure).ti,ab.
17 or 18 or 19
16 not 20
animal/
exp animal experiment/
nonhuman/
(rat or rats or mouse or mice or hamster or hamsters or animal or animals or dog or dogs or cat or cats or bovine or sheep).ti,ab,sh.
22 or 23 or 24 or 25
exp human/
human experiment/
27 or 28
26 not (26 and 29)
21 not 30
0959-8146.is.
(1469-493X or 1366-5278).is.
1756-1833.en.
32 or 33 or 34
31 not 35
conference abstract.pt.
36 not 37

*decision making/
*healthcare planning/
*resource allocation/
budget/
*"cost benefit analysis"/
("program$ budget$ marginal analys$" or PBMA).ti,ab.
("multi?criteria decision analys$" or MCDA).ti,ab.
"option appraisal".ti,ab.
("social return on investment" or SROI or "return on investment" or ROI).ti,ab.
("Priority-setting" adj2 method$).ti,ab.
39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48
38 or 49
(intoxica$ or beer or wine).ti,ab.
alcohol abuse/
*alcoholism/pc [Prevention]
*alcohol consumption/
*binge drinking/pc [Prevention]
*drinking behavior/pc [Prevention]
"alcohol use disorder"/pc [Prevention]
("Drink$ behavio$" or "binge drink$").ti,ab.
(Alcohol$ adj2 ("use disorder$" or abuse or beverage$ or addiction$ or consumption or drink$)).ti,ab.
51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59
50 and 60
limit 61 to (human and english language and yr="2015 -Current")
remove duplicates from 62

Table A.4. Search strategy for PsychINFO database

"Cost Containment"/
(economic adj2 evaluation$).ti,ab.
(economic adj2 analy$).ti,ab.
(economic adj2 (study or studies)).ti,ab.
(cost adj2 evaluation$).ti,ab.
(cost adj2 analy$).ti,ab.
(cost adj2 (study or studies)).ti,ab.
(cost adj2 effective$).ti,ab.
(cost adj2 benefit$).ti,ab.
(cost adj2 utili$).ti,ab.
(cost adj2 minimi$).ti,ab.
(cost adj2 consequence$).ti,ab.
(cost adj2 comparison$).ti,ab.
(cost adj2 identificat$).ti,ab.
(pharmacoeconomic$ or pharmaco-economic$).ti,ab.
or/1-16
(task adj2 cost$).ti,ab,id.
(switch$ adj2 cost$).ti,ab,id.
(metabolic adj cost).ti,ab,id.
((energy or oxygen) adj cost).ti,ab,id.
((energy or oxygen) adj expenditure).ti,ab,id.
or/18-22
(animal or animals or rat or rats or mouse or mice or hamster or hamsters or dog or dogs or cat or cats or bovine or sheep or ovine or pig or pigs).ab,ti,id,de.
editorial.dt.
letter.dt.
dissertation abstract.pt.
or/24-27
(0003-4819 or 0003-9926 or 0959-8146 or 0098-7484 or 0140-6736 or 0028-4793 or 1469-493X).is.
17 not (23 or 28 or 29)
("multi$ criteria decision analys$" or MCDA).ti,ab.
("program$ budget$ marginal analys$" or PBMA).ti,ab.
("Priority-setting" adj2 methods).ti,ab.
"option appraisal".ti,ab.
("social return on investment" or SROI or "return on investment" or ROI).ti,ab.
resource allocation/
*decision making/

or/31-37
30 or 38
alcohol abuse/
binge drinking/
alcohol drinking patterns/
drinking behavior/
alcoholic beverages/
alcoholism/
(intoxica$ or beer or wine).ti,ab.
("Drink$ behavio$" or "binge drink$").ti,ab.
(Alcohol$ adj2 ("use disorder$" or abuse or beverage$ or addiction$ or consumption or drink$)).ti,ab.
or/40-48
39 and 49
limit 50 to (human and english language and yr="2015 -Current")
remove duplicates from 51

Table A.5. Search strategy for CINAHL database

MH "Economics+"
MH "Financial Management+"
MH "Financial Support+"
MH "Financing, Organized+"
MH "Business+"
S2 OR S3 or S4 OR S5
S1 NOT S6
MH "Health Resource Allocation"
MH "Health Resource Utilization"
S8 OR S9
S7 OR S10
TI (cost or costs or economic* or pharmacoeconomic* or price* or pricing*) OR AB (cost or costs or economic* or pharmacoeconomic* or price* or pricing*)
S11 OR S12
PT editorial
PT letter
PT commentary
S14 or S15 or S16
S13 NOT S17
MH "Animal Studies"
(ZT "doctoral dissertation") or (ZT "masters thesis")
S18 NOT (S19 OR S20)
MH “decision making”
MH “resource allocation”
TI (“multi*criteria decision analys*” or MCDA) or AB (“multi*criteria decision analys*” or MCDA)
TI (“program* budget* marginal analys*” or PBMA) or AB (“program* budget* marginal analys*” or PBMA)
TI (“priority setting” N2 method*) or AB (“priority setting” N2 method*)
TI (option appraisal) or AB (option appraisal)
TI (“social return on investment” or “return on investment” or ROI) or AB (“Social return on investment” or “return on investment” or ROI)
S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28
S21 OR S29
MH “alcohol abuse”
MH “alcoholic beverages”
MH “drinking behavior”
MH “alcoholism”
MH “binge drinking”

TI (“alcohol abuse” or “alcohol misuse” or “binge drink*”) or AB (“alcohol abuse” or “alcohol misuse” or “binge drink*”)
TI (beer or wine or intoxica*) or AB (beer or wine or intoxica*)
TI (drink* behavio*) or AB (drink* behavio*)
TI Alcohol* N2 (“use disorder*” or beverage* or addiction* or consumption or drink*) or AB Alcohol* N2 (“use disorder*” or beverage* or addiction* or consumption or drink*)
S31 OR S32 OR S33 OR S34 OR S35 OR S36 or S37 or S38 or S39
S30 AND S38 (with limiters: Published date: 20150101-20160531; English Language; Human)

Table A.6. Search strategy for Scopus 2006-2016 database search

TITLE-ABS-KEY (“return on investment” or ROI)
TITLE-ABS-KEY (“social return on investment” or SROI)
TITLE-ABS-KEY (“multi*criteria decision analys*” or MCDA)
TITLE-ABS-KEY (“option appraisal”)
TITLE-ABS-KEY (“Program* budget* marginal analys*” or PBMA)
TITLE-ABS-KEY ("priority setting" w/2 method*)
TITLE-ABS-KEY (“resource allocate*”)
TITLE-ABS-KEY (“multi*criteria decision aid”)
#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8
TITLE-ABS-KEY (“alcohol abuse” or “alcohol misuse” or “binge drink*”)
TITLE-ABS-KEY (drink* w/1 behavio*)
TITLE-ABS-KEY (Alcohol* w/2 (“use disorder*” or abuse or beverage* or addiction* or consumption or drink*))
Or/10-12
9 AND 13 (limited to English language and year 2006-2016)

Appendix B. Quality assessment of studies for systematic review

Quality assessment of each included study in the systematic review reported in Chapter Four was based on the CHEERS checklist (182). The completed quality assessment checklist for each study is reported below. Due to the number of included studies it was not possible to include them all in one table; therefore, quality assessment is split over two tables in alphabetical order of authors. Table B.1 covers authors from A-K and Table B.2 covers authors from L-Z. The SROI by Tanaree *et al.,* 2019 (233) and the MCDA by Rogeberg *et al.,* 2018 (193) are included in the CHEERS checklist for completeness. However, the CHEERS checklist is not designed for SROI or MCDA evaluations, therefore, the fields may not be appropriate to consider the methodological quality of these evaluations. The checklist was particularly inapplicable to the MCDA study (60% of the fields were "not applicable"). Therefore, it was not considered appropriate to include the proportion score of eligible items calculated for the MCDA in the discussion of quality scores in Chapter Four (see section 4.4.3).

Table B.1. Quality assessment checklist based on CHEERS checklist – Authors A-K

	Quality checklist	Angus *et al.* (a) 2014 (198)	Angus *et al.* (b) 2014 (216)	Angus *et al.* 2015 (207)	Angus *et al.* 2016 (227)	Angus *et al.* 2018 (215)	Angus *et al.* 2019 (223)	Barbosa *et al.* 2015 (199)	Barrett *et al.* 2006 (392)	Brennan *et al.* 2009 (214)	Byrnes *et al.* 2010 (218)	Cobiac *et al.* 2009 (200)	Cobiac *et al.* 2018 (206)	Cowell *et al.* 2012 (393)	Crawford *et al.* 2015 (234)	De Wit *et al.* 2016 (237)	Drummond *et al.* 2009 (209)	Havard et. al 2012 (230)	Holm *et al.* (a) 2014 (201)	Holm *et al.* (b) 2014 (202)	Ingels *et al.* 2013 (232)	Kapoor *et al.* 2009 (194)
1	Form of economic evaluation clearly reported	✓	✓	✓	✗	✗	✓	✓	✓	✓	✗	✓	✓	✓	✗	✓	✗	✗	✓	✓	✓	✓
1a	Reported form of economic evaluation is accurate	✓*	✓◊	✓*◊	N/A	N/A	✓	✓*	✓	✓◊	N/A	✓*	✓*	✓	N/A	✓◊	N/A	N/A	✓*	✓*	✓	✓*
2	Target population and subgroups are reported	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	Setting of evaluation is reported	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	Study perspective is reported	✓	✗	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
4a	Reported perspective is accurate	✓	N/A	✓	✓	N/A	✓	✓	✓	N/A	✓+	✓+	✓	✓	✓	✓	N/A	✓	✓+	✓+	✓	✓
5	Comparator interventions reported	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6	Time horizon for evaluation is reported	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
7	Discount rate is reported	✓	✓	✓	✓	✗	✓	N/A	N/A	✓	✓	✓	✓	N/A	N/A	✓	N/A	✗	✓	✓	N/A	✓

Table B.1 cont. Quality assessment checklist based on CHEERS checklist – Authors A-K

	Quality checklist	Angus *et al.* (a) 2014 (198)	Angus *et al.* (b) 2014 (216)	Angus *et al.* 2015 (207)	Angus *et al.* 2016 (227)	Angus *et al.* 2018 (215)	Angus *et al.* 2019 (223)	Barbosa *et al.* 2015 (199)	Barrett *et al.* 2006 (392)	Brennan *et al.* 2009 (214)	Byrnes *et al.* 2010 (218)	Cobiac *et al.* 2009 (200)	Cobiac *et al.* 2018 (206)	Cowell *et al.* 2012 (393)	Crawford *et al.* 2015 (234)	De Wit *et al.* 2016 (237)	Drummond *et al.* 2009 (209)	Havard et. al 2012 (230)	Holm *et al.* (a) 2014 (201)	Holm *et al.* (b) 2014 (202)	Ingels *et al.* 2013 (232)	Kapoor *et al.* 2009 (194)
8	Relevance of outcome measures is reported	✓	✗	✗	✗	✗	✗	✓	✗	✓	✗	✓	✓	✓	✗	✓	✗	✗	✗	✗	✓	✗
9	Measurement of effectiveness described	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	Methods of valuing preference-based outcomes is described	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	✓	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
11	Methods of estimating resource use are described	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	N/A
12	Methods of valuing resources in terms of unit costs are reported	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
13	Details of currency and price adjustments for inflation or currency conversion are given (where appropriate)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✗	✗	✗	✓

Table B.1 cont. Quality assessment checklist based on CHEERS checklist – Authors A-K

	Quality checklist	Angus *et al.* (a) 2014 (198)	Angus *et al.* (b) 2014 (216)	Angus *et al.* 2015 (207)	Angus *et al.* 2016 (227)	Angus *et al.* 2018 (215)	Angus *et al.* 2019 (223)	Barbosa *et al.* 2015 (199)	Barrett *et al.* 2006 (392)	Brennan *et al.* 2009 (214)	Byrnes *et al.* 2010 (218)	Cobiac *et al.* 2009 (200)	Cobiac *et al.* 2018 (206)	Cowell *et al.* 2012 (393)	Crawford *et al.* 2015 (234)	De Wit *et al.* 2016 (237)	Drummond *et al.* 2009 (209)	Havard et. al 2012 (230)	Holm *et al.* (a) 2014 (201)	Holm *et al.* (b) 2014 (202)	Ingels *et al.* 2013 (232)	Kapoor *et al.* 2009 (194)
14	Description and justification of decision-analytic model (or other type of model) provided	✓	✓	✓	✓	✓	✓	✓	N/A	✓	✓	✓	✓	N/A	N/A	✓	N/A	N/A	✓	✓	N/A	✓
15	Assumptions related to model are described and explained	✓	✓	✓	✓	✓	✓	✗	N/A	✓	✓	✓	✓	N/A	N/A	✓	N/A	N/A	✓	✓	N/A	✓
16	Analytic methods to support evaluation are reported	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
17	Values and ranges of each component of cost and outcome are reported	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
18	Incremental analysis is reported (mean values of each component and mean differences provided)	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓

Table B.1 cont. Quality assessment checklist based on CHEERS checklist – Authors A-K

Quality checklist	Angus *et al.* (a) 2014 (198)	Angus *et al.* (b) 2014 (216)	Angus *et al.* 2015 (207)	Angus *et al.* 2016 (227)	Angus *et al.* 2018 (215)	Angus *et al.* 2019 (223)	Barbosa *et al.* 2015 (199)	Barrett *et al.* 2006 (392)	Brennan *et al.* 2009 (214)	Byrnes *et al.* 2010 (218)	Cobiac *et al.* 2009 (200)	Cobiac *et al.* 2018 (206)	Cowell *et al.* 2012 (393)	Crawford *et al.* 2015 (234)	De Wit *et al.* 2016 (237)	Drummond *et al.* 2009 (209)	Havard et. al 2012 (230)	Holm *et al.* (a) 2014 (201)	Holm *et al.* (b) 2014 (202)	Ingels *et al.* 2013 (232)	Kapoor *et al.* 2009 (194)
19 Uncertainty characterised via sensitivity analysis on key parameters	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓
20 Heterogeneity characterised via discussion of results in relation to subgroups and/or other participant characteristics	✓	✓	✓	×	✓	×	×	×	✓	×	×	✓	×	×	✓	×	×	×	×	×	✓
Proportion score of eligible items	**90%**	**90%**	**95%**	**80%**	**63%**	**90%**	**90%**	**89%**	**95%**	**60%**	**86%**	**95%**	**94%**	**76%**	**95%**	**69%**	**61%**	**86%**	**86%**	**89%**	**90%**

✓ = reported, × = not reported, N/A = not applicable

*Where a CUA is described as a CEA it is considered correct since CUA is a specific type of CEA, however a higher-quality paper would be expected to explain the difference in economic evaluation type

+Partly accurate due to costs included that are broader than would usually be included in a health sector perspective

◊CBA benefit using monetised QALYs so not a welfarist grounded CBA

Table B.2. Quality assessment checklist based on CHEERS checklist – Authors K-Z

	Quality checklist	Lai *et al.* 2007 (203)	Li *et al.* 2017 (197)	Mansdotter *et al.* 2007 (195)	Meng *et al.* 2012 (212)	Meng *et al.* 2013 (213)	Miller & Hendrie 2008 (236)	Miller *et al.* 2007 (196)	Navarro *et al.* 2011 (210)	Neighbors *et al.* 2010 (219)	Purshouse *et al.* 2009 (208)	Purshouse *et al.* 2013 (204)	Rogeberg *et al.* 2018 (193)	Sassi *et al.* 2015 (211)	Shanahan *et al.* 2006 (231)	Smit *et al.* 2011 (225)	Solberg *et al.* 2008 (220)	Tanaree *et al.* 2019 (233)	Tariq *et al.* 2009 (205)	van den Berg *et al.* 2008 (228)	Watson *et al.* 2013 (191)	Zur & Zaric 2016 (229)
1	Form of economic evaluation clearly reported	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	N/A	✗	✓	✗	✓	✓	✓	✗	✓	✓
1a	Reported form of economic evaluation is accurate	✓*	✗	✗	N/A	✓◊	✓§	✗	✓	✓	✓*	✓*	N/A	N/A	✓	N/A	✓	✓	✓*	N/A	✓*	✓
2	Target population and subgroups are reported	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	Setting of evaluation is reported	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	Study perspective is reported	✓	✓	✓	✗	✗	✓	✓	✗	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓
4a	Reported perspective is accurate	✓+	✓	✗	N/A	N/A	✓	✓	N/A	✓γ	✓	✓	N/A	N/A	✓	✓	✓γ	✓	✓	✓	✓	✓
5	Comparator interventions reported	✓	✗	✗	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
6	Time horizon for evaluation is reported	✓	✗	✗	✓	✓	✓	✗	✓	✗	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓
7	Discount rate is reported	✓	✗	✓	✗	✗	✓	✓	N/A	✓	✓	✓	N/A	✓	✗	N/A	✓	✓	✓	✓	N/A	✓

Table B.2 cont. Quality assessment checklist based on CHEERS checklist – Authors K-Z

	Quality checklist	Lai *et al.* 2007 (203)	Li *et al.* 2017 (197)	Mansdotter *et al.* 2007 (195)	Meng *et al.* 2012 (212)	Meng *et al.* 2013 (213)	Miller & Hendrie 2008 (236)	Miller *et al.* 2007 (196)	Navarro *et al.* 2011 (210)	Neighbors *et al.* 2010 (219)	Purshouse *et al.* 2009 (208)	Purshouse *et al.* 2013 (204)	Rogeberg *et al.* 2018 (193)	Sassi *et al.* 2015 (211)	Shanahan *et al.* 2006 (231)	Smit *et al.* 2011 (225)	Solberg *et al.* 2008 (220)	Tanaree *et al.* 2019 (233)	Tariq *et al.* 2009 (205)	van den Berg *et al.* 2008 (228)	Watson *et al.* 2013 (191)	Zur & Zaric 2016 (229)
8	Relevance of outcome measures is reported	✗	✗	✓	✗	✗	✓	✗	✓	✓	✗	✓	✗	✗	✓	✓	✗	✗	✗	✗	✓	✗
9	Measurement of effectiveness described	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	N/A	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	Methods of valuing preference-based outcomes is described	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	✓	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	✓
11	Methods of estimating resource use are described	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	N/A	✗	✓	✓	✓	✓	✓	N/A	✓	✓
12	Methods of valuing resources in terms of unit costs are reported	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	N/A	✗	✓	✓	✓	✓	✓	✓	✓	✓
13	Details of currency and price adjustments for inflation or currency conversion are given (where appropriate)	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	N/A	✗	✓	✓	✓	✓	✓	✓	✓	✓

Table B.2 cont. Quality assessment checklist based on CHEERS checklist – Authors K-Z

Quality checklist	Lai *et al.* 2007 (203)	Li *et al.* 2017 (197)	Mansdotter *et al.* 2007 (195)	Meng *et al.* 2012 (212)	Meng *et al.* 2013 (213)	Miller & Hendrie 2008 (236)	Miller *et al.* 2007 (196)	Navarro *et al.* 2011 (210)	Neighbors *et al.* 2010 (219)	Purshouse *et al.* 2009 (208)	Purshouse *et al.* 2013 (204)	Rogeberg *et al.* 2018 (193)	Sassi *et al.* 2015 (211)	Shanahan *et al.* 2006 (231)	Smit *et al.* 2011 (225)	Solberg *et al.* 2008 (220)	Tanaree *et al.* 2019 (233)	Tariq *et al.* 2009 (205)	van den Berg *et al.* 2008 (228)	Watson *et al.* 2013 (191)	Zur & Zaric 2016 (229)
14 Description and justification of decision-analytic model (or other type of model) provided	✗	N/A	N/A	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	N/A	✓	✓	N/A	✓
15 Assumptions related to model are described and explained	✗	N/A	N/A	✓	✓	✓	✓	✓	✓	✓	✓	N/A	✓	✓	✓	✓	N/A	✓	✓	N/A	✓
16 Analytic methods to support evaluation are reported	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	N/A	✓	✗	✓	✓	✓	✓	✓	✓	✓
17 Values and ranges of each component of cost and outcome are reported	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	N/A	✗	✗	✓	✓	✓	✓	✗	✓	✓
18 Incremental analysis is reported (mean values of each component and mean differences provided)	✓	✓	✗	✗	✗	✗	✗	✓	✓	✗	✓	N/A	✗	✓	✓	✓	N/A	✓	✓	✓	✓

Table B.2 cont. Quality assessment checklist based on CHEERS checklist – Authors K-Z

	Quality checklist	Lai *et al.* 2007 (203)	Li *et al.* 2017 (197)	Mansdotter *et al.* 2007 (195)	Meng *et al.* 2012 (212)	Meng *et al.* 2013 (213)	Miller & Hendrie 2008 (236)	Miller *et al.* 2007 (196)	Navarro *et al.* 2011 (210)	Neighbors *et al.* 2010 (219)	Purshouse *et al.* 2009 (208)	Purshouse *et al.* 2013 (204)	Rogeberg *et al.* 2018 (193)	Sassi *et al.* 2015 (211)	Shanahan *et al.* 2006 (231)	Smit *et al.* 2011 (225)	Solberg *et al.* 2008 (220)	Tanaree *et al.* 2019 (233)	Tariq *et al.* 2009 (205)	van den Berg *et al.* 2008 (228)	Watson *et al.* 2013 (191)	Zur & Zaric 2016 (229)
19	Uncertainty characterised via sensitivity analysis on key parameters	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
20	Heterogeneity characterised via discussion of results in relation to subgroups and/or other participant characteristics	✗	✗	✓	✓	✓	✗	✗	✗	✓	✓	✓	✗	✓	✗	✗	✗	✓	✗	✗	✗	✓	✓
	Proportion score of eligible items	**71%**	**57%**	**67%**	**68%**	**80%**	**81%**	**71%**	**89%**	**90%**	**91%**	**95%**	**44%**	**58%**	**76%**	**89%**	**90%**	**89%**	**90%**	**79%**	**94%**	**95%**

✓ = reported, ✗ = not reported, N/A = not applicable

*Where a CUA is described as a CEA it is considered correct since CUA is a specific type of CEA, however a higher-quality paper would be expected to explain the difference in economic evaluation type

§No details are provided on how quality of life was monetised to ascertain the accuracy of the CBA categorisation

γProductivity costs have not been included in a stated societal perspective. Debate remains over whether these costs should be included

◇CBA benefit using monetised QALYs so not a welfarist grounded CBA

Appendix C. Data extraction form for systematic review

Data were extracted for each included study into a pre-designed data extraction form. Table C.1 displays the data extraction form template for the initial review conducted in May 2016. Table C.2 reports the reduced data extraction form template used for the updated review in March 2019.

Table C.1. Data extraction form template for original review (May 2016)

General Data						
Title	Author	Publication year	Publication type	Country of study	Source of funding	Journal

Study Characteristics								
Intervention	Comparator(s)	Population	Setting	Sample size	Type of study	Follow up length	Time horizon	Perspective (stated)

Outcomes Data				
Primary outcome	Secondary outcome	Tertiary outcome	Cost-effectiveness estimate	Outcome valuation

Methodological considerations					
Extrapolation	Type of economic evaluation stated (actual)		Method of priority-setting	Justification of EE methods	Equity considerations
Reported limitations of method	Discounting	Reported strengths of method	Reported implications for decision-making	Details on how to use results from evaluation	

Costs by sector											
Healthcare	Education	Criminal Justice	Law enforcement	Environment	Employment	Social care	Voluntary	Private	Out of pocket	Government	Other

Productivity	
Changes accounted for?	Method

Table C.2. Data extraction form template for review update (March 2019)

General Data						
Title	Author	Publication year	Publication type	Country of study	Source of funding	Journal

Study Characteristics								
Intervention	Comparator(s)	Population	Setting	Sample size	Type of study	Follow up length	Time horizon	Perspective (stated)

Outcomes Data		
Primary outcome	Secondary outcome	Outcome valuation

Methodological considerations				
Extrapolation	Type of economic evaluation stated (actual)	Method of priority-setting	Discounting	Equity considerations

Costs by sector											
Healthcare	Education	Criminal Justice	Law enforcement	Environment	Employment	Social care	Voluntary	Private	Out of pocket	Government	Other

Productivity	
Changes accounted for?	Method

Appendix D. SRQR checklist for qualitative interview study

The twenty-one item SRQR checklist below was completed for the qualitative study exploring PHDMs' understanding and use of health economic tools reported in Chapter Five. The checklist described the standards for reporting qualitative research (252).

Table D.1. Complete SRQR checklist for qualitative study

Topic	No.	Item	Evidence
Title/Abstract			
Title	1	Concise description of the nature and topic of the study. Identifying the study as qualitative or indicating the approach or data collection methods is recommended	Not applicable for a thesis chapter
Abstract	2	Summary of key elements of the study using the abstract format of the intended publication; typically includes background, purpose, methods, results, and conclusions	Not applicable for a thesis chapter
Introduction			
Problem formulation	3	Description and significance of the problem/phenomenon studied; review of relevant theory and empirical work; problem statement	Section 5.1
Purpose of research question	4	Purpose of the study and specific objectives or questions	Section 5.2
Methods			
Qualitative approach and research paradigm	5	Qualitative approach (e.g., ethnography, grounded theory, case study, phenomenology, narrative research) and guiding theory if appropriate; identifying the research is also recommended; rationale	Section 5.3
Researcher characteristics and reflexivity	6	Researchers' characteristics that may influence the research, including personal attributes, qualifications/experience, relationship with participants, assumptions, and/or presuppositions.	Section 5.5.3
Context	7	Setting/site and salient contextual factors; rationale	Section 5.3
Sampling strategy	8	How and why research participants, documents, or events were selected; criteria for deciding when no further sampling was necessary	Section 5.3.1

Table D.1 cont. Complete SRQR checklist for qualitative study

Topic	No.	Item	Evidence
Ethical issues pertaining to human subjects	9	Documentation of approval by an appropriate ethics review board and participant consent, or explanation for lack thereof	Section 5.3.4
Data collection methods	10	Types of data collected; details of data collection procedures including start and stop dates of data collection and analysis, iterative process and modification of procedures in response to evolving study findings	Sections 5.3.1 and 5.3.2
Data collection instruments	11	Description of instruments and devices used for data collection; if/how the instrument(s) changed over the course of the study	Section 5.3.2
Units of study	12	Number and relevant characteristics of participants, documents, or events included in the study	Sections 5.3.2 and 5.4
Data processing	13	Methods for processing data prior to and during analysis, including transcription, data entry, data management and security, verification of data integrity, data coding, and anonymization/de-identification of excerpts	Section 5.3.2
Data analysis	14	Process by which inferences, themes, etc., were identified and developed, including the researchers involved in data analysis	Section 5.3.3
Techniques to enhance trustworthiness	15	Techniques to enhance trustworthiness and credibility of data analysis (e.g., member checking, audit trail, triangulation)	Sections 5.3.3 and 5.4
Results/Findings			
Synthesis and interpretation	16	Main findings (e.g., interpretations, inferences, and themes); might include development of a theory of model.	Section 5.4
Links to empirical data	17	Evidence (e.g., quotes, field notes, text excerpts, photographs) to substantiate analytic findings	sections 5.4.1 – 5.4.5
Discussion			
Integration with prior work, implications, transferability, and contributions to the field	18	Short summary of main findings; explanation of how findings and conclusions connect to support, elaborate on, or challenge conclusions of earlier scholarship	Section 5.5

Table D.1 cont. Complete SRQR checklist for qualitative study

Topic	No.	Item	Evidence
Limitations	19	Trustworthiness and limitations of findings	Section 5.5.4
Other			
Conflicts of interest	20	Potential sources of influence or perceived influence on study conduct and conclusions	Not applicable for a thesis chapter
Funding	21	Sources of funding and other support; role of funders in data collection, interpretation, and reporting	Not applicable for a thesis chapter

Table adapted from O'Brien *et al.* (2014) (252)

Appendix E. Qualitative study information sheet

During the recruitment phase of the qualitative study reported in Chapter Five an information sheet (displayed below) was sent to potential interviewees. The information sheet explained the purpose of the study and gave some indication of what participation would entail.

Research study on decision-making in public health

You are being invited to take part in a research study. This leaflet explains why the research is being done and what taking part will involve. Please read the following information carefully and discuss it with others if you wish. Thank you for reading this.

What is the study about?

This study is trying to find out how decisions are made about alcohol policies and interventions in the North-East. We are particularly interested in finding out about the types of evidence that are used to inform decisions, including the use of economic evaluation evidence. This study is part of a PhD research project which is evaluating alternative methods of economic evaluation for use in public health. In order to identify the most suitable methods it is important to understand how information is used by those making decisions so that a tool can be developed to best aid decision-making.

What does taking part involve?

The study involves an interview, preferably in person, where you will be asked a few questions about your experiences of the decision-making process within your organisation. The interview will be conducted by one researcher, Sarah Hill in a location that is suitable for you. It is completely voluntary and if you are asked any questions which you would rather not answer you have no obligation to do so. The interview should last around one hour.

Is the study confidential?

Yes, your participation in the study is entirely confidential. Your name would not appear in any reports or publications arising from the study. The study has been approved by Newcastle University Research Office.

How will the information be used?

Your comments will be used by Sarah Hill to develop the final stage of her research project which will involve conducting an economic evaluation of an alcohol intervention. This interview study will help focus the research onto a particular method of economic evaluation to best meet the needs of decision-makers identified during the interviews.

Who is funding this research?

The PhD research project is funded by fuse, the Centre for Translational Research in Public Health (http://www.fuse.ac.uk/).

Contact us

If you have any further questions about the research study and what is involved in taking part in an interview please contact Sarah Hill via e-mail at s.r.hill2@newcastle.ac.uk and she will be able to answer any questions you may have.

What happens next?

You will be contacted in a few days by Sarah Hill to discuss the opportunity to take part in the interview study. If you are happy to take part arrangements will be made that are convenient for you.

Appendix F. Interview schedule and topic guide

A topic guide was developed to guide the semi-structured interviews with PHDMs during the qualitative study reported in Chapter Five. The topic guide was nested within an interview schedule which began with an introductory statement to introduce myself and to reiterate the purpose of the interview and was concluded with closing remarks to thank the participant for their time. The topic guide questions for topics 1 – 6 could be asked in any order. The full interview schedule is presented below.

I. Introduction to interview

- Introduce myself and the project and reiterate the purpose of interview
- Thank interviewee for offering to participate
- Confirm that interviewee has read the information sheet and ask if he/she has any questions about the interview before commencing
- Talk through consent form (explain anonymity, confidentiality, audio recording etc) and ask participant if they are happy to continue. If they are happy ask interviewee to sign consent form.
- Inform interviewee that I am going to turn on the audio recorder (turn on recorder)

II. Preliminary questions

1) *"Could you tell me briefly about your current role and responsibilities?"*

III. Topic guide questions to address objectives

Topic 1: Role in decision-making

1) *"Do you consider yourself part of the process that makes decisions on which interventions or services should be funded with the public health budget? And if so in what capacity?"*
2) *"Who is/else is involved in that process and in what capacity?"*

Probing points:

- *Who is key to the process?*
- *Who makes ultimate decision?*
- *What is the political involvement now that public health is in the remit of LAs?*

- *Is PHE involved? How?*

3) *"How has the integration of public health responsibilities into local government impacted your role?"*

Topic 2: Decision-making process

1) *"Since funding is limited for public health, how are the choices made over what should be funded?"*

Probing points:

- *What basis is an individual intervention/policy/service chosen on?*
- *What type of evidence is used to justify those decisions?*

Topic 3: Understanding of health economic evidence

1) *"If I were to ask you about 'economic evidence' what does that mean to you?"*
2) *"Do you consider this sort of information to help inform decisions on investing in specific interventions/programmes?"*
3) *"Are you familiar with the terms on the sheet, and if so to what extent?"* (Refer to glossary document)

Topic 4: Use and opinions of health economic tools

1) *"Do you regularly use any of the tools described to aid in the decision-making process we have spoken about so far today?"*
2) *"What is your opinion of the economic evaluation tools that you have used in terms of how user friendly they are and how well received the information they provide is?"*
3) *"From the descriptions on the sheet of the other tools, what would your opinion be of those?"* (offer time to re-read descriptions carefully)

Topic 5: Barriers to use of health economic tools

If limited or no use of health economic tools discussed explore barriers:

1) *"What barriers do you perceive to the use of the tools?"*

Probing points:

- Is the information they provide not useful or required?
- Is the information or the tools themselves too technical to understand or use?
- Is there a lack of available evidence?

Topic 6: Information required to inform decisions

1) *"What kind of information is important for making an argument to invest in a public health programme?"*
2) *"In terms of expected outcomes or returns, what is important to demonstrate?"*
3) *"Do local authority priorities besides health have to be addressed to make the case for investment in public health programmes or are predicted health gains sufficient?"*

Probing point:

- Does the impact on wider outcomes need to be considered to make a case for investment?

IV. Closing remarks

- Ask interviewee if they have anything else they would like to discuss
- Let interviewee know they can contact me after the interview if they think of anything after the interview that they would like to share (provide e-mail details)
- Ask interviewee if they would be willing to be part of any further related work
- Ask about further contacts for interviewing – any colleagues they could suggest?
- Thank again for their time
- Let interviewee know you will be in touch with results if they are interested when they are available.

Appendix G. Glossary document given to interviewees

A glossary of health economic tools (including economic evaluation and priority-setting tools) was given to interviewees to aid discussions about each of the health economic tools during the interview study. The glossary is presented below.

Below are descriptions of some key methods of economic evaluation and priority-setting. To aid discussion during the interview could you please read through them. You may be familiar with some of the terms; if so please think about how the descriptions compare to your current understanding of the terms.

Economic evaluation terms

Cost-analysis

Cost-analysis is a partial economic evaluation which calculates and compares alternative the costs of alternative interventions by analysing only their costs. This analysis method does not consider outcomes therefore cannot be classed as a full economic evaluation, but rather a partial economic evaluation. This method is usually used in cases where the outcomes of interventions are unknown.

Cost-minimisation analysis (CMA)

In cases where outcomes of alternative programmes are known to be identical their costs alone can be calculated and compared to identify the lowest cost intervention and incidentally the one which provides the greatest value.

Cost-effectiveness analysis (CEA)

A method of evaluation that compares the costs and consequences (outcomes/benefits) of two or more interventions. Strictly, CEA measures outcomes in natural units such as the number of drinks consumed or the number of alcohol-free days but it is commonly used to refer to any form of economic evaluation. In a CEA the costs and consequences of each intervention should be compared with each other incrementally (i.e. in ascending order according to cost). The incremental difference in costs and outcomes between each intervention is calculated to produce an incremental cost-effectiveness ratio (ICER). The results of different CEAs are comparable **only if** the same outcome measures are used in

each analysis, the method of measuring outcomes is the same and they all include at least one common comparison.

Cost-utility analysis (CUA)

A specific form of CEA where outcomes are measured in quality-adjusted life years (QALYs) which are a measure of health-related quality of life. CUAs have the advantage of being more readily comparable to other CUAs comparing very different interventions, However, QALYs are limited to measuring health-related quality of life and may not capture wider, social outcomes of an intervention.

Cost-consequence analysis (CCA)

CCA is a method of setting out all the relevant costs and outcomes of an intervention in a "balance sheet" format. It is commonly used in economic evaluations alongside another say a CUA or CEA as it can clearly present a wide scope of outcomes (health and non-health) alongside the costs that may be of interest to decision-makers. CCA alone, however is potentially less useful for making decisions as it does not aggregate the costs and outcomes to produce a value for the intervention.

Cost-benefit analysis (CBA)

CBA values the costs and consequences using a common unit, most commonly money. **All relevant** outcomes of an intervention (health and non-health) are given a monetary value and compared to the relevant costs of intervention. Results can be presented as a ratio of costs to benefits or as a "net benefit" where costs are subtracted from the value of the outcomes; a positive net benefit would indicate that the returns from the intervention are greater than its costs and therefore could be considered a good value investment. CBA can be used to compare the value of alternative interventions, including those in different sectors for example minimum pricing for alcohol compared to improved transport links, due to the use of a common metric (money). However, it can often be difficult to attach monetary values to all relevant outcomes. Most studies in healthcare that describe themselves as a CBA are actually just cost-analyses as they fail to measure or value benefits but focus on costs and savings in resources.

Social Return on Investment (SROI)

SROI can be described as an extension of CBA that also includes socio-economic and environmental outcomes. Similar to CBA, it provides a ratio of benefits to costs. However, this ratio is unique to the intervention being evaluated and is not typically compared with SROI ratios for any other intervention. SROI analyses focus on the outcomes and costs of the intervention under evaluation, therefore, incremental comparisons with alternative interventions are not commonly analysed.

Priority setting terms

Programme Budgeting Marginal Analysis (PBMA)

PBMA is a prioritisation tool that helps decision-makers maximally allocate scarce resources in order to meet the needs of a local population or any other specified objective. Programme budgeting involves analysing the current expenditure and activity of a set of chosen programmes to identify the existing use of resources in the area being examined. Marginal analysis investigates the added benefits and costs (or lost benefits and savings) from investment, disinvestment or service redesign of the listed programmes. The combination of action options are weighed against a set of criteria by a group of stakeholders who ultimately decide on a strategy of reallocation of resources within the set of programmes. This tool would not be considered a substitute to economic evaluation but rather complements it by providing a framework for decision makers to appraise the information on the effectiveness, cost-effectiveness and other factors of several programmes. PBMA can be used to decide which of a number of competing programmes should be implemented in order to represent the optimal allocation of resources.

Multi-Criteria Decision Analysis (MCDA)

MCDA is a method of combining both health economic and non-health economic outcomes of a programme. Similar to PBMA, it is a prioritisation tool that utilises existing evidence on cost-effectiveness alongside other factors important to decision-makers such as budget impact and burden of disease, in order to systematically prioritise different interventions.

Appendix H. Qualitative study consent form

Prior to each interview, written consent for participation was obtained from each interviewee. Each interviewee was given a unique study identification number which would be used to refer to interviewees anonymously for the remainder of the study. A copy of the consent form presented to study participants is displayed below.

Research study on decision-making in public health

Consent form

Participant Identification Number:

	Please initial box
I have read the information sheet for the above study, have been given a copy to keep and have had the opportunity to ask questions.	☐
I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason.	☐
I agree to take part in an interview. I understand that the interview will be audio-recorded.	☐
I understand that anonymous extracts from the interview may be used in reporting the project findings.	☐
I agree to take part in the above study.	☐

Name of participant	Date	Signature
Name of person taking consent	Date	Signature

Appendix I. The contingent valuation survey

The full text of the CV survey is presented below. The survey was administered online with each section of the survey displayed in order. The format of the survey displayed below is taken from the template provided to the market research company, *ResearchNow,* who were responsible for programming and hosting the survey.

Introduction

You are being invited to participate in a research study to find out your views on an intervention (i.e. a service) to help young people in the UK who have risky drinking behaviour. This study is being done by researchers from Newcastle University.

This survey will take you approximately 15 minutes to complete.

Your participation in this study is entirely voluntary and you can stop the survey at any time, however please be advised that any data you have given up to that point will be kept. The survey is anonymous and we will not ask any personal information that could tell us who you are.

We believe there are no known risks associated with this research study. This study was approved by Newcastle University's Research Ethics Committee.

There are no right or wrong answers, we are just interested in what you think.

Background Information

<u>What is the problem?</u>

In the UK, it is suggested that young people under the age of 18 should not drink alcohol. Research has shown links between young people drinking alcohol and a number of health and social problems. Health problems in young people linked to alcohol use include: liver damage, alcohol poisoning, harm the development of the brain, mental health issues and trouble sleeping. Social problems in young people linked to alcohol use include: increased risk of accidents and injury, risky sexual behaviour, increased risk of suicide, criminal behaviour and poor performance at school.

Questionnaires about how much and how often young people drink can be used to see if someone drinks an amount of alcohol that is considered harmful. This can be called “risky drinking behaviour”. For example, a young person who drinks 3 units of alcohol (i.e. 3

shots of vodka or 1 pint of cider) 4 or more times in six months would be considered to have “risky drinking behaviour”. Research has shown that approximately 25% (¼) of ‘Year 10’ students (14-15 years old) in England have been identified as having risky drinking behaviour.

Why is this survey being carried out?

There are a number of ways that young people can be helped to reduce risky drinking behaviour. We are interested in your views about how we might help young people with this. Most of the time, Public Health help to provide services to young people to reduce their risky drinking. Public health services are generally paid for by taxation. The views of the public are important to us because the public funds public health services through the taxes that we all pay.

In this study, the researchers are looking at one way of measuring how important the public think a service is for helping young people who have risky drinking behaviour. The researchers also want to find out how valuable the service is to the public.

We want you to think about how much you would be willing to pay for Public Health to provide a described intervention. The amount you are willing to pay tells us how important the intervention is to you. The information will only be used for research purposes by researchers at Newcastle University. You will not be asked to pay anything at the end of the survey and your answers will not be used to add or increase taxes.

Before we ask you questions about the alcohol intervention, we will take you through a practice question.

Practice Question

Imagine you are on holiday abroad in quite a remote place and you get an eye infection which is very itchy and uncomfortable – it feels like there is grit in your eye all the time. If you don't take any medication for it, suppose it will last for 3 days. However, someone can arrange to have medicine delivered to you that will cure your eye infection within 24 hours, saving you 2 of the 3 days of discomfort from the eye infection. Think about what would be the MAXIMUM you would be willing to pay as a one-off payment to get the medicine delivered.

Would you be willing to pay something to get this medicine delivered which will reduce your eye infection from 3 days to 1 day? Remember the money you spend on medicine cannot be spent on other things during your holiday.

[If "yes" to payment selected]

You have said you would be willing to pay as a one-off payment for medication for your eye infection.

What is the maximum you would be willing to pay? In order to help you decide you will be shown different amounts of money. For each amount please decide if you "definitely WOULD pay", "definitely WOULD NOT pay" and "would MAYBE pay".

[Payment values offered in random order]

£1, £5, £10, £20, £30, £50, £100

From the list of amounts, you said the highest amount you "definitely WOULD pay" as a one-off payment is **[XX]** and the lowest amount you "definitely WOULD NOT pay" as a one-off payment is **[XX].**

What is the MAXIMUM value you would be willing to pay as a one-off payment for the medicine? It could be one of these amounts or something in between.

Maximum willingness to pay: £_____

[If "No" to payment selected]

You have said that you are not willing to pay anything for the medicine to reduce your eye infection from 3 days to 1 day.

Please select from the options below WHY you are not willing to pay:

- The infection only lasts for 3 days so it is not worth paying for the treatment
- The symptoms of the infection are not too bad, I could live with it
- I cannot afford to spend money on medicine
- I don't think I should have to pay for healthcare

Other (please state) _____

You have completed the practice question. We will now move on to the main survey questions.

Instructions

We will now ask you some questions about an alcohol intervention to help young people who have risky drinking behaviour. You will be shown information about the intervention. After that you will be asked questions about 3 different scenarios to do with the intervention. Each scenario is separate to the others so please treat each scenario separately when you answer the questions.

The Intervention

A brief alcohol intervention is carried out in a school setting with 'Year 10' students aged 14-15. The intervention involves an alcohol screening questionnaire and a **30-minute personalised interactive worksheet-based session** for students who are identified as having a risky drinking behaviour. The session contains structured and detailed feedback about the student's drinking behaviour and advice about the **health and social consequences** of continued risky alcohol consumption, such as: **weight gain, accidents, violence and impact on relationships.**

[Scenario description displayed – see Chapter Six for the scenarios]

Payment

Thinking about the intervention and outcomes that have just been shown to you, would you be willing to pay anything for the brief alcohol intervention to be provided to 'Year 10' students in schools in the UK? A payment would be made in the form of extra monthly taxation for one year which would be used to directly fund the intervention.

[If "Yes" to payment selected]

You said you would be willing to pay something through extra taxation that would be used to directly fund the intervention.

What is the maximum amount that you would be willing to pay every month for the next year? In order to help you decide you will be shown different amounts of money. For each amount please decide if you "definitely WOULD pay", "definitely WOULD NOT pay" and "would MAYBE pay".

When you are thinking about this, please think about what you would be prepared to pay, given your actual income and savings.

[Random card sort exercise begins – Payment values presented]

£0.50 per month (equivalent to £6 over one year)

£1 per month (equivalent to £12 over one year)

£1.50 per month (equivalent to £18 over one year)

£2 per month (equivalent to £24 over one year)

£3.50 per month (equivalent to £42 over one year)

£5 per month (equivalent to £60 over one year)

£7.50 per month (equivalent to £90 over one year)

£10 per month (equivalent to £120 over one year)

£12.50 per month (equivalent to £150 over one year)

£15 per month (equivalent to £180 over one year)

£25 per month (equivalent to £300 over one year)

£50 per month (equivalent to £600 over one year)

From the list of amounts, you said the highest amount you "definitely WOULD pay" as a one-off payment is **[XX]** and the lowest amount you "definitely WOULD NOT pay" as a one-off payment is **[XX].**

What is the MAXIMUM value you would be willing to pay as a one-off payment for the intervention? It could be one of these amounts or something in between.

Maximum willingness to pay: £_____ per month

[If "No" to payment selected]

You have said that you are not willing to pay anything as a one-off payment to directly fund the brief alcohol intervention.

Please select from the options below WHY you are not willing to pay:

- Other interventions are more valuable
- I am not concerned about the issue of risky drinking in young people
- I think the intervention is valuable but I cannot afford it
- Parents/guardians of 'Year 10' students should pay for it
- Other (please state) _____

[Following completion of payment task]

Please state below **why** you have said that you are willing to pay/not willing to pay for the brief alcohol intervention with outcomes described in this scenario.

I.1 Alternative wording of open-ended WTP questions

The wording displayed to respondents differed depending on the placement of payment values into the “definitely WOULD pay”, “definitely WOULD NOT pay” and “would MAYBE pay” boxes during the random card sort. The text displayed to respondents who did not place values in all three boxes is displayed below.

[If no value is placed in the “definitely WOULD NOT pay” box]

From the list of amounts, you said the highest amount you “definitely WOULD pay” per month is **[XX].**

What is the MAXIMUM amount you would be willing to pay as extra monthly taxation for the intervention? It could be this amount or something higher than this.

Maximum willingness to pay: £_____ per month

[If no value is placed in the “definitely WOULD pay” box]

From the list of amounts, you said the highest amount you “<u>maybe</u> WOULD pay” per month is **[XX]** and the lowest amount you “definitely WOULD NOT pay” per month is **[XX].**

What is the MAXIMUM value you would be willing to pay as extra monthly taxation for the intervention? It could be one of these amounts or something in between.

Maximum willingness to pay: £_____ per month

I.2 **Demographic questions asked at the beginning of the survey**

Prior to the survey taking place, demographic questions were asked of respondents to i) ensure eligibility for the survey and ensure a representative sample of the UK population was obtained, and ii) to provide respondent characteristics for the examination of WTP predictors during the analysis of the CV survey data (Chapter Seven). The demographic questions and multiple choice responses are presented below.

Which of the following describes how you think of yourself?

- Male
- Female
- Prefer not to say

How old are you?

- Under 18
- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65+

Which region do you live in?

- East Anglia
- East Midlands
- London
- North East
- North West
- Scotland
- South East
- South West
- Wales
- West Midlands
- Yorkshire & Humberside
- Northern Ireland
- I do not live in the UK

What is your marital status?

- Married/Cohabitating
- Single
- Divorced/Widowed
- Prefer not to say

Which of the following best describes your employment status? (tick one box)

- Employed Full-Time
- Employed Part-Time
- Self-Employed
- Unemployed
- Retired
- Full-time student
- Part-time student
- Other (please specify): _____

What is your occupation? (please specify):

_____

What is your highest educational qualification?

- Degree or equivalent / NVQ level 4 or 5 (or a higher level qualification)
- Higher education below a degree
- GCE A-Level or AS-Level / NVQ level 3
- GCSE grade A*-C / GCE O-Level / NVQ level 2
- GCSE grade D-G / CSE / NVQ level 1
- Foreign qualifications / other (please specify):

  _____
- No formal qualifications

*Which of the following bands does your **annual** household income from all sources, **before tax**, fit into?*

- Less than £11,850
- £11,850 - £19,999
- £20,000 - £29,999
- £30,000 - £39,999

- £40,000 - £49,999
- £50,000 - £69,999
- £70,000 - £100,000
- Above £100,000
- Prefer not to say
- Unknown

Are you the parent / guardian / grandparent of a child under the age of 16?

- Yes
- No

How often do you drink alcohol?

- Never
- Less than once a month
- 1-2 times a month
- 1-2 times a week
- 3-5 times a week
- Every day
- Prefer not to say

Appendix J. Histograms of WTP data

The choice of model used for the regression analysis of the WTP data reported in Chapter Seven was based largely on the distribution of the WTP data. The histograms of the WTP data for each of the CV study scenarios are displayed below. The histograms evidence the skewed distribution of the data, with a spike at £0, which influenced the choice to use a two-part model for the base-case analysis in Chapter Seven.

Figure J.1 Histogram of WTP values (in £) for Scenario 1

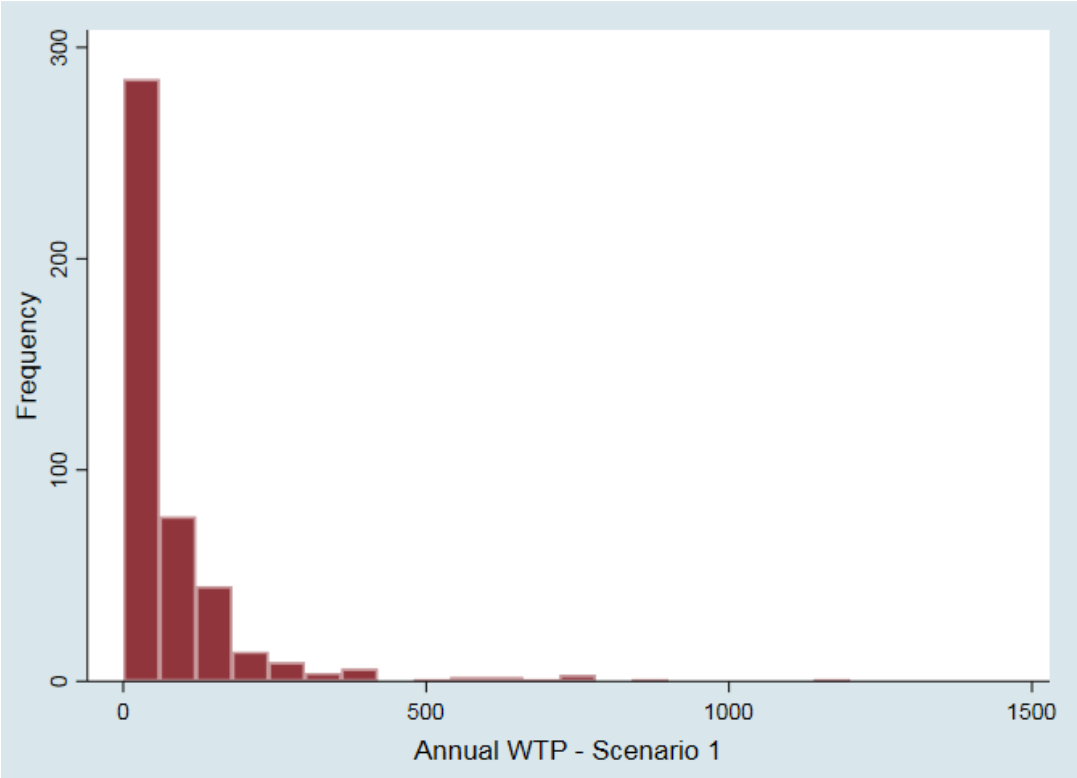


Figure J.2 Histogram of WTP values (in £) for Scenario 2

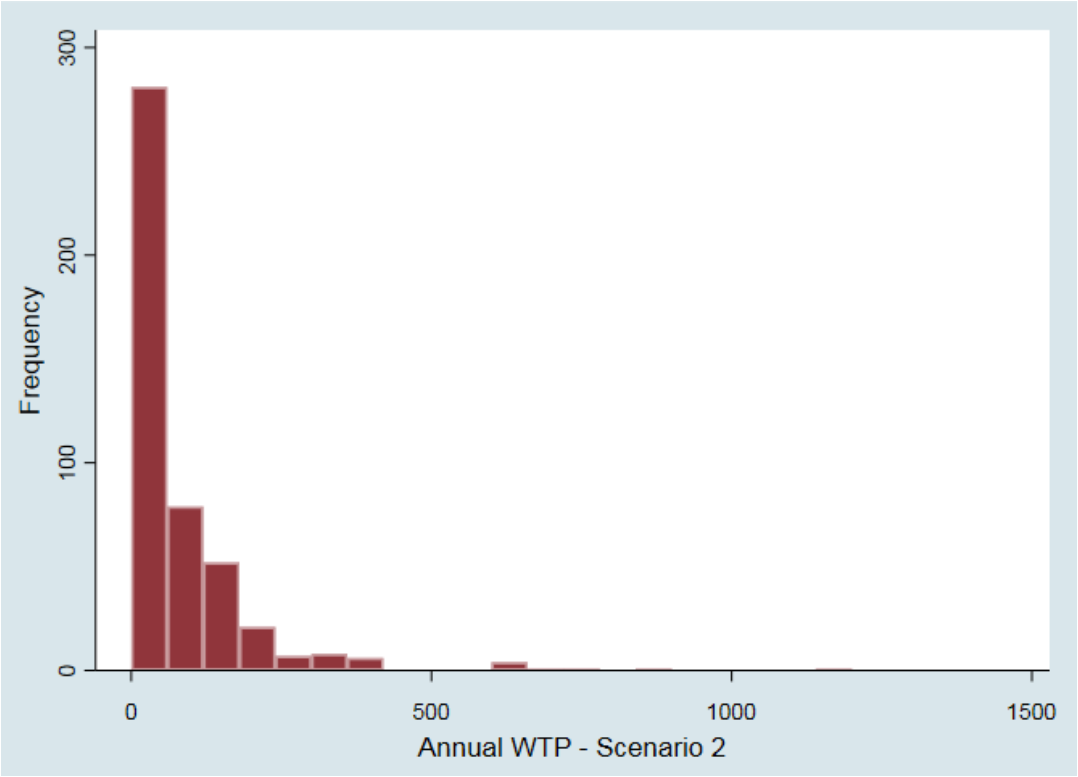
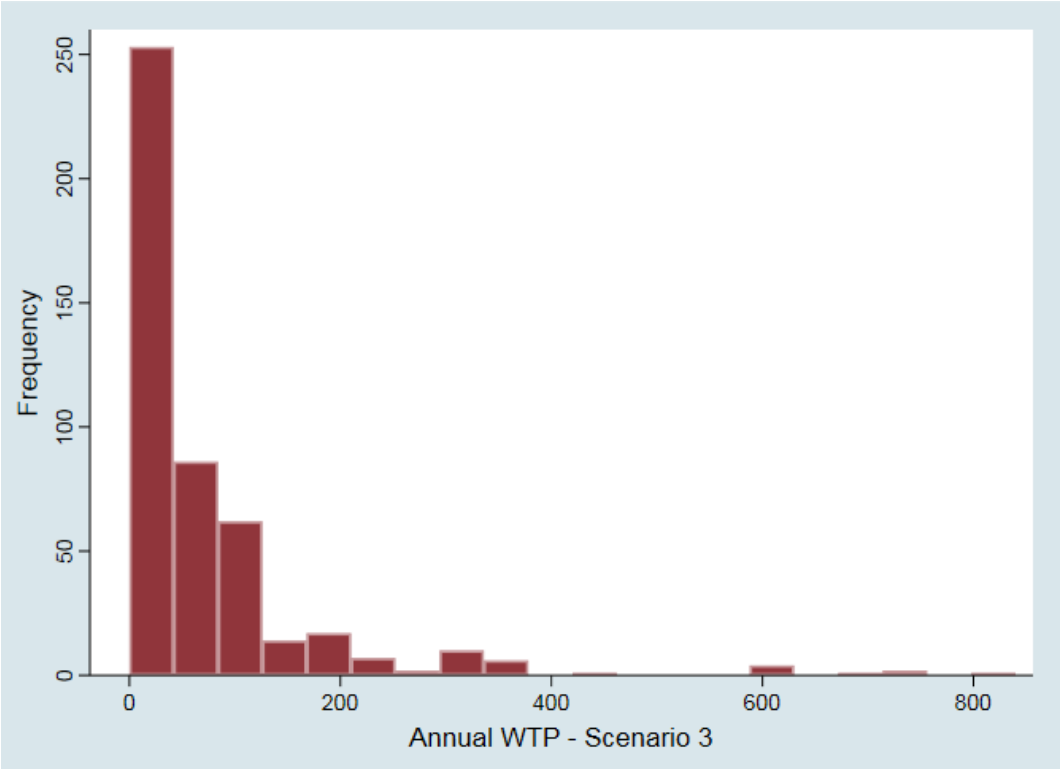


Figure J.3 Histogram of WTP values (in £) for Scenario 3



Appendix K. Regression outputs for sensitivity analyses of WTP predictors

Sensitivity analysis was conducted on the regression analysis using two alternative estimators, which are commonly used for the analysis of WTP data. The output from a Tobit (Table K.1) and log OLS (Table K.2) regression, respectively, with the value of WTP (in £) as the dependent variable are displayed below. The results for all three scenarios are displayed. Statistically significant predictors of the value of WTP are highlighted in bold text. Asterisks denote the level of statistical significance.

Table K.1. Tobit model results - dependent variable WTP, £

Variable[A]	Scenario 1		Scenario 2		Scenario 3	
	Coefficient (S.E.)	P>|t|	Coefficient (S.E.)	P>|t|	Coefficient (S.E.)	P>|t|
EDUCATION						
No formal qualifications	37.87 (57.57)	0.511	3.04 (54.32)	0.955	7.39 (46.70)	0.874
Higher education	3.75 (21.61)	0.862	9.84 (19.20)	0.608	14.90 (16.94)	0.380
NON_ENGLISH						
Non-English UK nations	-14.64 (27.79)	0.599	-36.65 (24.67)	0.138	-20.44 (21.26)	0.337
AGE						
Under 35	38.86 (24.92)	0.120	30.75 (21.84)	0.160	**33.61** (19.69)	**0.089***
Over 65	-27.28 (30.05)	0.364	-4.20 (26.61)	0.875	-18.69 (23.56)	0.428
MALE						
Male	**44.43** (20.92)	**0.034****	-6.30 (18.57)	0.735	8.87 (16.59)	0.593
INCOME						
Less than £20,000	-32.99 (27.13)	0.225	-33.08 (23.87)	0.167	-7.51 (21.24)	0.724
Over £40,000	41.15 (25.32)	0.105	**68.34** (22.52)	**0.003*****	**64.23** (20.01)	**0.001*****
PARENT_GUARDIAN						
Parent	15.76 (23.73)	0.507	-5.73 (21.48)	0.790	-17.09 (18.91)	0.367
MARRIED						
Married/cohabiting	40.78 (26.06)	0.118	**52.46** (22.78)	**0.022****	**52.43** (20.43)	**0.011****
NONDRINK						
No alcohol	1.08 (21.62)	0.960	-12.24 (19.15)	0.523	14.11 (17.16)	0.411
CONSTANT	-61.93 (33.59)	0.066	-26.88 (29.19)	0.358	-41.84 (27.05)	0.123
Observations	414		422		426	

[A] Base factors excluded from regression: (education) school level qualifications, (location) England, (age) 35-65, (gender) female, (income) £20,000-£40,000, (parent/guardian) non-parent, (marriage status) single/widowed, (drinking frequency) alcohol consumer

P>|t| = significance level of coefficient

*=p<0.1, **=p<0.05, ***=p<0.01

Table K.2. Log OLS regression model results - dependent variable log WTP, £

Variable[A]	Scenario 1		Scenario 2		Scenario 3	
	Coefficient (S.E.)	P>|t|	Coefficient (S.E.)	P>|t|	Coefficient (S.E.)	P>|t|
EDUCATION						
No formal qualifications	0.428 (0.361)	0.236	0.060 (0.380)	0.874	0.064 (0.345)	0.852
Higher education	-0.158 (0.133)	0.236	-0.031 (0.128)	0.808	-0.116 (0.119)	0.332
NON_ENGLISH						
Non-English UK nations	-0.232 (0.167)	0.165	-0.088 (0.165)	0.595	-0.135 (0.149)	0.366
AGE						
Under 35	**0.376** (0.159)	**0.019****	0.053 (0.151)	0.724	0.178 (0.143)	0.215
Over 65	-0.164 (0.181)	0.367	**-0.329** (0.169)	**0.053***	-0.163 (0.162)	0.315
MALE						
Male	**0.222** (0.128)	**0.083***	-0.011 (0.125)	0.932	0.124 (0.118)	0.295
INCOME						
Less than £20,000	-0.073 (0.173)	0.673	-0.015 (0.167)	0.927	0.124 (0.158)	0.433
Over £40,000	-0.031 (0.153)	0.840	0.085 (0.145)	0.558	0.169 (0.137)	0.221
PARENT_GUARDIAN						
Parent	0.087 (0.143)	0.542	0.206 (0.143)	0.152	0.165 (0.133)	0.214
MARRIED						
Married/cohabiting	0.128 (0.165)	0.438	-0.097 (0.159)	0.544	-0.092 (0.150)	0.541
NONDRINK	-0.033	0.806	-0.192	0.142	-0.092	0.452
No alcohol	(0.135)		(0.130)		(0.122)	
CONSTANT	4.166 (0.214)	0.000	4.435 (0.198)	0.000	4.240 (0.196)	0.000
Observations	231		248		262	
Adjusted R^2	0.021		0.009		0.007	

[A] Base factors excluded from regression: (education) school level qualifications, (location) England, (age) 35-65, (gender) female, (income) £20,000-£40,000, (parent/guardian) non-parent, (marriage status) single/widowed, (drinking frequency) alcohol consumer

P>|t| = significance level of coefficient

*=p<0.1, **=p<0.05

Appendix L. Cost-benefit planes from validation analysis excluding outliers

Figure L.1 and L.2 display the cost-benefit planes generated using the bootstrap estimated from the sensitivity analysis examining truncated costs for the CBA validation analysis in Chapter Eight. Truncating costs at the 95th percentile removed the extreme outlier values of cost difference between control and intervention, which are visible in the cost-benefit planes for the base-case analyses in Chapter Eight (Figure 8.3 and Figure 8.4). Figure L.1 and Figure L.2 demonstrate that in the absence of the extreme outliers, the cost-benefit planes of the validation analysis are largely comparable with those of the primary analysis (see Figure 8.1 and Figure 8.2 in Chapter Eight).

Figure L.1 Cost-benefit plane of national level sensitivity analysis using bootstrap estimates of cost-difference truncated at the 95th percentile

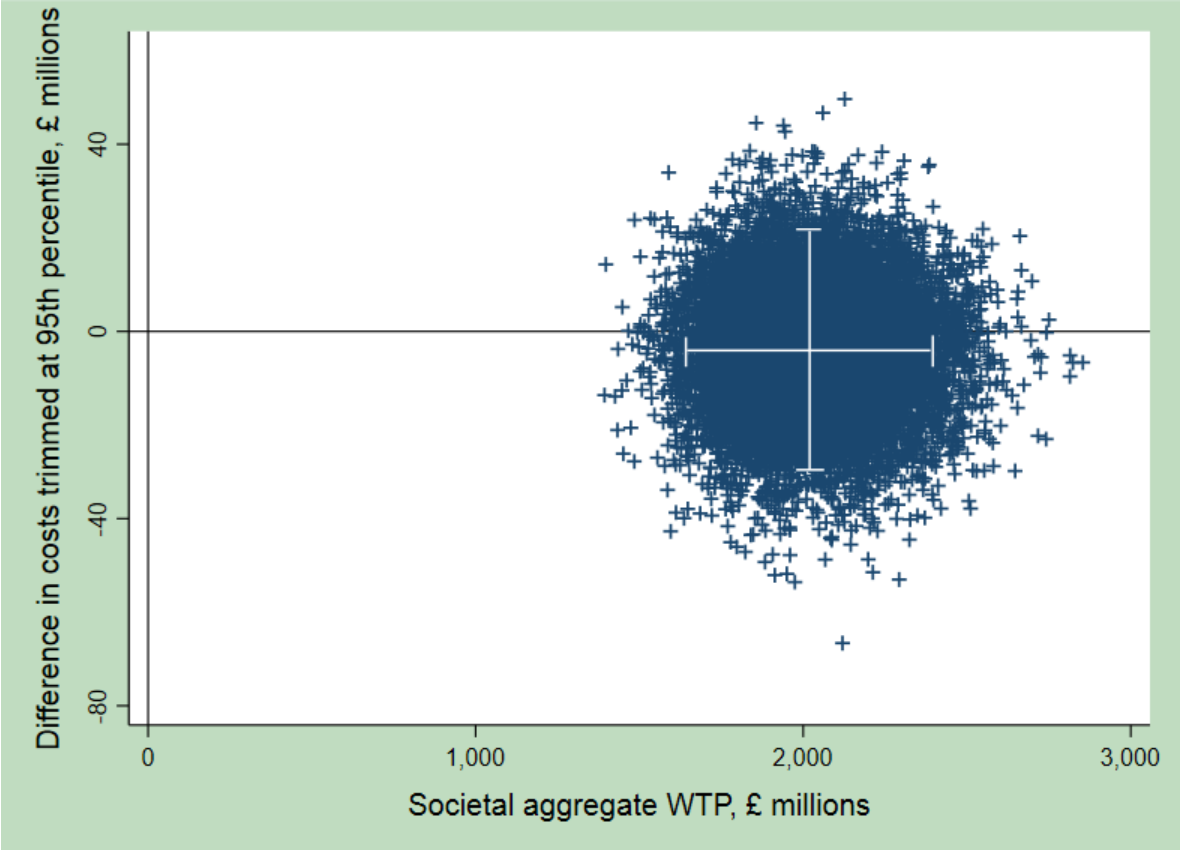
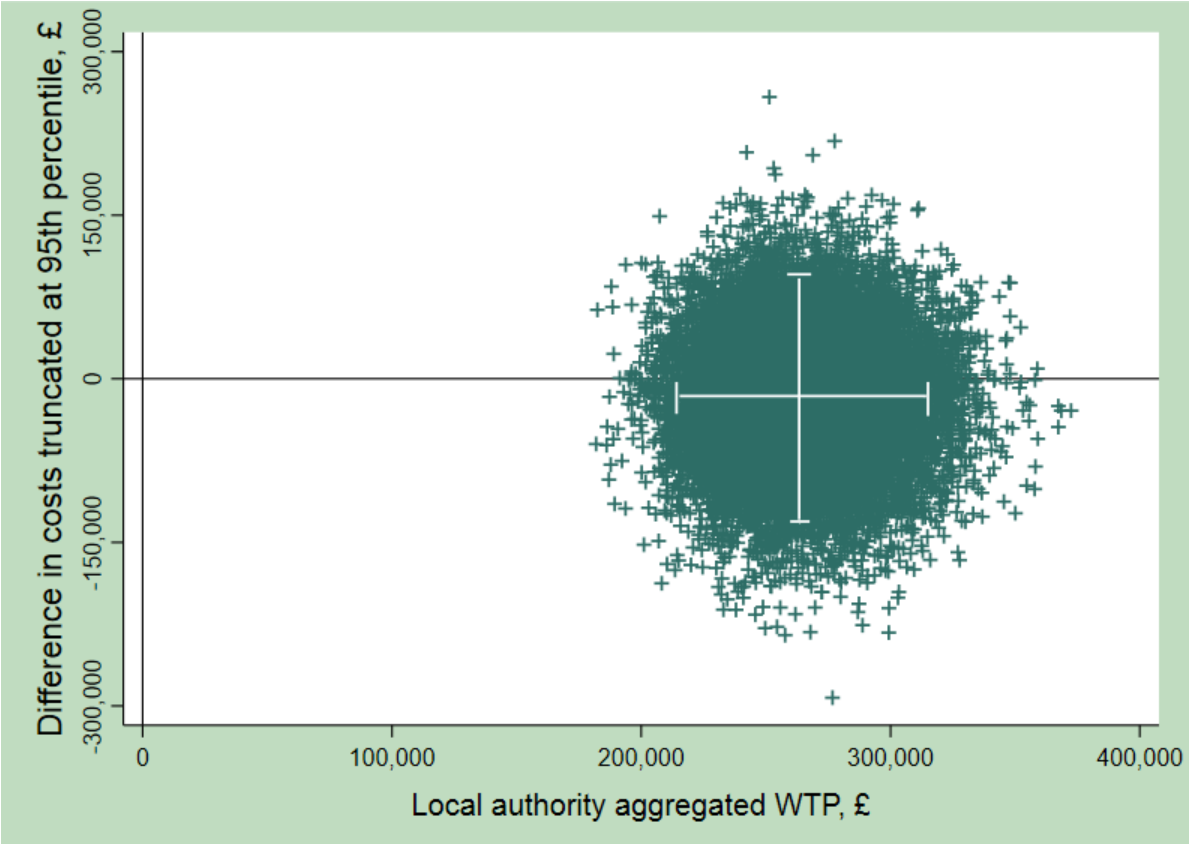


Figure L.2 Cost-benefit plane of local authority level sensitivity analysis using bootstrap estimates of cost-difference truncated at the 95th percentile



Appendix M. SROI validation analysis

In order to conduct a validation analysis of the SROI using adjusted values, all resource use was combined into a single outcome variable. Unlike the CBA, the SROI examines both resource and QALY outcomes. In clinical trial data a common method of adjusting costs and QALY outcomes is to use a bivariate regression which allows for different covariates to influence the estimation of costs and effects (348). The benefits of using this approach have been argued by Willan *et al.,* 2004 (394) who promote the use of 'seemingly unrelated regression' (SUREG) estimation, which accounts for the correlation between the errors across equations for each individual in a sample to improve the efficiency of estimation (395). This is particularly relevant in the case of cost and effect estimates in clinical trials where each outcome is elicited from the same trial participant, thus the errors for both cost and QALY estimates would be expected to be correlated.

A SUREG was employed to estimate the mean financial proxy values for both the combined resource outcomes and QALYs. The SUREG estimates multiple linear regressions simultaneously, each with a different dependent variable (i.e. resources and QALYs) and a different selection of covariates, whilst accounting for the correlated errors for each given individual in the sample (395). The selection of covariates necessarily differs for each regression when using seemingly unrelated estimation, otherwise estimation reduces to ordinary least squares, negating the benefit of using the SUREG model (395). Both QALYs and total resources were adjusted by trial arm, gender, location, baseline alcohol use disorder identification test (AUDIT) score and a dummy variable indicating that 12-month follow-up was either 30 days earlier or later than expected. Total resources were additionally adjusted for the value of resources reported at baseline, baseline smoking status and race. QALYs were additionally adjusted for baseline EQ-5D-3L score and baseline score for emotional wellbeing (WEMWBS).

The mean financial values of the impact and deadweight for both total resources and QALYs were obtained from the coefficients reported by the SUREG. Nonparametric bootstrapping with 5000 repetitions of the regression results was used to estimate confidence intervals. The same process described in Chapters Seven (footnote 11) and Eight (footnote 21) was used to determine the appropriate number of bootstrap repetitions required to generate robust confidence intervals.

The results of the adjusted analysis are reported below. Table M.1 outlines the calculation of the net financial reduction in resource use. The net financial reduction of resources is positive, indicating a saving in total resources from implementation of the ASBI. This outcome reflects the result of the unadjusted analysis reported in Chapter Eight. However, the 95% confidence interval around the net financial reduction crosses zero, indicating substantial uncertainty over the outcome. There is a considerable chance that the ASBI could actually result in an increase in financial value of resource use. This uncertainty reflects the findings of the CBA (see Chapter Eight, sections 8.2.4 and 8.2.5).

Table M.1. Calculation of net financial value of reduced combined resource use outcomes

	Impact			Deadweight			
	Mean financial proxy value per student	**N**	**Total**	**Mean financial proxy value per student**	**N**	**Total**	**Net financial reduction in resource use from intervention, (95% CI[a], £)**
All resource use combined	£1,906	210	£400,312	£4,365	210	£916,752	£516,440 (-852,875 to 2,156,147)

[a] 95% confidence interval (CI) calculated using 5000 bootstrap repetitions
N = number of students

Table M.2 outlines the calculation of the net financial value of QALY gain from the intervention using the results from the SUREG. As in the primary analysis using unadjusted values reported in Chapter Eight the net financial QALY gain is negative, however, the 95% confidence intervals again provide evidence of uncertainty around this outcome. The confidence interval crosses zero, which suggests the net financial QALY gain could be positive. The point estimate reported in Table M.2 (approximately -£63,000), however, is of a similar magnitude to that reported in the primary analysis (approximately -£50,000).

Table M.2. Calculation of financial value of health improvements

	Impact			Deadweight			
Health improve-ment	**Mean financial value of QALY gain per student**	**N**	**Total**	**Mean financial value of QALY gain per student**	**N**	**Total**	**Net financial value of health improvement (95% CI[a], £)**
QALYs	£21,455	210	£4,505,654	£21,759	210	£4,569,375	-£63,721 (-265,549 to 138,108)

[a] 95% confidence interval (CI) calculated using 5000 bootstrap repetitions
N = number of students

There is no change to the investment value compared to the primary analysis, therefore, the same value of £4,666 is used to calculate the SROI ratio for the adjusted analysis. Equation M.1 displays the SROI calculation using the total adjusted impact calculated from summing the net financial reduction in resources and net financial value of health improvement (£516,440 + -£63,721 = £452,719). The resulting ratio is £97:1, which is larger than the ratio obtained from the primary analysis. However, given the uncertainty surrounding the health and resource impact values, this outcome should be considered with caution. The imprecision around the SROI ratio can be examined using the confidence intervals estimated for each component of the investment value. Using the lower bound of the 95% confidence intervals from Table M.1 and Table M.2 (i.e. -£852,875 + -£265,549), the SROI ratio would be approximately -£196:1, whilst using the upper bound (i.e. £2,156,147 + £138,108) the ratio would be approximately £491:1. The SROI ratio, therefore, ranges from -£196:1 to £491:1.

$$SROI = \frac{Net\ present\ impact\ value}{Net\ present\ investment\ value} = \frac{£452,719}{£4,666} = £97 \qquad \textbf{(M.1)}$$

The sensitivity analyses conducted in the primary analysis (see Chapter Eight, section 8.3.2) were replicated with the adjusted data. Table M.3 displays the analysis of alternative financial values of a QALY. As in the primary analysis, the sensitivity analyses report an increase in the SROI ratio due to a reduction in the negative net financial value of health improvement.

Table M.3. Sensitivity analysis for alternative financial values of a QALY

	Impact		**Deadweight**				
	Mean financial value of QALY gain per student	**Total for 210 students**	**Mean financial value of QALY gain per student**	**Total for 210 students**	**Net financial value of health improvement**	**Net present impact value**	**SROI ratio, £**
£20,000 (NICE)	£7,152	£1,501,885	£7,253	£1,523,125	-£21,240	£90,510	**106:1**
£12,936 (Claxton *et al.*)	£4,626	£971,419	£4,691	£985,157	-£13,738	£96,444	**108:1**

Table M.4 displays the re-analysis of the SROI excluding the financial value associated with reduced school absenteeism. As in the primary analysis, the net financial value of resources

saved from implementing the ASBI is negative, indicating that when the financial savings from reduced school absenteeism are excluded from the impact value the intervention results in greater incremental spending on resources. However, as in the base-case analysis of resource savings, the 95% confidence interval crosses zero, thus there is considerable uncertainty over this outcome. Table M.4 reports a SROI ratio of -£16:1 for the ASBI when financial savings from reduced school absenteeism are removed from consideration of the impact on resource outcomes i.e. a loss of £16 for every £1 invested.

Table M.4. Sensitivity analysis excluding missed school days from impact

Total impact value	**Total deadweight value**	**Net financial value of reduction in resource use (96% CI, £)**	**Net financial value of health improve-ment**	**Total impact**	**SROI ratio, £**
£139,654	£127,410	-£12,244 (-52,157 to 28,418)	-£63,721	-£75,965	**-16:1**

Overall, the results of the SROI using financial proxy values that have been adjusted for baseline variation amongst participants in the SIPS Jr HIGH trial reflect the results of the SROI evaluation using unadjusted values reported in Chapter Eight, section 8.3.3. However, the adjusted results indicate significant uncertainty around the impact parameters, a finding that was also identified in the CBA for intervention costs and in the within trial CUA conducted for the SIPS Jr HIGH trial (59).

The results reported in this appendix suggest that the primary analysis presented in Chapter Eight is a conservative evaluation. However, given the uncertainty surrounding the parameter values reported in the adjusted analysis, a conservative analysis would be recommended so as not to risk over-estimating the SROI from the ASBI.

Appendix N. The case study information presented to PHDMs

The case study: Alcohol screening and brief intervention in schools

A randomised controlled trial (RCT), SIPS Jr HIGH, was recently conducted to assess the effectiveness and cost-effectiveness of alcohol screening and brief interventions provided in schools for young people compared to a control group. The intervention was aimed at Year 10 students (aged 14-15) who display risky drinking behaviour, identified as those who screened positively on a single alcohol screening question. The intervention covered four locations in the UK (North-East, North-West, London, and Kent). Students who screened positive to displaying risky drinking behaviour, who were otherwise eligible for the trial and consented to taking part were randomised into either the control group or intervention group.

The intervention involved an alcohol screening questionnaire and a 30-minute personalised interactive worksheet-based session for students who were identified as having risky drinking behaviour. The session contains structured and detailed feedback about the student's drinking behaviour and advice about the health and social consequences of continued risky alcohol consumption, such as: weight gain, accidents, violence and impact on relationships.

The control group received a healthy lifestyle information leaflet which contained general advice on healthy living but contained to no information about alcohol consumption. The control group were not given feedback on their alcohol screening result.

Data was collected from trial participants at the start of the trial and at the end of the trial (after 12 months). The data collected as part of the trial has been used to conduct the economic evaluations you will be presented with today.

Appendix O. Cost-utility analysis report presented to PHDMs

Each of the economic evaluation evidence reports presented to the public health decision-makers at the workshop are reported in the following four appendices[36].

Methods

A cost-utility analysis (CUA) was carried out to estimate and compare the costs and effectiveness of the brief alcohol intervention against usual practice. The CUA calculated the additional cost per child for those who were in the intervention group compared to those who were in the control group of the trial and compared those to the outcomes of the trial which, for the CUA, are Quality Adjusted Life Years (QALYs) which capture the effect on both length of life and quality of life. The CUA outcome is incremental (additional) costs per QALY, which is commonly reported as an incremental cost-effectiveness ratio (ICER), which provides an estimate of the additional cost associated with achieving one more QALY.

The perspective of the analyses was the UK public sector (NHS, educational, social, and criminal services). The time horizon for the analysis was 12-months, therefore, only costs and outcomes of the trial after 12-months are considered in this analysis.

Costs

Costs were calculated based on the cost of delivering the brief intervention and the costs (and cost savings) in terms of UK public sector resources. The resources assessed were: GP visits, social worker visits, Accident & Emergency visits, non-A&E hospital visits, school nurse visits, absence from school and arrests.

Outcomes

QALYs were calculated by estimating quality of life using the EQ-5D-3L which is a questionnaire used commonly in economic evaluations to assess generic health-related quality of life.

[36] Formatting of the reports has been minimally adapted to maintain consistency with the formatting of the thesis.

Results

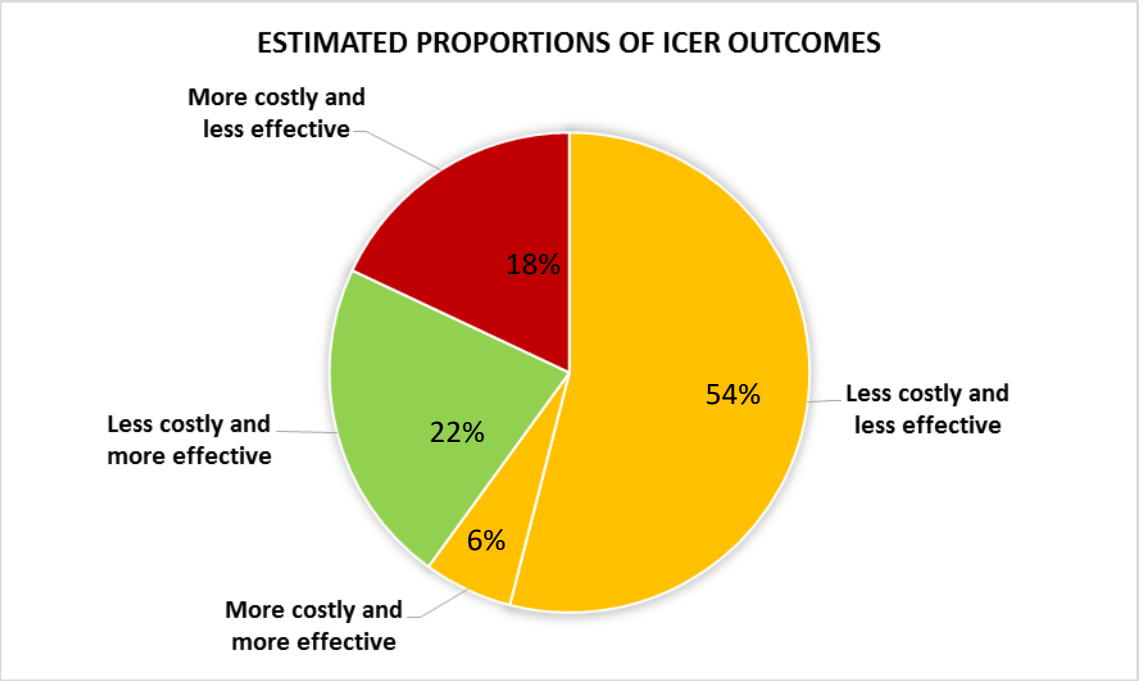
Table 1 Incremental cost effectiveness

Option	Average Total Cost[a]	Average Total QALYs	Comparison	Incremental Cost (95% CI)[b]	Incremental QALYs (95% CI)	Cost per QALY gained
Control	£9077	0.367	-	-	-	-
Intervention	£6212	0.363	Intervention vs. Control	-£2865 (-£1,272 - £2707)	-0.004 (-0.019 - 0.011)	£723,048

[a]Costs over a period of 12-months. Costs and QALYS adjusted for baseline data and participant characteristics.
[b]95% CI stands for 95% confidence interval which is an estimated range of intervals which contains the true average value 95% of the time. I.e. one can be 95% confident that the true average value lies between the range of values stated in the 95% CI.

Statistical analyses were conducted to calculate costs and QALYs, adjusted for baseline data and participant characteristics which might affect outcomes. Table 1 shows that the intervention group both had lower costs and fewer QALYs at the end of the intervention period on average. In other words, the intervention was less effective but also less costly than usual care. The ICER associated with the more effective and costlier control is **£723,048 per QALY**. The National Institute for Health and Care Excellence (NICE) typically considers interventions which have an ICER of £20,000 per QALY or less acceptable to provide in the UK. The ICER for the control (usual care) compared to intervention (brief alcohol intervention) is much greater than this acceptable threshold, suggesting that, on average, the control is not cost-effective compared to the intervention.

Figure 1 Chart showing the proportion of ICER outcomes in statistical analysis results



A statistical method was used to the estimate the possible values of average incremental costs and QALYs that might be found if we had a much larger sample. This provides an idea of how certain one could be that the results from the analysis of the trial reflect the costs and effects of the intervention if data were re-estimated from a much larger sample (which would give more certain results). After completing the additional statistical analysis, it was found that **54% of the estimated values agreed with our analysis of the trial data** that intervention is less costly and less effective than usual care. Figure 1 shows the outcome of the statistical analysis. The green segment depicts outcomes where the intervention is definitively cost-effective, the yellow segment depicts outcomes which may be cost-effective if either the extra effect is worth the additional cost or if the cost-saving is worth the loss of some effectiveness, and the red segment shows outcomes that are not cost-effective.

The statistical analysis also considered the probability that the intervention is cost-effective. The analysis suggests there was a 76% probability that the brief alcohol intervention is cost-saving compared to usual practice. At values considered by NICE to represent cost-effectiveness of £20,000 - £30,000 for an additional QALY gained, the probability that the intervention could be cost-effective compared to usual practice was 73%. These results were driven by the likelihood that the brief alcohol intervention is cost-saving as the average difference in QALYs is so small.

Sensitivity Analysis

The outcomes of CUA are crucially dependent on the estimated costs of the intervention. In order to assess the certainty of the results of the CUA, sensitivity analysis is conducted to assess how sensitive the results are to changes in certain costs. Two sensitivity analysis were carried out: 1) trimming 5% of the data from the top to remove extremely large values, 2) removing the missed school costs as these are based on future lifetime income lost and can be considered a type of indirect cost, as opposed to the direct realisable costs of service use from being arrested, going to hospital, etc.

Table 2 shows the results of both sensitivity analyses. Trimming the data did not have much effect on the ICER, whereas removing the costs associated with missed school days reduced the ICER by ~£400,000/QALY to £334,537/QALY. Removing costs associated with missing school reduces the mean net savings of the intervention by about half, however the original conclusion that the intervention is less costly and less effective than control is not changed. The ICER for control compared to intervention is still far higher than what would be considered cost-effective by NICE thresholds.

Table 2 Sensitivity analysis results

Category	**Original analyses**	**Sensitivity analysis: Extreme values**	**Sensitivity analysis: Missed school days**
Incremental costs	-£2,865 (-£11,272 - £2707)	-£2,911 (-£9,900 - £4,077)	-£1,324 (-£5,277 - £1,727)
Incremental QALYs	-0.004 (-0.019, 0.011)	-0.004 (-0.019, 0.011)	-0.004 (-0.019, 0.011)
Cost per QALY gained	£723,048	£734,804	£334,537

Appendix P. Cost-consequence analysis report presented to PHDMs

Methods

The cost-consequence analysis (CCA) presents the cost data and outcomes data of the intervention in the form of a balance sheet. Cost and effects are presented in a disaggregated manner to enable inclusion of all potentially relevant costs and outcomes from the intervention. No attempt is made to aggregate the data as is done in other forms of economic evaluation, the intention is to provide a spectrum of relevant information.

The costs reported in the CCA are taken from the within-trial cost-utility analysis. Outcomes include primary and secondary outcomes from the trial. The time horizon considered for the analysis is 12 months.

Results

Table 1 displays primary and secondary outcomes from the trial alongside the estimated costs of the intervention and control groups. None of the outcomes are statistically significant, as can be seen by the confidence intervals crossing zero, indicating it is unclear whether there is any meaningful difference in effect between the two arms at the end of the trial.

Please turn over to view Table 1.

Table 1 Cost-consequence balance sheet

	Intervention[a] Average	Control[a] Average	Adjusted difference[b] (95% CI)[d]
Costs			
Intervention delivery	£22.20	0	£22.20 (£21.80 - £22.60)
Total resource costs – including: GP visits A&E attendances Hospital visits Social worker visits School nurse visits Arrests >5 days missing from school	£6212	£9077	-£2865 (-£11,272 - £2707)
Outcomes (at the end of the trial)			
QALY gains	0.367	0.363	-0.004 (-0.019 - 0.011)
	Median (Inter-quartile range)		
Alcohol intake (in units) measured by 28 day follow-back	7.3 (1.9-18.5)	7.7 (0-18.0)	0.8 (-2.5 - 4.0)
Drinks per drinking day in past 28 days	4.2 (1.5-7.8)	3.9 (0-7.6)	-0.5 (-1.6 - 0.6)
Mental wellbeing score measured on WEMWBS (score out of 70, higher indicates better mental wellbeing)	50 (43.1-55)	49 (41-55)	1.7 (-0.7 - 4.1)
AUDIT score (score >4 indicates hazardous in adolescents)	5 (3-8)	5 (2-8)	-0.1 (-1.0 - 0.8)
	Percentage %		
A-SAQ results: Proportion reducing drinking Proportion drinking same Proportion drinking more	 60 26.11 13.9	 59 28.2 12.8	N/A
Smoking behaviours[c]: Smoking more Smoking less Started smoking Smoking same	 11.1 13.9 12.8 20	 8.7 14.4 14.4 20	N/A
Energy drinks consumption[c]: Consuming more Consuming less Stopped consuming Consuming same amount	 5.6 30.6 20.6 15.6	 10.8 26.2 15.9 25.6	N/A

[a]Values reported for intervention and control are the raw values collected from the trial. They have not been adjusted for baseline values of participant characteristics (with the exception of the costs and QALY gains).

[b]Difference measured as intervention compared to control. Adjusted by baseline values and participant characteristics

[c]Compared to baseline

[d]95% CI stands for 95% confidence interval which is an estimated range of intervals which contains the true average value 95% of the time. I.e. one can be 95% confident that the true average value lies between the range of values stated in the 95% CI.

Appendix Q. Cost-benefit analysis report presented to PHDMs

Methods

The cost-benefit analysis compares the costs and benefits of the brief alcohol intervention in monetary terms. The total difference in costs between the intervention group and control group in the trial are subtracted from the monetarised benefits associated with the intervention group outcomes compared to control outcomes. The outcome presented is in the form of a net societal benefit (NSB).

Monetised benefit is derived from a willingness to pay (WTP) survey in which a representative sample of the UK population was provided with information on trial outcomes and asked to place a monetary value on the intervention. The monetary value in the WTP survey was asked in the form of extra monthly taxation (i.e. how much would survey respondents be willing to pay in extra monthly taxation for the intervention to be provided in schools across the UK). The outcomes valued in the WTP survey include: average difference in alcohol consumption, total number of arrests and total number of days missing from school between the intervention and control groups.

The costs included in the CBA are estimated as the costs associated with the difference in average resource use at 12-month follow-up between the intervention and control groups. The resources measured include: delivery of intervention, GP visits, social worker visits, Accident & Emergency visits, non-A&E hospital visits, school nurse visits and arrests.

The perspective of the analysis is the UK public sector and the time horizon considered for costs and outcomes is 12 months.

Results

Two sets of analyses are presented below: 1) at a UK societal level, and 2) at a Newcastle Local Authority level. The average costs and benefits reported in Table 1 are used for both analyses.

Table 1 Average cost and benefit

	Average (S.D.)[a]
Annual WTP	£65 (126)
Difference in costs per child (intervention compared to control)	£118 (84)

[a]S.D. stands for standard deviation which shows the spread of data in relation to the average value. Large values indicate a large spread in the data whereas small values show that the data points are close to the average value.

1. UK societal level

The total welfare on a UK societal level was estimated by aggregating the cost and benefit on a UK societal level. In order to do this, the mean cost difference (Table 1) was multiplied by the estimated number of Year 10 students with risky-drinking behaviour in the UK (23.5% of Year 10 students according to the SIPS Jr HIGH trial findings) and the annual WTP in extra taxes (Table 1) was multiplied by the number of UK taxpayers.

This analysis assumes that the intervention would be funded by all UK taxpayers at an average rate commensurate to the average value placed on the intervention from the WTP survey (this assumption does not necessitate that the value contributed by each individual tax payer is equal, only that the average across tax payers equates to the average WTP value identified in the WTP survey). The costs represent the costs to UK public sector from implementing the intervention across the UK. In the UK there are approximately 30 million taxpayers and approximately 165,000 Year 10 students with risky drinking behaviour (calculated as 23.5% of all Year 10 students in the UK).

Table 2 UK societal level of Net Societal Benefit

Category	Average (95% CI)[a]
Incremental costs	£19.50 million (-£8.05 million - £46.98 million)
Benefit	£2.01 billion (£1.65 billion - £2.36 billion)
Net Societal Benefit	1.98 billion (£1.975 billion - £1.989 billion)

[a]95% CI stands for 95% confidence interval which is an estimated range of intervals which contains the true average value 95% of the time. I.e. one can be 95% confident that the true average value lies between the range of values stated in the 95% CI.

Table 2 shows the result of the NSB which is large and positive, indicating that the benefits to society of the intervention outweigh the costs to society. The confidence interval for NSB does not cross zero indicating that we can be 95% confident that the true NSB lies within the range of £1.975 billion to £1.989 billion, therefore, it is very likely that the true NSB is positive.

A statistical method was used to estimate the possible values of the average costs and average benefits of the intervention. In the sample of estimated average costs and average benefits, **92.8% of the estimates had a monetized benefit greater than zero and showed the intervention to be more costly than the control**. The minority remainder of the estimates (7.2%) showed the intervention to be less costly than the intervention but still remained an average monetized benefit greater than zero.

On a UK societal level assuming all UK tax payers contributing towards funding the intervention, if the costs of the intervention remain unchanged, in order for societal benefit to be less than the costs the average WTP for the intervention would have to be less than **£1 per year** (compared to £65 identified in the WTP survey). Alternatively, assuming the intervention is valued at the current rate of £65 over one year, the average cost of the intervention compared to usual care would have to be over **£12,000 per risky-drinking Year 10 student in the UK** in order for the costs to outweigh the benefits.

2. Newcastle Local Authority level

The societal level results indicate that the intervention is net beneficial if it were funded by society and implemented across the UK. However, it may be more relevant to calculate the NSB aggregated on an individual local authority level. Analysis here is conducted for Newcastle local authority. It is assumed that the intervention will be funded via an addition to council tax payments for the highest bands of council tax (bands F-H) and the costs will be those associated with the number of risky drinking Year 10 students in Newcastle. The estimated proportion of risky drinking students in the North-East (27.4%) was higher than the SIPS Jr HIGH trial average used in the UK societal analysis.

The number of band F-H households in Newcastle is 4020 and the number of risky drinking Year 10 students is 725 (27.4% of all Year 10 students in Newcastle).

Table 3 Newcastle local authority NSB

Category	Average (95% CI)
Incremental costs	£85,442 (-£35,362 - £206,245)
Benefit	£261,930 (£214,846 - £309,015)
Net Societal Benefit	£177,000 (£174,532 - £179,468)

Table 3 shows that the average NSB in Newcastle LA is still positive but not as large as the UK societal level. The confidence interval again does not cross zero indicating that we can be 95% confident that the true NSB lies within the range of £174,532 to £179,468, therefore it is very likely that the true NSB is positive.

As in the previous analysis, the same statistical method is used to estimate the possible values of the average costs and average benefits of the intervention. Again, **all estimated values for benefit were greater than zero and the large majority (92.8%) estimated the intervention to be more costly than control.**

In the local authority setting where council tax from households in bands F-H fund the intervention, assuming costs remain unchanged, in order for societal benefit to be less than the costs the average WTP for the intervention would have to be less than **£21 per year**

(compared to £65 identified in the WTP survey). Alternatively, assuming the intervention is valued at the current rate of £65 over one year, the average cost associated with the intervention compared to usual care would have to be over **£361 per risky-drinking Year 10 student in the Newcastle** in order for the costs to outweigh the benefits.

Sensitivity analysis

The results of the CBA are sensitive to a number of factors: the WTP value, the cost value and the level in which these values are aggregated. Sensitivity analyses were conducted to assess the effect on NSB of aggregating the WTP so that only parents and guardians of Year 10 students pay for the intervention, trimming the top 5% of WTP values, and trimming the top 5% of costs in order to reduce the effect of extremely large values. The sensitivity analyses were conducted for both the UK societal and Newcastle LA levels.

Tables 4 and 5 below show the sensitivity results for the UK and local authority levels, respectively. In both cases the NSB is reduced significantly on the assumption that only parent and guardians of Year 10 students would fund the intervention. Even in this case where the proportion of individuals funding the intervention is fraction of the original analysis (~2% in the UK analysis), the net benefit is still positive. Overall, all sensitivity analyses demonstrate positive NSB with confidence intervals that do not cross, or get close to, zero. Therefore, it could be said with a large degree of confidence that the NSB is positive for this intervention, on whichever level it is considered.

Table 4 Sensitivity analysis for the UK societal level

Category	Parents and guardians pay	WTP values trimmed	Costs trimmed
Incremental costs (95% CI)	£19.5 million (-£8.1 million - £47.0 million)	£19.5 million (-£8.1 million - £47.0 million)	£1.18 million (-£14.8 million - £17.2 million)
Benefit (95% CI)	£45.8 million (£37.6 million - £54.5 million)	£1.319 billion (£1.164 billion -£1.494 billion)	£2.01 billion (£1.65 billion - £2.36 billion)
Net Societal Benefit (95% CI)	£26.3 million (£25.7 million - £26.8 million)	£1.298 billion (£1.295 billion - £1.301 billion)	£2.00 billion (£1.99 billion - £2.01 billion)

Table 5 Sensitivity analysis for the local authority level

Category	Parents and guardians pay	WTP values trimmed	Costs trimmed
Incremental costs (95% CI to nearest £1,000)	£85,442 (-£35,000 - £206,000)	£85,442 (-£35,000 - £206,000)	£5,216 (-£65,000 - £76,000)
Benefit (95% CI to nearest £1,000)	£172,404 (£143,000 - £205,000)	£172,169 (£150,000 - £194,000)	£261,930 (£215,000 - £309,000)
Net Societal Benefit (95% CI to nearest £1,000)	£86,774 (£84,000 - £89,000)	£86,773 (£84,000 - £89,000)	£256,023 (£254,000 - £258,000)

Appendix R. Social return on investment report presented to PHDMs

Methods

A social return on investment (SROI) was conducted to assess the value of the brief alcohol intervention compared to the investment required to implement it. The values of the investment and the return are taken from the SIPS Jr HIGH trial data.

The value of investment is equivalent to the intervention delivery costs which include materials for use during the alcohol screening and brief intervention (ASBI) interview, training costs for those implementing the ASBI with the year 10 students in the trial and the costs associated with the time taken to provide the intervention.

The return of the intervention includes: financial savings from reduced resource use associated with the intervention compared to the control (usual care) and a monetarised value of the associated health improvements from the intervention. The SROI considers a broad perspective of returns, including those attributed to the healthcare sector, social care sector, education sector, criminal justice sector and those to the students exhibiting risky drinking behaviour.

Impacts and investments to be included in the SROI calculation are estimated as the financial value of each component with any deadweight (the financial value of what would happen anyway without the intervention) and attribution (the value that can be attributed to factors other than the intervention) subtracted. As the values for this SROI are taken from a randomised controlled trial which is designed with the intention of eliminating external factors affecting the outcomes, it is assumed that attribution is zero in this analysis. Deadweight is calculated using values from the control group of the trial which should represent what happens in the absence of the intervention.

Results

Investment

Table 1 shows the calculation of the investment component of the SROI. The average unit price of each element contributing to the investment value is multiplied by the total number of Year 10 students who were in the intervention group (n=210).

Table 1 Calculation of investment value

	Financial proxy			**Deadweight**		
	Intervention group			Control group		
	Unit price	Students	Total	Unit price	Students	Total
Interview materials	£1.58	210	£331.80	£0	210	£0
Training costs for learning mentors and screening	£14.20	210	£2,982.00	£0	210	£0
Learning mentor time	£6.44	210	£1,352.40	£0	210	£0
Total			£4,666.20			£0
Total investment value*						£4,666

***value rounded to nearest £**

Impact

Calculating the impact of the intervention is divided into two parts. Firstly, calculating the financial value of health improvement and secondly the financial value of reduced service use (i.e. cost savings). The financial value of health improvement is estimated using a value of £60,000 per QALY gained, which is the value used by the Department of Health and is derived from the value society places on a life.

Table 2 Calculation of financial value of health improvements

	Financial proxy			**Deadweight**			
	Intervention			Control			Difference
	Value per student	Students	Total	Value per student	Students	Total	
Difference in QALYs at intervention follow-up	£21,720	210	£4,561,200	£21,960	210	£4,611,600	-£50,400
Total health improvement impact							-£50,400

Table 2 outlines the calculation of the financial value of health improvement. The value of the impact is negative due to the larger value of health improvement attributed to the control group compared to the intervention group.

Table 3 Calculation of financial value of reduced resource use

		Financial proxy			Deadweight			
Resource sector		Intervention			Control			Financial reduction in resources from intervention
		Value per student	Students	Total	Value per student	Students	Total	
Health services	GP visits	£98	210	£20,616	£125	210	£26,261	£5,645
	Secondary care	£293	210	£61,510	£227	210	£47,565	-£13,945
	School nurse visits	£83	210	£17,366	£54	210	£11,281	-£6,086
Social services	Social worker visits	£27	210	£5,678	£9	210	£1,805	-£3,874
School absenteeism	>5 days missed	£1,134	210	£238,105	£2,083	210	£437,533	£199,428
Local Authority	Arrests	£0	210	£26	£1	210	£214	£189
Total impact on reduction in services								£181,357

Table 3 outlines the calculation of the financial value of resource use. As the interest is in savings from implementing the intervention, the total impact is calculated as the costs of resource use associated with the intervention group subtracted from the control group. The total outcome is positive since of the costs of resource use are larger for the control group than the intervention group.

Calculating the SROI

The total value if the impacts is **£130,957.**

$$\text{SROI} = \frac{£130,957}{£4,666} = £28.07$$

Every £1 invested in the brief alcohol intervention generates ~£28 in social value

Sensitivity Analysis

The SROI outcome is heavily dependent on certain parameters and assumptions. Sensitivity analysis looks at how robust the outcome is to change in these assumptions and parameters.

Removing missed school days from analysis

The financial value associated with absenteeism from school is considered an indirect outcome dependent on long term outcomes in terms of wages and productivity. Therefore, to assess the effect on SROI of just the direct outcomes of the intervention a sensitivity analysis was conducted where the resource associated with absenteeism is removed from the intervention impact.

Table 4 Sensitivity analysis of removing school absenteeism

Impact without missed school days	
Resource use reduction	-£18,071
Health improvement	-£50,400
Total impact	-£68,471

Assuming the investment value is unchanged the SROI ratio is = **-£15:1**

Every £1 invested in the brief alcohol intervention results in a loss of societal value of ~£15

Using different values to estimate the financial value of health improvement

The financial value for health gain is determined by multiplying QALY gain by £60,000 per QALY. However, there are several alternative values of a QALY used in economic evaluation which could be used to value the health gains associated with the intervention. The National Institute of Health and Care Excellence (NICE) consider a value of £20,000 per QALY or less to be considered cost-effective and a recent estimation of the value of a QALY by Claxton *et al.* (2015) stands at £12,936 per QALY. A sensitivity analysis using both of these values was conducted.

Table 5 Sensitivity analysis using different values of a QALY

QALY values	**Intervention**			**Control**			**Difference**	**SROI**
	Value per student	Students	Total	Value per student	Students	Total		
£20,000 (NICE)	£7,240	210	£1,520,400	£7,320	210	£1,537,200	-£16,800	£35:1
£12,936 (Claxton *et al.*)	£4,683	210	£983,395	£4,735	210	£994,261	-£10,866	£37:1

Assuming the resource costs and investment values remain unchanged the SROI ratios are just over £35 for both scenarios.

The sensitivity analyses demonstrate that the SROI ratio is not greatly affected by the choice of financial value of a QALY but the return on the investment is heavily driven by the resource savings associated by school absenteeism.

References

CLAXTON, K., MARTIN, S., SOARES, M., RICE, N., SPACKMAN, E., HINDE, S., DEVLIN, N., SMITH, P. C. & SCULPHER, M. 2015. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess, 19.

Appendix S. Advertisement for workshop

The flyer distributed via e-mail in order to advertise the workshop with PHDMs (reported in Chapter Ten) is inserted below.

Fuse health economics workshop

Critical appraisal of health economic decision tools for public health decision-makers

Wednesday 5th September 2018, 1.00pm – 4.00pm

Baddiley Clark seminar room, The Baddiley Clark building, Newcastle upon Tyne, NE2 4AX

To register for the workshop please follow this link **to a registration form**

What is the workshop about?

Health economic decision tools can play an important role in public health decision-making. Economic evaluations of public health interventions provide information to aid prioritisation decisions. Economic evaluation evidence, however is not always designed with public health decision-makers in mind. This workshop provides guidance on how to identify relevant evidence, to critically appraise the evidence for quality and relevance, and how to interpret the evidence for use in decision-making.

The workshop aims to further develop public health decision-makers' knowledge of the different tools and methods available and how to incorporate these into decision-making. It also aims to seek feedback from decision-makers on the usefulness and usability in practice of the tools discussed.

Who is this workshop for?

The event is intended for local authority officers and anyone involved in public health decision-making, from public health teams and broader departments.

Who is organising the workshop?

This workshop is being organised by Sarah Hill, a fuse PhD candidate. The resources provided are informed by her research with public health decision-makers to date. Feedback from the interactive group work session will feature in her final thesis in order to demonstrate the opinions of public health decision-makers around the usefulness and usability of available health economics tools to aid decision-making. All data collected at the workshop will be kept anonymous and no personal identifiable data will be used. Approval for this research project has been granted by Newcastle University Ethics Committee. If you have any questions about Sarah's PhD research or how the workshop will feature in her thesis please contact her at s.r.hill2@newcastle.ac.uk.

Programme

1.00pm – 2.15pm Critical appraisal training session

This session will introduce different methods of evaluating public health interventions using health economic tools, discuss how to appraise evidence and how to interpret the evidence effectively to aid decision-making.

2.15pm – 2.30pm Coffee break

Refreshments will be provided

2.30pm – 4.00pm Interactive group work session

The second part of the event will be an interactive session allowing participants to work with economic evidence of a local case study intervention to practice the skills developed in the first part of the event. Attendees will also be invited to provide feedback to the organisers on how useable the evidence is and how useful each tool is for decision-making in practice.

Appendix T. Consent form used at the workshop

Prior to the commencement of the interactive workshop with PHDMs (reported in Chapter Ten) written consent was obtained from workshop attendees for their feedback to be used in this thesis. The consent form that was used is displayed below.

PhD research study on health economic decision tools for public health

Consent form

	Please initial box
I have been provided with information on this research study and have had the opportunity to ask questions.	☐
I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason.	☐
I agree that notes may be taken from group discussions and will be kept by the researcher for use in her PhD.	☐
I understand that any data used from the group discussions in the researcher's PhD will be done so anonymously.	☐
I agree to take part in the above study.	☐

_____	_____	_____
Name of participant	Date	Signature
_____	_____	_____
Name of person taking consent	Date	Signature

References

1. Bateman I, Carson RT, Day B, Hanemann M, Hanley N, Hett T, et al. *Economic valuation with stated preference techniques : a manual*. Cheltenham: Edward Elgar; 2002.

2. Smith RD, Petticrew M. Public health evaluation in the twenty-first century: time to see the wood as well as the trees. *J Pub Health*. 2010;32:2-7.

3. Acheson D. *Public health in England: the report of the Committee of Inquiry into the Future Development of the Public Health Function*. London: HMSO; 1988.

4. Department of Health and Social Care. *New focus for public health - The Health and Social Care Act 2012*. gov.uk; 2012, 15/06/2012.

5. Local Government Association. *Public health transformation nine months on: bedding inand reaching out*. Local Government Association; 2014.

6. Department of Health. Local Authority Circular on the public health ring-fenced grant 2017/18. Gov.uk: Department of Health; 2016.

7. Department of Health. Public Health Ring-fenced Grant 2019/20 Circular. Department of Health; 2018.

8. Local Government Association. *Local government funding: moving the conversation on*. 2018.

9. Drummond M, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Fourth Edition ed. New York: Oxford University Press; 2015.

10. National Institute of Health and Care Excellence. *Developing NICE guidelines: the manual*. National Institute of Health and Care Excellence; 2014 [Accessed 04/11/2015]. Available from: https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview.

11. National Institute of Health and Care Excellence. *Guide to the methods of technology appraisal 2013*. NICE; 2013 [Accessed 20/10/2015]. Available from: https://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf.

12. Lal A, Moodie M, Peeters A, Carter R. Inclusion of equity in economic analyses of public health policies: systematic review and future directions. *Australian and New Zealand Journal of Public Health*. 2018;42(2):207-13.

13. Boujaoude MA, Mirelman AJ, Dalziel K, Carvalho N. Accounting for equity considerations in cost-effectiveness analysis: a systematic review of rotavirus vaccine in low- and middle-income countries. *Cost Effectiveness and Resource Allocation*. 2018;16.

14. Dukhanin V, Searle A, Zwerling A, Dowdy DW, Taylor HA, Merritt MW. Integrating social justice concerns into economic evaluation for healthcare and public health: A systematic review. *Social Science & Medicine*. 2018;198:27-35.

15. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Effectiveness and Resource Allocation*. 2006;4(1):1-9.

16. Local Government Association. *Public health, prevention and health improvement* 2019.

17. 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.

18. Russell LB, Sinha A. Strengthening Cost-Effectiveness Analysis for Public Health Policy. *American Journal of Preventive Medicine*. 2016;50(5):S6-S12.

19. Phillips CJ, Fordham R, Marsh K, Bertranou E, Davies S, Hale J, et al. Exploring the role of economics in prioritization in public health: what do stakeholders think? *Eur J Public Health*. 2011;21(5):578-84.

20. Payne K, McAllister M, Davies LM. Valuing the economic benefits of complex interventions: when maximising health is not sufficient. *Health Econ*. 2013;22(3):258-71.

21. Owen L, Fischer A. The cost-effectiveness of public health interventions examined by the National Institute for Health and Care Excellence from 2005 to 2018. *Public Health*. 2019;169:151-62.

22. Deidda M, Geue C, Kreif N, Dundas R, McIntosh E. A framework for conducting economic evaluations alongside natural experiments. *Soc Sci Med*. 2019;220:353-61.

23. Chalkidou K, Culyer A, Naidoo B, Littlejohns P. Cost-effective public health guidance: asking questions from the decision-maker's viewpoint. *Health Economics*. 2008;17(3):441-8.

24. Brazier J, Tsuchiya A. Improving Cross-Sector Comparisons: Going Beyond the Health-Related QALY. *Applied Health Economics and Health Policy*. 2015;13(6):557-65.

25. Lorgelly PK, Lawson KD, Fenwick EAL. Outcome Measurement in Economic Evaluations of Public Health Interventions: a Role for the Capability Approach? *Int J Environ Res Public Health*. 2010;7.

26. Cookson R, Mirelman AJ, Griffin S, Asaria M, Dawkins B, Norheim OF, et al. Using Cost-Effectiveness Analysis to Address Health Equity Concerns. *Value in Health*. 2017;20(2):206-12.

27. Asaria M, Griffin S, Cookson R. Distributional Cost-Effectiveness Analysis: A Tutorial. *Medical Decision Making*. 2016;36(1):8-19.

28. Frew E. Aligning Health Economics Methods to Fit with the Changing World of Public Health. *Applied Health Economics and Health Policy*. 2017;15(3):287-9.

29. HM Treasury. *The Green Book Appraisal and Evaluation in Central Government*. London; 2003.

30. HM Treasury. *The Green Book Central Government Guidance on Appraisal and Evaluation*. London; 2018.

31. National Institute of Health and Care Excellence. *Methods for the development of NICE public health guidance (third edition)*. 2012 [Accessed 20/10/2015]. Available from: https://www.nice.org.uk/process/pmg4/chapter/introduction.

32. 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(1):1-14.

33. Charles JM, Jones A, Lloyd-Williams H. Return on investment, social return on investment, and the business case for prevention. In: Edwards RT, McIntosh E, editors. *Applied Health Economics for Public Health Practice and Research*. Oxford: Oxford University Press; 2019. p. 279-300.

34. Masters R, Anwar E, Collins B, Cookson R, Capewell S. Return on investment of public health interventions: a systematic review. *Journal of Epidemiology and Community Health*. 2017;71(8):827-34.

35. Public Health England. *Local Health and Care Planning: Menu of preventative interventions*. London, UK; 2016.

36. Brousselle A, Benmarhnia T, Benhadj L. What are the benefits and risks of using return on investment to defend public health programs? *Prev Med Rep*. 2016;3:135-8.

37. Ruta D, Mitton C, Bate A, Donaldson C. Programme budgeting and marginal analysis: bridging the divide between doctors and managers. *BMJ*. 2005;330(7506):1501-3.

38. Goodwin E, Frew EJ. Using programme budgeting and marginal analysis (PBMA) to set priorities: Reflections from a qualitative assessment in an English Primary Care Trust. *Social Science & Medicine*. 2013;98:162-8.

39. Edwards RT, Charles JM, Thomas S, Bishop J, Cohen D, Groves S, et al. A national Programme Budgeting and Marginal Analysis (PBMA) of health improvement spending across Wales: disinvestment and reinvestment across the life course. *BMC public health*. 2014;14:837.

40. Marsh K, Dolan P, Kempster J, Lugon M. Prioritizing investments in public health: A multi-criteria decision analysis. *Journal of Public Health (United Kingdom)*. 2013;35(3):460-6.

41. Youngkong S. Incorporating cost-effectiveness data in a fair process for priority setting efforts. *International Journal of Health Policy and Management*. 2015;4(7):483-5.

42. The European Observatory on Health Systems and Policies. *Intersectoral Governance for Health in All Policies*: World Health Organization; 2012.

43. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health*. 2002;56(2):119-27.

44. Medical Research Council. *MRC Guidance on developing and evaluating complex interventions*. Medical Research Council; 2008.

45. Petticrew M. When are complex interventions 'complex'? When are simple interventions 'simple'? *European Journal of Public Health*. 2011;21(4):397-8.

46. Shiell A, Hawe P, Gold L. Complex interventions or complex systems? Implications for health economic evaluation. *BMJ*. 2008;336(7656):1281-3.

47. Kelly MP, McDaid D, Ludbrook A, Powell JE. *Economic Appraisal of Public Health Interventions*. Health Development Agency; 2005.

48. Weatherly H, Drummond M, Claxton K. Methods for assessing the cost-effectiveness of public health interventions: Key challenges and recommendations. *Health Policy*. 2009;93.

49. Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, et al. The need for a complex systems model of evidence for public health. *Lancet*. 2017;390(10112):2602-4.

50. Williams I, Bryan S. Understanding the limited impact of economic evaluation in health care resource allocation: a conceptual framework. *Health Policy*. 2007;80(1):135-43.

51. Williams I, McIver S, Moore D, Bryan S. The use of economic evaluations in NHS decision-making: a review and empirical investigation. *Health Technology Assessment*. 2008;12(7):1-196.

52. Hoffmann C, Stoykova BA, Nixon J, Glanville JM, Misso K, Drummond MF. Do Health-Care Decision Makers Find Economic Evaluations Useful? The Findings of Focus Group Research in UK Health Authorities. *Value in Health*. 2002;5(2):71-8.

53. Marks L, Hunter DJ, Scalabrini S, Gray J, McCafferty S, Payne N, et al. The return of public health to local government in England: Changing the parameters of the public health prioritization debate? *Public Health*. 2015;129(9):1194-203.

54. Hunter DJ, Marks L, Brown J, Scalabrini S, Salway S, Vale L, et al. The potential value of priority-setting methods in public health investment decisions: qualitative findings from three English local authorities. *Critical Public Health*. 2016:1-10.

55. Edwards RT, McIntosh E, editors. *Applied Health Economics for Public Health Practice and Research*. First ed. Oxford: Oxford University Press; 2019.

56. Frew E, Breheny K. Health economics methods for public health resource allocation: a qualitative interview study of decision makers from an English local authority. *Health Economics, Policy and Law*. 2019:1-13.

57. Frew E, Breheny K. Methods for public health economic evaluation: A Delphi survey of decision makers in English and Welsh local government. *Health Econ*. 2019;28(8):1052-63.

58. Brown J, Hunter, D.J., Marks, L., Salway, S., Gray, J.,, McCafferty S, Payne, N., Peckham, S., Thokala, P., Vale, L. *Shifting the gravity of spending? Exploring methods for supporting public health commissioners in priority-setting to improve population health and address health inequalities.*: Durham University; 2017.

59. Giles EL, McGeechan GJ, Coulton S, Deluca P, Drummond C, Howel D, et al. Brief alcohol intervention for risky drinking in young people aged 14–15 years in secondary schools: the SIPS JR-HIGH RCT. *Public Health Res*. 2019;7:9.

60. Department of Health. *Guidance on the consumption of alcohol by children and young people. A report by the Chief Medical Officer*. 2009, 17/12/2009.

61. Newbury-Birch D, Walker J, Avery L, Beyer F, Brown N, Jackson K, et al. *Impact of Alcohol Consumption on Young People: A Systematic Review of Published Reviews*. Department for children, school and families; 2009. Report No.: DCSF-RR067

62. Patton R, Deluca P, Kaner E, Newbury-Birch D, Phillips T, Drummond C. Alcohol Screening and Brief Intervention for Adolescents: The How, What and Where of Reducing Alcohol Consumption and Related Harm Among Young People. *Alcohol and Alcoholism*. 2013;49(2):207-12.

63. Gravelle H, Rees R. *Microeconomics*. 3rd ed. Essex, England: Pearson Education Limited; 2004.

64. Sloman J. *Economics*. 8th ed. ed. Wride A, Garratt D, editors. Harlow, England: Pearson; 2012.

65. Morris S, Devlin N, Parkin D, Spencer A. Welfarist and non-welfarist foundations of economic evaluation. *Economic Analysis in Health Care*. West Sussex, UK: John Wiley & Sons, Ltd; 2012. p. 204-31.

66. Hurley J. An overview of the normative economics of the health sector. In: Culyer AJ, Newhouse JP, editors. *Handbook of Health Economics*. 1: Elsevier; 2000. p. 55-118.

67. Brouwer WBF, Culyer AJ, van Exel NJA, Rutten FFH. Welfarism vs. extra-welfarism. *Journal of Health Economics*. 2008;27(2):325-38.

68. Tsuchiya A, Williams A. Welfare economics and economic evaluation. In: Drummond M, McGuire A, editors. *Economic Evaluation in Health Care: Merging Theory with Practice*. Oxford, UK: Oxford University Press; 2001.

69. Hicks J. *Value and Capital: An Inquiry Into Some Fundamental Principles of Economic Theory*: Clarendon Press; 1939.

70. Kaldor N. Welfare Propositions of Economics and Interpersonal Comparisons of Utility. *The Economic Journal*. 1939;49(195):549-52.

71. Rice T. An alternative framework for evaluating welfare losses in the health care market. *J Health Econ*. 1992;11(1):86-92.

72. Mooney G, Russell E. Equity in health care: The need for a new economics paradigm? In: Scott A, Maynard A, Elliott R, editors. *Advances in Health Economics*. West Sussex, England: John Wiley & Sons; 2003. p. 205-21.

73. Sen A. Personal Utilities and Public Judgements: Or What's Wrong with Welfare Economics? *Econ J*. 1979;89(355):537-58.

74. Culyer AJ. The normative economics of health care finance and provision. *Oxford Review of Economic Policy*. 1989;5(1):34-58.

75. Culyer AJ. Commodities, characteristics of commodities, characteristics of people, utilities, and the quality of life. In: Culyer AJ, editor. *The humble economist: Tony Culyer on health, health care and social decision making*. York: University of York; 2012. p. 9-27.

76. Birch S, Donaldson C. Valuing the benefits and costs of health care programmes: where's the 'extra' in extra-welfarism? *Social Science & Medicine*. 2003;56(5):1121-33.

77. Olsen J, Smith R, Harris A. *Economic Theory and the Monetary Valuation of Health Care: An Overview of the Issues as Applied to the Economic Evaluation of Health Care Programs*. Melbourne, Australia; 1999.

78. Kahneman D, Ritov I, Jacowitz KE, Grant P. Stated Willingness-to-Pay for Public-Goods - a Psychological Perspective. *Psychol Sci*. 1993;4(5):310-5.

79. Johnson FR. Why Not Real Economics? *Pharmacoeconomics*. 2012;30(2):127-31.

80. Morris S, Devlin N, Parkin D, Spencer A. Principles of economic evaluation in health care. *Economic Analysis in Health Care*. West Sussex, U.K.: John Wiley & Sons, Ltd; 2012. p. 232-52.

81. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Cost-benefit analysis. *Methods for the Economic Evaluation of Health Care Programmes* Third ed. Oxford, England: Oxford University Press; 2005. p. 211-45.

82. Coast J. Is economic evaluation in touch with society's health values? *BMJ*. 2004;329(7476):1233-6.

83. Nicholls J. Social return on investment-Development and convergence. *Evaluation and Program Planning*. 2017;64:127-35.

84. Nicholls J, Lawlor E, Neitzert E, Goodspeed T. *A Guide to Social Return on Investment*. London; 2012.

85. Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. *Cost-Effectiveness in Health and Medicine*. 1 ed: Oxford University Press; 1996.

86. Neumann PJ, Sanders GD, Russell LB, Ganiats TG, Siegel JE. *Cost-effectiveness in Health and Medicine*: Oxford University Press; 2017.

87. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care*. 1989;5(4):559-75.

88. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Medical decision making : an international journal of the Society for Medical Decision Making*. 1989;9(2):142-9.

89. Mehrez A, Gafni A. The Healthy-years Equivalents:How to Measure Them Using the Standard Gamble Approach. *Medical Decision Making*. 1991;11(2):140-6.

90. Buckingham K. A note on HYE (healthy years equivalent). *Journal of Health Economics*. 1993;12(3):301-9.

91. Towers I, Spencer A, Brazier J. Healthy year equivalents versus quality-adjusted life years: the debate continues. *Expert Review of Pharmacoeconomics & Outcomes Research*. 2005;5(3):245-54.

92. Drummond M, Brixner D, Gold M, Kind P, McGuire A, Nord E, et al. Toward a consensus on the QALY. *Value in Health*. 2009;12:S31-S5.

93. Murray CJL, Lopez AD. *The Global burden of disease : a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge, MA; 1996.

94. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. *Applied health economics and health policy*. 2017;15(2):127-37.

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. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes*. 2003;1:54-.

97. Goodwin E, Green C. A Systematic Review of the Literature on the Development of Condition-Specific Preference-Based Measures of Health. *Applied Health Economics and Health Policy*. 2016;14(2):161-83.

98. Lorgelly PK, Doble B, Rowen D, Brazier J, Thomas DM, Fox SB, et al. Condition-specific or generic preference-based measures in oncology? A comparison of the EORTC-8D and the EQ-5D-3L. *Quality of Life Research*. 2017;26(5):1163-76.

99. Mulhern B, Rowen D, Jacoby A, Marson T, Snape D, Hughes D, et al. The development of a QALY measure for epilepsy: NEWQOL-6D. *Epilepsy & behavior : E&B*. 2012;24(1):36-43.

100. Goodwin E, Green C, Spencer A. Estimating a Preference-Based Index for an Eight-Dimensional Health State Classification System for Multiple Sclerosis. *Value in Health*. 2015;18(8):1025-36.

101. Mathes T, Jacobs E, Morfeld J-C, Pieper D. Methods of international health technology assessment agencies for economic evaluations- a comparative analysis. *BMC Health Services Research*. 2013;13(1):371.

102. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Cost-effectiveness analysis. *Methods for the Economic Evaluation of Health Care Programmes* 3rd edition ed. Oxford, England: Oxford University Press; 2005. p. 103-36.

103. Sculpher M, Claxton K. Real Economics Needs to Reflect Real Decisions Response. *Pharmacoeconomics*. 2012;30(2):133-6.

104. Paulden M. Recent amendments to NICE's value-based assessment of health technologies: implicitly inequitable? *Expert Rev Pharmacoecon Outcomes Res*. 2017;17(3):239-42.

105. Coast J, Smith R, Lorgelly P. Should the capability approach be applied in Health Economics? *Health Economics*. 2008;17(6):667-70.

106. Sen A. *Commodities and Capabilities*. Amsterdam: North-Holland; 1985.

107. Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research*. 2012;21(1):167-76.

108. Coast J, Flynn TN, Natarajan L, Sproston K, Lewis J, Louviere JJ, et al. Valuing the ICECAP capability index for older people. *Social science & medicine*. 2008;67(5):874-82.

109. Lorgelly PK, Lorimer K, Fenwick EAL, Briggs AH, Anand P. Operationalising the capability approach as an outcome measure in public health: The development of the OCAP-18. *Social Science & Medicine*. 2015;142:68-81.

110. Simon J, Anand P, Gray A, Rugkåsa J, Yeeles K, Burns T. Operationalising the capability approach for outcome measurement in mental health research. *Social science & medicine*. 2013;98:187-96.

111. Mitchell PM, Roberts TE, Barton PM, Coast J. Assessing sufficient capability: A new approach to economic evaluation. *Social Science & Medicine*. 2015;139:71-9.

112. McIntosh E. Economic evaluation of guideline implementation strategies. In: Thorsen T, Mäkelä M, editors. *Changing Professional Practice: Theory and Practice of Clinical Guidelines Implementation*. Denmark: Danish Institute for Health Services Research and Development; 1999.

113. Mauskopf JA, Paul JE, Grant DM, Stergachis A. The Role of Cost—Consequence Analysis in Healthcare Decision—Making. *PharmacoEconomics*. 1998;13(3):277-88.

114. Hill SR, Vale L, Hunter D, Henderson E, Oluboyede Y. Economic evaluations of alcohol prevention interventions: Is the evidence sufficient? A review of methodological challenges. *Health Policy*. 2017;121(12):1249-62.

115. Charles J, Edwards RT. *A Guide To Health Economics For Those Working in Public Health* Bangor; 2016.

116. Keating B, Keating MO. Chapter 1: An Introduction to Cost Benefit Analysis. *Basic cost benefit analysis for assessing local public projects*. New York: Business Expert Press; 2014.

117. Keating B, Keating MO. Cost-Benefit Analysis and Individual Preferences. *Basic cost benefit analysis for assessing local public projects*. First edition. ed. New York: Business Expert Press; 2014.

118. Johannesson M, Jonsson B. Economic evaluation in healthcare: is there a role for cost-benefit analysis? *Health Policy*. 1998;17(1):1-23.

119. Robinson R. Cost-benefit analysis. *British Medical Journal*. 1993;307(6909):924-6.

120. Mishan EJ. Evaluation of Life and Limb: A Theoretical Approach. *Journal of Political Economy*. 1971;79(4):687-705.

121. Horowitz JK, McConnell KE. A Review of WTA/WTP Studies. *J Environ Econ Manag*. 2002;44(3):426-47.

122. Hanemann WM. Willingness to Pay and Willingness to Accept: How Much Can They Differ? *The American Economic Review*. 1991;81(3):635-47.

123. Tversky A, Kahneman D. Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*. 1991;106(4):1039-61.

124. Mitchell RC, Carson RT. *Using Surveys to Value Public Goods: The Contingent Valuation Method*: Taylor & Francis; 2013.

125. Davis RK. Recreation Planning as an Economic Problem. *Nat Resources J*. 1963;3(2).

126. Jones-Lee MW, Loomes G, Philips PR. Valuing the Prevention of Non-Fatal Road Injuries: Contingent Valuation vs. Standard Gambles. *Oxford Economic Papers*. 1995;47(4):676-95.

127. Mooney GH. *The valuation of human life*. London: Macmillan; 1977.

128. Acton JP. *Evaluating Public Programs To Save Lives: The Case of Heart Attacks*. Santa Monica, California; 1973. Report No.: R950RC.

129. Diener A, O'Brien B, Gafni A. Health care contingent valuation studies: a review and classification of the literature. *Health Econ*. 1998;7(4):313-26.

130. McIntosh E, Clarke P, Frew E. *Applied Methods of Cost-Benefit Analysis in Health Care*: OUP Oxford; 2010.

131. Carson RT, Flores NE, Martin KM, Wright JL. Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods. *Land Economics*. 1996;72(1):80-99.

132. Kennedy CA. Revealed preference valuation compared to contingent valuation: radon-induced lung cancer prevention. *Health Economics*. 2002;11(7):585-98.

133. Ryan M, Mentzakis E, Jareinpituk S, Cairns J. External Validity of Contingent Valuation: Comparing Hypothetical and Actual Payments. *Health Economics*. 2017;26(11):1467-73.

134. Klose T. The contingent valuation method in health care. *Health Policy*. 1999;47(2):97-123.

135. Carson R. *Contingent Valuation: A Comprehensive Bibliography and History*: Edward Elgar Publishing, Incorporated; 2012.

136. Thayer MA. Contingent valuation techniques for assessing environmental impacts: Further evidence. *J Environ Econ Manag*. 1981;8(1):27-44.

137. Hanemann M, Loomis J, Kanninen B. Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *Am J Agr Econ*. 1991;73(4):1255-63.

138. Mitchell RC, Carson RT. *An Experiment in Determining Willingness to Pay for National Water Quality Improvements* Washington, D.C.: U.S. Environmental Protection Agency; 1981, June 1981. Report No.: EE-0011.

139. Covey J, Loomes G, Bateman IJ. Valuing risk reductions: Testing for range biases in payment card and random card sorting methods. *J Environ Plann Man*. 2007;50(4):467-82.

140. Smith RD. The discrete-choice willingness-to-pay question format in health economics: should we adopt environmental guidelines? *Medical decision making : an international journal of the Society for Medical Decision Making*. 2000;20(2):194-206.

141. Shackley P, Dixon S. The random card sort method and respondent certainty in contingent valuation: an exploratory investigation of range bias. *Health Econ*. 2014;23(10):1213-23.

142. Smith RD. It's not just what you do, it's the way that you do it: the effect of different payment card formats and survey administration on willingness to pay for health gain. *Health Economics*. 2006;15(3):281-93.

143. Bateman IJ, Langford IH, Turner RK, Willis KG, Garrod GD. Elicitation and Truncation Effects in Contingent Valuation Studies. *Ecological Economics*. 1995;12(2):161-79.

144. Desvousges WH, Smith VK, Fisher A. Option price estimates for water quality improvements: A contingent valuation study for the monongahela river. *J Environ Econ Manag*. 1987;14(3):248-67.

145. Rowe RD, D'Arge RC, Brookshire DS. An experiment on the economic value of visibility. *J Environ Econ Manag*. 1980;7(1):1-19.

146. Daubert JT, Young RA. Recreational Demands for Maintaining Instream Flows - a Contingent Valuation Approach. *Am J Agr Econ*. 1981;63(4):666-76.

147. Greenley DA, Walsh RG, Young RA. Option Value - Empirical-Evidence from a Case-Study of Recreation and Water-Quality. *Q J Econ*. 1981;96(4):657-73.

148. Carson RT. Contingent valuation: Theoretical advances and empirical tests since the NOAA panel. *Am J Agr Econ*. 1997;79(5):1501-7.

149. Wiser RH. Using contingent valuation to explore willingness to pay for renewable energy: A comparison of collective and voluntary payment vehicles. *Ecological Economics*. 2007;62(3-4):419-32.

150. Smith RD. Construction of the contingent valuation market in health care: a critical assessment. *Health Econ*. 2003;12(8):609-28.

151. Torras M. The total economic value of Amazonian deforestation, 1978–1993. *Ecological economics*. 2000;33(2):283-97.

152. Brun F. Multifunctionality of mountain forests and economic evaluation. *Forest Policy and Economics*. 2002;4(2):101-12.

153. Ryan M, Gerard K, Amaya-Amaya M. *Using Discrete Choice Experiments to Value Health and Health Care*. Netherlands: Springer 2007.

154. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics*. 2014;32(9):883-902.

155. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ*. 2012;21(2):145-72.

156. Louviere JJ, Flynn TN, Carson RT. Discrete Choice Experiments Are Not Conjoint Analysis. *Journal of Choice Modelling*. 2010;3(3):57-72.

157. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy*. 2003;2(1):55-64.

158. Hauber AB, Gonzalez JM, Groothuis-Oudshoorn CGM, Prior T, Marshall DA, Cunningham C, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in Health*. 2016;19(4):300-15.

159. Johnson FR, Lancsar E, Marshall D, Kilambi V, Muhlbacher A, Regier DA, et al. Constructing Experimental Designs for Discrete-Choice Experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value in Health*. 2013;16(1):3-13.

160. Hanley N, Mourato S, Wright RE. Choice modelling approaches: A superior alternative for environmental valuation? *J Econ Surv*. 2001;15(3):435-62.

161. Millar R, Hall K. Social return on investment (SROI) and performance measurement: The opportunities and barriers for social enterprises in health and social care. *Public Manag Rev*. 2013;15(6):923-41.

162. van Voorn GA, Vemer P, Hamerlijnck D, Ramos IC, Teunissen GJ, Al M, et al. The Missing Stakeholder Group: Why Patients Should be Involved in Health Economic Modelling. *Appl Health Econ Health Policy*. 2016;14(2):129-33.

163. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 2011;14(4):403-13.

164. Musto V. *A guide to social return on investment for alcohol and drug treatment commissioners*. London: Public Health England; 2015, February 2015.

165. Jones A. *Nottingham City Homes: Decent Homes Impact Study*. Nottingham City Homes; 2013.

166. Wilson ECF, Peacock SJ, Ruta D. Priority setting in practice: what is the best way to compare costs and benefits? *Health Economics*. 2009;18(4):467-78.

167. Mitton C, Dionne F, Donaldson C. Managing healthcare budgets in times of austerity: the role of program budgeting and marginal analysis. *Appl Health Econ Health Policy*. 2014;12(2):95-102.

168. Thokala P, Marsh K, Devlin N, van Til J, Reddy B, Baltussen R, et al. Multi Criteria Decision Analysis Methods in Health Care: Current Status, Good Practice and Future Recommendations. *Value in Health*. 2014;17(3):A34-A.

169. Thokala P, Duenas A. Multiple Criteria Decision Analysis for Health Technology Assessment. *Value in Health*. 2012;15(8):1172-81.

170. Nutt DJ, Phillips LD, Balfour D, Curran HV, Dockrell M, Foulds J, et al. Estimating the Harms of Nicotine-Containing Products Using the MCDA Approach. *European Addiction Research*. 2014;20(5):218-25.

171. Kelly MP. Public health guidance and the role of new NICE. *Public Health*. 2005;119(11):960-8.

172. Wagstaff A, Culyer A. Four decades of Health Economics through a bibliometric lens. *Journal of Health Economics*. 2012;31(2):406-39.

173. Public Health England. *From Evidence into Action: opportunities to improve and protect the nation's health*. Public Health England; 2014.

174. Greenfield TK, Ye Y, Kerr W, Bond J, Rehm J, Giesbrecht N. Externalities from alcohol consumption in the 2005 US National Alcohol Survey: implications for policy. *Int J Environ Res Public Health*. 2009;6(12):3205-24.

175. Copello AG, Velleman RDB, Templeton LJ. Family interventions in the treatment of alcohol and drug problems. *Drug and Alcohol Review*. 2005;24(4):369-85.

176. Hill SR, Oluboyede Y, Vale L, Hunter D, Henderson E. *Economic evaluation and priority-setting in public health: protocol for a review of the methods used to evaluate interventions for the prevention and reduction of excessive alcohol consumption*. PROSPERO: International prospective register of systematic reviews; 2016. Report No.: CRD42016039063.

177. Thielen FW, Van Mastrigt GAPG, Burgers LT, Bramer WM, Majoie HJM, Evers SMAA, et al. How to prepare a systematic review of economic evaluations for clinical practice guidelines: database selection and search strategy development (part 2/3). *Expert Review of Pharmacoeconomics & Outcomes Research*. 2016;16(6):705-21.

178. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *J Clin Epidemiol*. 2015;68(6):617-26.

179. Higgins J, Green S. Selecting studies and collecting data. 2011. In: *Cochrane Handbook of Systematic Reviews of Interventions* [Internet]. The Cochrane Collaboration. Version 5.1.0. [Accessed 14/08/2017]. Available from: http://handbook.cochrane.org.

180. Centre for Reviews and Dissemination. *Systematic reviews: CRD's guidance for undertaking reviews in health care.* York; 2009, January 2009.

181. Drummond M, Weatherly H, Claxton K, Cookson R, Ferguson B, Godfrey C, et al. *Assessing the Challenges of Applying Standard Methods of Economic Evaluation to Public Health Interventions*. Public Health Research Consortium; 2006.

182. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, 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):231-50.

183. 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.

184. 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.

185. Chiou CF, Hay JW, Wallace JF, Bloom BS, Neumann PJ, Sullivan SD, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care*. 2003;41(1):32-44.

186. Watts RD, Li IW. Use of Checklists in Reviews of Health Economic Evaluations, 2010 to 2018. *Value in Health*. 2019;22(3):377-82.

187. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity costs measurement through quality of life? A response to the recommendation of the Washington Panel. *Health Econ*. 1997;6(3):253-9.

188. Deeks JJ, Higgins J, Altman D. Analysing data and undertaking meta-analyses. 2011. In: *Cochrane Handbook for Systematic Reviews of Interventions* [Internet]. The Cochrane Collaboration. Version 5.1.0. [Accessed 14/08/2017]. Available from: http://handbook.cochrane.org.

189. Valentine JC, Wilson SJ, Rindskopf D, Lau TS, Tanner-Smith EE, Yeide M, et al. Synthesizing Evidence in Public Policy Contexts:The Challenge of Synthesis When There Are Only a Few Studies. *Evaluation Review*. 2017;41(1):3-26.

190. Coulton S, Bl, M, Crosby H, Dale V, Drummond C, et al. Effectiveness and Cost-effectiveness of Opportunistic Screening and Stepped-care Interventions for Older Alcohol Users in Primary Care. *Alcohol & Alcoholism*. 2017;52(6):655-64.

191. Watson J, Crosby H, Dale V, Tober G, Wu Q, Lang J, et al. AESOPS: a randomised controlled trial of the clinical effectiveness and cost-effectiveness of opportunistic screening and stepped care interventions for older hazardous alcohol users in primary care. *Health Technology Assessment*. 2013;17(25):1-157.

192. Sheffield Alcohol Research Group. *The Sheffield Alcohol Policy Model*. 2019 [updated 2019; Accessed 08/04/2019]. Available from: https://www.sheffield.ac.uk/scharr/sections/ph/research/alpol/research/sapm.

193. Rogeberg O, Bergsvik D, Phillips LD, van Amsterdam J, Eastwood N, Henderson G, et al. A new approach to formulating and appraising drug policy: A multi-criterion decision analysis applied to alcohol and cannabis regulation. *International Journal of Drug Policy*. 2018;56:144-52.

194. Kapoor A, Kraemer KL, Smith KJ, Roberts MS, Saitz R. Cost-effectiveness of screening for unhealthy alcohol use with %carbohydrate deficient transferrin: results from a literature-based decision analytic computer model. *Alcoholism: Clinical and Experimental Research*. 2009;33(8):1440-9.

195. Mansdotter AM, Rydberg MK, Wallin E, Lindholm LA, Andreasson S. A cost-effectiveness analysis of alcohol prevention targeting licensed premises. *European Journal of Public Health*. 2007;17(6):618-23.

196. Miller TR, Zaloshnja E, Spicer RS. Effectiveness and benefit-cost of peer-based workplace substance abuse prevention coupled with random testing. *Accident Analysis & Prevention*. 2007;39(3):565-73.

197. Li T, Waters TM, Kaplan EK, Kaplan CM, Nyarko KA, Derefinko KJ, et al. Economic Analyses of an Alcohol Misconduct Prevention Program in a Military Setting. *Military Medicine*. 2017;182(1):e1562-e7.

198. Angus C, Scafato E, Ghirini S, Torbica A, Ferre F, Struzzo P, et al. Cost-effectiveness of a programme of screening and brief interventions for alcohol in primary care in Italy. *BMC Family Practice*. 2014;15(1):26.

199. Barbosa C, Cowell A, Bray J, Aldridge A. The cost-effectiveness of alcohol screening, brief intervention, and referral to treatment (SBIRT) in emergency and outpatient medical settings. *Journal of Substance Abuse Treatment*. 2015;53:1-8.

200. Cobiac L, Vos T, Doran C, Wallace A. Cost-effectiveness of interventions to prevent alcohol-related disease and injury in Australia. *Addiction*. 2009;104(10):1646-55.

201. Holm AL, Veerman L, Cobiac L, Ekholm O, Diderichsen F. Cost-effectiveness of changes in alcohol taxation in Denmark: a modelling study. *Cost Effectiveness and Resource Allocation*. 2014;12(1).

202. Holm AL, Veerman L, Cobiac L, Ekholm O, Diderichsen F. Cost-effectiveness of preventive interventions to reduce alcohol consumption in Denmark. *Plos One*. 2014;9(2):e88041.

203. Lai T, Habicht J, Reinap M, Chisholm D, Baltussen R. Costs, health effects and cost-effectiveness of alcohol and tobacco control strategies in Estonia. *Health Policy*. 2007;84:75-88.

204. Purshouse RC, Brennan A, Rafia R, Latimer NR, Archer RJ, Angus CR, et al. Modelling the cost-effectiveness of alcohol screening and brief interventions in primary care in England. *Alcohol and Alcoholism*. 2013;48(2):180-8.

205. Tariq L, van den Berg M, Hoogenveen RT, van Baal PH. Cost-effectiveness of an opportunistic screening programme and brief intervention for excessive alcohol use in primary care. *Plos One*. 2009;4(5):e5696.

206. Cobiac LJ, Mizdrak A, Wilson N. Cost-effectiveness of raising alcohol excise taxes to reduce the injury burden of road traffic crashes. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention*. [E-pub ahead of print 18 October, 2018].

207. Angus C, Gillespie D, Ally A, Brennan A. *Modelling the impact of Minimum Unit Price and Identification and Brief Advice policies using the Sheffield Alcohol Policy Model Version 3*. University of Sheffield, ScHARR; 2015, March 2015.

208. Purshouse RC, Brennan A, Latimer N, Meng Y, Rafia R, Jackson R, et al. *Modelling to assess the effectiveness and cost-effectiveness of public health related strategies and*

interventions to reduce alcohol attributable harm in England using the Sheffield Alcohol Policy Model version 2.0. University of Sheffield, ScHARR; 2009.

209. Drummond C, Coulton S, James D, Godfrey C, Parrott S, Baxter J, et al. Effectiveness and cost-effectiveness of a stepped care intervention for alcohol use disorders in primary care: pilot study. *British Journal of Psychiatry*. 2009;195(5):448-56.

210. Navarro HJ, Shakeshaft A, Doran CM, Petrie DJ. The potential cost-effectiveness of general practitioner delivered brief intervention for alcohol misuse: evidence from rural Australia. *Addictive Behaviors*. 2011;36(12):1191-8.

211. Sassi F, Cecchini M, Devaux M, Astolfi R. Health and economic impacts of key alcohol policy options. In: Sassi F, editor. *Tackling Harmful Alcohol Use*: OECD Publishing; 2015.

212. Meng Y, Hill-McMannus D, Brennan A, Meier P. *Model-based appraisal of alcohol minimum pricing and off-licensed trade discount bans in Scotland using the Sheffield Alcohol Policy Model (v.2): Second update based on newly available data*. University of Sheffield, ScHARR; 2012.

213. Meng Y, Brennan A, Holmes J, Hill-McMannus D, Angus C, Purshouse RC, et al. *Modelled income group-specific impacts of alcohol minimum unit pricing in England 2014/15: Policy appraisals using new developments to the Sheffield Alcohol Policy Model (v2.5)*. University of Sheffield, ScHARR; 2013.

214. Brennan A, Purshouse RC, Taylor K, Rafia R, Meier P, Booth A, et al. *Modelling the Potential Impact of Pricing and Promotion Policies for Alcohol in England: Results from the Sheffield Alcohol Policy Model Version 2008 (1-1)*. University of Sheffield, ScHARR; 2009.

215. Angus C, Holmes J, Brennan A, Meier P. *Model-based appraisal of the comparative impact of Minimum Unit Pricing and taxation policies in Wales: Final report*. Cardiff: Welsh Government, ScHARR; 2018, February 2018.

216. Angus C, Meng Y, Ally A, Holmes J, Brennan A. *Model-based appraisal of minimum unit pricing for alcohol in the Republic of Ireland*. University of Sheffield, ScHARR; 2014.

217. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, 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):1093-103.

218. Byrnes JM, Cobiac LJ, Doran CM, Vos T, Shakeshaft AP. Cost-effectiveness of volumetric alcohol taxation in Australia. *Medical Journal of Australia*. 2010;192(8):439-43.

219. Neighbors CJ, Barnett NP, Rohsenow DJ, Colby SM, Monti PM. Cost-effectiveness of a motivational intervention for alcohol-involved youth in a hospital emergency department. *Journal of Studies on Alcohol and Drugs*. 2010;71(3):384-94.

220. Solberg LI, Maciosek MV, Edwards NM. Primary care intervention to reduce alcohol misuse: ranking its health impact and cost effectiveness. *American Journal of Preventive Medicine*. 2008;34(2):143-52.e3.

221. Tilling C, Kro M, Tsuchiya A, Brazier J, van Exel J, Brouwer W. Does the EQ-5D Reflect Lost Earnings? *PharmacoEconomics*. 2012;30(1):47-61.

222. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care*. 2003;12(1):47-52.

223. Angus C, Li J, Romero-Rodriguez E, Anderson P, Parrott S, Brennan A. Cost-effectiveness of strategies to improve delivery of brief interventions for heavy drinking in primary care: results from the ODHIN trial. *Eur J Public Health*. 2019;29(2):219-25.

224. van Baal PH, van den Berg M, Hoogenveen RT, Vijgen SM, Engelfriet PM. Cost-effectiveness of a low-calorie diet and orlistat for obese persons: modeling long-term health gains through prevention of obesity-related chronic diseases. *Value in Health*. 2008;11(7):1033-40.

225. Smit F, Lokkerbol J, Riper H, Majo MC, Boon B, Blankers M. Modeling the cost-effectiveness of health care systems for alcohol use disorders: how implementation of eHealth interventions improves cost-effectiveness. *Journal of Medical Internet Research*. 2011;13(3):e56.

226. Brennan A, Meier P, Purshouse R, Rafia R, Meng Y, Hill-Macmanus D, et al. The Sheffield alcohol policy model - A mathematical description. *Health Economics (United Kingdom)*. 2015;24(10):1368-88.

227. Angus C, Thomas C, Anderson P, Meier PS, Brennan A. Estimating the cost-effectiveness of brief interventions for heavy drinking in primary health care across Europe. *European Journal of Public Health*. 2016;27(2):345-51.

228. van den Berg M, van Baal PHM, Tariq L, Schuit AJ, de Wit GA, Hoogenveen RT. The cost-effectiveness of increasing alcohol taxes: a modelling study. *BMC Medicine*. 2008;6(1):1-11.

229. Zur RM, Zaric GS. A microsimulation cost-utility analysis of alcohol screening and brief intervention to reduce heavy alcohol consumption in Canada. *Addiction*. 2016;111(5):817-31.

230. Havard A, Shakeshaft AP, Conigrave KM, Doran CM. Randomized controlled trial of mailed personalized feedback for problem drinkers in the emergency department: the short-term impact. *Alcoholism: Clinical and Experimental Research*. 2012;36(3):523-31.

231. Shanahan M, Shakeshaft A, Mattick RP. Modelling the costs and outcomes of changing rates of screening for alcohol misuse by GPs in the Australian context. *Applied Health Economics and Health Policy*. 2006;5(3):155-66.

232. Ingels JB, Corso PS, Kogan SM, Brody GH. Cost-effectiveness of the strong African American families-teen program: 1-year follow-up. *Drug and Alcohol Dependence*. 2013;133(2):556-61.

233. Tanaree A, Assanangkornchai S, Isaranuwatchai W, Thavorn K, Coyte PC. Integrated treatment program for alcohol related problems in community hospitals, Songkhla province of Thailand: A social return on investment analysis. *PLoS One*. 2019;14(1):e0209210.

234. Crawford MJ, Sanatinia R, Barrett B, Byford S, Dean M, Green J, et al. The clinical and cost-effectiveness of brief advice for excessive alcohol consumption among people attending sexual health clinics: a randomised controlled trial. *Sexually Transmitted Infections*. 2015;91(1):37-43.

235. Canadian Agency for Drugs and Technologies in Health. *Guidelines for the Economic Evaluation of Health Technologies: Canada*. Canada: CADTH; 2017.

236. Miller TR, Hendrie D. *Substance Abuse Prevention Dollars and Cents: A Cost-Benefit Analysis*. Center for Substance Abuse Prevention, Substance Abuse and Mental Health Services Administration, DHHS US; 2008. Report No.: (SMA) 07-4298.

237. De Wit A, Van Gils P, Over E, Lokkerbol J, Smit F, Mosca I, et al. *Social cost benefit analysis (SCBA) of three alcohol policy measures*. 2017.

238. Dubourg R, Hamed J, Thorns J. *The economic and social costs of crime against individuals and households 2003/04*. Home Office; 2005.

239. McKie J, Kuhse H, Richardson J, Singer P. Allocating Healthcare By QALYs: The Relevance of Age. *Cambridge Quarterly of Healthcare Ethics*. 2009;5(4):534-45.

240. Harris J. QALYfying the value of life. *Journal of Medical Ethics*. 1987;13(3):117-23.

241. Carthy T, Chilton S, Covey J, Hopkins L, Jones-Lee M, Loomes G, et al. On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part 2 - The CV/SG "Chained" Approach. *Journal of Risk and Uncertainty*. 1998;17(3):187-214.

242. Glover D, Henderson J. *Quantifying health impacts of government policies: A how-to guide to quantifying the health impacts of government policies* Department of Health; 2010.

243. Byford S, Raftery J. Perspectives in economic evaluation. *BMJ*. 1998;316(7143):1529-30.

244. Östlin P, Schrecker T, Sadana R, Bonnefoy J, Gilson L, Hertzman C, et al. Priorities for Research on Equity and Health: Towards an Equity-Focused Health Research Agenda. *PLOS Medicine*. 2011;8(11):e1001115.

245. Griffin S, Love-Koh J, Pennington B, Owen L. Evaluation of Intervention Impact on Health Inequality for Resource Allocation. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2019;39(3):171-82.

246. Willmott M, Womack J, Hollingworth W, Campbell R. Making the case for investment in public health: experiences of Directors of Public Health in English local government. *J Public Health (Oxf)*. 2016;38(2):237-42.

247. Oliver KA, de Vocht F. Defining 'evidence' in public health: a survey of policymakers' uses and preferences. *European Journal of Public Health*. 2017;27:112-7.

248. Wye L, Brangan E, Cameron A, Gabbay J, Klein JH, Pope C. Evidence based policy making and the 'art' of commissioning - how English healthcare commissioners access and use information and academic research in 'real life' decision-making: an empirical qualitative study. *Bmc Health Services Research*. 2015;15.

249. Kneale D, Rojas-García A, Raine R, Thomas J. The use of evidence in English local public health decision-making: a systematic scoping review. *Implementation Science*. 2017;12(1):53.

250. Denford S, Lakshman R, Callaghan M, Abraham C. Improving public health evaluation: a qualitative investigation of practitioners' needs. *BMC Public Health*. 2018;18(1):190.

251. Atkins L, Smith JA, Kelly MP, Michie S. The process of developing evidence-based guidance in medicine and public health: A qualitative study of views from the inside. *Implementation Science*. 2013;8(1).

252. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for Reporting Qualitative Research: A Synthesis of Recommendations. *Acad Med*. 2014;89(9):1245-51.

253. Patton MQ. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*: SAGE Publications; 2014.

254. Morgan D. Sampling Frame. In: Given L, editor. *The SAGE Encyclopedia of Qualitative Research Methods*. 1st ed. London: SAGE Publications; 2008. p. 800-1.

255. Trotter RT. Qualitative research sample design and sample size: Resolving and unresolved issues and inferential imperatives. *Preventive Medicine*. 2012;55(5):398-400.

256. Bernard HR. *Research methods in anthropology: qualitative and quantitative methods*: AltaMira Press; 2002.

257. Turner DW. Qualitative Interview Design: A Practical Guide for Novice Investigators. *The Qualitative Report*. 2010;15(3):754-60.

258. Longhurst R. Semi-structured Interviews and Focus Groups. In: Clifford N, Cope M, Gillespie T, French S, editors. *Key Methods in Geography*. Third ed: SAGE Publications; 2016. p. 143-56.

259. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess B, editors. *Analyzing Qualitative Data*: Taylor & Francis; 2002. p. 173-94.

260. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*. 2013;13:117-25.

261. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006;3(2):77-101.

262. Boyatzis RE. *Transforming Qualitative Information: Thematic Analysis and Code Development*: SAGE Publications; 1998.

263. Marks L, Jehu LM, Visram S, Mason A, Liu D, Davis H, et al. *Commissioning public health services: the impact of the health reforms on access, health inequalities and innovation in service provision* 2017, December 2017. Report No.: PR-R6-1113-25002

264. Eddama O, Coast J. Use of economic evaluation in local health care decision-making in England: a qualitative investigation. *Health Policy*. 2009;89(3):261-70.

265. Public Health England. *Overview of the Prioritisation Framework*. Public Health England; 2018 [updated 19/06/18; Accessed 11/07/18]. Available from: https://www.gov.uk/government/publications/the-prioritisation-framework-making-the-most-of-your-budget/overview-of-the-prioritisation-framework.

266. Hunter DJ, Gray J, Maniatopoulos G. *Shifting the Gravity of Spending – Mark 2: Final Evaluation Report* 2019, February 2019.

267. Barbour R. Interviews. *Introducing Qualitative Research: a student's guide*. second ed: SAGE Publications Ltd; 2014. p. 111-32.

268. Collins M, Shattell M, Thomas SP. Problematic interviewee behaviors in qualitative research. *Western J Nurs Res*. 2005;27(2):188-99.

269. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin*. 2010;96(1):5-21.

270. Giles EL, Coulton S, Deluca P, Drummond C, Howel D, Kaner E, et al. Multicentre individual randomised controlled trial of screening and brief alcohol intervention to prevent

risky drinking in young people aged 14–15 in a high school setting (SIPS JR-HIGH): study protocol. *BMJ Open*. 2016;6(12).

271. Screening and Intervention Programme for Sensible Drinking (SIPS). *Modified-Single Alcohol Screening Question (M-SASQ)*. [Accessed 03/06/2019]. Available from: www.sips.iop.kcl.ac.uk/msasq.php.

272. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Cost analysis. *Methods for the Economic Evaluation of Health Care Programmes* Third ed. Oxford, England: Oxford University Press; 2005. p. 55-101.

273. Hanemann WM. Valuing the Environment through Contingent Valuation. *Journal of Economic Perspectives*. 1994;8(4):19-43.

274. Arrow K, Solow R, Portney P, Leamer E, Radner R, Schuman H. *Report of the NOAA panel on Contingent Valuation*: National Oceanic and Atmospheric Administration; 1993.

275. Carson RT. Contingent Valuation: A User's Guide. *Environmental Science & Technology*. 2000;34(8):1413-8.

276. Somers C, Grieve E, Lennon M, Bouamrane MM, Mair FS, McIntosh E. Valuing Mobile Health: An Open-Ended Contingent Valuation Survey of a National Digital Health Program. *Jmir Mhealth Uhealth*. 2019;7(1).

277. Bala MV, Mauskopf JA, Wood LL. Willingness to pay as a measure of health benefits. *Pharmacoeconomics*. 1999;15(1):9-18.

278. Vaughan WJ, Russell CS, Darling AH. Determining the Optimal Sample Size for Contingent Valuation Surveys. [Working Paper]. In press 2000.

279. Alayli-Goebbels AF, van Exel J, Ament AJ, de Vries NK, Bot SD, Severens JL. Consumer willingness to invest money and time for benefits of lifestyle behaviour change: an application of the contingent valuation method. *Health Expectations*. 2015;18(6):2252-65.

280. Epstein AJ, Barry CL, Fiellin DA, Busch SH. Consumers' Valuation of Primary Care-Based Treatment Options for Mental and Substance Use Disorders. *Psychiatr Serv*. 2015;66(8):772-4.

281. Pennington M, Gomes M, Donaldson C. Handling Protest Responses in Contingent Valuation Surveys. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2017;37(6):623-34.

282. Olsen JA, Donaldson C. Helicopters, hearts and hips: Using willingness to pay to set priorities for public sector health care programmes. *Social Science & Medicine*. 1998;46(1):1-12.

283. Psacharopoulos G, Patrinos HA. Returns to investment in education: a further update. *Education Economics*. 2004;12(2):111-34.

284. Wilson V, Malcolm H, Edward S, Davidson J. 'Bunking off': the impact of truancy on pupils and teachers. *British Educational Research Journal*. 2008;34(1):1-17.

285. Sälzer C, Heine J-H. Students' skipping behavior on truancy items and (school) subjects and its relation to test performance in PISA 2012. *International Journal of Educational Development*. 2016;46:103-13.

286. Tanner-Smith EE, Lipsey MW. Brief Alcohol Interventions for Adolescents and Young Adults: A Systematic Review and Meta-Analysis. *Journal of Substance Abuse Treatment*. 2015;51:1-18.

287. McIntosh S, Vignoles A. Measuring and assessing the impact of basic skills on labour market outcomes. *Oxford Economic Papers*. 2001;53(3):453-81.

288. Balfanz R, Byrnes V. *The Importance of Being There: A Report on Absenteeism in the Nation's Public Schools* Baltimore; 2012.

289. Hyla E. The Long-Term Economic Impact of Juvenile Criminal Activity. *Park Place Economist*. 2016;24(1):74-9.

290. Colman RA, Kim DH, Mitchell-Herzfeld S, Shady TA. *Long-Term Consequences of Delinquency: Child Maltreatment and Crime in Early Adulthood* New York State Office of Children and Family Services 2009.

291. Kirk DS, Sampson RJ. Juvenile Arrest and Collateral Educational Damage in the Transition to Adulthood. *Sociology of education*. 2013;88(1):36-62.

292. Meyerhoff J, Liebe U. Do protest responses to a contingent valuation question and a choice experiment differ? *Environmental and Resource Economics*. 2008;39(4):433-46.

293. Frey UJ, Pirscher F. Distinguishing protest responses in contingent valuation: A conceptualization of motivations and attitudes behind them. *Plos One*. 2019;14(1).

294. Jorgensen BS, Syme GJ, Bishop BJ, Nancarrow BE. Protest responses in contingent valuation. *Environ Resour Econ*. 1999;14(1):131-50.

295. Pennington M, Baker R, Brouwer W, Mason H, Hansen DG, Robinson A, et al. Comparing WTP Values of Different Types of QALY Gain Elicited from the General Public. *Health Economics*. 2015;24(3):280-93.

296. Gyldmark M, Morrison GC. Demand for health care in Denmark: results of a national sample survey using contingent valuation. *Social Science & Medicine*. 2001;53(8):1023-36.

297. Aumann I, Treskova M, Hagemann N, von der Schulenburg JM. Analysis of Driving Factors of Willingness to Use and Willingness to Pay for Existing Pharmacological Smoking Cessation Aids Among Young and Middle-Aged Adults in Germany. *Applied Health Economics and Health Policy*. 2016;14(4):441-52.

298. Donaldson C. Valuing the benefits of publicly-provided health care: does 'ability to pay' preclude the use of 'willingness to pay'? *Social Science & Medicine*. 1999;49(4):551-63.

299. Donaldson C. Eliciting patients' values by use of 'willingness to pay': letting the theory drive the method. *Health Expectations*. 2001;4(3):180-8.

300. Donaldson C, Birch S, Gafni A. The distribution problem in economic evaluation: income and the valuation of costs and consequences of health care programmes. *Health Econ*. 2002;11(1):55-70.

301. Jung S. Does education affect risk aversion? Evidence from the British education reform. *Applied Economics*. 2015;47(28):2924-38.

302. Halstead JM, Luloff AE, Stevens TH. Protest bidders in contingent valuation. *Northeastern Journal of Agricultural and Resource Economics*. 1992;21(2):160-9.

303. Diamond PA, Hausman JA. Contingent Valuation - Is Some Number Better Than No Number. *Journal of Economic Perspectives*. 1994;8(4):45-64.

304. Adem KG. Contingent valuation technique: A review of literature. *ISABB Journal of Health and Environmental Sciences*. 2011;1(1):8-16.

305. StataCorp LLC. Stata 14.1. 2016.

306. Gould W, Pitblado J. *How large should the bootstrapped samples be relative to the total number of cases in the dataset?* : Stata; [Accessed 24/07/2018]. Available from: https://www.stata.com/support/faqs/statistics/bootstrapped-samples-guidelines/.

307. Chatfield M, Mander A. The Skillings-Mack test (Friedman test when there are missing data). *The Stata journal*. 2009;9(2):299-305.

308. Frew E. Benefit assessment for cost-benefit analysis studies in health care using contingent valuation methods. In: McIntosh E, Clarke P, Frew E, Louviere J, editors. *Applied methods of cost-benefit analysis in health care*. Oxford: Oxford University Press; 2010.

309. Donaldson C. The (near) equivalence of cost-effectiveness and cost-benefit analyses. Fact or fallacy? *Pharmacoeconomics*. 1998;13(4):389-96.

310. Breusch T, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*. 1979;47:1287-94.

311. Cook RD, Weisberg S. Diagnostics for heteroscedasticity in regression. *Biometrika*. 1983;70(1):1-10.

312. Belotti F, Deb P, Manning WG, Norton EC. Twopm: Two-Part Models. *The Stata Journal*. 2015;15(1):3-20.

313. Deb P, Norton EC, Manning W. Generalised Linear Models. *Health Econometrics using Stata*. Texas: Stata Press; 2017. p. 71-92.

314. Deb P, Norton EC, Manning W. *Health Econometrics Using Stata*. Texas, USA: Stata Press; 2017.

315. Smith R, Olsen JA, Harris AH. A Review of Methodological Issues in the Conduct of WTP Studies in Health Care II: Administration of a CV Survey. [Working Paper]. In press 1999.

316. Bobinac A, van Exel NJA, Rutten FFH, Brouwer WBF. GET MORE, PAY MORE? An elaborate test of construct validity of willingness to pay per QALY estimates obtained through contingent valuation. *Journal of Health Economics*. 2012;31(1):158-68.

317. Bateman I, Munro A, Rhodes B, Starmer C, Sugden R. Does part-whole bias exist? An experimental investigation. *Econ J*. 1997;107(441):322-32.

318. Borzykowski N, Baranzini A, Maradan D. Scope Effects in Contingent Valuation: Does the Assumed Statistical Distribution of WTP Matter? *Ecological Economics*. 2018;144:319-29.

319. Hausman J. Contingent Valuation: From Dubious to Hopeless. *Journal of Economic Perspectives*. 2012;26(4):43-56.

320. Home Office. *Police powers and procedures, England and Wales, year ending 31 March 2018*. Statistics at Home Office; 2018, 25th October 2018.

321. Donaldson C, Jones AM, Mapp TJ, Olson JA. Limited dependent variables in willingness to pay studies: applications in health care. *Applied Economics*. 1998;30(5):667-77.

322. Meyerhoff J, Liebe U. Determinants of protest responses in environmental valuation A meta-study. *Ecological Economics*. 2010;70(2):366-74.

323. Strazzera E, Scarpa R, Calia P, Garrod GD, Willis KG. Modelling zero values and protest responses in contingent valuation surveys. *Applied Economics*. 2003;35(2):133-8.

324. Jeanrenaud C, Pellegrini S. Valuing Intangible Costs of Alcohol Dependence : a Contingent Valuation Study. *Revue d'économie politique*. 2007;117(5):813-25.

325. Strazzera E, Genius M, Scarpa R, Hutchinson G. The effect of protest votes on the estimates of WTP for use values of recreational sites. *Environ Resour Econ*. 2003;25(4):461-76.

326. Morris S, Devlin N, Parkin D, Spencer A. Measuring and valuing health care output. *Economic Analysis in Health Care*. West Sussex, UK: John Wiley & Sons, Ltd; 2012. p. 253-85.

327. Haefeli M, Elfering A, McIntosh E, Gray A, Sukthankar A, Boos N. A Cost-Benefit Analysis Using Contingent Valuation Techniques: A Feasibility Study in Spinal Surgery. *Value in Health*. 2008;11(4):575-88.

328. van der Pol M, McKenzie L. Costs and benefits of tele-endoscopy clinics in a remote location. *Journal of Telemedicine and Telecare*. 2010;16(2):89-94.

329. Eisen-Hecht JI, Kramer RA. A cost-benefit analysis of water quality protection in the Catawba basin. *JAWRA Journal of the American Water Resources Association*. 2002;38(2):453-65.

330. Kolimenakis A, Bithas K, Richardson C, Latinopoulos D, Baka A, Vakali A, et al. Economic appraisal of the public control and prevention strategy against the 2010 West Nile Virus outbreak in Central Macedonia, Greece. *Public Health*. 2016;131:63-70.

331. Andronis L, Barton P, Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. *Health Technol Assess*. 2009;13(29).

332. Doshi JA, Glick HA, Polsky D. Analyses of cost data in economic evaluations conducted alongside randomized controlled trials. *Value Health*. 2006;9(5):334-40.

333. 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. *J Public Health*. 2009;31(3):423-33.

334. Sibbald B, Roland M. Understanding controlled trials: Why are randomised controlled trials important? *BMJ*. 1998;316(7126):201.

335. Glick H, Doshi J, Sonnad S, Polsky D. *Economic Evaluation in Clinical Trials*. Gray A, Briggs A, editors. Ofxford: Oxford University Press; 2010.

336. Kriström B. Valuing environmental benefits using the contingent valuation method : an econometric analysis [Doctoral thesis, monograph]. Umeå: Umeå University; 1990.

337. HM Revenues & Customs. *Number of individual income taxpayers by marginal rate, gender and age, 1990-91 to 2018-19*. National statistics; 2018 [Accessed 03/07/2018]. Available from: https://www.gov.uk/government/statistics/number-of-individual-income-taxpayers-by-marginal-rate-gender-and-age.

338. Department for Education. *Schools, Pupils and their Characteristics: January 2018 - National Tables*. National Statistics; 2018 [updated August 2018; Accessed 04/07/18].

Available from: https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2018.

339. Scottish Government Education Directorate. *List of Secondary Schools with pupil roll information as at September 2017*. Scottish Government; 2017 [updated August 2018. Available from: https://www2.gov.scot/Topics/Statistics/Browse/School-Education/RollsByStage/PupilNumbers2017.

340. Department of Education. *School Census - post primary - enrolments 2017/18*. OpenDataNI; 2018 [updated February 2017; Accessed 04/07/18]. Available from: https://www.opendatani.gov.uk/dataset/school-census-post-primary/resource/4531eb68-119b-4ead-9fda-248b46a10ada?inner_span=True.

341. StatsWales. *Pupils by school and year group, 2018*. Welsh government; 2018 [updated 2018; Accessed 04/07/18]. Available from: https://statswales.gov.wales/Catalogue/Education-and-Skills/Schools-and-Teachers/Schools-Census/Pupil-Level-Annual-School-Census/Pupils/pupils-by-school-yeargroup.

342. Briggs A, Gray A. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess*. 1999;3(2).

343. Efron B, Tibshirani R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statist Sci*. 1986;1(1):54-75.

344. Poi B. From the help desk: Some bootstrapping techniques. *The Stata journal*. 2004;4(3):312–28.

345. Valuation Office Agency. *Council Tax: stock of properties 2017*. Official Statistics; 2017 [updated August 2018; Accessed 06/07/2018]. Available from: https://www.gov.uk/government/statistics/council-tax-stock-of-properties-2017.

346. Independent schools council. *Find a school*. 2018 [Accessed 15/07/2018]. Available from: https://www.isc.co.uk/schools/.

347. Newcastle City Council. *Independent schools*. Newcastle City Council; 2014 [Accessed 06/08/2018]. Available from: https://www.newcastle.gov.uk/education-and-learning/find-a-school-or-academy/independent-schools.

348. Briggs A. Statistical methods for cost-effectiveness analysis alongside clinial trials. In: Jones A, editor. *The Elgar companion to Health Economics*. Glouscester, UK: Edward Elgar Publishing Limited; 2006. p. 503-13.

349. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Critical Assessment of economic evaluation. *Methods for the Economic Evaluation of Health Care Programmes* Oxofrd, England: Oxford University Press; 2005. p. 27-54.

350. Sanghera S, Frew E, Gupta JK, Kai J, Roberts TE. Exploring the Use of Cost-Benefit Analysis to Compare Pharmaceutical Treatments for Menorrhagia. *PharmacoEconomics*. 2015;33(9):957-65.

351. Ramsay CR, Clarkson JE, Duncan A, Lamont TJ, Heasman PA, Boyers D, et al. Improving the Quality of Dentistry (IQuaD): a cluster factorial randomised controlled trial comparing the effectiveness and cost-benefit of oral hygiene advice and/or periodontal instrumentation with routine care for the prevention and management of periodontal disease in dentate adults attending dental primary care. *Health Technol Assess*. 2018;22(38):1-144.

352. Office for National Statistics. *CPI INDEX*. 2019 [Accessed 26/07/2018]. Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/d7bt/mm23.

353. Curtis L, Burns A. *Unit Costs of Health and Social Care 2017*. Canterbury: University of Kent; 2017.

354. Home Office. *Crime, Justice and the law: Arrests*. 2019 [updated 18/03/19; Accessed 25/06/2019]. Available from: https://www.ethnicity-facts-figures.service.gov.uk/crime-justice-and-the-law/policing/number-of-arrests/latest.

355. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. *Value Health*. 2015;18(2):161-72.

356. Malehi AS, Pourmotahari F, Angali KA. Statistical models for the analysis of skewed healthcare cost data: a simulation study. *Health economics review*. 2015;5:11-.

357. Manning W. Dealing with skewed data on costs and expenditures In: Jones A, editor. *The Elgar Companion to Health Economics*. UK: Edward Elgar Publishing Limited; 2006. p. 439-46.

358. Saunders JB, Aasland OG, Babor TF, De La Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*. 1993;88(6):791-804.

359. Kind P, Brooks R, Rabin R. *EQ-5D concepts and methods: A developmental history*. Netherlands: Springer; 2005.

360. International Monetary Fund. *Report for Selected Countries and Subjects*. 2019 [Accessed 17/07/19]. Available from: https://tinyurl.com/yxcxrhmb.

361. Department of Health. Local Authority Circular on the public health ring-fenced grant 2018/19. GOV.UK: Department of Health; 2017.

362. Stone PW. Return-on-investment models. *Appl Nurs Res*. 2005;18(3):186-9.

363. Malcolm H, Wilson V, Davidson J, Kirk S. *Absence from School: A study of its causes and effects in seven LEAs*. Research Report. University of Glasgow; 2003. Report No.: 424.

364. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19(14).

365. Buscha F, Conte A. The Impact of Truancy on Educational Attainment during Compulsory Schooling: a Bivariate Ordered Probit Estimator with Mixed Effects. *The Manchester School*. 2014;82(1):103-27.

366. Hex N, Tatlock S. *Altogether Better: Social Return on Investment Case Studies* York: York Health Economics Consortium; 2011.

367. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD, et al. Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value in Health*. 2012;15(6):835-42.

368. Hartfiel N, Edwards RT. Cost-consequence analysis of public health interventions. In: Edwards RT, McIntosh E, editors. *Applied health economics for public health practice and research*. Oxford: Oxford University press; 2019. p. 233-47.

369. Jayatunga W. *Contraception: Economic Analysis Estimation of the Return on Investment (ROI) for publicly funded contraception in England*. London, UK; 2018, June 2018.

370. Optimity Advisors Ltd. *Cost-effectiveness and Return on Investment (ROI) of interventions associated with the Best Start in Life*. London, UK; 2018, September 2018.

371. Oxford English Dictionary. *Bottom line*. 2016 [Accessed 02/08/2019]. Available from: https://www.oed.com/view/Entry/36569594.

372. Owen L, Morgan A, Fischer A, Ellis S, Hoy A, Kelly MP. The cost-effectiveness of public health interventions. *Journal of Public Health*. 2012;34(1):37-45.

373. Owen L, Pennington B, Fischer A, Jeong K. The cost-effectiveness of public health interventions examined by NICE from 2011 to 2016. *J Public Health (Oxf)*. 2018;40(3):557-66.

374. CEMIPP. *Review of Cost-Effectiveness Methodology for Immunisation Programmes & Procurements*. 2016.

375. Lomas J, Claxton K, Martin S, Soares M. Resolving the "Cost-Effective but Unaffordable" Paradox: Estimating the Health Opportunity Costs of Nonmarginal Budget Impacts. *Value in Health*. 2018;21(3):266-75.

376. Martin S, Lomas JRS, Claxton K. Is an ounce of prevention worth a pound of cure? Estimates of the impact of English public health grant on mortality and morbidity. [Discussion paper]. In press 2019.

377. Musto V. *Estimating the social return on investment of treating substance-misusing parents: a guide to collecting local data*. London, UK; 2016.

378. Buck D. The King's Fund [Internet]: The King's Fund. 2018. [Accessed 26/07/18]. Available from: https://www.kingsfund.org.uk/blog/2018/04/return-investment-public-health#comments-top.

379. Hunter DJ. Meeting the Challenge of the "Know-Do" Gap; Comment on "CIHR Health System Impact Fellows: Reflections on 'Driving Change' Within the Health System". *International Journal of Health Policy and Management*. 2019;8(8):498-500.

380. Wilson P, Sheldon TA. Using evidence in health and healthcare. In: Boaz A, Davies H, Fraser A, Nutley S, editors. *What Works Now?: Evidence-Informed Policy and Practice*. Bristol: Policy Press; 2019.

381. McAteer J, Di Ruggiero E, Fraser A, Frank JW. Bridging the academic and practice/policy gap in public health: perspectives from Scotland and Canada. *Journal of Public Health*. 2018.

382. Cheetham M, Wiseman A, Khazaeli B, Gibson E, Gray P, Van der Graaf P, et al. Embedded research: a promising way to create evidence-informed impact in public health? *Journal of Public Health*. 2018;40(suppl_1):i64-i70.

383. Schuman H, Presser S. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*: SAGE Publications; 1996.

384. Cochrane M, Watson PM, Timpson H, Haycox A, Collins B, Jones L, et al. Systematic review of the methods used in economic evaluations of targeted physical activity and sedentary behaviour interventions. *Social Science & Medicine*. 2019;232:156-67.

385. Zanganeh M, Adab P, Li B, Frew E. A Systematic Review of Methods, Study Quality, and Results of Economic Evaluation for Childhood and Adolescent Obesity Intervention. *International Journal of Environmental Research and Public Health*. 2019;16(3).

386. National Institute of Health and Care Excellence. *Methods for the development of public health guidance (second edition)*. NICE; 2009 [Accessed 20/10/2015]. Available from: http://www.wphna.org/htdocs/downloadsmay2012/2009%20NICE%20public%20health%20guidance%20pdf.pdf.

387. Oliver K, Innvar S, Lorenc T, Woodman J, Thomas J. A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Services Research*. 2014;14(1):1-12.

388. Charles JM, Brown G, Thomas K, Johnstone F, Vandenblink V, Pethers B, et al. Use of programme budgeting and marginal analysis as a framework for resource reallocation in respiratory care in North Wales, UK. *Journal of Public Health*. 2016;38(3):e352-e61.

389. Sullivan SM, Wells G, Coyle D. What Guidance are Economists Given on How to Present Economic Evaluations for Policymakers? A Systematic Review. *Value in Health*. 2015;18(6):915-24.

390. McIntosh E, Baba C, Botha W. Cost-benefit analysis for applied public health economic evaluation. In: Edwards RT, McIntosh E, editors. *Applied health economics for public health practice and research*. Handbooks in Health Economic Evaluation. First ed. Oxford: Oxford University Press; 2019.

391. Fujiwara D, Campbell R. *Valuation Techniques for Social Cost-Benefit Analysis: Stated Preference, Revealed Preference and Subjective Well-Being Approaches*. HM Treasury, Department for Work and Pensions; 2011.

392. Barrett B, Byford S, Crawford MJ, Patton R, Drummond C, Henry JA, et al. Cost-effectiveness of screening and referral to an alcohol health worker in alcohol misusing patients attending an accident and emergency department: a decision-making approach. *Drug and Alcohol Dependence*. 2006;81(1):47-54.

393. Cowell AJ, Brown JM, Mills MJ, Bender RH, Wedehase BJ. Cost-effectiveness analysis of motivational interviewing with feedback to reduce drinking among a sample of college students. *Journal of Studies on Alcohol and Drugs*. 2012;73(2):226-37.

394. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Economics*. 2004;13(5):461-75.

395. Cameron AC, Trivedi PK. *Microeconometrics Using Stata*. Texas: Stata Press; 2010.